ENHANCED GENERIC FOURIER DESCRIPTORS FOR OBJECT-BASED IMAGE RETRIEVAL

Dengsheng Zhang and Guojun Lu
Gippsland School of Computing and Info Tech
Monash University
Churchill, Victoria 3842
dengsheng.zhang, guojun.lu@infotech.monash.edu.au

Copyright 2002 IEEE. Published in the 2002 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002), scheduled for May 13-17, 2002 in Orlando, Florida, USA. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works, must be obtained from the IEEE. Contact: Manager, Copyrights and Permissions / IEEE Service Center / 445 Hoes Lane / P.O. Box 1331 / Piscataway, NJ 08855-1331, USA. Telephone: + Intl. 908-562-3966.

ENHANCED GENERIC FOURIER DESCRIPTORS FOR OBJECT-BASED IMAGE RETRIEVAL

Dengsheng Zhang and Guojun Lu
Gippsland School of Computing and Info Tech
Monash University
Churchill, Victoria 3842
dengsheng.zhang, guojun.lu@infotech.monash.edu.au

ABSTRACT

Shape (object) description consists a key part of image content description in MPEG-7. Most of existing shape descriptors are usually either application dependent or non-robust, making them undesirable for generic shape description. In this paper, an Enhanced Generic Fourier Descriptor (EGFD) is presented to overcome the drawbacks of existing shape representation techniques. The EGFD is obtained based on our previously proposed Generic Fourier Descriptor (GFD). It is acquired by deriving GFD from the rotation and scale normalized shape. Experimental results show that the proposed EGFD outperforms GFD and Zernike moments descriptors (ZMD) significantly.
Keywords: Generic Fourier descriptors, shape, CBIR, retrieval.

1. INTRODUCTION

Due to the tremendous increase of multimedia information, there is an urgent need of multimedia content description so that automatic searching is possible. Two multimedia applications: CBIR (Content-Based Image Retrieval) and MPEG-7 have emerged to address this urgent issue. Since shape is a fundamental property of an object, shape description consists a key part of image description in CBIR and MPEG-7.

Many shape descriptors exist in the literature. Generally there are two types of shape descriptors: contour-based shape descriptors and region-based shape descriptors.

Contour-based shape descriptors include Fourier descriptor (FD) [8][13], wavelet descriptors [10], curvature scale space descriptors [7] and shape signatures [2]. Since contour-based shape descriptors exploit only boundary information, they cannot capture shape interior content. Besides, these methods cannot deal with disjoint shapes where contour information is not available. As a result, contour-based methods have limited applications

In region based techniques, shape descriptors are derived using all the pixel information within a shape region. Region-based shape descriptors can be applied to general applications. Region-based methods include moment descriptors [3][5][9], grid descriptors [6]. It has been shown in [9][5][11] that Zernike moments descriptors (ZMD) outperform many other shape descriptors in terms of representation effectiveness. Basically, ZMD is acquired by applying Zernike moment transform on a polar image. The normalized coefficients of the transform are used as ZMD to describe shape. Although ZMD can capture shape interior features by examining shape in circular directions, it does not effectively capture shape features in radial directions. As a result, a generic Fourier descriptor has been proposed to overcome this problem [12]. GFD is acquired by applying a modified polar Fourier transform on a rectangularized polar shape image. The acquired GFD captures shape features well in both radial and circular directions, it is more robust than ZMD. Since GFD is acquired from polar image, the concentric circular scanning in the feature extraction process causes inaccurate description for severely skewed or stretched shapes (which are common in natural shapes). To solve this problem, we propose a enhanced GFD which is acquired from a rotation and scale normalized shape. The enhancement improves GFD's retrieval performance significantly.

The rest of the paper is organized as following. In section 2, we briefly describe GFD. In Section 3, the proposed EGFD is described in details. Section 4 show the experimental results. Section 5 concludes the paper.

