Generic Fourier Descriptor for Shape-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 description is one of the key parts 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, a generic Fourier descriptor (GFD) is proposed to overcome the drawbacks of existing shape representation techniques. The proposed shape descriptor is derived by applying 2-D Fourier transform on a polar raster sampled shape image. The acquired shape descriptor is application independent and robust. Experimental results show the GFD outperforms Zernike moment descriptor (ZMD) which has been proposed to MPEG-7 as shape descriptor.
Keywords: Fourier descriptors, shape, CBIR, retrieval.

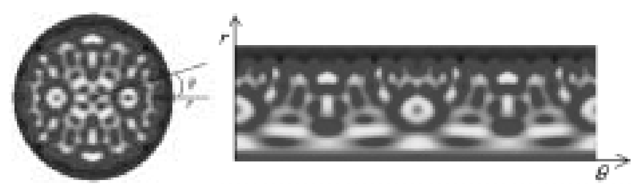
1. INTRODUCTION

Due to the rapid increase of multimedia information, there is an urgent need of multimedia content description to facilitate automatic searching. The newly emerging multimedia application MPEG-7 is to address this issue. In MPEG-7, shape is one of the key components for describing digital image along with other features such as texture and color. Many shape descriptors exist in the literature, these descriptors are broadly categorized into two groups, i.e., contour-based shape descriptors and region-based shape descriptors. 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. 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. Common region based methods use moments to describe shape [2, 3, 5, 6, 7, 8, 9, 10, 11, 12]. These include geometric moments, Legendre moments, Zernike moments and pseudo Zernike moments. It has been shown in [10] that Zernike moment descriptor outperforms other moment methods in terms of overall performance. Several researches also report promising retrieval results using Zernike moments [5, 12]. Basically, Zernike moments are computed from a polar raster sampled shape image. However, shape description using Zernike moments has two drawbacks. One is that Zernike moments have repetitions at each order, which means, at each order there are several closely similar Zernike moments which are not much useful for shape description. In other words, there are redundant features in the selected Zernike moments. The other is that Zernike moments cannot capture spectral feature in radial directions and does not allow multi-resolution feature analysis along radial directions [5, 12, 13]. In other words, Zernike moments cannot examine shape details in radial directions.

In this paper, we propose a generic Fourier descriptor (GFD). The GFD is extracted from spectral domain by applying 2-D Fourier transform on polar raster sampled shape image. Compared with Zernike moments, GFD has no redundant features and allows multi-resolution feature analysis in both radial and angular directions. The rest of the paper is organized as following. In section 2, the proposed GFD is described. In Section 3, retrieval experiments and comparison between GFD and ZMD are given. Section 4 concludes the paper.

2. GENERIC FOURIER DESCRIPTOR

It is well known that Fourier transform (FT) is very useful for pattern analysis. Shape analysis using FT is backed by well developed and well understood Fourier theory. However, it is not desirable to acquire shape features using FT directly, because, the acquired features are not rotation invariant. Furthermore, the acquired features are not compact (Fig. 2(a)(b)). Therefore, a modified polar FT (MPFT) is proposed by treating the polar image in polar space as a normal two-dimensional rectangular image in Cartesian space. Fig. 1 demonstrates the rectangular polar image. Fig. 1(a) is the original shape image in polar space, Fig. 1(b) is the rectangular polar image plotted into Cartesian space.



(a) (b)

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

The polar image of Figure 1(b) is the normal rectangular image. Therefore, if we apply 2-D FT on this rectangular image, the polar FT has the similar form to the conventional 2-D discrete FT in Cartesian space. Consequently, for a given shape image $f(x, y)$, the MPFT is defined 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 $\theta_i = i(2\pi/T)$ $(0\le i<T)$; (x_c, y_c) is the center of mass of the shape; $0\le\rho<R$, $0\le\phi<T$. R and T are the radial and angular resolutions. The physical meanings of ρ and ϕ are clear, they are the ρth radial frequency and the ϕth angular frequency respectively. 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. Fig. 2(c)(d) show the polar images and the polar Fourier spectra of the two patterns in Fig. 2(a)(b). It can be observed from Fig. 2(c)(d) that rotation of pattern in Cartesian space results in circular shift in polar space. The circular shift does not change the spectra distribution on polar space. The polar Fourier spectra is more concentrated around the origin of the polar space. This is particularly well-suited for shape representation, because for efficient shape representation, the number of spectra features selected to describe the shape should not be large. Since *f*(*x*, *y*) is a real function, the spectra is circularly symmetric, only one quarter of the spectra features are needed to describe the shape. The acquired polar Fourier coefficients are translation invariant. Rotation and scaling invariance are achieved by the following normalization:

