A Comparative Study of Fourier Descriptors for Shape Representation and Retrieval

Dengsheng Zhang and Guojun Lu
Gippsland School of Computing and Information Technology
Monash University
Churchill, Victoria 3842, Australia
Email: dengsheng.zhang, guojun.lu@infotech.monash.edu.au

Abstract

Shape is one of the primary low level image features in Content Based Image Retrieval (CBIR). Many shape representations and retrieval methods exist. However, most of those methods either do not well capture shape features or are difficult to do normalization (making matching difficult). Among them, methods based Fourier descriptors (FDs) achieve both good representation (perceptually meaningful) and easy normalization. Besides, FDs are easy to derive and compact in terms of representation. Design of FDs focuses on how to derive Fourier invariants from Fourier coefficients and how to obtain Fourier coefficients from shape signatures. Different Fourier invariants and shape signatures have been exploited to derive FDs. In this paper, we study different FDs and build a Java retrieval framework to compare shape retrieval performance using different FDs in terms of computation complexity, robustness, convergence speed and retrieval performance. The retrieval performance of the different FDs is compared using a standard shape database.

Keywords*: CBIR, Shape, Fourier descriptors, Retrieval.*

1. Introduction

In the newly emerged multimedia application CBIR, shape is exploited as one of the several primary low level image features for image retrieval [8]. Many shape representations and retrieval methods exist. However, most of those methods either do not well capture shape features or are difficult to do normalization (making matching difficult). Among them, methods based Fourier descriptors (FDs) achieve both good representation and easy normalization. Besides, FDs are easy to derive and compact in terms of representation.

Various FD methods have been reported in the literature, these include using FD for shape analysis [10][12], character recognition [2][7], shape classification [5] and shape retrieval [3][9][11][13]. In these methods, different shape signatures and different Fourier invariants have been exploited to obtain FDs. However, FDs derived from different signatures has significant different performance on shape retrieval. In this paper, we compare shape retrieval using FDs derived from different shape signatures and from different Fourier invariants in terms of computation complexity, robustness, convergence speed and retrieval performance. The rest of the paper is organized as following. In Section 2, we introduce FD and affine Fourier invariants. Section 3 discusses different shape signatures used to derive FD and in Section 4, we analyze the convergence speed of the Fourier series of each shape signatures. Section 5 gives retrieval performance of each FDs. We concludes the paper in Section 6.

2. Fourier Descriptors

For a given shape defined by a closed curve C which in turn is represented by a one dimensional function $u(t)$, called shape signature. At every time t, there is a complex $u(t)$, $0<t<T$, T is the period of t. Since $u(t)$ is periodic, we have $u(t+nT) = u(t)$. The discrete Fourier transform is given by

$$a_n = \frac{1}{N}\sum_{t=0}^{N-1} u(t)\exp(-j2\pi nt/N) \quad n \in \mathbb{Z}$$

The coefficients a_n, $n = 0, 1, \ldots, N-1$, are used to derive Fourier descriptors (FDs) of the shape.

2.1 Fourier Descriptors

The general form for the Fourier coefficients of a contour generated by translation, rotation, scaling and change of start point from an original contour is given by:

$$a_n = \exp(jn\tau)\cdot\exp(j\phi)\cdot s \cdot a_n^{(o)}$$

where $a_n^{(o)}$ is the nth Fourier coefficient of the original shape. To achieve translation, rotation invariance, phase information of the a_n are ignored and only the magnitudes $|a_n|$ are used. Scale invariance is achieved by dividing the magnitudes by the *DC* component, i.e., $|a_0|$. The normalized Fourier coefficients are called FD. The similarity between a query shape *Q* and a target shape *T* is measured by the Euclidean distance between their FD representations. Different $u(t)$ have been used to derive FD, they will be discussed in Section 3.

