Chapter 6

Generic Fourier Descriptor

6.1 Introduction

In the previous two chapters, contour-based shape descriptors and region-based shape descriptors have been studied. The study has found that contour-based shape descriptors are usually suitable for describing contour shape without sophisticated boundary. Region-based shape descriptors can be applied to more general applications and are more robustness in describing shape with sophisticated boundary compared with contour-based shape descriptors, because they exploit more shape information to derive shape features. Among them Zernike moment descriptor (ZMD) proves to be preferred to other descriptors. In this chapter, we demonstrate that there are several shortcomings inherent with ZMD and propose a new shape descriptor which is more powerful than ZMD for shape description and retrieval. The rest of the chapter is organized as following. In Section 6.2, related work to the proposed shape descriptor is briefly analyzed. The theory of the proposed generic Fourier descriptor (GFD) will be introduced in Section 6.3. In Section 6.4 and 6.5, the derivation and implementation of GFD are described respectively. In Section 6.6, the retrieval effectiveness of the proposed GFD is tested and GFD is compared with other shape descriptors. The retrieval results will be reported and analyzed. The enhanced GFD will be presented in Section 6.7 to solve the affine sensitivity problem. Discussions of the proposed shape descriptor will be given in Section 6.8. The chapter is summarized in Section 6.9.

6.2 Motivations of Proposing a New Method

In this section, we give a brief review of two previous work and show how the idea of the proposed shape descriptor is formulated.

6.2.1 One Dimensional Fourier Descriptor

One dimensional FD has been successfully applied to many shape representation applications. The nice characteristics of FD, such as simple derivation, simple normalization, simple to do matching, robust to noise, perceptually meaningful feature, compact and hierarchical coarse to fine representation, make it a popular shape descriptor. However, 1-D FD assumes the knowledge of shape boundary information which may not be available in general situations. For example, it is difficult to derive 1-D FD for the shape in Figure 6.1(a), because the contour of the shape is not available. Furthermore, 1-D FD cannot capture shape interior content which is important for shape discrimination. For example, 1-D FD is not able to discriminate the shape in Figure 6.1(b) from the shape in Figure 6.1(c). These drawbacks limit the application of 1-D FD for generic shape description.

(a) (b) (c)

Figure 6.1. (a) A shape without contour; (b)(c) two shapes with same contour but with different interior content.

6.2.2 Zernike Moment Descriptor

The application of Zernike moments on shape overcomes the aforementioned drawbacks of 1-D FD. Results from Chapter 5 show that ZMD is a promising shape descriptor for image retrieval. An examination into ZMD reveals that it is derived from radial-angular transform of shape on polar space. The theory of ZMD is similar to FD. However, a more careful analysis on ZMD reveals that there are several shortcomings in ZMD. In the next, these shortcomings are explained in details. Section 5.3 of Chapter 5 is referred to for the following discussions.

The first shortcoming is that the bases of ZMD contain only angular frequency, they do not capture spectral features in radial directions. Among the two terms in the kernel of Zernike

moments $R_{nm}(\rho)\exp(jm\theta)$, $\exp(jm\theta)$ ($m \in \mathbb{Z}$) are the conventional sinusoid bases. They capture shape spectral features in angular directions when shape is sampled circularly. However, the real parts $R_{nm}(\rho)$ which are used to capture radial features are polynomials, they do not sum to zero energy. This indicates that they do not have characteristics of spectral bases. Consequently, the use of Zernike polynomials does not capture shape spectral features in radial directions. $R_{nm}(\rho)$ behaves similarly to the kernel of conventional moments. But unlike geometric moments which have physical meaning such as mean, variance, skew etc. for lower order moments (Chapter 5), moments derived from $R_{nm}(\rho)$ do not have physical meaning in radial directions because the radial moments derived are not central moments. The second shortcoming is that Zernike moments calculate different number of moments at each order. Specifically, Zernike moments calculate (n/2) +1 moments at order n, the moments compute (n/2) +1 odd circular frequencies at odd order and compute (n/2) +1 even circular frequencies at even order (Table 5.1). For example, at order 4, Zernike moments only compute moments of $A_{4,0}$, $A_{4,2}$, $A_{4,4}$, reflecting the 0^{th}, 2^{nd} and 4^{th} circular frequency respectively; at order 5, Zernike moments only compute moments of $A_{5,1}$, $A_{5,3}$, $A_{5,5}$, reflecting the 1^{st}, 3^{rd} and 5^{th} circular frequency respectively. Therefore, the circular spectral features are not evenly captured at each order. The uneven feature distribution at each order can result in loss of significant circular features, consequently, can affect the accuracy of shape description. The third shortcoming is that Zernike polynomials are complex to compute, and shape has to be scale normalized during the implementation due to Zenike moments being defined in unit disk.

