ACCV2002: The 5th Asian Conference on Computer Vision, 23—25 January 2002, Melbourne, Australia

An Integrated Approach to Shape Based Image Retrieval

Dengsheng Zhang and Guojun Lu
Gippsland School of Computing and Information Technology
Monash University
Churchill, Victoria 3842
Australia
Tel: 03-99026772
Fax: 03-99026842
dengsheng.zhang, guju.lu@infotech.monash.edu.au

Abstract

Shape is one of the primary low level image features in the newly emerged Content Based Image Retrieval (CBIR). Many shape representations have been proposed, and they are generally classified into contour-based methods and region-based methods. Contour-based methods capture shape boundary features while ignore shape inner content. Region-based methods capture shape inner features while do not emphasize on boundary features. It is known neither methods can produce ideal retrieval results in all situations. The shortcoming of both types of methods can be overcome by combine the strength of two methods. In this paper, a contour shape descriptor and a region shape descriptor are studied. Then a integrated shape descriptor, which combines a contour shape descriptors and the region shape descriptors, is presented. The integrated shape descriptors provide interactivity between users and the retrieval system and can improve retrieval performance. Retrieval results are given to show the comparison and the improvement.
Keywords: *Fourier descriptors, Zernike moments, CBIR, shape*

1. Introduction

Shape is one of the several primary low level image features. Many shape representation methods, or shape descriptors, exist in the literature. These methods can be broadly classified into two categories: region based and contour based.

In region based techniques, all the pixels within a shape region are taken into account to obtain the shape representation. Common region based methods use moment descriptors to describe shape [3][12][11][7][13]. It has been shown in [11] Zernike moments and pseudo Zernike moments outperform the other moments in terms of noise sensitivity, redundancy and reconstruction error.

Contour shape representations only exploit shape boundary information and ignore the interior information. Contour shape representations include Fourier descriptors [16][10][[14], curvature scale space (CSS) [9] and elastic matching [2]. Among these contour techniques, methods based upon Fourier descriptors (FD) prove to be more advantageous than other techniques in terms of computation complexity, robustness, easy normalization and retrieval performance [15].

Region-based methods combine information across an entire object, however, they do not emphasize shape boundary features which are equal important to interior features. Contour-based methods capture shape boundary features while fail to extract shape interior content. It is known neither methods can produce ideal retrieval results in all situations. All these methods assume the type of shapes in the database is known, i.e., either region shape or contour shape. In reality, the type of shapes in the database may be unclear, i.e., some shapes may be best described using region descriptors while other shapes may be best described using contour descriptors. This is also true for the queries presented by the users. Therefore, it is practical for a retrieval system to provide certain flexibility to the users so that the users are able to employ either methods or combine both methods to produce best retrieval result.

In this paper, we first study the Fourier descriptors and Zernike moments descriptors (ZMD), then a integrated shape descriptors which combine the strengths of both types of descriptors is presented. The integrated shape descriptors also provides interactivity between the users and the retrieval system and can improve the retrieval performance. The rest of the paper is organized as following. In Section 2, Fourier descriptors and Zernike moments descriptors are described and discussed. In Section 3, we present the integrated approach and the experiments on retrieval performance of the three shape descriptors on two different shape databases. We conclude the paper in Section 4.

2. Fourier Descriptors and Zernike Moments Descriptors

It has been shown in [15] that FD outperforms CSS descriptors which is adopted by MPEG-7 [4] and ZMD outperforms the grid method [8]. ZMD also outperforms other moments methods [11]. It has been proposed to MPEG-7 as region shape descriptor [4][5]. In this section, FD and ZMD are discussed.

2.1 Fourier Descriptors

In general, FDs are obtained by applying Fourier transform on a shape signature, the Fourier transformed coefficients are called the Fourier descriptors of the shape. The shape signature is a one dimensional function which is derived from shape boundary coordinates. Different shape signatures have been exploited to obtain FDs, complex coordinates, curvature function, cumulative angular function, centroid distance are the commonly used shape signatures. However, FDs derived from different signatures has significant different performance on shape retrieval. As has been shown in [14] that FDs derived from centroid distance outperforms FDs derived from other shape signatures in terms of overall performance.