2.2 Affine Fourier Invariants

The above normalization generates shape descriptors invariant under translation, rotation, and scaling. It is also desirable for a shape representation to be invariant to affine transform to address shapes obtained from different views of objects. Arbiter [1] has proposed the use of the following affine invariants as Fourier descriptors.

$$Q_k = \frac{X_k Y_p^* - Y_k X_p^*}{X_p Y_p^* - Y_p X_p^*} \quad k = \pm 1, \pm 2, \ldots$$

where $\mathbf{U}_k = (X_k, Y_k)^T$ and X_k, Y_k are the Fourier coefficients of $x(t)$, $y(t)$ respectively, p is a constant and $p \neq 0$. However, Q_k cannot be derived from the shape signatures described in Section 3. To derive affine Fourier invariants, the Fourier transform derived from an analytic solution originally used by Arbert [1] is used.

3. Shape Signatures

Fourier descriptors are derived from a *shape signature*. In general, a *shape signature* $u(t)$ is any 1-D function representing 2-D areas or boundaries. Different shape signatures have been used to derive FD, in this section we describe these shape signatures in details. In the following, we assume the shape boundary coordinates $(x(t), y(t))$, t = 0, 1, …, N-1, have been extracted in the preprocessing stage, t usually means arclength. In our implementation, the shape boundary points are extracted through a 8-connectivity contour tracing technique [6]. For notation convenience, different function names will be employed to denote different shape signatures.

3.1 Complex Coordinates

A *complex coordinates function* is simply the complex number generated from the boundary coordinates:

$$z(t) = [x(t) - x_c] + i[y(t) - y_c]$$

where (x_c, y_c) is the centroid of the shape, which is the average of the boundary coordinates

$$x_c = \frac{1}{N}\sum_{t=0}^{N-1} x(t) \quad y_c = \frac{1}{N}\sum_{t=0}^{N-1} y(t)$$

$z(t)$ is a straightforward representation of shape boundary. $z(t)$ is a translation invariant signature. The advantage of using complex coordinates function is it involves no extra computation in derive shape signature.

3.2 Centroid Distance

The *centroid distance* function is expressed by the distance of the boundary points from the centroid (x_c, y_c) of the shape

$$r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2}$$

$r(t)$ is invariant to translation. Computation of $r(t)$ is low. Figure 1 shows the centroid distance signatures of the an apple shape (referred to as the apple in this paper).

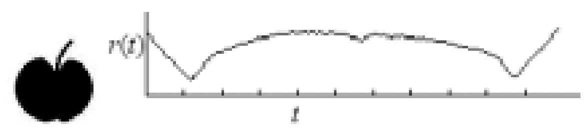


Figure 1. The centroid distance signatures of an apple.

3.3 Chord Length Signature

The *chord length function* $r^*(t)$ is derived from shape boundary without using any reference point. For each boundary point $\mathbf{p}$, its $r^*(t)$ is the distance between $\mathbf{p}$ and another boundary point $\mathbf{p}'$ such that $\mathbf{pp}'$ is perpendicular to the tangent vector at $\mathbf{p}$. This definition can cause ambiguities when the line of $\mathbf{pp}'$ pass through more than one boundary points. To solve this problem, another constraint is added by limiting $\mathbf{pp}'$ within the shape. For example, in Figure 2(a), $\mathbf{p}_1$ is one of the candidates for $\mathbf{p}'$, however by checking the middle point $\mathbf{p}_2$ of $\mathbf{pp}_1$, $\mathbf{p}_1$ should be eliminated because $\mathbf{p}_2$ is not within the shape. If $\mathbf{p}_2$ is still within the shape, the middle point of $\mathbf{pp}_2$ and the middle of $\mathbf{p}_1\mathbf{p}_2$ are checked by the second constraint. This process is repeated until a middle point is found to be outside the shape. If after this recursive checking process, no outer middle point is found, then $\mathbf{p}' = \mathbf{p}_1$. The $r^*(t)$ overcomes the biased reference point (which means the centroid is often biased by boundary noise or defections) problems, however, the non-reference-point representation can cause problem when shape is traced in different directions. For example, for point $\mathbf{q}$ on shape (b), its chord length will be $\mathbf{qq}_1$ when tracing in counter clockwise, however, it will be $\mathbf{qq}_2$ when tracing in clockwise. This will result in quite different shape signatures for mirrored shapes. In addition, $r^*(t)$ is very sensitive to noise, there may be drastic burst in the signature of even smoothed shape boundary. To reduce noise sensitivity, a post-processing using an average filter (width=7) is attempted to eliminate false features (Figure 2(b)(c)). $r^*(t)$ is invariant to translation. The computation to derive $r^*(t)$ is expensive.