In summary, Zernike moments has three shortcomings: (i) Zernike moments capture spatial moment features rather than spectral features in radial directions, they do not allow multi-frequency examination in radial directions; (ii) circular spectral features are not captured evenly at each order, this can result in loss of significant features which are useful for shape description. (iii) the extraction of Zernike moment features is complex.

It is known from the previous chapters that features captured from spatial domain are not as robust as features captured from spectral domain. Higher order moments do not have physical meaning as spectral features have. The uneven circular feature distribution at each order of Zernike moments can be easily overcome by using Fourier transform. Besides, Fourier transform is simpler and more convenient. The spectral features captured by Fourier coefficients are clearer. With these ideas in mind, we attempt to use 2-D polar FT (PFT) instead of Zernike moments to extract shape features. 2-D polar FT allows multi-frequency examination in both radial and angular directions. The elegance of deriving shape features using PFT is that both radial features

and angular features captured are coherent, that is, they are all spectral features. Spectral features are well understood and studied, which means they are clearer and more reliable. Since the transform kernel is much simpler and there is no constraint inherent in the definition, the computation of features is more efficient. In the next section, PFT is introduced from formal FT theory.

6.3 Polar Fourier Transform

Fourier transform has been widely used for image processing and analysis. The advantage of analyzing image in spectral domain over analyzing shape in spatial domain is that it is easy to overcome the noise problem which is common to digital images. Besides, the spectral features of an image are usually more concise than the features extracted from spatial domain. One dimensional FT has been successfully applied to contour shape (which is usually represented by a shape signature derived from the shape boundary coordinates) to derive FD. The application of one dimensional FT on shape assumes the knowledge of shape boundary information. There is no reported work on region based FD. In this section, generic FD is formally introduced from 2-D PFT.

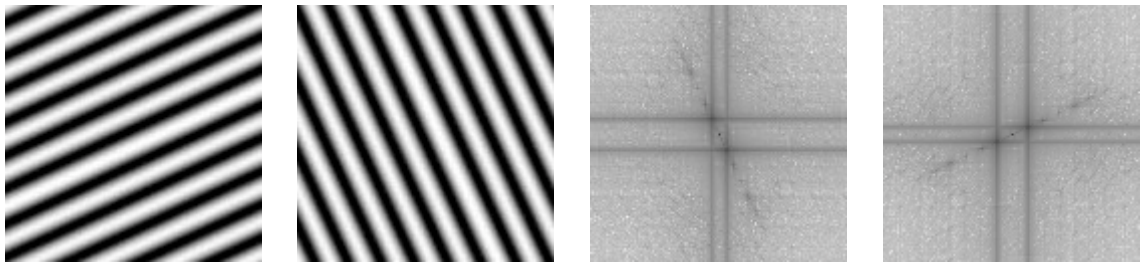
The continuous and discrete 2-D Fourier transform of a shape image $f(x, y)$ ($0 \le x < M$, $0 \le y < N$, M and N are the width and height of the image) are given by (6.1) and (6.2) respectively.

$$F(u,v) = \int_x \int_y f(x,y) \exp[-j2\pi(ux+vy)]\,dx\,dy \tag{6.1}$$

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \exp[-j2\pi(ux/M + vy/N)] \tag{6.2}$$

The u and v in (6.1) and (6.2) are the uth and vth spatial frequency in horizontal and vertical direction respectively. 2-D FT can be directly applied to any shape image without assuming the knowledge of boundary information. However, direct applying 2-D FT on a shape image in Cartesian space to derive FD is not practical because the features captured by 2-D FT are not rotation invariant. Rotation invariance of a shape is important because similar shapes can be under different orientations. For example, the two patterns (shapes) in Figure 6.2(a) and (b) are similar patterns (shapes), however, their Fourier spectra distributions (Figure 6.2(c) and (d)) on frequency plane are different. The difference of feature distributions makes it impractical to match the two patterns, especially online. Therefore, we consider shape image in polar space and applying polar Fourier transform (PFT) on shape image. The PFT produces rotation-invariant

data particularly well-suited for accurate extraction of orientation features. In the following, the two PFTs are studied and described. The study is necessary, because theoretically sound method may not be suitable for practical application.