2. GENERIC FOURIER DESCRIPTOR

Basically, GFD is derived by applying a *modified polar Fourier transform* (MPFT) on shape image. In order to apply MPFT, the polar shape image is treated as a normal rectangular image. Figure 1 demonstrates the rectangular polar image. Figure 1(a) is the original shape image in polar space, Figure 1(b) is the rectangular polar image plotted into Cartesian space.

The polar image of Figure 1(b) is the normal rectangular image. Therefore, if we apply 2-D Fourier transform (FT) on this rectangular image, the polar FT has the similar form to the normal 2-D discrete FT in Cartesian space. Consequently, for a given shape image $f(x, y)$, the modified polar FT is obtained as

$$pf(\rho,\phi)=\sum_{r}\sum_{i} f(r,\theta_i)\exp[j2\pi(\frac{r}{R}\rho+\frac{2\pi i}{T}\phi)]$$

where $0\le r=[(x-x_c)^2+(y-y_c)^2]^{1/2}<R$ and $\theta_i = i(2\pi/T)$ $(0\le i<T)$; $0\le\rho<R$, $0\le\phi<T$. (x_c, y_c) is the center of mass of the shape; R and T are the radial and angular resolutions. The physical meaning of ρ and ϕ is clear. The ρ and ϕ are respectively the ρth radial frequency and the ϕth angular frequency selected to describe shape. The determination of the number of ρ and ϕ for shape description is physically achievable, because shape features are normally captured by the few lower frequencies.

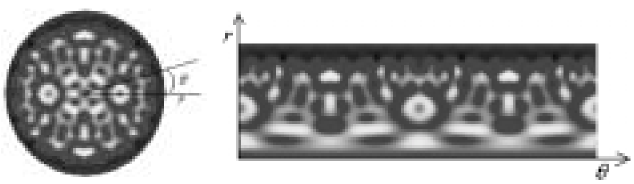
The acquired Fourier coefficients are translation invariant. Rotation and scaling invariance are achieved by the following normalization:

$$\mathbf{GFD} = \{\frac{|pf(0,0)|}{area}, \frac{|pf(0,1)|}{|pf(0,0)|}, \ldots, \frac{|pf(0,n)|}{|pf(0,0)|}, \ldots, \frac{|pf(m,0)|}{|pf(0,0)|}, \ldots, \frac{|pf(m,n)|}{|pf(0,0)|}\}$$

where *area* is the area of the bounding circle in which the polar image resides. m is the maximum number of the radial frequencies selected and n is the maximum number of angular frequencies selected. m and n can be adjusted to achieve hierarchical coarse to fine representation requirement.

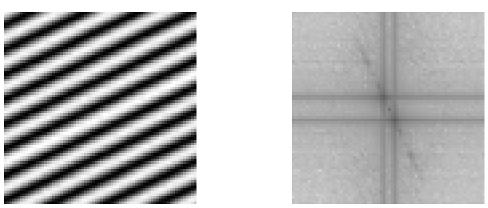
For efficient shape description, only a small number of the acquired GFD features are selected for shape representation. In our implementation, 36 GFD features reflecting 4 radial frequencies and 9 angular frequencies are selected to index the shape. For two shapes represented by their GFDs, the similarity between the two shapes is measured by the Euclidean distance between the two feature vectors of the shapes.

The advantage of MPFT over FT is that the acquired spectra is rotation invariant and more concentrated to the origin (Figure 2). MPFT is also more advantageous than the conventional polar FT, because both the radial features and the circular features captured by the coefficients are physically meaningful [12].

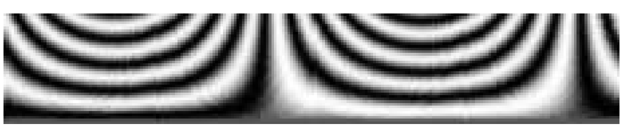


(a) (b)

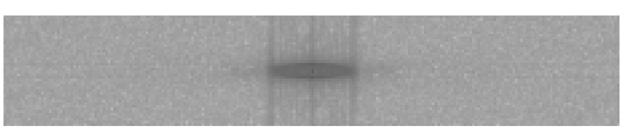
Figure 1. (a) original shape image in polar space; (b) polar image of (a) plotted into Cartesian space.



