Novel Fourier Descriptor Based on Complex Coordinates Shape Signature

Emir Sokic, Samim Konjicija
Department of Automatic Control and Electronics
Faculty of Electrical Engineering, University of Sarajevo
Sarajevo, Bosnia and Herzegovina
esokic@etf.unsa.ba

***Abstract*—Shape, color and texture are the most important discriminative elements for content based image retrieval. Fourier descriptors are widely used in shape based image retrieval problems. This paper presents a novel method of extracting Fourier descriptors from the simplest shape signature - complex coordinates. Instead of the commonly used scale normalization with the magnitude of the first harmonic, normalization with the sum of magnitudes of all harmonics is used. This leads to an improved shape scale normalization. All the experimental results indicate that the proposed method outperforms many other state-of-the-art Fourier descriptors based methods, both in terms of retrieval performance and computational time.**

***Index Terms*—Content based image retrieval, Fourier descriptors, shape signature, complex coordinates, scale normalization**

I. Introduction

A rapid increase in volume of multimedia collections is noticeable in the recent years. This requires means of efficient and effective multimedia indexing and retrieval. In the last decades, content-based image retrieval (CBIR) emerged as a promising tool for retrieving images and browsing large images databases, and has been a topic of intensive research.

Although color and texture are the most obvious features of an image or represented object, shape is the most important content for image understanding. Shape description methods may be classified into two groups: contour based and region based [1]. Contour based approaches are generally more compact, faster, and often even perform better than region based methods.

Shape signatures are widely used as contour-based method for shape description [2]. In general, they are very sensitive to noise and distortions, and rarely invariant to rotation, translation and scale (RTS). To overcome these problems, transformations such as Fourier transform [3], Wavelet transform [4], combination of both [2], [5], Radon transform [6] or other, are conducted over shape signatures.

Fourier descriptors (FD) are obtained by applying the discrete Fourier transform (DFT) over a shape signature. After simple normalization, these descriptors are invariant to rotation, translation, scale and change of the starting point of the contour. They also show good retrieval accuracy, compactness, insensitivity to noise, and have a hierarchical representation, which makes them good shape descriptors [7].

Fourier descriptors have been derived from several shape signatures: Complex coordinates, Centroid/Radial distance, Tangent angle [8], Curvature function, Area function, Triangle-area representation [9], Triangular centroid area, Chord length, Polar coordinates, Farthest point distance [10], Perimeter area function [11], Improved arc-height function [12], Rectangle centroid distance [13], and many others.

In this paper, a novel method for extracting Fourier descriptors will be presented. The proposed descriptor is originally based on the complex coordinates signature, which is the simplest possible signature existing in literature. The main contribution of the paper exists in FD scale normalization phase. Instead of normalization only with the magnitude of the first harmonic, sum of magnitudes of all harmonics is used. Although the common FD based on complex coordinates signature achieve modest results in image retrieval problems, it will be shown that our improved Fourier descriptors outperform FD extracted from other shape signatures, both in terms of retrieval performance and computational speed. All descriptors were tested on two different image databases.

The paper is organized as follows. Five shape signatures used for comparison are chosen in Section II. In Section III, our proposed descriptor is introduced and discussed. Experimental setup and results are discussed in Sections IV and V. Conclusion is given at the end of the paper.

II. Shape signatures

The shapes that are analyzed in the paper are outline shapes, which can be described as single plane closed curves. In preprocessing stage, the coordinates of the boundary are extracted from the image. Every contour of the shape is then re-sampled with the same number of points N, using equal arc-length distance between them. For future analysis, it may be assumed that a shape contour is given with N boundary points $P_n = (x_n, y_n)$ where $n = 0, 1, ..., N-1$. Using these boundary points, different shape signatures Z_n can be derived.

Five shape signatures will be used for comparison with our proposed method:

1) Complex coordinates signature (CC) [3],
2) Radial/Centroid distance (RD) [8],
3) Farthest point distance (FPD) [10],
4) Combined perimeter area function (CPAF) [11],

5) Improved arc-height function (IARH) [12].

The first one is chosen since the proposed signature and descriptors are based on CC signature. RD is the most frequently used signature, because it is inherently invariant to rotation. The signatures 3)-5) are chosen because the authors claim that they outperform many other signatures and shape description methods such as: Area function, Curvature signature, Wavelet-Fourier descriptor, Zernike moments, Curvature scale space, Chord length distance, Angular function, Triangular centroid area, Triangular area representation, Polar coordinates etc.