(a) (b) (c) (d)

Figure 6.2. (a) a pattern; (b) pattern (a) rotated by 90 degree; (c) Fourier spectra of (a); (d) Fourier spectra of (b).

To derive PFT, both the data $f(x, y)$ and the spectra $F(u, v)$ are put into polar space, that is, let

$$\begin{aligned} x &= r \cdot \cos\theta, \quad y = r \cdot \sin\theta \\ u &= \rho \cdot \cos\psi, \quad v = \rho \cdot \sin\psi \end{aligned} \tag{6.3}$$

(r, θ) is the polar coordinates in image plane and (ρ, ψ) is the polar coordinates in frequency plane. The differentials of x and y are:

$$\begin{aligned} dx &= \cos\theta\, dr - r\sin\theta\, d\theta \\ dy &= \sin\theta\, dr - r\sin\theta\, d\theta \end{aligned} \tag{6.4}$$

The Jacobian of (6.4) is r. By substituting (6.3) and (6.4) into (6.1) we have the polar Fourier transform (PFT1):

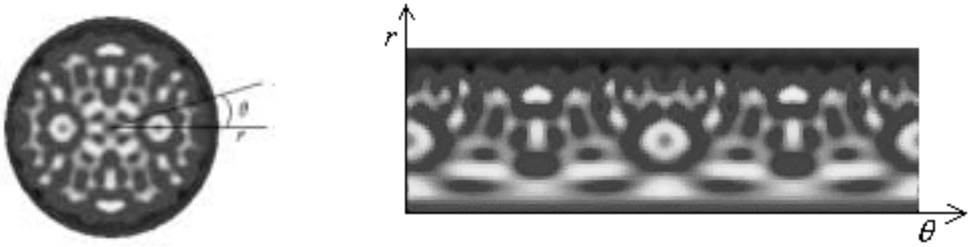
$$PF_1(\rho,\psi) = \int_r \int_\theta rf(r,\theta) \exp[-j2\pi r\rho \sin(\theta+\psi)]\, dr\, d\theta \tag{6.5}$$

The discrete PFT1 is then obtained as

$$PF_1(\rho_l,\psi_m) = \sum_p \sum_i f(r_p,\theta_i) \cdot r_p \cdot \exp[-j2\pi r_p \rho_l \sin(\theta_i+\psi_m)] \tag{6.6}$$

where $r_p = p/R$, $\theta_i = i(2\pi/T)$ ($0 \le i < T$); $\rho_l = l$ ($0 \le l <$ R) and $\psi_m = m\theta_i$. R and T are the resolution of radial frequency and angular frequency respectively. The acquired polar Fourier coefficients $PF_1(\rho_l, \psi_m)$ are used to derive normalized FD for shape representation.

PFT1 is the direct result from the polar transform of (6.1). However, due to the presence of ψ_m within *sine* function $sin(\theta_i + \psi_m)$, the physical meaning of ψ_m is not the mth angular frequency. The features captured by the PFT1 lose physical meaning in circular direction. To overcome the problem, a modified polar FT (PFT2) is derived by treating the polar image in polar space as a normal two-dimensional rectangular image in Cartesian space. Figure 6.3 demonstrates the rectangular polar images. For example, Figure 6.3(a) is the original shape image in polar space, Figure 6.3(b) is the rectangular polar image plotted into Cartesian space.



(a) (b)

(c)

(d)

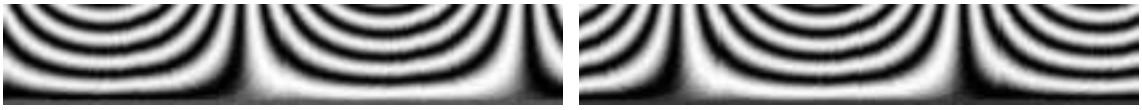
Figure 6.3. (a) original shape image in polar space; (b) polar image of (a) plotted into Cartesian space; (c) a heart image and its polar image; (d) a trade mark shape and its polar image.