The first step of computing FD is to obtain the boundary coordinates $(x(t), y(t))$, t = 0, 1, …, N-1, where N is the number of boundary points. In our implementation, the shape boundary points are extracted in a preprocessing stage which consists of binarization, denoising, a *m-connectivity* connection and a *8-connectivity* contour tracing technique [Pavlidis82]. The *centroid distance* function is expressed by the distance of the boundary points from the centroid (x_c, y_c) of the shape

$$r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2},\ t = 0, 1, ..., N-1$$

where

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

The discrete Fourier transform of $r(t)$ is then given by

$$a_n = \frac{1}{N}\sum_{t=0}^{N-1} r(t)\exp(\frac{-j2\pi nt}{N}),\ n = 0, 1, ..., N-1$$

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

The Fourier coefficients acquired are translation invariant due to the translation invariance of centroid distance. To achieve 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., $|a_0|$. Since centroid distance is a real value function, only half of the a_n are needed to index the shapes. Finally, the following feature vector are used as the Fourier descriptors to index the shape

$$\mathbf{f} = [\frac{|a_1|}{|a_0|}, \frac{|a_2|}{|a_0|}, ..., \frac{|a_{N/2}|}{|a_0|}]$$

the similarity measure of the query shape and a target shape in the database is simply the Euclidean distance between the query and the target shape feature vectors. It has been shown in [14] that 60 FDs are sufficient in need to represent a shape.

The nice properties of Fourier descriptors are its robustness, being able to capture perceptual characteristics of the shape boundary and easy to derive. With Fourier descriptors, coarse shape features or global shape features are captured by lower order coefficients and the finer shape features are captured by higher order coefficients. Noise and irregularities are not problems with Fourier descriptors, for noise and irregularities only appear in very high frequencies which are truncated out. Since slight deformations around the shape boundary doesn't cause significant structural change on the shape signature, they do not significantly affect in the final representation. The computation efficiency can be improved by using the fast Fourier transform (FFT). Because usually, only a small number of low order coefficients are sufficient to capture the overall shape features, the representation is also compact. However, because FD only employs shape boundary information, it does not capture shape interior features, besides, it also fails when there are large boundary indentations or protrusions.

2.2 Zernike Moments Descriptors

Teague [13] has proposed the use of orthogonal moments to recover the image from moments based on the theory of orthogonal polynomials, and has introduced Zernike moments, which allow independent moment invariants to be constructed to an arbitrarily high order. The complex Zernike moments are derived from Zernike polynomials:

$$V_{nm}(x, y) = V_{nm}(\rho\cos\theta, \rho\sin\theta) = R_{nm}(\rho)\exp(jm\theta)$$

and

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!} \rho^{n-2s}$$

where n and m are subject to n-$|m|$ = *even*, $|m|\leq n$. Zernike polynomials are a complete set of complex-valued function orthogonal over the unit disk, i.e., $x^2 + y^2 = 1$. Then the complex Zernike moments of order n with repetition m are defined as:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V_{nm}^*(x, y),\ x^2 + y^2 \leq 1$$

The theory of Zernike moments is similar to that of Fourier transform, to expand a signal into series of orthogonal basis. The precision of shape representation depends on the number of moments truncated from the expansion, the first 36 moments up to order 10 are used in our implementation.

Since Zernike basis functions take the unit disk as their domain, this disk must be specified before moments can be calculated. In our implementation, all the shapes are normalized into a unit circle of fixed radius of 64. The unit disk is then centered on the shape centroid. This makes the obtained moments scale and translation invariant. Rotation invariance is achieved by only using magnitudes of the moments. The magnitudes are then normalized into [0, 1] by dividing them by the mass of the shape. Rotation invariance is achieved by only using magnitudes of the moments. The magnitudes are then normalized by dividing them by the mass of the shape. The similarity between two shapes indexed with Zernike moments descriptors is measured by the Euclidean distance between the two Zernike moments vectors.