Discrete Fourier transform is applied over all used shape signatures:

$$a_n = \frac{1}{N}\sum_{k=0}^{N-1} Z_k e^{-j2\pi nk/N} \tag{1}$$

where $n = 0, 1, ..., N-1$. Fourier coefficients a_n are used to derive Fourier descriptors $\mathbf{F}$.

In order to judge the similarity between different objects represented by shape descriptors, a distance metric is used. City block distance (also called Manhattan) is used for Complex coordinates and Radial distance (because it gives better performance), while Euclidean distance is used for the other shape signatures.

III. Proposed descriptor

We propose a different signature based on the complex coordinates signature, called Normalized complex coordinates signature (NCC). "Normalized" is used to describe the difference in FD normalization phase. Points of the contour $P_n = (x_n, y_n)$ are written in form of complex numbers:

$$NCC_n = Z_n = x_n + jy_n \tag{2}$$

Unlike in CC signature which is given with $CC_n = (x_n - x_c) + j(y_n - y_c)$, the centroid point $P_C = (x_c, y_c) = \left(\frac{1}{N}\sum_{n=0}^{N-1} x_n, \frac{1}{N}\sum_{n=0}^{N-1} y_n\right)$ does not have to be calculated in advance. In order to achieve the invariance on translation, rotation, starting point and scale, discrete Fourier transform is applied as in (1).

a) Invariance under translation : Suppose that whole contour is translated for $T = x_T + jy_T$. Then the resulting Fourier coefficients are:

$$\begin{aligned} a_n^{(T)} &= \frac{1}{N}\sum_{k=0}^{N-1}(NCC_k + T)e^{-j2\pi nk/N} && (3)\\ &= \frac{1}{N}\sum_{k=0}^{N-1} NCC_k e^{-\frac{j2\pi nk}{N}} + \frac{T}{N}\sum_{k=0}^{N-1} e^{-\frac{j2\pi nk}{N}} && (4)\\ &= \begin{cases} a_n + \frac{T}{N}\frac{1-(e^{-j2\pi n/N})^N}{1-e^{-j2\pi n/N}} = a_n, & n \neq 0 \\ a_0 + \frac{T}{N}N = a_0 + T, & n = 0 \end{cases} && (5) \end{aligned}$$

This means that if $P_T = 0$, the resulting DC component (a_0) is actually the centroid of the shape. Hence, by eliminating the DC component a_0, invariance under translation may be achieved.

b) Invariance under rotation : Rotation of the contour will affect only the phase of the Fourier coefficients a_n. Suppose that the initial contour is rotated for angle ϕ around the origin (if not rotated around the origin, the transformation may be represented as translation+rotation, and translation invariance is already explained). This means that the new signature is $NCC_n e^{j\phi}$. The resulting Fourier coefficients are:

$$\begin{aligned} a_n^{(R)} &= \frac{1}{N}\sum_{k=0}^{N-1} e^{j\phi} NCC_k e^{-j2\pi nk/N} && (6)\\ &= e^{j\phi} a_n && (7) \end{aligned}$$

It is easy to see that by using only the modulus of the Fourier coefficients, invariance under rotation is achieved.

c) Invariance to starting point change : If a different point P_l is used instead of the initial point P_0, then the Fourier coefficients become:

$$\begin{aligned} a_n^{(SP)} &= \frac{1}{N}\sum_{k=0}^{N-1} NCC_{k+l} e^{-j2\pi nk/N} && (8)\\ &= \frac{1}{N}\sum_{k=l}^{N-1+l} NCC_k e^{-j2\pi n(k-l)/N} && (9)\\ &= e^{j2\pi nl/N} a_n && (10) \end{aligned}$$

The previous transformations were possible since NCC_n is periodic with period N. It is clear that using only the modulus of a_n, invariance to starting point change is obtained.

d) Scale invariance: The critical part is scale normalization. All earlier implementations of CC ([8], [10] and many others), used $|a_1|$ as the normalization coefficient. This is numerically justified because if the object is scaled with α then its CC (or NCC) signature becomes:

$$a_n^{(SC)} = \frac{1}{N}\sum_{k=0}^{N-1} \alpha NCC_k e^{-j2\pi nk/N} = \alpha a_n \tag{11}$$

since scaling the object with coefficient α will scale all coefficients a_i ($i = 2, ..., N-1$) linearly to αa_i. This means that $a_i^{(SC)}/a_1^{(SC)} = \alpha a_i/\alpha a_1 = a_i/a_1$ so the descriptors will be invariant to scale. Actually, this is true for objects from similar shape classes, which have similar "contribution" of the first harmonic a_1 in total representation. However, when comparing different objects using similarity measures, if one object has a larger a_1 and other smaller one, it may lead to inadequate comparison on other (higher) frequencies. This is illustrated in Figure 1 a) and b). It is clear that when the apple shape and the classic car shape are normalized to different "unit" sizes using only the first harmonics, it may lead to inadequate comparison in frequency domain.

