

A Compact Shape Descriptor Based on the Beam Angle Statistics

Nafiz Arica and Fatos T. Yarman-Vural

Computer Engineering Department,
Middle East Technical University,
06531 Ankara / Turkey ,
{nafiz , vural}@ceng.metu.edu.tr

Abstract. In this study, we propose a compact shape descriptor, which represents the 2-D shape information by 1-D functions. For this purpose a two-step method is proposed. In the first step, the 2-D shape information is mapped into 1-D moment functions without using a predefined resolution. The mapping is based on the beams, which are originated from a boundary point, connecting that point with the rest of the points on the boundary. At each point, the angle between a pair of beams is taken as a random variable to define the statistics of the topological structure of the boundary. The second order statistics of all the beam angles is used to construct 1-D Beam Angle Statistics (BAS) functions. In the second step, the 1-D functions are further compressed by using Discrete Fourier Transforms applied on the BAS functions of the shape boundary. BAS function is invariant to translation, rotation and scale. It is insensitive to distortions. Experiments are done on the dataset of MPEG 7 Core Experiments Shape-1. It is observed that proposed shape descriptor outperforms the popular MPEG 7 shape descriptors.

1 Introduction

Shape descriptors are important tools in content-based image retrieval systems, which allow searching and browsing images in a database with respect to the shape information. The goal of the shape descriptors is to uniquely characterize the object shape, in a large image database with large within-class-variances and small between-class-variances. A vigorous shape descriptor should be invariant under affine transform and insensitive to noise. It should contain sufficient information to resolve distinct images and compact enough to ignore the redundancies in the shapes. Additionally, it should give results consistent to human visual system.

The shape description methods can be divided into three main categories; contour based [11], [8], [4], image based [7] and skeleton based descriptors [9]. In the Core Experiment CE-Shape-1 for the MPEG-7 standard, the performance of above referenced shape descriptors are evaluated. Among them, the descriptors proposed in [11] and [8] outperform the others in most of the experiments. In [11], the scale space approach is applied to shape boundary. The simplified contours are obtained by the

scale-space curve evolution, based on contour smoothing with a Gauss function. The scale space image of the curvature function is then used as hierarchical shape descriptor. In [8], the scale of shapes are reduced by a process of curve evolution and represented in tangent space. To compute the similarity measure, the best possible correspondence of visual parts is established. Unfortunately, the techniques mentioned above rely on the threshold values corresponding to the resolution of the shape representation, which depends on the level of the smoothness of the images in the database. The methods have, also, limited performance under noise.

The motivation of this paper is to find a compact representation of the two-dimensional shape information without a predefined resolution scale. For this purpose a two-step method is proposed. In the first step, a mapping, which transforms the 2-D shape information into a set of 1-D functions, is formed. This mapping is consistent with the human visual system, preserving all the convexities and concavities of the shape. In the second step the 1-D function is further compressed by using Discrete Fourier Transforms.

The proposed descriptor is based on the beams, which are the lines connecting a point with the rest of the points on the boundary. The characteristics of each boundary point can be extracted by using the beams. The angle between each pair of beams is taken as the random variable at each point on the boundary. Then, the moment theorem provides the statistical information, which is used to obtain the desired 1-D Beam Angle Statistics (BAS) functions, where the valleys and hills of the first moment correspond to concavities and convexities of the object shape. After the representation of 2-D shape boundary as a set of 1-D BAS functions, further compression is realized by using the Spectral Coefficients of the functions.

The main contribution of this study is to eliminate the use of a heuristic rule or empirical threshold value in representation of shape boundaries in a predefined resolution scale. The proposed method also gives globally discriminative features to each boundary point by using all other boundary points. Another advantage of this representation is its consistency with human perception through preserving the concave and convex parts of the shapes. It provides a compact representation of shapes, which is insensitive to noise and affine transform. It is invariant to size, orientation and position of the object.

2 Mapping 2-D shape information into a set of 1-D Functions

Curvature function is a popular 1-D representation of 2-D shape information and has been an inspiration of many studies related to shape analysis [10], [5]. The curvature function can be computed as the derivative of the contour's slope function. In a discrete grid, this can be performed by K-slope method. K-slope at a boundary point is defined as the slope of a line connecting that point with its K^{th} right neighbor. Then, the K-curvature at a boundary point is defined as the difference between the K-slope at that pixel and the K-slope of its K^{th} left neighbor.