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

where *area* is the area of the bounding circle the shape 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 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. The selected GFD features form a feature vector which is used for indexing the shape. For two shapes represented by their GFDs, the similarity between the two shapes is measured by the *city block distance* between the two feature vectors of the shapes. Therefore, the online matching is efficient and simple.

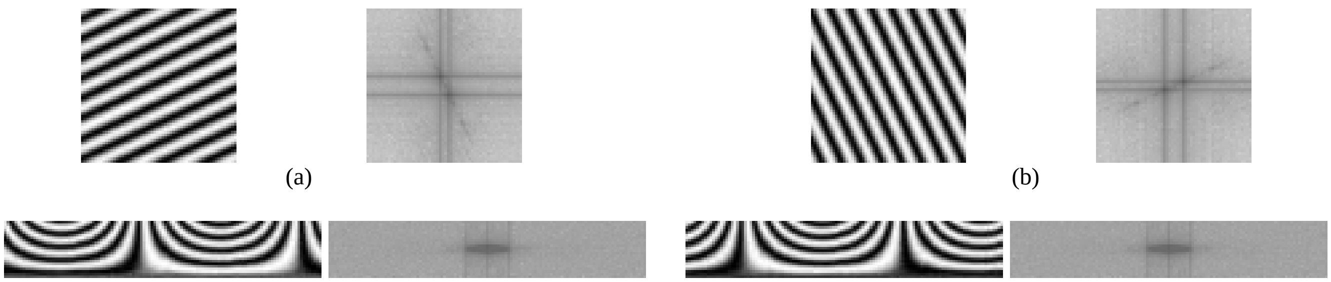


Fig. 2. (a) A pattern and its Fourier spectra; (b) rotated pattern of (a) and its Fourier spectra; (c) polar image of (a) and its Fourier spectra; (d) polar image of (b) and its Fourier spectra.

3. TEST OF RETRIEVAL EFFECTIVENESS

In order to test the retrieval performance of the proposed GFD and compare with ZMD, five sets of tests are carried out on MPEG-7 region shape database. MPEG-7 region shape database CE-2 consists of 3621 shapes of mainly trademarks. It has been organized as 6 sets for different types of tests. The use of the database is briefly described here. Set A1 consists of 2881 shapes from the whole database, it is for test of scale invariance. 100 shapes in Set A1 has been classified into 20 groups (5 similar shapes in each group) which are designated as queries. Set A2 consists of 2921 shapes from the whole database, it is for test of rotation invariance. 140 shapes in Set A2 has been classified into 20 groups (7 similar shapes in each group) which are designated as queries. Set A4 consists of 3101 from the whole database, it is for test of robustness to perspective transform. 330 shapes in Set A4 has been classified into 30 groups (11 similar shapes in each group) which are designated as queries. Set B consists of 2811 shapes from the whole database, it is for subjective test. 682 shapes in Set B have been manually classified into 10 groups by MPEG-7. For the whole database, 651 shapes have been classified into 31 groups (21 similar shapes in each groups) which can be used as queries. Among the 21 similar shapes in each group, there are 10 perspective transformed shapes, 5 rotated shapes and 5 scaled shapes. The 31 groups of shapes reflect overall shape distortions, and they test the overall robustness of a shape descriptor. The whole database is 17-29% larger in size than the individual sets.