(a) (b)



(c)



(d)

Figure 2. (a) a shape image; (b) FT spectra of shape (a); (c) polar image of (a) plotted into Cartesian space; (d) MPFT spectra of shape (a).

3. ENHANCED GENERIC FOURIER DESCRIPTOR

The above derived GFD generally has a very good retrieval performance. It has been tested on MPEG-7 region shape database and compared with ZMD which has been adopted by MPEG-7 as region shape descriptor [4]. GFD has better retrieval performance than ZMD. It has very high performance on rotation invariance test (overall precision is 98%), scale invariance test (overall precision is 89.6%) and rotation-scale invariance test (overall precision is 98.8%). However, comparing with rotation and scale invariance test, the retrieval performance on perspective transform test (overall precision 72.5%) and general distortions test (overall precision 80.2%) are significantly lower.

The reason for causing this low performance is because when severe skew or stretching occur, the shape region distribution within the circle the shape resides is changed largely. For example, in Figure 3, perceptually, shape (a) and shape (d) are homogeneities. However, due to severe skew, shape (a) only occupies half of the circle it resides while shape (b) occupies fully the circle it resides. When extracting GFD, the shape is scanned circularly on concentric circles. The scan expands from the center towards periphery. When the scan reaches the periphery of shape (a), it meets more and more positions without shape information. Therefore, the GFD extracted from shape (a) will be much different from the GFD extracted from shape (b) where all scanned positions contain shape information. To solve this problem, a shape normalization process is applied before the feature extraction. The normalization involves two steps: (i) rotation normalization; (ii) scale normalization.

First, in order to do the rotation normalization, the *major axis* (MA) of the shape is found. The MA is the line segment connects the two shape points of the furthest apart. Since the shape under consideration is region shape or generic shape, it is impractical to find MA by computing distance between all shape points, the computation would be prohibitive. Therefore, an optimized *major axis algorithm* (MAA) is proposed. The MAA involves two steps: (i) finding the pair of boundary points in a number of directions (360 in our case); (ii) finding the pair of points of the furthest apart in the found boundary points. The pair of boundary points is found by traversing the line (through the shape centroid) from shape outer boundary (defined by the maximum radius) towards the centroid at both ends. Figure 3(c) illustrates the finding of the pair of boundary points in a particular direction θ. The two blobs on the line of angle θ are the pair of boundary points found. After all the boundary points (consists of all the pairs of boundary points found above) have been found, the next step is to find the two points of the furthest distance among the boundary points. This two points of the furthest distance define the MA.

Once the MA is found, the shape is rotated so that the MA is horizontal. The width (w) and the height (h) of the shape are then found. The shape is then scaled with horizontal scale ratio of $128/w$ and vertical scale ratio of $128/h$ to fit into a square of size of 128×128. Figure 3(e) and (f) are the normalized shapes of (d) and (a) respectively. They are much more similar. The noise and irregularities resulted from the normalization poise no problems for the representation because the extracted spectral features are extremely robust to noise and irregularities.

The MPFT described in Section 2 is then applied on the normalized shape to obtain the GFD of the shape. The acquired GFD is the enhanced GFD. The similarity between two shapes described by their EGFDs is also measured by the Euclidean distance between the two EGFD feature vectors.

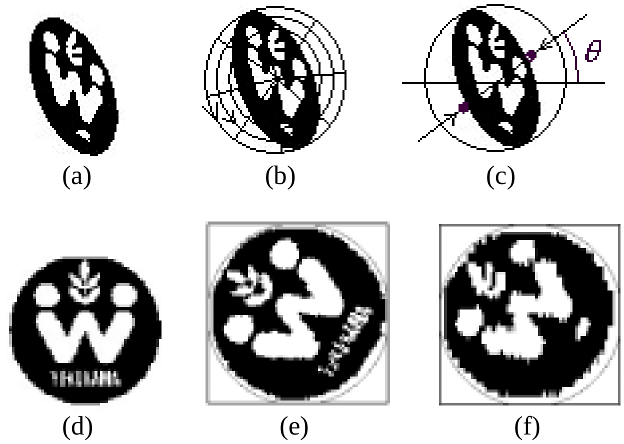


Fig. 3. (a) A skewed shape; (b) polar scanning of image (a); (c) searching boundary points of shape (a) in one particular direction; (d) a homogeneity of shape (a); (e) normalized shape of (d); (f) normalized shape of (a).