The K-curvature function may be exploited to form an appropriate shape descriptor, if one could identify an optimal value for the parameter K, which extracts the concavities and convexities of the shape at a predefined scale settled by K. At this point, we need a rigorous technique to identify the scale parameter K, which discriminates a wide range of shapes in large image databases. In [1], the curvature function is obtained by using a fixed K value and then it is filtered in order to stress the main features. Another study [12] uses an adaptive K value by changing it according to the distance between relevant points. However, choosing a single K value cannot capture the exact curvature information for all varieties of shapes in a large database.

Figure 1 b, c and d indicate the plot of K-curvature function for a sample shape with various K values. Note that the plots for small K capture fine details. By increasing K, it is possible to plot fine to coarse representation of a closed boundary as a one-dimensional function. Examining these figures shows that each peak and valley of the curvature function plot corresponds to a convexity, and a concavity of the shape. For this particular example, the shape information is preserved in the peaks corresponding to the head, tail and fins. The remaining peaks and valleys of the curvature plot are rather redundant details that result from the noise in the shape, leading mismatches in the image database. Therefore, one needs to select an appropriate K to avoid this redundant information. Selection of K defines the amount of smoothing of the shape and highly depends on the context of the image. If K smoothes the ripples corresponding to some context information, this will result in important information loss. On the other hand, if the ripples correspond to noise, keeping them will increase the number of convexities and concavities, which may carry superfluous shape information. As a result, the problem of selecting a generic K, which resolves the necessary and sufficient information for all the images, has practically no solution, due to the diversity of the shape context in large databases.

The above discussion leads us to somehow find a representation, which employs the information in K-curvature function for all values of K. Superposition of the plots of K-curvature function for all K, yields impractically large data with no formal way of representation, as indicated in figure 1.e.

In this study, we attack this problem by modeling the shape as the outcomes of a stochastic process, which is generated by the same source at different scales. In this model, Figure 1.e shows the possible outcomes of the shape curvature plot, which generates the fish shape. At a given boundary point $p(i)$, the value of K-curvature function is assumed to be a function of a random variable K and may take one of the 1-D function indicated in the figure, depending on the values of K. Therefore, K-curvature function for each K, can be considered as the output of a stochastic process.



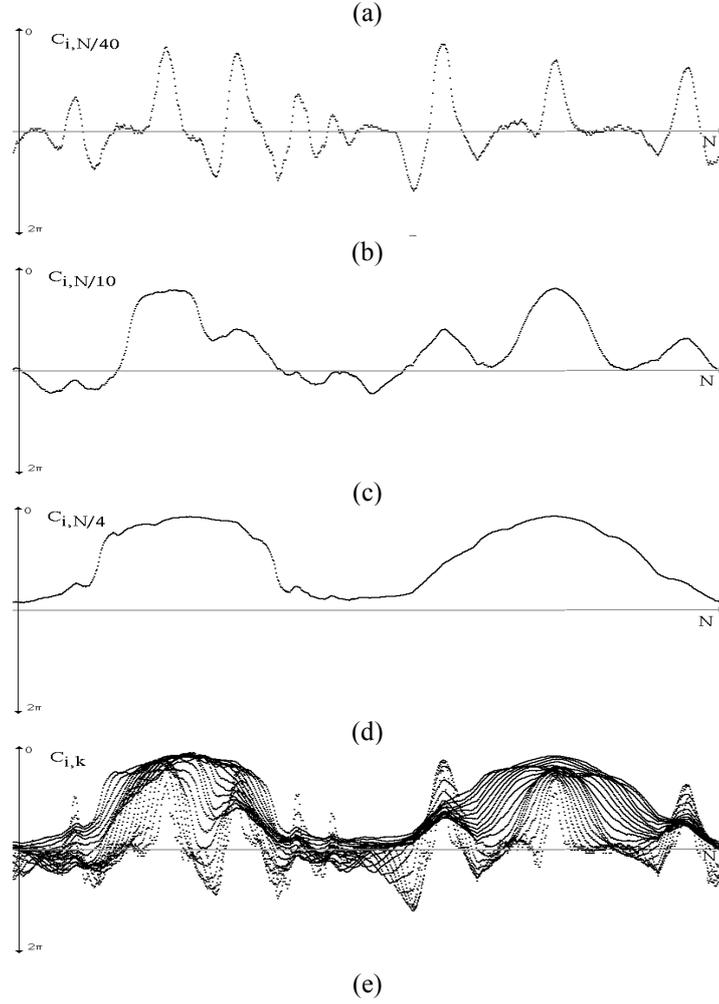


Fig. 1. a) A sample shape boundary and K-Curvatures for b) $K=N/40$, c) $N/10$, d) $N/4$ and e) $C_{i,k}$ Plots for $N/40 \dots N/4$ from Top to Bottom Respectively.