Since the member IDs of each query group are known, the retrieval is done automatically. However, the retrieval system is also put online to test real time retrieval. For online retrieval, the indexed data and the shape databases are put in a web server, user can do online retrieval by visiting the retrieval site using either common browsers or Java appletviewer at http://gofaster.gscit.monash.edu.au/~dengs/Regionn/ gfd_src/ query.html.

For Set A1, A2, A4 and the whole database, the commonly used precision and recall pair [1] is used as the performance measurement. 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-recall of retrieval using the two shape descriptors on each set are shown in Fig. 3(a)-(d). Some screen shots of retrieval are shown in Fig. 4 and Fig. 5. In all the screen shots, the top left shape is the query shape. The retrieved shapes are ranked in descending order of similarity to the query shape. For Set B, because the number of members in each group is different, the Bull's eye performance (BEP) is used for the evaluation of retrieval effectiveness. The BEP is measured by the correct retrievals among the top 2*N* retrievals, where *N* is the number of relevant (or similar shapes) shapes to the query in the database. The BEP of Set B is given in Table 1.

It can be seen from Fig. 3 that there is only slight difference (overall precision is less than 2% different) of retrieval performance between GFD and ZMD on Set A2. Both GFD and ZMD have very high performance on this set. However, the difference between GFD and ZMD on Set A1, A4 and Set B is substantial (difference of overall precision on each set is over 4%-5%) and the difference between GFD and ZMD on the whole database is significant (difference of overall precision is over 12%). The reasons are explained as following.

Scaling, especially large scaling, can cause shape content or spatial distribution substantially changed. ZMD meets problem in dealing this type of situation because it is only able to examine shape features in circular direction. However, GFD can successfully deal with this type of situation by examining shape more carefully on radial directions (Fig. 4(a)).

Perspective deformations can also result in scaling effect, as a result, shape spatial distribution can be changed substantially. Parts of shape can be lost due to the transform. GFD can cope with this type of situation by examining shape features in radial directions (Fig.4(b)).

Due to the capturing of shape features in both radial and circular directions, the retrieved shapes are more perceptually acceptable. For example, in Fig. 5(a), GFD not only retrieves those similar shapes to the query, but also retrieves perceptually relevant shapes such as the members in group 1002. Example retrievals from Set B (Fig. 5(b)(c)(d)) also demonstrate retrievals using GFD are more perceptually acceptable than ZMD.

GFD is more robust than ZMD when the size of the shape database is increased. This is reflected in the retrieval performance on the whole database (Fig. 3(d)).

The computation of extracting GFD is simpler than ZMD. First, it does not need to normalize shape into an unit disk as is required in extracting ZMD (because Zernike moments is defined within a unit disk). Furthermore, the proposed polar Fourier transform is simpler than the Zernike moments calculation. PFT avoids the complex computation of Zernike polynomials [5]. The computations of online matching using GFD and ZMD are the same, because both methods use city block distance for similarity measurement and the number (36) of GFD features used to index the shape is the same as the number of ZMD used to index the shape [4].

4. CONCLUSIONS

In this paper, we have proposed a generic Fourier descriptor desirable for image retrieval. It has been tested on MPEG-7 region shape database. Comparisons have been made between GFD and MPEG-7 shape descriptor ZMD, results show that the proposed GFD outperforms ZMD. Compared with ZMD, GFD has four advantages: **i**) it captures spectral features in both radial and circular directions; **ii**) it is simpler to compute; **iii**) it is more robust and perceptually meaningful; **iv**) the physical meaning of each feature is more clear. The proposed GFD satisfies all the six requirements set by MPEG-7 for shape representation, that is, good retrieval accuracy, compact features, general application, low computation complexity, robust retrieval performance and hierarchical coarse to fine representation.

REFERENCES:

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

[2] S. O. Belkasim, M. Shridhar and M. Ahmadi. Pattern recognition with moment invariants: a comparative study and new results. Pattern Recognition, Vol. 24, No. 12, pp. 1117-1138, 1991.

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

[7] D. Mohamad, G. Sulong and S. S. Ipson. Trademark Matching Using Invariant Moments. Proceedings of the Second Asian Conference on Computer Vision, pp. I439-III444, Singapore, 5-8 December 1995.

[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.