4. RETRIEVAL EFFECTIVENESS

In order to test the retrieval effectiveness of the EGFD and compare with GFD and ZMD, 2 sets of experiments are conducted. The first test is conducted on Set A4 of MPEG-7 region shape database, Set A4 is for test of shape descriptor's invariance to perspective transform. The second test is conducted on MPEG-7 region shape database CE-2, CE-2 is for test of shape descriptor's robustness to general distortions including perspective transform, scaling and rotation distortions. The use of the two databases are described in the following.

- Set A4 consists of 3101 from the whole database, it is for test of robustness to perspective transform. 330 shapes in Set A4 are organized into 30 groups (11 similar shapes in each group) which are designated as queries for test of retrieval. In our experiment, all the 330 shapes from the 30 groups are used as queries to test the retrieval.
- The whole database (CE-2) consists of 3621 shapes, 651 shapes of the 3621 shapes are organized into 31 groups (21 similar shapes in each groups). The 31 groups of shapes reflect general shape distortions, and they test the overall robustness of a shape descriptor. In our experiment, all the 651 shapes from the 31 groups are used as queries to test the retrieval.

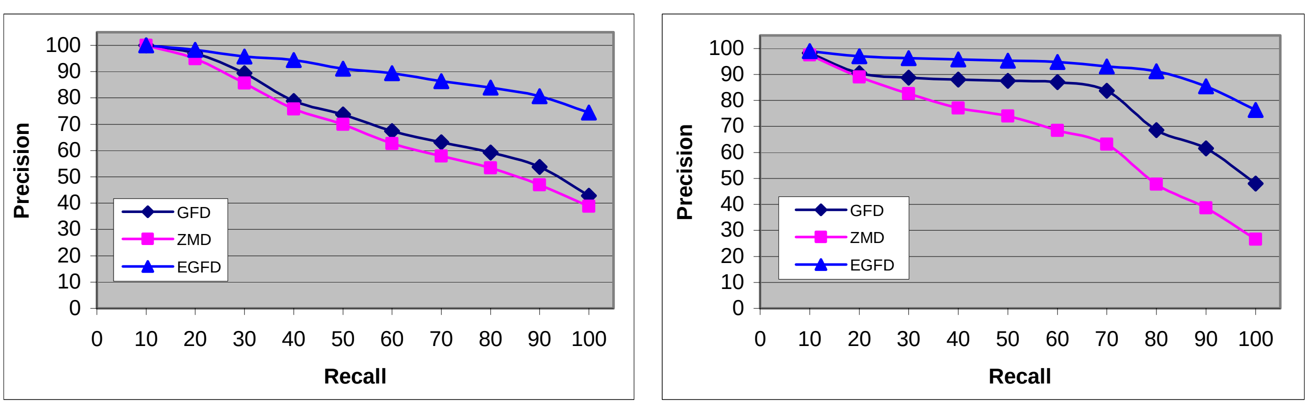
Since the member IDs in each query group are known, the retrieval is conducted automatically. The retrieval has also been put online through a Java-based client-server retrieval framework. It can be visited by using Java appletviewer at http://gofaster.gscit.monash.edu.au/~dengs/Regionn/egfd_src/query.html.

Common evaluation method, i.e., precision-recall [1], is used for the evaluation of retrieval effectiveness. 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 measures 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 measures the robustness of the retrieval. For each query, the precision of the retrieval at each level of the recall is obtained. The result precision of retrieval is the average precision of all the query retrievals. The average precision and recall of the retrieval using the three type of shape descriptors on Set A4 and CE-2 are shown in Figure 4(a)(b).