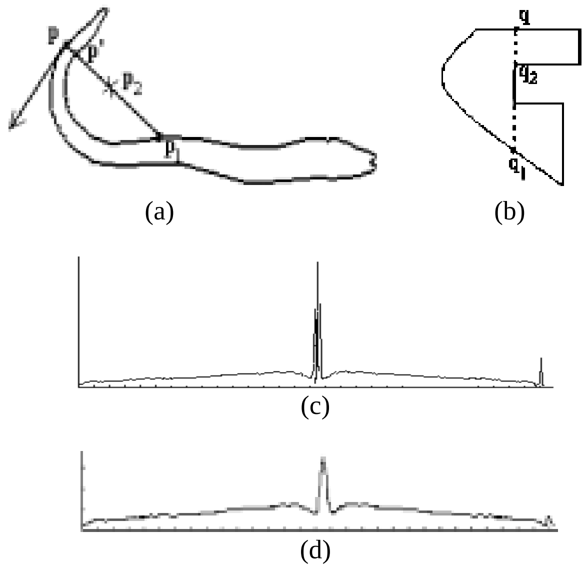


Figure 2. (a) chord length at **p**; (b) multiple chord length at **q**; (c) $r^*(t)$ of (a); (d) average smoothed $r^*(t)$ of (c).

3.4 Cumulative Angular Function

Intuitively, the tangent angles of the shape boundary indicate the change of angular directions of the shape boundary. The change of angular directions is important to human perception. Therefore, shape can be represented by its boundary tangent angles:

$$\theta(t) = \arctan \frac{y(t) - y(t-w)}{x(t) - x(t-w)} \qquad (3.4.1)$$

where w, a integer, is a jump step used in practice. However the tangent angle function $\theta(t)$ can only assume values in a range of length 2π, usually in the interval of $[-\pi, \pi]$ or $[0, 2\pi]$. Therefore $\theta(t)$ in general contains discontinuities of size 2π. Because of this, a cumulative angular function is introduced to overcome this problem. The *cumulative angular function* $\varphi(t)$ is the net amount of angular bend between the starting position $z(0)$ and position $z(t)$ on the shape boundary

$$\varphi(t) = [\theta(t) - \theta(0)] \bmod (2\pi) \qquad (3.4.2)$$

$\varphi(t)$ is continuous at places where $\theta(t)$ is multiples of 2π. $\varphi(t)$ is 2π periodic, it is suitable for Fourier transform. The normalized variant of $\varphi(t)$ is defined by Zahn and Roskies [10] using normalized arclength (assuming boundary is traced counter clock-wise)

$$\psi(t) = \varphi\left(\frac{Lt}{2\pi}\right) - t \qquad (3.4.3)$$

The subtraction of t from the cumulative angles makes $\psi(t) \equiv 0$ for a circle and $\psi(t) \neq 0$ for other shapes. This conforms to human intuition that a circle is "shapeless". An example of $\psi(t)$ is given in Figure 3.

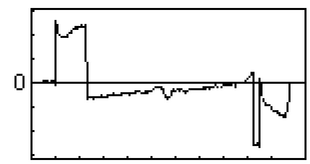


Figure 3. $\psi(t)$ of the apple.