The polar image in Figure 6.3(b) is the normal rectangular image. Therefore, if 2-D FT is applied on this rectangular image, the polar FT has the similar form to the normal 2-D discrete FT of (6.2) in Cartesian space. Consequently, the modified polar FT is defined as

$$PF_2(\rho,\phi)=\sum_r\sum_i f(r,\theta_i)\exp[j2\pi(\frac{r}{R}\rho+\frac{2\pi i}{T}\phi)] \tag{6.7}$$

where $0\le r<R$ and $\theta_i = i(2\pi/T)$ $(0\le i<T)$; $0\le\rho<R$, $0\le\phi<T$. R and T are the radial frequency resolution and angular frequency resolution respectively. PFT2 has a simpler form than ZMD and PFT1. There is no need to constrain the shape into a unit circle (the constraint requires a extra scale normalization in spatial domain) as required in the implementation of ZMD (because Zernike moment is defined in a unit circle). And the physical meaning of ρ and ϕ in (6.7) is similar to u and v in (6.1) and (6.2). The ρ and ϕ are simply the radial frequency and the angular frequency respectively. The determination of the number of ρ and ϕ is physically achievable, because significant shape features are usually captured by a few lower frequencies.

Figure 6.4(a)(b) shows the polar images of the two patterns in Figure 6.2(a)(b) and their polar Fourier spectra are shown in Figure 6.4(c) and (d). It can be observed from Figure 6.4 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. Because from Chapter 2, shift or translation in spatial domain results in phase change in spectral domain. If phase is ignored, the spectra is shift or translation invariant. This is demonstrated in Figure 6.4(c) and (d). 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 $F(\rho, \phi)$ are used to derive normalized FD for shape representation.



(a) (b)

(c) (d)

Figure 6.4. (a)(b) polar images of the two patterns in Figure 6.2(a) and (b); (c) Fourier spectra of (a); (d) Fourier spectra of (b).

6.4 Derivation of Generic FD

In this section, the derivation of FD using the above described PFTs is given in details. The two polar FTs: PFT1 and PFT2 are both implemented in the experiments to derive FD for the purpose of determining which one is the more desirable for shape retrieval.

Given a shape image $I = \{f(x, y);\ 0 \le x < M,\ 0 \le y < N\}$. To apply PFT, the shape image is converted from Cartesian space to polar space $I_p = \{f(r, \theta);\ 0 \le r < R,\ 0 \le \theta < 2\pi\}$, R is the maximum radius of the shape. The origin of the polar space is set to be the centroid of the shape, so that the shape is translation invariant. The centroid (x_c, y_c) is given by

$$x_c = \frac{1}{M}\sum_{x=0}^{N-1} x, \qquad y_c = \frac{1}{N}\sum_{y=0}^{M-1} y \tag{6.8}$$

and (r, θ) is given by

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2}, \quad \theta = \arctan\frac{y - y_c}{x - x_c} \tag{6.9}$$

The PFTs are applied on I_p. The acquired coefficients of the two PFTs are translation invariant due to the use of centroid as polar space origin. Rotation invariance is achieved by ignoring the phase information in the coefficients and only retaining the magnitudes of the coefficients. To achieve scale invariance, the first magnitude value is normalized by the area of the circle (*area*) in which the polar image resides or the mass of the shape (*mass*), and all the other magnitude values are normalized by the magnitude of the first coefficient. The translation, rotation and scale normalized PFT coefficients are used as the shape descriptors. To summarize, the shape descriptor derived from PFT1 and PFT2 are **FD1** and **FD2** respectively, they are shown as following

$$\mathbf{FD1} = \{\frac{|PF_1(0,0)|}{mass}, \frac{|PF_1(0,1)|}{|PF_1(0,0)|}, ..., \frac{|PF_1(0,n)|}{|PF_1(0,0)|}, ..., \frac{|PF_1(m,0)|}{|PF_1(0,0)|}, ..., \frac{|PF_1(m,n)|}{|PF_1(0,0)|}\}$$

$$\mathbf{FD2} = \{\frac{|PF_2(0,0)|}{area}, \frac{|PF_2(0,1)|}{|PF_2(0,0)|}, ..., \frac{|PF_2(0,n)|}{|PF_2(0,0)|}, ..., \frac{|PF_2(m,0)|}{|PF_2(0,0)|}, ..., \frac{|PF_2(m,n)|}{|PF_2(0,0)|}\}$$

where 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. Normally, the first coefficient, or the *DC* component is used as the

normalization factor and is discarded after normalization. However, this component is used as an additional feature, because it reflects the average energy (scale) of the shape which is useful for shape description.

For efficient shape description, only a small number of the acquired FD features are selected for shape representation. The selected FD features form a feature vector which is used for indexing the shape. For two shapes represented by their FD features, 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.

6.5 Implementation of GFD

The implementation of GFD can be summarized into 4 major steps, translation normalization, polar Fourier transform, rotation normalization and scale normalization. The algorithm for deriving GFD using PFT2 is given in the following. The algorithm for deriving GFD using PFT1 is similar, with the difference in the basis calculation of polar Fourier transform step.