It is observed from Figure 4 that on both databases, EGFD outperforms GFD and ZMD significantly. Comparing with GFD, the improvement on Set A4 is 17%, the overall precision is increased from 72.5% to 89.5%. The improvement on CE-2 is 12.2%, the overall precision is increased from 80.2% to 92.4%.

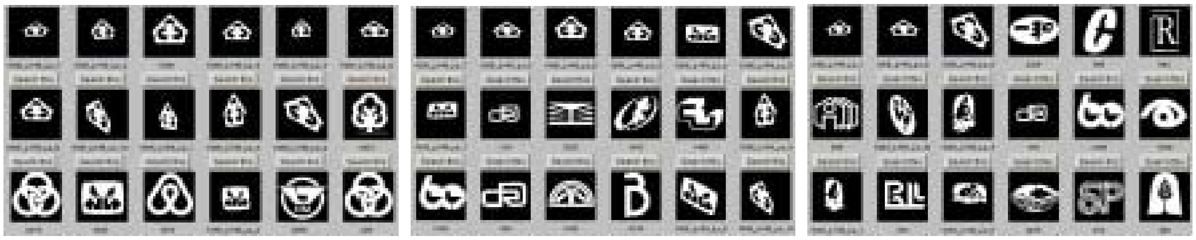
Figure 5 shows two cases of retrievals on the two databases using the three shape descriptors discussed in this paper. In both cases, EGFD retrieves all the shapes in the query group in the first screen. While both GFD and ZMD either miss shapes with large distortions or retrieve them but with much lower ranking.

Although the incorporation of the shape normalization process increases the computation time for the extraction of EGFD, due to the use of optimized MAA, the increase is not dramatic. The average extraction time taken for each shape is 1490ms on Windows platform of a PC-III of 866MHZ, while this is 1245ms for GFD and 1193 for ZMD. The online retrieval time for all the three shape descriptors is the same, since we use the same number (36) of descriptors for shape description.



(a) (b)

Figure 4. (a) Average precision-recall of 330 queries from Set A4 of MPEG-7 region shape database; (b) Average precision-recall of 651 queries from CE-2 of MPEG-7 region shape database.



(a) Retrieval of query 1006_p144_pa_7 using EGFD (left), GFD (middle) and ZMD (right)

(b) Retrieval of query 1004_p144_pa_3 using EGFD (left), GFD (middle) and ZMD (right)

Figure 5. Screen shots of example retrievals on (a) Set A4; (b) CE-2. In all the screen shots, the top left shape is the query shape, and all the other retrieved shapes are ranked in descending order of similarity to the query shape.

5. CONCLUSIONS

In this paper, we have presented a enhanced generic Fourier descriptor (EGFD) for object-based image retrieval. The EGFD outperforms GFD and ZMD significantly on both perspective transform test and general distortion test. The main contributions of the paper are in the following three aspects.

- The proposed EGFD improves GFD significantly. It solves GFD's low retrieval performance on severely skewed and stretched shapes. It also improves GFD's robustness to general shape distortions.
- A shape normalization method is presented. The shape normalization method can be exploited for general shape representation purposes.
- An optimized major axis algorithm (MAA) is proposed. MA is a common normalization mechanism in shape modeling and representation. Common MAA is only for finding MA of contour shape. The proposed optimized MAA can be used for finding MA of generic shapes.

REFERENCES:

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

[2] E. R. Davies. Machine Vision: Theory, Algorithms, Practicalities. Academic Press, 1997.

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

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

[8] 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.

[9] 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

[10] H. S. Yang, S. U. Lee, K. M. Lee. Recognition of 2D Object Contours Using Starting-Point-Independent Wavelet Coefficient Matching. Journal of Visual Communication and Image Representation, 9(2): 171-181, 1998.

[11] D. S. Zhang and G. Lu Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study. In Proc. IEEE International Conference on Multimedia and Expo (ICME2001), August 22-25, 2001, Tokyo, Japan, pp.317-320.

[12] D. S. Zhang and G. Lu. "Shape Based Image Retrieval Using Generic Fourier Descriptors". Submitted to International Journal of Signal Processing: Image Communication, August, 2001.

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