Mathematically speaking, let the shape boundary $B = \{ p(1), \dots, p(N) \}$ is represented by a connected sequence of points,

$$p(i) = (x(i), y(i)) \quad i = 1, \dots, N, \quad (1)$$

where N is the number of boundary points and $p(i) = p(i+N)$. For each point $p(i)$, the *beams* of $p(i)$ is defined as the set of vectors;

$$L[p(i)] = \{ V_{i+j}, V_{i-j} \} \quad (2)$$

where V_{i+j} and V_{i-j} are the forward and backward vectors connecting $p(i)$ with the points, $p(i+j)$ and $p(i-j)$ in the boundary, for $j=1, \dots, N/2$. Figure 2 indicates the beams and beam angle at point $p(i)$, respectively.

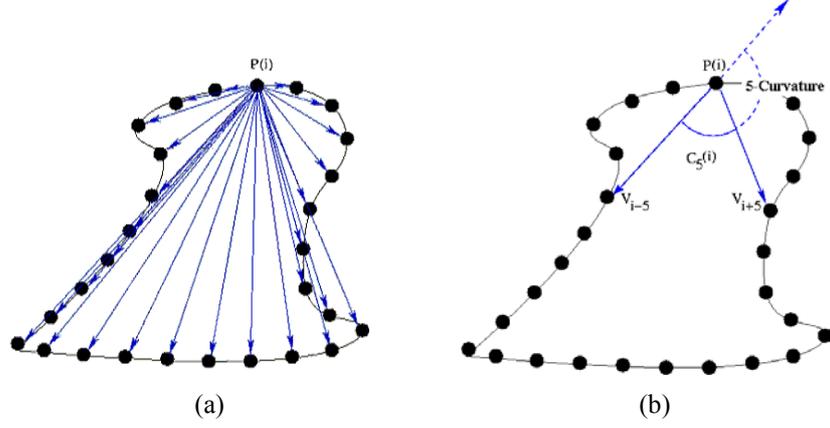


Fig. 2. a) The beams of point $p(i)$, b) The 5-Curvature and the beam angle for $K=5$ at the boundary point $p(i)$

The slope of each beam, V_{i+K} is then calculated as,

$$\theta_{V_{i+l}} = \tan^{-1} \frac{\Delta y_{i+l}}{\Delta x_{i+l}}, \quad l = \pm K. \quad (3)$$

For the point $p(i)$ and for a fixed K , the beam angle between the forward and backward beam vectors, is then computed as (see figure 2.b).

$$C_K(i) = (\theta_{V_{i-K}} - \theta_{V_{i+K}}). \quad (4)$$

Note that, beam angle for a fixed K is nothing but the K -curvature function (which takes values between 0 and 2π).

Now, for each boundary point $p(i)$ of the curve Γ , the beam angle $C_K(i)$ can be taken as a *random variable* with the probability density function $P_K(C_K(i))$ and $C_K(i)$ vs. i plot for each K becomes an outcome of the stochastic process which generates the shape at different scales. Therefore, Beam Angle Statistics (BAS), may provide a stochastic representation for a shape descriptor. For this purpose, m^{th} moment of the random variable $C_K(i)$ is defined as follows:

$$E[C_K^m(i)] = \sum_K C_K^m P_K(C_K(i)) \quad m = 0, 1, 2, 3, \dots \quad (5)$$

In the above formula E indicates the expected value operator and $P_K(C_K(i))$ is the probability density function of $C_K(i)$. Note that the maximum value of K is $N/2$, where N represents the total number of boundary points. Note also that, the value of $C_K(i)$ approaches to 0 as K approaches to $N/2$. During the implementations, $P_K(C_K(i))$ is approximated by the histogram of $C_K(i)$ at each point $p(i)$.