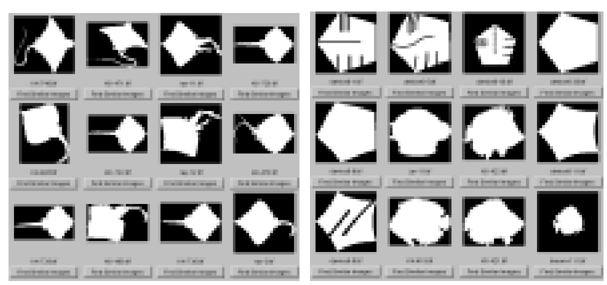
The computation of ZMD does not need to know boundary information, making it suitable for more complex shape representation. Like Fourier descriptors, Zernike moments descriptors can be constructed to arbitrary order, this overcomes the drawback of geometric moments in which higher order moments are difficult to construct. However, Zernike moments descriptors lose the perceptual meanings as those reflected in Fourier descriptors and geometric moments. Besides, ZMD does not emphasize shape boundary features which are important features of a shape.

3. Integrated Approach

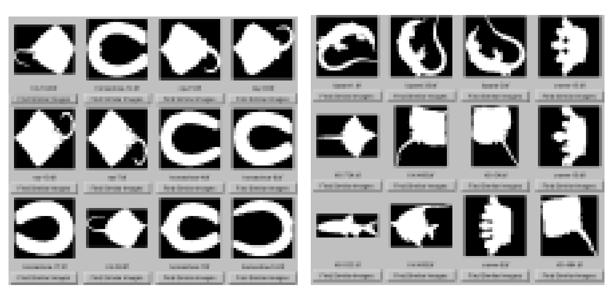
In this section, the proposed integrated shape descriptor and the comparison of its retrieval performance with other shape descriptors are described in details.

3.1 Integrated Shape Descriptors

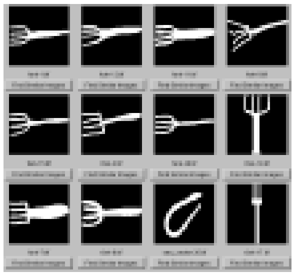
In the above section, Fourier descriptors and Zernike momnets descriptors are studied. It has been known from the above study that Fourier descriptors captures shape boundary features, however, it cannot capture shape interior content. Zernike moments descriptors can be applied to both region shapes and contour shapes, however, it does not emphasize shape boundary features. Furthermore, from our experiments, it has been found that although FD is robust to slight or acceptable boundary variations, it does not perform well in situations where very large indentations or protrusions occur. Examples are the ray fish shape in Figure 1(a) and the device shape in Figure 1(a). The device shape is perceptually a pentagonal shape, however, due to the deep indentations around the boundary, it is difficult to find out similar pentagonal shapes from the database using FD. In this situation, the retrieval performance of ZMD is quite satisfactory as can be seen in Figure 1(b). The ray fish shape has a long thin tail attached to the main body of the shape. The tail makes FD unable to distinguish it from hook-like shapes (Figure 1(c)) or a lizard shape (Figure 1(d)). ZMD performs extremely well in this situation (Figure 1(a)). The above two examples show although the device shape and the ray fish shape are contour shapes, they are better described using region descriptors than using contour descriptors. However, ZMD has poor performance on such shapes as the fork shape and the octopus shape (Figure 1(g)(h)) where protrusions and indentations consists the main body of the shapes. FD has high performs in these situations (Figure 1(e)(f)).