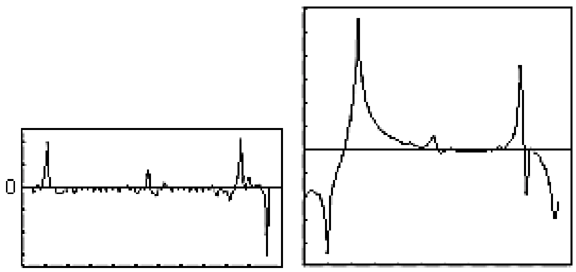
$\psi(t)$ is invariant under translation, rotation and scaling. Cumulative angular signature uniquely describe a shape. However, boundary noise can cause much bigger change in the representation than the change in centroid distance therefore, the structure of $\psi(t)$ is usually much more rugged than $r(t)$ (Figure 3). Since cumulative angular signature is derived from boundary tangents which are actually the first derivatives of the boundary coordinates, it usually contains discontinuities in the representation. As can be expected, its Fourier series converge rather slowly (see Section 5).

3.5 Curvature signature

Curvature is a very important boundary feature for human to judge similarity between shapes. It is not surprise that many researchers use curvature for shape representation. *Curvature function* is given by

$$\kappa(t) = d\theta / dt \qquad (3.5.1)$$

where θ is defined in (3.4.1). Although curvature is important curve feature, there is a problem for using curvature as shape representation. For many digital curve, especially for polygonal curve, $\theta(t)$ is a step function, so $\kappa(t)$ is zero almost everywhere and infinite at the discrete jumps of $\theta(t)$. This makes $\kappa(t)$ a poor candidate for shape representation. In order to use $\kappa(t)$ for shape representation, a smooth curvature function should be derived. It is appealing to use Fourier reconstructed shape to derive curvature as the reconstructed shape is an approximation to the original shape and is smooth. The reconstructed $\psi(t)$ is used to derive the smoothed curvature $k(t)$ [10]. Figure 4(a)(b) show the different curvature signatures of the apple. It is clear that (b) is more stable than (a). Although $k(t)$ works well for polygon shapes, however, for non-polygon shapes, shape features may be lost in the smoothed curvature representation.

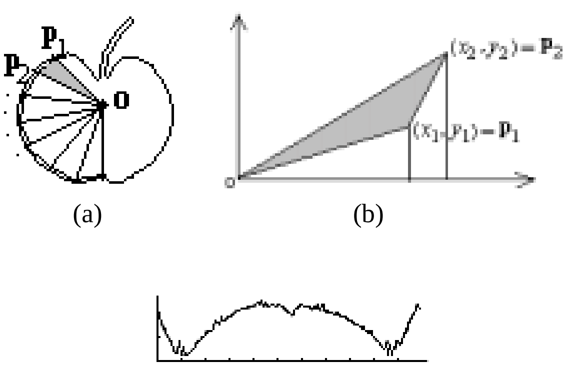


(a) (b)

Figure 4. (a) $\kappa(t)$ of the apple; (b) $k(t)$ of the apple.

3.6 Area Function

When the boundary points change along the shape boundary, the area of the triangle formed by the two boundary points and the center of gravity also changes (Figure 5(a)). This forms an *area function* which can be exploited as shape representation.



(a) (b)

(c)

Figure 5. (a) area function of the apple; (b) area of a triangle; (c) $A(t)$ of the apple.

For the triangle formed by **o**, $\mathbf{p}_1$ and $\mathbf{p}_2$ in Figure 5(b), its area is given by [35]

$$A(t) = \frac{1}{2} | x_1(t) y_2(t) - x_2(t) y_1(t) |$$

For each boundary points, the area of the triangle with 5 degree angle at vertex **o** is calculated. Figure 5(c) shows the $A(t)$ of the apple.