Algorithm of computing GFD:

1. Input shape image data $f(x, y)$;
2. Get centroid of the shape (x_c, y_c);
3. Set the centroid as the origion; /* translation normalization */
4. Get the maximum radius of the shape image (*maxRad*);
5. Polar Fourier transform
   For radial frequency (*rad*) from zero to maximum radial frequency (*m*)
      For angular frequency (*ang*) from zero to maximum angular frequency (*n*)
         For *x* from zero to width of the shape image
            For *y* from zero to height of the shape image
            {
               radius = square root$[(x-x_c)^2 + (y-y_c)^2]$;
               theta = arctan2$[(y-x_c)/(x-y_c)]$; /* *theta* falls within $[-\pi, +\pi]$ */
               if(theta<0) theta += 2π; /* convert *theta* to $[0, 2\pi]$ */
               FR[*rad*][*ang*] += $f(x,y) \times \cos[2\pi \times rad \times (radius/maxRad) + ang \times theta]$; /* real part */
               FI[*rad*][*ang*] –= $f(x,y) \times \sin[2\pi \times rad \times (radius/maxRad) + ang \times theta]$; /* imaginary part */
            }
6. Calculate GFD
For *rad* from zero to *m*
   For *ang* from zero to *n*

{

/* rotation and scale normalization */

If (rad=0 & ang=0)

DC= square root[(FR^2[0][0] + FR^2[0][0])];

GFD[0] = DC/($\pi \times maxRad^2$);

Else

GFD[$rad \times n + ang$] = square root[(FR^2[rad][ang] + FI^2[rad][ang])]/DC;

}

7. Output feature vector **GFD**.

6.6 Test of Retrieval Effectiveness

In order to test retrieval effectiveness of the proposed methods, 3 sets of experiments are conducted. The first experiment is to compare the two proposed methods to determine which is the most suitable for shape retrieval. The other two experiments are to compare the proposed GFD with contour shape descriptors and region shape descriptor.

6.6.1 Comparison of FD1 and FD2

To test the retrieval effectiveness of the two FDs derived from PFT1 and PFT2, a Java-based indexing and retrieval framework is implemented. The framework runs on Windows platform of a Pentium III-866 PC. The retrieval effectiveness of the two types of FD described in Section 6.3 is tested on the region-based shape database of MPEG-7. MPEG-7 region shape database CE-2 consists of 3621 shapes of mainly trademarks. 651 shapes from 31 classes of shapes are selected as queries. The 31 classes of shapes reflect general variations (rotation, scaling and perspective transform) of shapes. Each class has 21 members generated through scaling, rotation and perspective transformation. Since the IDs of all the similar shapes to each query in the classes 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) (Figure 6.5).

Precision and recall of the retrieval are used as the evaluation of the query result. For each query, the precision of the retrieval at each level of the recall is obtained. The final precision of retrieval using a shape descriptor is the average precision of all the query retrievals using of shape

descriptor. The average precision and recall of the 651queries using the two derived shape descriptors are shown in Figure 6.6.

It is clear from Figure 6.6 that FD derived from PFT2 outperforms FD derived from PFT1. The fact that FD2 outperforms FD1 significantly indicates that FD derived from PFT2 is more suitable for shape description. Therefore, FD derived from PFT2 is selected as the generic FD (GFD) as shape representation. Hereafter, GFD refers to FD derived using PFT2, that is, FD2 is referred as GFD in the following sections.

60 GFD features (reflecting 5 radial frequencies and 12 angular frequencies) are selected as shape descriptors. However, different number of GFD features with different parameters are tested to determine which is the most appropriate number of GFD features to describe the shape. The test results are given in Table 6.1. From Table 6.1, it is observed that retrieval effectiveness is improved by increasing radial frequency resolution. However, retrieval effectiveness does not improve significantly when the radial resolution is greater than 3. It is also observed from the table that retrieval effectiveness does not improve significantly when the angular resolution is greater than 9. The observation indicates that for efficient retrieval, 36 GFD features (reflecting 4 radial frequencies and 9 angular frequencies) or 48 GFD features (reflecting 4 radial frequencies and 12 angular frequencies) is the suitable number of GFD features for shape description.