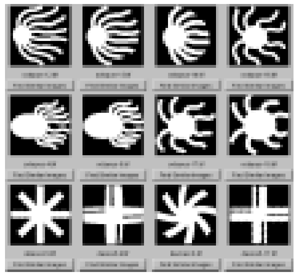


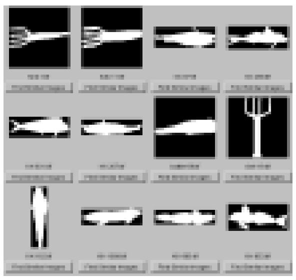
(a) (b)



(c) (d)


(e)


(f)


(g)

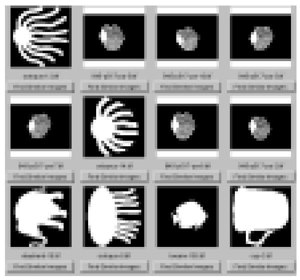

(h)

Figure 1. Screen shots of (a) retrieval of ray fish shape using ZMD; (b) retrieval of device shape using ZMD; (c) retrieval of ray fish shape using FD; (c) retrieval of lizard shape using FD; (e) retrieval of fork shape using FD; (f) retrieval of octopus shape using FD; (g) retrieval of fork shape using ZMD; (h) retrieval of octopus shape using ZMD. In all these screen shots, the top left shape is the query shape.

It is clear from the experiments that both methods can fail in some situations, however, the two methods can complement each other in these situations.

From the above analysis, it is appealing to combine both methods to form a integrated shape descriptors which can capture both shape boundary information and shape inner content. For this reason, FD and Zernike moments are selected to create a integrated shape descriptors. Assume the FD: $\mathbf{f} = \{FD_1, FD_2, \ldots, FD_m\}$ and the Zernike moments: $\mathbf{z} = \{z_1, z_2, \ldots, z_n\}$ are both acquired from a shape using the two methods described in Section 2.1 and Section 2.2, the new shape descriptors is created as:

$$\mathbf{fz} = \{\alpha \cdot FD_1, \alpha \cdot FD_2, \ldots, \alpha \cdot FD_m, \beta \cdot z_1, \beta \cdot z_2, \ldots, \beta \cdot z_n\}$$

where α, β are weight factors. α and β can be adjusted according to the type of query shape presented by the user and the type of shapes presented in the database. When $\beta = 0$, **fz** is simply the FD, and when $\alpha = 0$, **fz** is the Zernike moments descriptors. The new shape descriptors let users interact with the retrieval system according to their preference, so that more accurate retrieval can be carried out. By providing two adjusting parameters, when a retrieval using contour-based method fails, the user is given the choice to do that retrieval using region-based method, vice versa. When both methods fail, the user has the choice to use combined shape descriptors for retrieval. When the database is composed of pure region shapes, α is set to zero so that shapes are described by Zernike moments descriptors. However, when the database is composed of contour shapes or mixed types of shapes, α and β can be adjusted to produce the best result.

3.2 Retrieval Performance of the Three Shape Descriptors

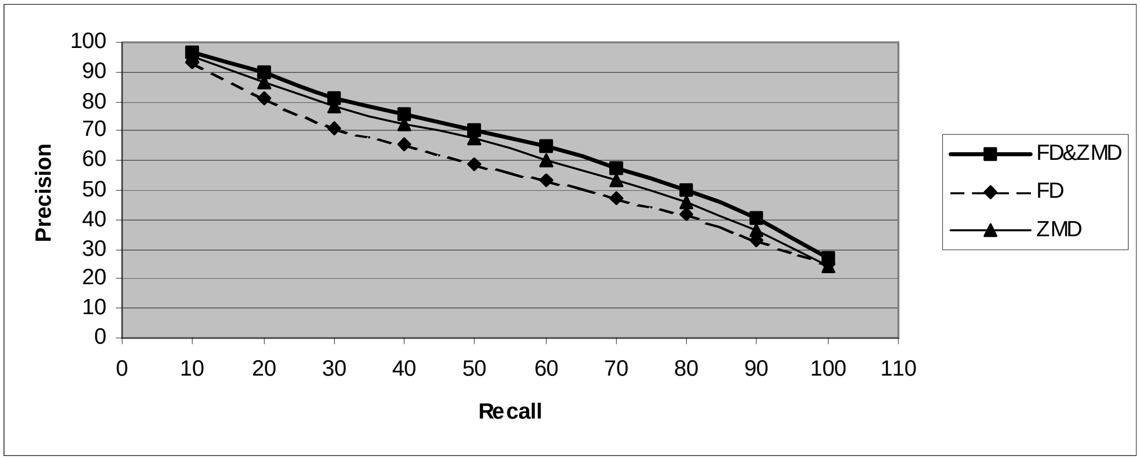
To compare the integrated shape descriptors with FD and Zernike moments descriptors, two databases are created from MPEG-7 contour and region shape databases. One is the contour shape database and the other is the mixed shape database which consists of both contour shapes and region shapes. The contour-based shape database used for test is Set B of MPEG-7 contour shape database. The contour shape database consists of 1400 shapes of 70 classes. The mixed shape database consists of the above contour shape database and 415 boundary connected region shapes from MPEG-7 region shape database. MPEG-7 region shape database consists of 3621 shapes of mainly trademarks.