It has been found that $A(t)$ is very similar to $r(t)$. However $A(t)$ is more computation expensive than $r(t)$. Due to the numerical error in calculating angles, $A(t)$ is more rugged than $r(t)$. The derivation of $A(t)$ involves more computation than that of $r(t)$. $A(t)$ is linear under affine transform, this property can be exploited to generate affine invariant FDs from $A(t)$. However, this linearity only works for polygon shape sampled at its vertices. For polygon shape which is sampled at every point, the $A(t)$ may not be linear under affine transform due to the change of boundary perimeter. The sampled points on the original shape may not exist at where they are on the affined shape. The vertices are usually kept after affine transform, but finding vertexes and polygon approximation are themselves non-trivia issues.

4. Convergence Speed

In the above, different shape signatures have been described in details. The shape signatures are used to derive FDs, which are the normalized Fourier coefficients of Fourier transform on a shape signature (Section 2). In practice, to represent a signal, only finite number of coefficients are used to approximate the signal. To use limited Fourier coefficients to represent a shape, the Fourier series has to be truncated to finite elements. For shape retrieval application, the number of coefficients to represent a shape should not be large, therefore, the convergence speed of the Fourier series derived from the signature function is crucial. The faster the convergence, the less number of Fourier coefficients is needed to represent the shape. In the following we study the convergence speed of the Fourier series derived from each shape signature function.

Ten very complex shapes (Figure 6) are selected to simulate the worst convergence cases in the database. The average convergence speed for each signature function is calculated. For $z(t)$, $r(t)$, $r^*(t)$ and $A(t)$, their spectra are normalized by a_0, or the *DC* component. For $\kappa(t)$, $\psi(t)$, $k(t)$, their spectra are normalized by a_1, or the first Fourier coefficients. Q_k is itself a normalized invariants. The number of normalized spectra greater than 0.01 is given in table 1.

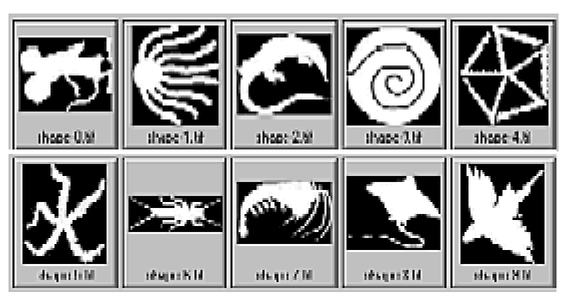


Figure 6. 10 complex shapes used to simulate worst convergence speed of FD.

Table 1. Comparing convergence speed of Fourier series of each shape signature

Signature functions	Number of normalized spectra greater than 0.1	Number of normalized spectra greater than 0.01
$r(t)$	15	120
$r^*(t)$	40	360
$A(t)$	20	210
$z(t)$	10	50
$\psi(t)$	40	280
$\kappa(t)$	∞	∞
$k(t)$	100	600
Q_k	20	100

From the table, it clear that $r(t)$, $z(t)$, and Q_k converge rather fast, $r^*(t)$, $A(t)$, and $\psi(t)$ converge satisfactorily, $k(t)$ converges rather slowly, while $\kappa(t)$ converges so slowly that very large number of coefficients is needed to represent the shape effectively, which is not practical in shape retrieval application. If we set Fourier coefficients with normalized spectra energy greater than 10^{-2} as significant shape features, then 60—600 Fourier coefficients will represent shape well in most cases. If we note that except $z(t)$, all the other signature functions are real functions, therefore, only half of the coefficients are distinct. Then for $r(t)$, $z(t)$, and Q_k, 60 Fourier coefficients; for $r^*(t)$, $A(t)$, and $\psi(t)$, 150 Fourier coefficients and for $k(t)$ 200 Fourier coefficients are reasonable choice of number of coefficients as shape representation, taking into consideration of the tradeoff between accuracy and efficiency.