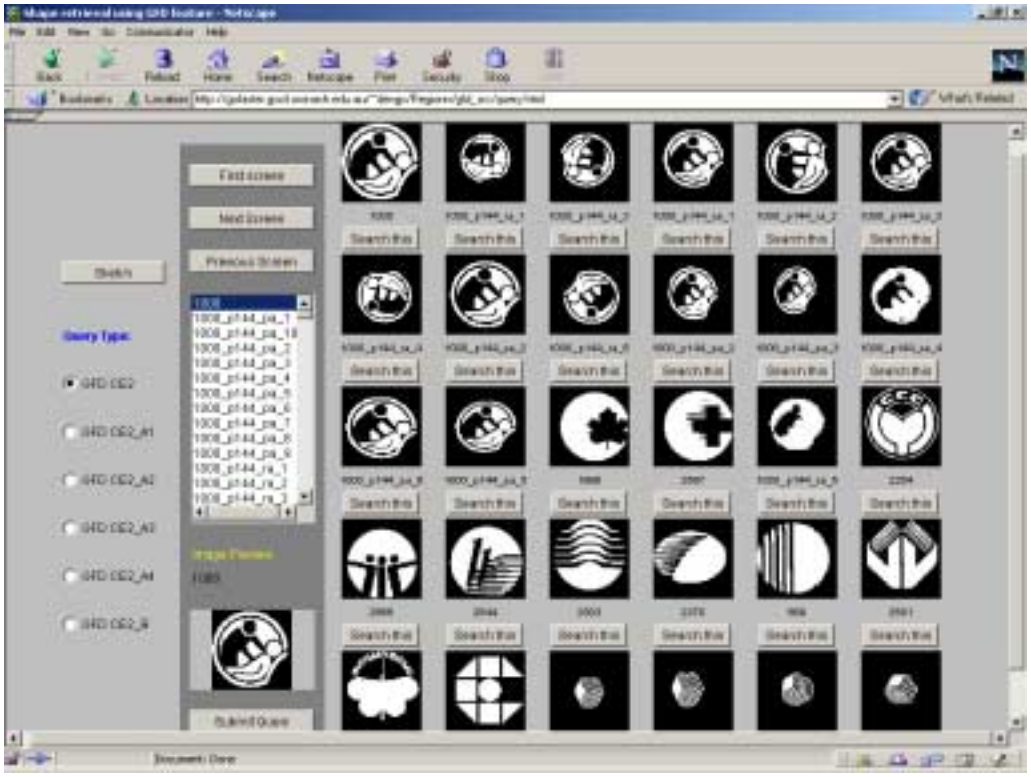


Figure 6.5. Retrieval interface.

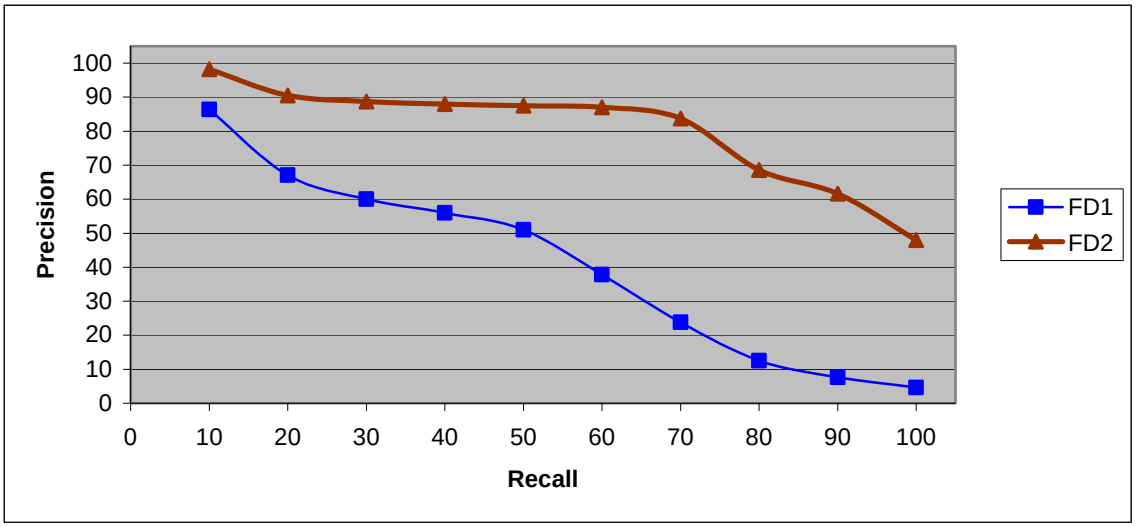


Figure 6.6. Average precision-recall of 651 retrievals using FD1 and FD2 on MPEG-7 region shape database CE-2.

Table 6.1. Retrieval performance of FDs with different number of radial and angular frequencies. *r*: number of radial frequencies selected; *t*: number of angular frequencies selected.