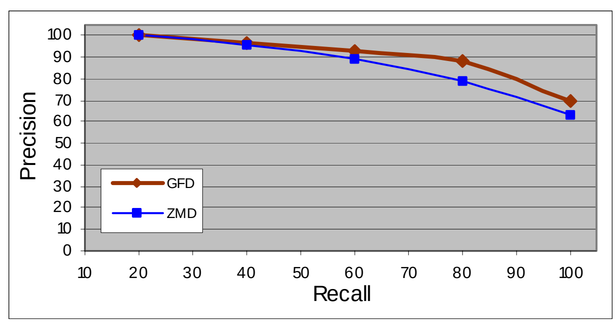
[9] M. R. Teague. Image Analysis Via the General theory of Moments. Journal of Optical Society of America, 70(8):920-930, 1980.

[10] 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

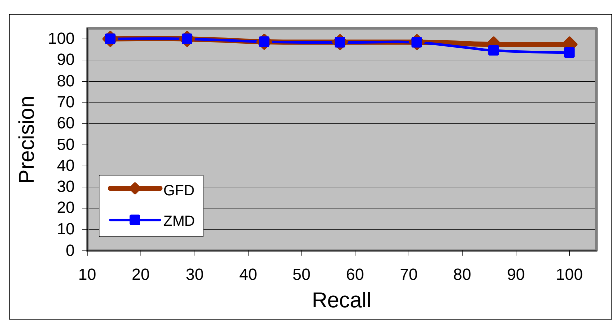
[11] G. Taubin and D. B. Cooper. Recognition and Positioning of Rigid Objects Using Algebraic Moment Invariants. SPIE Conf. On Geometric Methods in Computer Vision, Vol. 1570, pp.175-186, 1991.

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

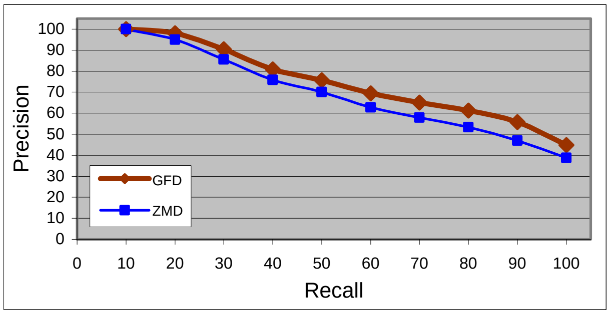
[13] D. S. Zhang and G. Lu. "Shape Based Image Retrieval Using Generic Fourier Descriptors". Submitted to International Journal of Signal Processing: Image Communication, August, 2001.



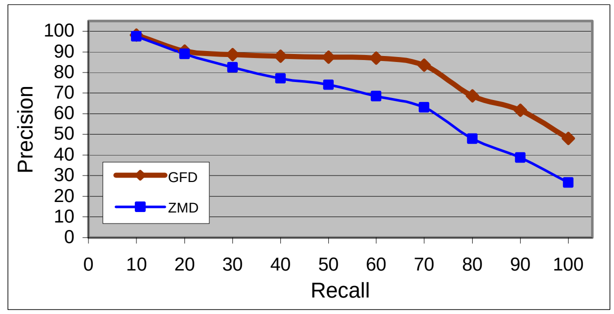
(a)



(b)



(c)

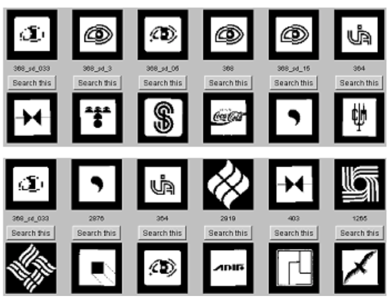


(d)

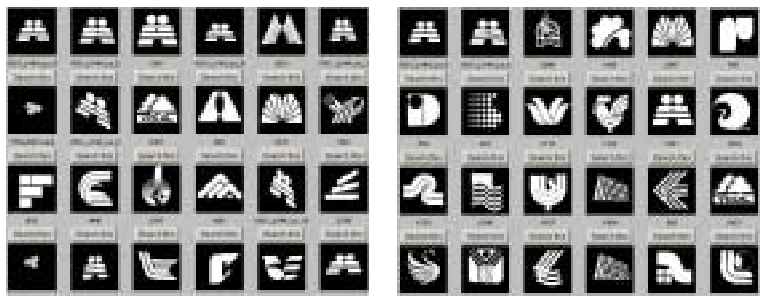
Fig. 3. Average precision-recall of (a) 100 queries in Set A1; (b) 140 queries in Set A2; (c) 330 queries in Set A4; (d) 651 queries in CE-2.