In order to achieve scale normalization properly, and to avoid these problems, we introduce a different scaling coefficient:

$$Sc = \sum_{i=1}^{N-1} |a_i| \tag{12}$$

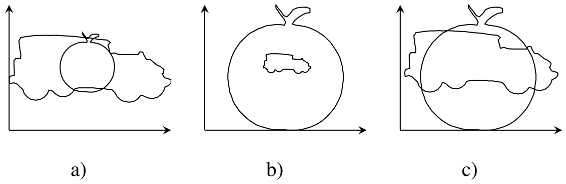

Figure 1. a) Original shape sizes, b) Normalized shape sizes after scaling with the magnitude of the first harmonic, c) Normalized shape sizes after scaling with the sum of magnitudes of all harmonics

It is easy to show that $|a_i^{(SC)}|/Sc = |a_i^{(SC)}|/\sum_{i=1}^{N-1}|a_i^{(SC)}| = \alpha|a_i|/\sum_{i=1}^{N-1}\alpha|a_i| = \alpha|a_i|/\alpha Sc = |a_i|/Sc$, hence the descriptor is scale invariant. Now, scale is not dependent only on first harmonic, but on all other harmonics. Also, the first harmonic in the representation is kept, which is good since it contains a lot of information of the shape. Also, it is assured that $\sum_{i=1}^{N-1} f_i = 1$. This way, representations of different shapes have equal and comparable (unity) size, or “energy”. Only the M lower frequency descriptors are used for representation $(M \leq N-1)$:

$$\mathbf{F} = \left\{ \frac{|a_1|}{Sc}, \frac{|a_2|}{Sc}, ..., \frac{|a_{M/2}|}{Sc}, \frac{|a_{N-M/2}|}{Sc}, ..., \frac{|a_{N-1}|}{Sc} \right\} \quad (13)$$

City block distance is used as similarity measure. It will be shown that this extremely simple technique outperforms all existing Fourier descriptors techniques in terms of retrieval performance and speed.

IV. Methodology

In order to investigate the performance of the proposed algorithm, two parameters were analyzed: retrieval performance (effectiveness), and computational time (efficiency).

The retrieval performance is analyzed using precision and recall (PR) diagrams, a commonly adopted method in CBIR [8], [10], [12], [11]. Precision is defined as the ratio of the number of the relevant shapes to the total number of retrieved shapes, while recall is defined as the ratio of the number of retrieved relevant shapes to the total number of relevant shapes in the entire database.

First, each shape in the database is used as a query shape. The shape descriptor of the query shape is compared to the descriptors of other shapes, and the results are ranked according to the metrics described in Section II. For each query, the precision of the retrieval at each level of the recall is obtained. Then the average precision for all recall values for all query shapes in the database is calculated, and presented by PR diagram. If a retrieval algorithm has better precision for the same recall values, it is considered better.

The retrieval performance is tested on two shape databases, commonly used MPEG7 CE-1 Set B database [14], and Swedish leaf database used by Xu et al. [12], showed in Figure 2 a) and b) respectively.

MPEG7 CE-1 Set B consists of 1400 shapes representing real life objects, classified into 70 classes with 20 similar shapes for each class. This database is widely accepted for shape retrieval testing purposes, and it is convenient because it includes rotation, scaling, skew, stretching, defection and indentation of shapes which may aid to test for robustness of the shape descriptors.

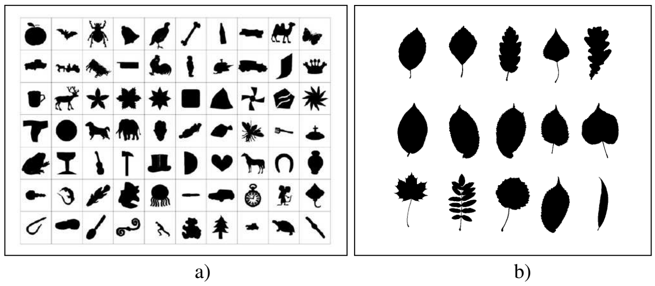

Figure 2. a) MPEG7 CE-1 Set B dataset representative shapes (70 classes with 20 variations per class), b) Leaf dataset representative shapes (15 classes with 75 variations per class)

The Swedish leaf database is freely available database generated by Linkoping University (http://www.isy.liu.se/cvl/ImageDB/public/blad/), but preprocessed and binarized by Xu et al. It has 15 species of leafs, with 75 leaves per species.