Recall (%) / Parameters	10	20	30	40	50	60	70	80	90	100	Overall precision on full recall
r=1, *t*=20	94.4	85.5	83.4	82.3	81.0	76.4	70.5	59.7	53.0	37.6	72.4
r=2, *t*=15	96.3	89.0	87.2	86.5	86.0	85.0	82.0	66.5	60.8	46.5	77.6
r=3, *t*=12	97.6	90.6	88.9	88.0	87.6	87.0	84.0	68.5	62.0	48.8	80.1
r=4, *t*=9	97.8	90.7	89.0	88.2	87.7	87.2	84.3	68.6	62.5	48.9	80.5
r=4, *t*=12	98.2	90.8	89.2	88.4	88.0	87.4	84.1	69.0	62.7	48.3	80.6
r=4, *t*=20	98.3	91.0	89.4	88.5	88.1	87.5	84.5	69.1	63.0	48.8	80.8
r=5, *t*=12	98.3	90.8	88.9	88.0	87.7	87.1	84.0	68.8	62.0	48.2	80.4
r=5, *t*=20	98.3	91.0	89.1	88.3	87.9	87.3	84.3	68.9	62.3	48.7	80.6
r=6, *t*=6	97.4	88.6	86.8	86.0	85.7	84.7	81.0	66.4	58.0	44.0	77.9
r=8, *t*=8	97.8	89.6	87.7	87.0	86.7	85.7	82.2	68.1	60.2	46.8	79.2
r=10, *t*=6	97.4	88.7	86.7	86.0	85.6	84.6	80.8	66.7	58.3	44.7	78.0
r=8, *t*=15	98.3	90.6	88.7	87.8	87.3	86.9	83.1	68.7	62.1	48.5	80.2
r=10, *t*=15	98.3	90.7	88.7	87.8	87.3	87.0	83.2	68.7	62.1	48.5	80.2

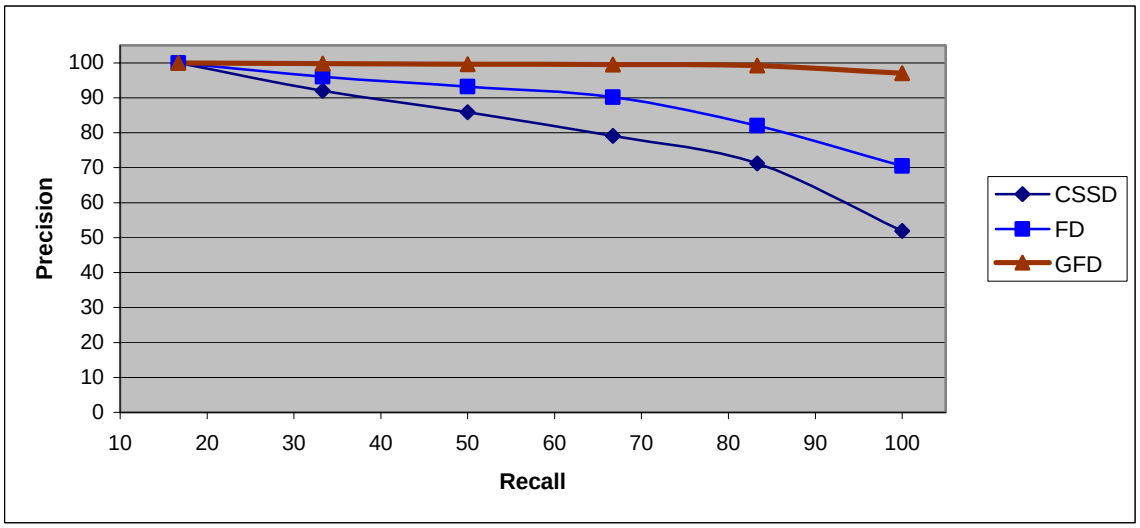
6.6.2 Comparison of GFD and Contour-based Shape Descriptors

Since the proposed GFD is a generic shape descriptor, we want to compare it with both contour-based shape descriptors and region-based shape descriptors. In this section, GFD is compared with contour FD and curvature scale space descriptor (CSSD) which has been adopted as contour shape descriptor in MPEG-7. The technique details of contour FD and CSSD have been described in Chapter 4.