The moments describe the statistical behavior of the beam angle at the boundary point $p(i)$. Each boundary point is, then, represented by a vector whose components are the moments of *the beam angles*:

$$\Gamma(i) = [E[C^1(i)], E[C^2(i)], \dots] = [\Gamma^1(i), \Gamma^2(i), \dots] \quad (6)$$

The shape boundary is finally represented by plotting the moments of $C_K(i)$'s for all the boundary points. In the proposed representation, the first moment $\Gamma^1(i)$ preserves the most significant information of the object. Figure 3, indicates the first three moments of boundary points for the sample shape of figure 1.a. For this particular example, it should be noted that the first moment suffice to represent the convexities and concavities of the shape. The second moment $\Gamma^2(i)$ increases the discriminative power of the representation. The third moment $\Gamma^3(i)$ on the other hand, does not bring any additional significant information to the representation. The order of the statistics naturally depends on the characteristics of the probability density function, $P_K(C_K(i))$. Central limit theorem provides us a strong theoretical basis to assume Gaussian distribution for each point $p(i)$ for large N . This implies that second order statistics is sufficient for representing most of the shape information provided that we have enough samples on shape boundary.

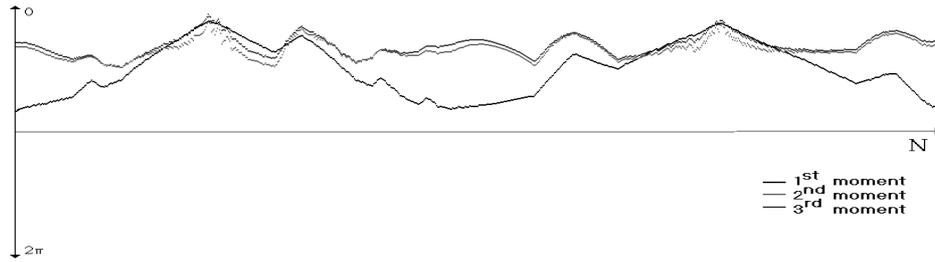


Fig 3. Third order statistics of Beam Angle for the shape in figure 1.a.

The details of BAS function, explained above, can be found in [2].

3 Compression of BAS functions

After mapping 2-D shape information into the 1-D BAS functions, the next step is to construct a compact feature vector, which retains the information in the BAS functions. For this purpose, two different approaches are proposed. In the first approach, Fourier descriptors are used as the entries of feature vector. The second approach identifies the Nyquist rate for the overall image database in frequency domain and performs sampling on 1-D BAS functions in the space domain.

3.1 Fourier Descriptors

Fourier Descriptors (FD) are commonly used technique for characterizing the shape boundary. Basic advantages of FD method in addition to the well-established theory of Fourier Transformations, include the ability to characterize a shape boundary with a small size descriptors. Besides, FDs are easy to derive and achieve good representation.

For a given m^{th} moment BAS function $\Gamma^m(i)$, the spectral coefficients is calculated by

$$a_n^{(m)} = \frac{1}{N} \sum_{i=0}^{N-1} \Gamma^m(i) \exp(-j2\pi i / N) \quad m = 1, \dots, M \quad (7)$$

In order to achieve invariance in starting point of boundary extraction, phase information is ignored and only the magnitudes $|a_n^{(m)}|$ are used. The coefficients are also normalized by dividing the magnitudes with the DC component, $|a_0^{(m)}|$. Then, the T lowest frequency Fourier coefficients are used to construct the feature vector. T is taken as a variable in the experiments and used to identify the Nyquist rate. Finally, the feature vector for a given BAS function is formed as follows:

$$F^{(m)} = \left[\frac{|a_1^{(m)}|}{|a_0^{(m)}|}, \frac{|a_2^{(m)}|}{|a_0^{(m)}|}, \dots, \frac{|a_T^{(m)}|}{|a_0^{(m)}|} \right] \quad m = 1, \dots, M \quad (8)$$

The shape boundary is then represented by concatenating $F^{(m)}$'s for each BAS functions, $\Gamma^1(i)$ and $\Gamma^2(i)$. The similarity between two shapes is measured by Euclidean distance between features $[F^{(1)}, \dots, F^{(M)}]$ extracted from the BAS moment functions.

3.2 Sampling By Fourier Transformation

The easiest method for feature extraction from BAS functions is to perform sampling with equal distance. However, there is a trade off between the sampling rate and the accuracy of the representation. As the sampling rate decreases, the method loses information about the visual parts (convexities and concavities). In order to find an optimal sampling rate, Fourier analysis is utilized in this study.