5. Retrieval Performance

To test the retrieval performance of the FDs derived from the discussed shape signatures and affine invariants, a Java online indexing and retrieval frame work is implemented and a database consisted of 1400 shapes is created from Set B of the MPEG-7 contour shape database. The shapes in Set B are grouped into 70 classes of perceptually similar shapes. Most of the shape images in set B are quite large (400×400 to 800×800), to save computation cost in the indexing, either sampling or scaling can be exploited. In the implementation, all the shape images in the database are scaled to 128×128, the signatures are derived from all the points on the shape boundaries.

All the 1400 shapes in the database are used as queries. The common retrieval performance measure – precision and the recall [14] – are used as the evaluation of the query results. For each query, the precision of the retrieval at each level of the recall is obtained. The result precision of retrieval using a type of FD is the average precision of all the query retrievals using the type of FD. The average retrieval performance of each FD is shown in Figure 7.

It is clear from Figure 7 that the retrieval performance of the 7 methods falls into three groups. The first group which includes centroid distance and the area function has the highest performance. The performances of affine invariants, chord length and the position function are comparable and are significantly lower than that of the first group. The precision of the position function drops so sharply that it can only recall nearly half of the similar shapes. The curvature and the *psi* function gives the lowest performance, this indicates that the first and the second derivatives of the shape boundary are very unreliable. On average, retrieval using FDs derived from the centroid distance has significantly higher performance than that of the other methods. It has a nicely smooth degradation of precision, showing it's robust in terms of precision and recall. Although area function and chord length show some advantages over centroid distance in that the area is invariant under affine transform and the chord length does not need any reference point, the vertices finding problem in the area representation and the symmetry and noise sensitivity problems in the chord length representation reduce the retrieval performance. The affine invariants is designed for polygon shape representation, its retrieval performance for generic shapes is not desirable. The position function and the *psi* function representations have been widely adopted in shape recognition, especially character recognition, however, their use for generic shape retrieval is not desirable compared with the other representations. Although a smoother curvature $k(t)$ is used as shape representation in place of the very versatile curvature $\kappa(t)$, the retrieval performance proves it is a poor shape representation.

In order to test that how the increase and the decrease of number of FDs will actually affect the retrieval performance, retrievals using different number of FDs are tested. In this test, FDs derived from $r(t)$ is used as shape descriptors. From the test, it has been found that the increase of the number of FDs over 60 to describe the shape does not significantly improve the retrieval performance. The actual retrieval performance does not degenerate significantly when the number of FDs is reduced to 10. This means that for efficient indexing and retrieval, 10 FDs are sufficient for shape representation.

6. Conclusions

In this paper we have made a study of Fourier descriptors generated from different shape signatures and using different Fourier invariants. The study focuses on shape retrieval application. It has been found that centroid distance $r(t)$ is the best shape signature among the other shape signatures discussed in this paper in terms of robustness, computation complexity, convergence speed of its Fourier series and retrieval performance of its FDs. Although position function $z(t)$ and cumulative angular function $\psi(t)$ are dominantly used in deriving FDs in the literature, $r(t)$ outperforms them in shape retrieval application. The advantages of area function $A(t)$ and chord length $r^*(t)$ over $r(t)$ are overcome by the non-robustness nature within them, no gain on retrieval performance is obtained by those advantages. The affine Fourier invariants Q_k does not perform well for generic shapes or non-polygon shapes. Curvature is not suitable for deriving FDs due to the very slow convergence nature of its Fourier series. The study also finds that the actual number of FDs to describe a shape is quite small, i.e., 10 FDs are sufficient to describe a shape. This low dimensions of FD feature of a shape makes FD very suitable for shape indexing and retrieval. In the future, the combination of FDs with region based method will be studied, to deal with very complex shapes and to improve its distinguishability of structurally similar but perceptually dissimilar shapes.