The average time needed to calculate the descriptor from the same number N of previously extracted contour points was used as the measure of computational speed. Time was calculated as average of 100.000 iterations. All experiments were conducted using the same computer, running MATLAB 2009 on Ubuntu Linux 9.10, with Intel Core2duo 2.0GHz processor and 2GB RAM.

V. Experimental results and discussion

All shape contours have been sampled with the same number of points (N=256). Then, for fair comparison, the same number (M=32) of descriptors were extracted for all signatures.

First, the retrieval performance was analyzed. For each signature, a precision and recall diagram is calculated on both test sets. The results for MPEG7 dataset are given in Figure 3, and for Leaf dataset in Figure 4. As it may be seen in Figure 3, three different “groups” of similar retrieval performance are present. The CC signature has lower performance, FPD and RD have comparable performance, while IARH,CPAF and NCC show the best results. Clearly, NCC has the best retrieval performance for MPEG7 dataset, in comparison with all other presented signatures.

Diagrams are more distinct in Figure 4, using the Leaf dataset. This is because IARH slightly underperformed, and CC and NCC over-performed on the dataset. This dataset has less difference between classes than the MPEG set, and transformations of the contour (IARH, CPAF, RD and FPD) fail to capture fine details important for discrimination. NCC does not distort the shape information of the contour as other signatures do, hence it preserves most of the perceptual features of the shape. As visible from PR diagrams, our method generates the best precision for all recall rates, on both test sets.

Next, the computational time for extraction of the Fourier descriptors using different number of contour points N was

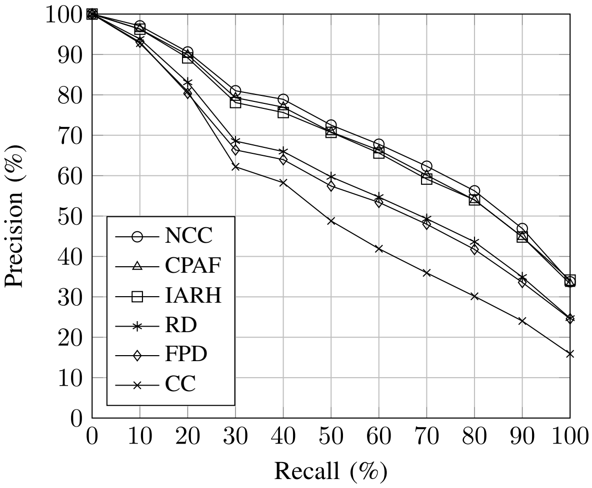

Figure 3. Precision and recall diagram for MPEG7-CE1 B dataset

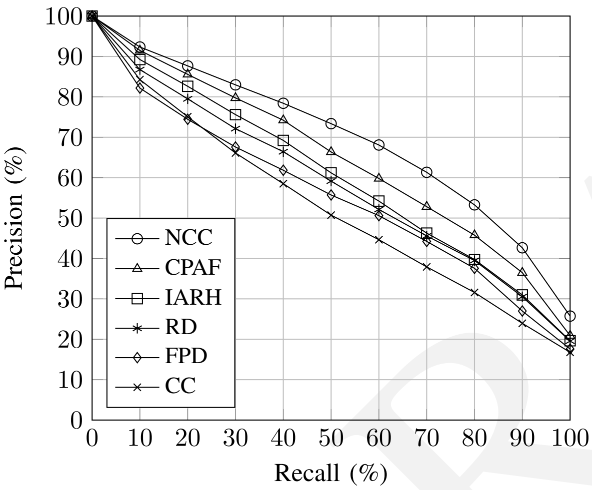

Figure 4. Precision and recall diagram for Leaf dataset

analyzed. The results are presented in Table I. Calculation time of similarity measures were not included in total time, since they are relatively independent of the shape descriptors. The computational complexity of FFT ($O(N\log(N))$) is affecting each algorithm equally. All presented signatures have computational complexity $O(N)$, except the slowest FPD which has $O(N^2)$. IARH and CPAF have similar computational time, while CPAF being slightly longer because its arc-length is changing from point to point. IARH uses constant arc-length, so points of the triangle are calculated in a simpler way if the contour points are distributed by equal arc-lengths.