The feature extraction is performed in two steps. In the first step, the Fourier Transform of the BAS functions $\Gamma^m(i)$, are calculated. In the second step, Inverse Fourier transformation of the first T coefficients is calculated. In other words, low-pass filter on the frequency domain is performed. T is taken as variable during the experiments to find the optimal rate for the images in the database.

The similarity between the features is measured by Optimal Correspondent Subsequence (OCS) algorithm [13]. The objective in similarity measurement is to minimize the distance between two vectors by allowing deformations. This is achieved by first solving the correspondence problem between items in feature vectors and then computing the distance as a sum of matching errors between corresponding items. Dynamic programming technique is employed for minimizing the total distance between

corresponding items by building a minimum distance table, which accumulates the information of correspondence.

4 Experiments

The performance of the BAS descriptor is tested in the data set of MPEG 7 Core Experiments Shape-1 Part B, which is the main part of the Core Experiments. The total number of images in the database is 1400. There are 70 classes of various shapes, with 20 images in each class. Each image is used as query and the number of similar images, which belong to the same class, was counted in the top 40 matches. Since the maximum number of correct matches for a single query image is 20, the total number of correct matches is 28000.

In the first set of experiments the BAS descriptor is compared with the popular descriptors of MPEG7. For this purpose, the boundaries of objects are mapped into mean and variance BAS functions. Then, 100 points from each BAS function with equal distances are used as the feature vector. C_i values between 0 and 2π for the boundary points are taken as it is. Then, elastic matching algorithm is employed for similarity measurement. The proposed descriptor correctly matches 81.4% of the images. In Table 1, the comparison of the BAS function with the recently reported results of [3] (Shape Context), [8] (Tangent Space), [11] (Curvature Scale Space), [7] (Zernika Moments), [4] (Wavelet) and [9] (Directed Acyclic Graph) is provided. As it is seen from the table, the proposed descriptor performs better than the best-performance descriptors available in the literature, for the data set of MPEG CE Shape-1 part B.

Table 1: Best Performances on MPEG 7 CE Shape-1 Part B.

	Shape Context	Tangent Space	CSS	Zernika Moment	Wavelet	DAG	BAS
Similarity Results (%)	76.51	76.45	75.44	70.22	67.76	60	81.4

Using the row BAS functions as the shape descriptor is computationally expensive because of the dimension of the feature space and the complexity of the elastic matching algorithm. The first step is to further compress the size of feature vector to a “reasonable” dimension. At this point, Fourier theorems provide us convenient tools to represent the BAS functions in a more compact form. In the second set of experiments, the Fourier transform of BAS functions are used to find an optimal sampling rate. In the transform domain the Fourier spectrum is truncated after the T coefficients. The descriptor is, then, defined by using the T components of the inverse transform. It is observed that, for $T > 64$, no further improvement is achieved. The reason for this, is depicted in figure 4, where the power spectrum of BAS mean function extracted from a sample image is plotted. The plot shows the rapid decay of the spectral coefficients which becomes zero after $n=60$. Table 2 indicates the results of

using inverse Fourier transform as a shape descriptor for various T values. Note that even for $T = 32$ the descriptor yields better results than the other MPEG-7 descriptors reported in the literature.

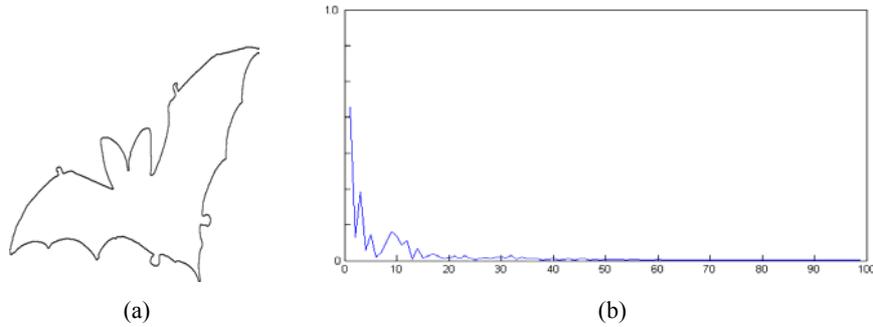


Fig 4. a) A Sample shape b) Power Spectrum of its BAS mean function.

Table 2: Comparison of row BAS Function and its inverse Fourier Transform with various T values.