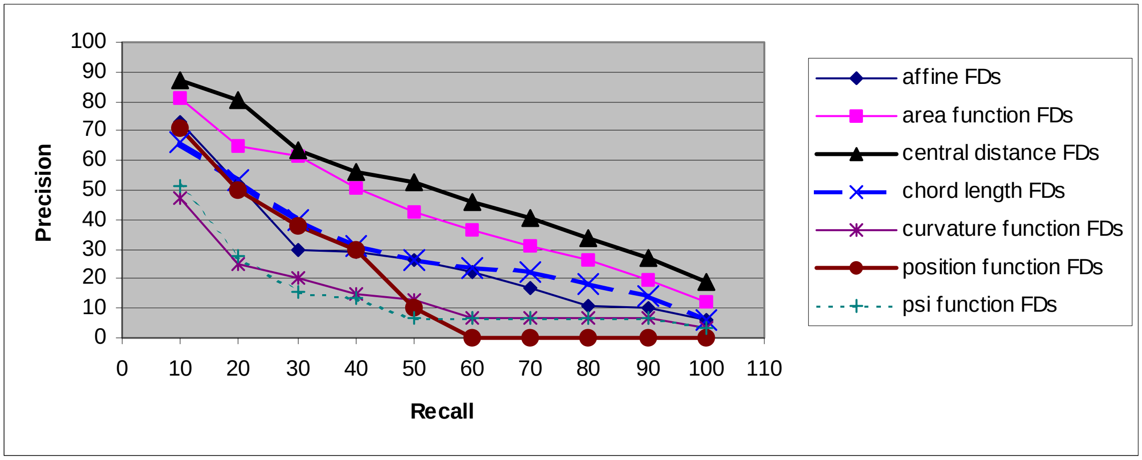


Figure 7. Average retrieval performance of different FD.

Reference:

[1] K. Arbter. Affine-invariant Fourier Descriptors. In J.C. Simon editor: From Pixels to Features, Science Publishers 1989, pp.153-165.

[2] G. Granlund. Fourier Preprocessing for hand print character recognition. IEEE Trans. Computers, Vol. 21, 1972, pp.195-201.

[3] C.-L. Huang and D.-H. Huang. A Content-based image retrieval system. Image and Vision Computing, 16:149-163, 1998.

[4] Ming-Kuei Hu. Visual pattern Recognition by Moment Invariants. IRE Transactions on Information Theory, IT-8:179-187, 1962.

[5] Hannu Kauppinen, Tapio Seppanen and Matti Pietikainen. An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification. IEEE Trans. PAMI-17(2):201-207, 1995

[6] T. Pavlidis. Algorithms for Graphics and Image Processing. Computer Science Press, 1982, pp.143.

[7] E. Persoon and K.-S. Fu. Shape Discrimination Using Fourier Descriptors. IEEE Trans. On Systems, Man and Cybernetics, SMC-7(3):170-179, 1977.

[8] W. Niblack et al. The QBIC Project: Querying Images By Content Using Color, Texture and Shape. SPIE Conf. On Storage and Retrieval for Image and Video Databases, vol 1908, San Jose, CA, pp.173-187, 1993.

[9] Guojun Lu and Atul Sajjanhar. Region-based shape representation and similarity measure suitable for content-base image retrieval. Multimedia Systems, 7:165-174, 1999.

[10] C. T. Zahn and R. Z. Roskies. Fourier Descriptors for Plane closed Curves. IEEE Trans. On Computer, c-21(3):269-281, 1972.

[11] B. M. Mehtre, M. S. Kankanhalli and W. F. Lee. Shape Measures for Content based Image Retrieval: A Comparison. Information Processing & Management, 33(3):319-337, 1997.

[12] P. J. van Otterloo. A contour-Oriented Approach to Shape Analysis. Prentice Hall International (UK) Ltd. C1991, pp.90-108.

[13] D. Zhang and G. Lu. Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study. In Proc. of IEEE International Conference on Multimedia and Expo (ICME'01), August 22-25 Tokyo, Japan, 2001, pp.317-320.

[14] A. Del Bimbo. Visual Information Retrieval. pp. 56-57, Morgan Kaufmann Publishers, Inc. San Francisco, USA, 1999.