Nevertheless, our method is computationally faster than the other methods for the same number of points. The scaling process is more complicated for NCC than the fast CC, but the centroid of the shape does not need to be calculated, which makes NCC slightly faster than CC.

NCC, CPAF and IARH have shown similar retrieval performance, but the execution time for NCC is almost 50 times shorter than for CPAF and IARH, which makes NCC the best choice among the presented methods.

Table I
AVERAGE SHAPE DESCRIPTOR EXTRACTION TIME IN MICROSECONDS

Nr.points (N)	CC	RD	FPD	IARH	CPAF	NCC
64	21.3	24.6	482.4	569.6	786.5	19.1
128	29.5	38.8	1184.9	1099.2	1430.0	26.7
256	49.3	60.8	3206.7	2127.7	2790.9	42.3
512	78.5	107.1	10003.7	4370.4	5703.8	71.9

VI. CONCLUSION

The novel Fourier descriptor based on NCC signature outperforms other FD based image retrieval approaches, both in terms of effectiveness and efficiency. This makes it a promising shape descriptor for image retrieval problems. The inovative FD scale normalization should open new insights in scale normalization for other shape description methods. The main drawback of the proposed descriptor is that it is essentially contour based, and may fail in tasks that require region-based approaches.

REFERENCES

[1] T. Sikora, “The MPEG-7 Visual standard for content description-an overview,” *Circuits and Systems for Video Technology, IEEE*, vol. 11, no. 6, pp. 696–702, 2001.

[2] R. Yadav, N. Nishchal, a. Gupta, and V. Rastogi, “Retrieval and classification of shape-based objects using Fourier, generic Fourier, and wavelet-Fourier descriptors technique: A comparative study,” *Optics and Lasers in Engineering*, vol. 45, no. 6, pp. 695–708, Jun. 2007.

[3] D. Zhang and G. Lu, “Study and evaluation of different Fourier methods for image retrieval,” *Image and Vision Computing*, vol. 23, no. 1, pp. 33–49, Jan. 2005.

[4] M. Khalil, “Affine invariants for object recognition using the wavelet transform,” *Pattern recognition letters*, vol. 23, pp. 57–72, 2002.

[5] I. Kunttu, L. Lepisto, J. Rauhamaa, and a. Visa, “Multiscale Fourier descriptors for defect image retrieval,” *Pattern Recognition Letters*, vol. 27, no. 2, pp. 123–132, Jan. 2006.

[6] S. Tabbone, L. Wendling, and J. Salmon, “A new shape descriptor defined on the Radon transform,” *Computer Vision and Image Understanding*, vol. 102, no. 1, pp. 42–51, Apr. 2006.

[7] M. Yang, K. Kpalma, J. Ronsin, and Others, “A survey of shape feature extraction techniques,” *Methods*, vol. 2008, 2008.

[8] D. Zhang and G. Lu, “A comparative study on shape retrieval using Fourier descriptors with different shape signatures,” in *Proc. of international conference on intelligent multimedia and distance education (ICIMADE01)*, 2001, pp. 1–9.

[9] N. Alajlan, I. El Rube, M. Kamel, and G. Freeman, “Shape retrieval using triangle-area representation and dynamic space warping,” *Pattern Recognition*, vol. 40, no. 7, pp. 1911–1920, 2007.

[10] A. El-ghazal, O. Basir, and S. Belkasim, “Farthest point distance: A new shape signature for Fourier descriptors,” *Signal Processing: Image Communication*, vol. 24, no. 7, pp. 572–586, Aug. 2009.

[11] B. Wang, “Shape retrieval using combined Fourier features,” *Optics Communications*, vol. 284, no. 14, pp. 3504–3508, Jul. 2011.

[12] G.-Q. Xu, Z.-C. Mu, and B.-F. Nan, “Shape retrieval using improved arc-height function,” in *Wavelet Analysis and Pattern Recognition (ICWAPR), 2012 International Conference on*. IEEE, 2012, pp. 73–77.

[13] S. Eini and A. Chalechale, “Shape retrieval using smallest rectangle centroid distance.” *International Journal of Signal Processing, Image Processing & Pattern Recognition*, vol. 6, no. 4, 2013.

[14] L. Latecki, R. Lakamper, and T. Eckhardt, “Shape descriptors for non-rigid shapes with a single closed contour,” in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1. IEEE, 2000, pp. 424–429.