Dimension	Similarity Results (%)	
	Sampling with Equal Distance	Sampling By Fourier Transform
$T=8$	55.2	64.3
$T=16$	69	72.8
$T=32$	77.7	80.4
$T=64$	81.3	81.3

Finally, in order to avoid the complexity of elastic matching algorithm, the Fourier descriptors of BAS function are used and the similarity is simply measured by the Euclidean distance. This approach is indeed very fast and simple with the considerable tradeoff in the retrieval rate as indicated in Table 3. The decrease in the retrieval rates is expected due to the loss of the phase information. However, when compared to the available FD methods, it is observed that the BAS function FD's outperforms the centroid distance function FD's, which is reported as the best FD method in [6]. Note that the BAS FD's give considerably better results even only the first order BAS is used.

Table 3: Comparison of Fourier Descriptors on BAS Function and Centroid Distance Function with various T values.

Dimension	Similarity Results (%)		
	Centroid Distance Function FD's	BAS FD's	
		1 st moment	1 st + 2 nd Moment
$T=8$	65.46	67.51	70.66
$T=16$	66.61	68.91	72.05
$T=32$	66.72	69.12	72.21
$T=64$	66.73	69.13	72.23

5 Conclusion

This study introduces a compact shape descriptor for identifying the similar objects in an image database. The two-dimensional object silhouettes are represented by one-dimensional BAS moment functions, which capture the perceptual information using the statistics of the beam angles of individual points. The BAS functions avoid smoothing and preserve the available information in the shape. It also avoids the selection of a threshold value to represent the resolution of the boundary, thus eliminates the context-dependency of the representation to the data set. Therefore, rather than using a single representation of the boundary, at a predefined scale, the BAS functions gather the information at all scales. It gives globally discriminative features to each boundary point by using all other boundary points. It is consistent with human perception through preserving the concave and convex parts of the shapes. The computational cost of BAS function is reduced by using the Discrete Fourier Transforms.

References

1. Agam, G., Dinstein, I., Geometric Separation Of Partially Overlapping Nonrigid Objects Applied to Automatic Chromosome Classification. *IEEE Trans. PAMI*, 19, (1997) 1212-1222
2. Arica, N., Yarman-Vural F. T., BAS: A Perceptual Shape Descriptor Based On The Beam Angle Statistics, *Pattern Recognition Letters*, vol: 24/9-10, (2003) 1627-1639
3. Belongie, S., Malik, J., Puzicha, J., Shape Matching and Object Recognition Using Shape Contexts. *IEEE Trans. PAMI*, 24, 4, (2002) 509-522
4. Chuang, G., Kuo, C. -C., Wavelet Descriptor of Planar Curves: Theory and Applications. *IEEE Trans. Image Processing*, 5, (1996) 56-70
5. Cohen, F. S., Huang, Z., Yang, Z., Invariant Matching and Identification of Curves Using B-Splines Curve Representation. *IEEE Trans. Image Processing*, 4, (1995) 1-10
6. Zhang, D., Lu, G., A Comparative Study Of Fourier Descriptors for Shape Representation and Retrieval. *ACCV2002, Asian Conference on Computer Vision*, (2002) 646-651
7. Khotanzad, A., Hong, Y. H., Invariant Image Recognition By Zernike Moments. *IEEE Trans. PAMI*, 12, (1990) 489-497
8. Latecki, L. J., Lakamper, R., Shape Similarity Measure Based on Correspondence of Visual Parts. *IEEE Trans. PAMI*, 22, 10, (2000) 1185-1190
9. Lin, L. -J., Kung, S. Y.,. Coding and Comparison of Dags as a Novel Neural Structure With Application To Online Handwritten Recognition. *IEEE Trans. Signal Processing*, (1996)
10. Loncaric, S., A survey of Shape Analysis Techniques. *Pattern Recognition*, 31, (1998) 983-1001
11. Mokhtarian, F., Abbasi, S., Kittler, J., Efficient and Robust Retrieval By Shape Content Through Curvature Scale Space. *Image Databases and Multimedia Search*, A. W. M. Smeulders and R. Jain ed., 51-58 World Scientific Publication (1997)
12. Urdiales, C., Bandera, A., Sandoval, F., Non-parametric Planar Shape Representation Based on Adaptive Curvature Functions. *Pattern Recognition*, 35, (2002) 43-53
13. Wang Y. P., Pavlidis T., Optimal Correspondence of String Subsequences. *IEEE Trans. PAMI*, 12, (1990) 1080-1087