Table 1. Bull's eye performance of the 10 class (682) of queries in Set B of CE-2

class	1	2	3	4	5	6	7	8	9	10	
No. of shapes	68	248	22	28	17	22	45	145	45	42	Average
GFD (%)	47.0	66.4	55.6	50.0	50.0	24.8	30.4	50.8	55.6	29.0	46.0
ZMD (%)	37.0	58.0	55.0	41.2	42.6	22.6	33.6	52.0	41.4	34.0	41.7

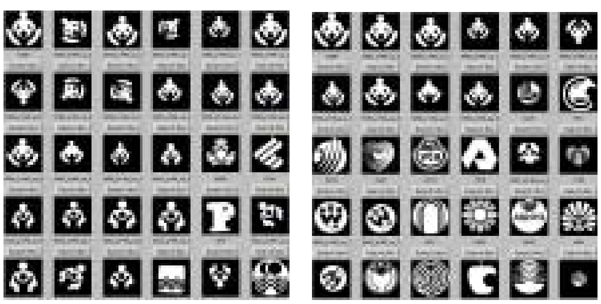


(a)

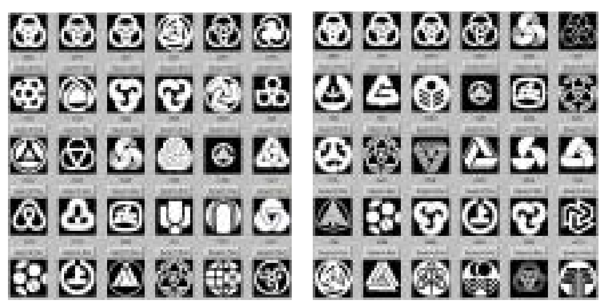


(b)

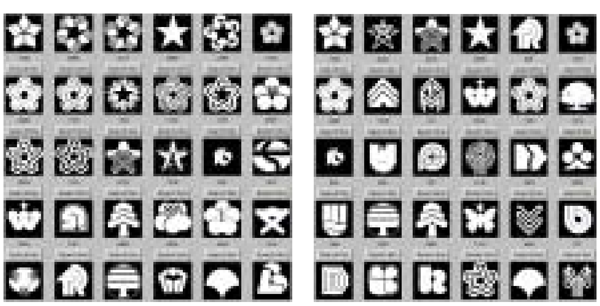
Fig. 4. (a) A retrieval on Set A1 using GFD (top) and ZMD (bottom); (b) a retrieval on Set A4 using GFD(left) and ZMD (right).



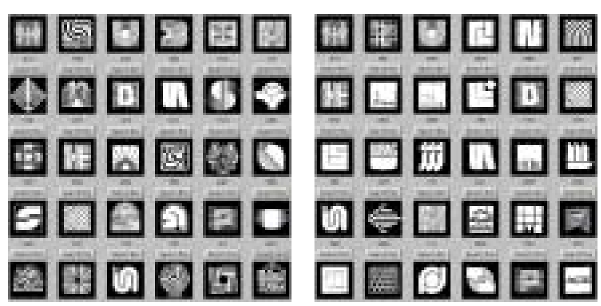
(a)



(b)



(c)



(d)

Fig. 5. Retrieve shape (a)1009 on CE-2 using GFD (left) and ZMD (right); (b)2992 using GFD (left) and ZMD (right); (c)1180 using GFD (left) and ZMD (right); (c)1011using GFD (left) and ZMD (right). (b)(c)(d) are retrievals on Set B of CE-2.