The common retrieval performance measure – precision and the recall [1] – are used as the evaluation of the query results. Precision P is defined as the ratio of the number of retrieved relevant shapes r to the total number of retrieved shapes n, i.e. $P = r/n$. Precision P indicates the accuracy of the retrieval. Recall R is defined as the ratio of the number of retrieved relevant images r to the total number m of relevant shapes in the whole database, i.e. $R = r/m$. Recall R indicates the robustness of the retrieval performance. 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 shape descriptors is the average precision of all the query retrievals using the type of shape descriptors.

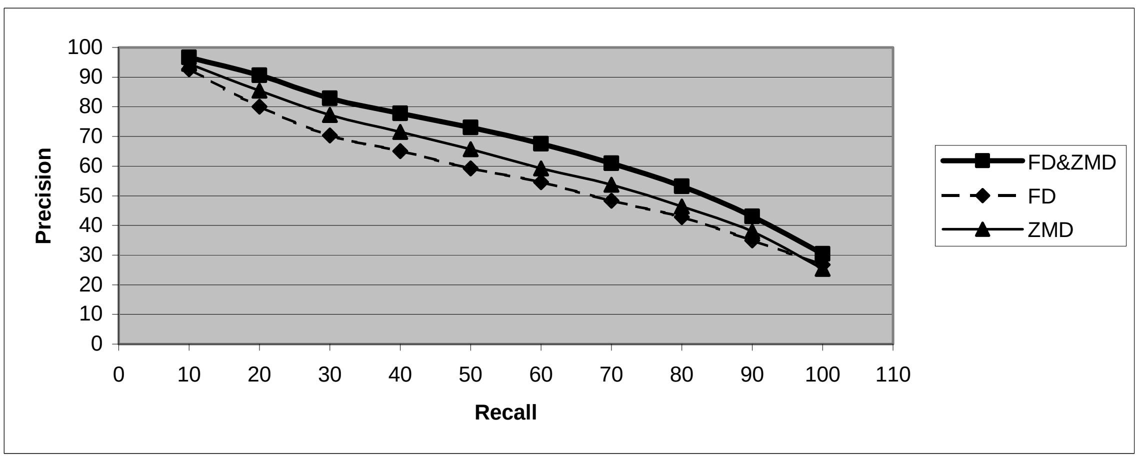
For the test on the contour shape database, All the 1400 shapes in the database are used as queries. The average retrieval performance of each shape descriptor for the 1400 queries is shown in Figure 2(a).

For the mixed shape database, all the 70 classes (1400) of contour shapes plus 15 classes (315 shapes) of region shapes are used as queries. Different α, β values differing from 0.2~0.8 are tested for the integrated shape descriptors, it turns out that $\alpha = 0.7$, $\beta = 0.3$ produce the best retrieval result in the contour shape database and in the mixed database, $\alpha = 0.3$, $\beta = 0.7$ and $\alpha = 0.4$, $\beta = 0.6$ produce the about equal best result. The average retrieval performance of each shape descriptor for the 85 groups (1715) of shapes is shown in Figure 2 (b).

It has been shown in Figure 2(a)(b) that the integrated descriptors outperforms FD and Zernike moments descriptors on both the contour shape database and the mixed shape database. It has been observed from the experiments that the poor performance of FD on shapes with large boundary indentations, protrusions and large boundary distortions (such as the ray fish shapes, the device shapes, the tree shapes) has been significantly improved by the new descriptors. FD's poor distinguishability of shapes of similar contour but different interior content has been overcome by the new descriptors. The new descriptors also overcomes the poor performance of Zernike moments descriptors on such shapes as the fork shapes, the octopus shapes. As the result, the integrated shape descriptor is more suitable for generic shape retrieval.



(a)



(b)

Figure 2. (a) average retrieval performance of the three shape descriptors on contour-based shape database; (b) average retrieval performance of the three shape descriptors on mixed shape database.

4. Conclusions

In this paper we have made a study on Fourier descriptors and Zernike moments descriptors. A integrated shape descriptors is presented. The main advantages of the new shape descriptors is that it provides interactivity between the users and the retrieval system. It combines the strengths of both contour shape descriptors and region shape descriptors. The two weight factors α and β can be adjusted by the users to suit different type of queries and different type of shapes presented in the database. When one type of query fails, the user is provided with a choice to try another type of query; when both types of queries fail, the user is provided with a choice to use combined query. In this sense, the integrated shape descriptors is adaptive to both the query type presented by the user and the type of shapes in the shape database. It overcomes the limitation in common shape retrieval approaches, which assume the shape database is either contour-based or region-based. It overcomes the disadvantages inherited in both the contour shape descriptors and the region shape descriptors. The integrated shape descriptors is tested on both the contour shape database and the mixed shape database composed of contour shapes and region shapes. Results show that the integrated shape descriptors has better retrieval performance than the retrieval performance of both the contour and the region shape descriptors studied in this paper.

References:

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

[2] A. Del Bimbo, P. Pala, Visual Image Retrieval by Elastic Matching of User Sketches, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 2, 1997, pp. 121-132.

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

[4] S. Jeannin Ed. MPEG-7 Visual part of experimentation Model Version 5.0. ISO/IEC JTC1/SC29/WG11/N3321, Nordwijkerhout, March, 2000.

[5] H. Kim and J. Kim. Region-based shape descriptor invariant to rotation, scale and translation. Signal Processing: Image Communication 16:87-93, 2000.

[6] 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

[7] Simon X. Liao and Miroslaw Pawlak. On Image Analysis by Moments. IEEE Trans. On Pattern Analysis and Machine Intelligence. 18(3):254-266, 1996.

[8] G. J. Lu and A. Sajjanhar. Region-based shape representation and similarity measure suitable for content-based image retrieval. Multimedia System 7:165-174, 1999.

[9] F. Mokhtarian, S. Abbasi and J. Kittler. Robust and Efficient Shape Indexing through Curvature Scale Space. Proc. British Machine Vision Conference, pp.53-62, Edinburgh, UK, 1996.

[10] Eric Persoon and King-sun Fu. Shape Discrimination Using Fourier Descriptors. IEEE Trans. On Systems, Man and Cybernetics, Vol.SMC-7(3):170-179, 1977.

[11] C.-H. Teh and R. T. Chin. On image analysis by the methods of moments. IEEE Trans. On Pattern Analysis and Machine Intelligence, 10(4):496-513, 1988

[12] G. Taubin and D. B. Cooper. Recognition and Positioning of Rigid Objects Using Algebraic Moment Invariants. SPIE Conf. On Geometric Methods in Computer Vision, Vol. 1570, pp.175-186, 1991.

[13] Michael Reed Teague. Image Analysis Via the General theory of Moments. Journal of Optical Society of America, 70(8):920-930, 1980.

[14] D. S. Zhang and G. J. Lu. Shape Retrieval Using Fourier Descriptors. In Proc. of Int. Conference on Multimedia and Distance Education. Fargo, ND, USA, June 2001, pp.1-9.

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

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