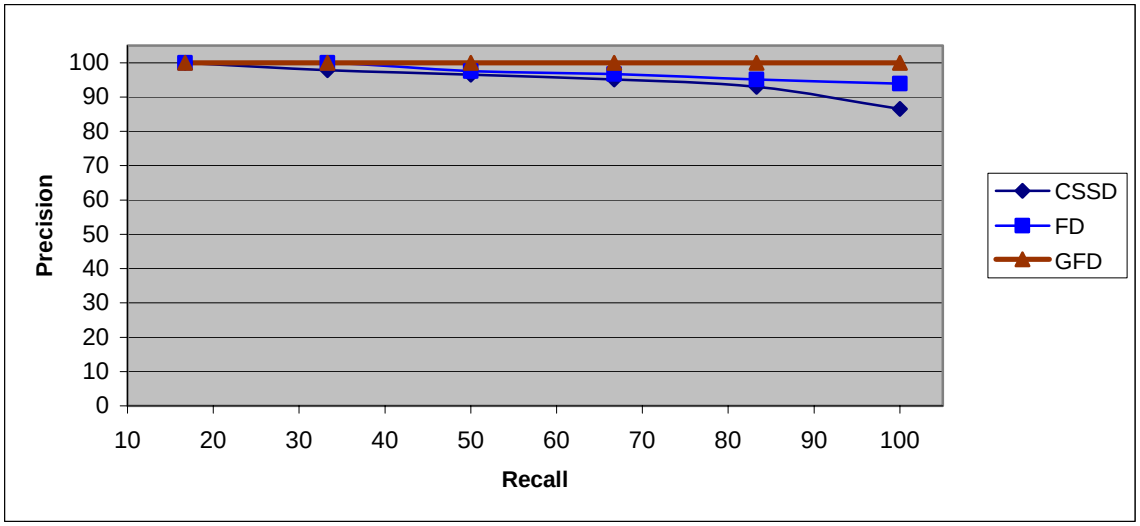
6.6.2.1 Test Setup and Retrieval Results

The retrieval tests are conducted on MPEG-7 contour shape database CE-1. CE-1 has been introduced in Chapter 4. It consists of Set A1, A2, B and C which are designed to test shape descriptor's behavior under scaling transformation, rotation transformation, arbitrary distortion and affine distortion respectively.

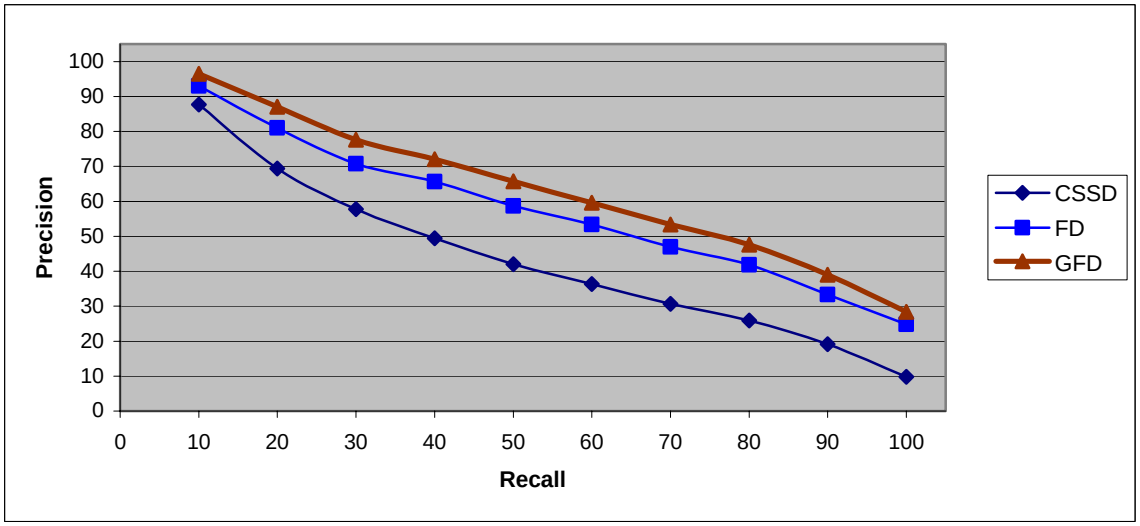
All the 420 shapes in A1 and A2, and all the 1400 shapes in B are used as queries. For Set C, the 200 bream fishes are used as queries. The precision-recall is used for evaluation of the retrieval effectiveness. The average precision and recall of the retrieval using the three shape descriptors on each set are shown in Figure 6.7(a)-(d). Some screen shots of retrievals on MPEG-7 contour shape database are shown in Figure 6.8(a)-(g). 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, they are arranged in left to right and top to bottom order.



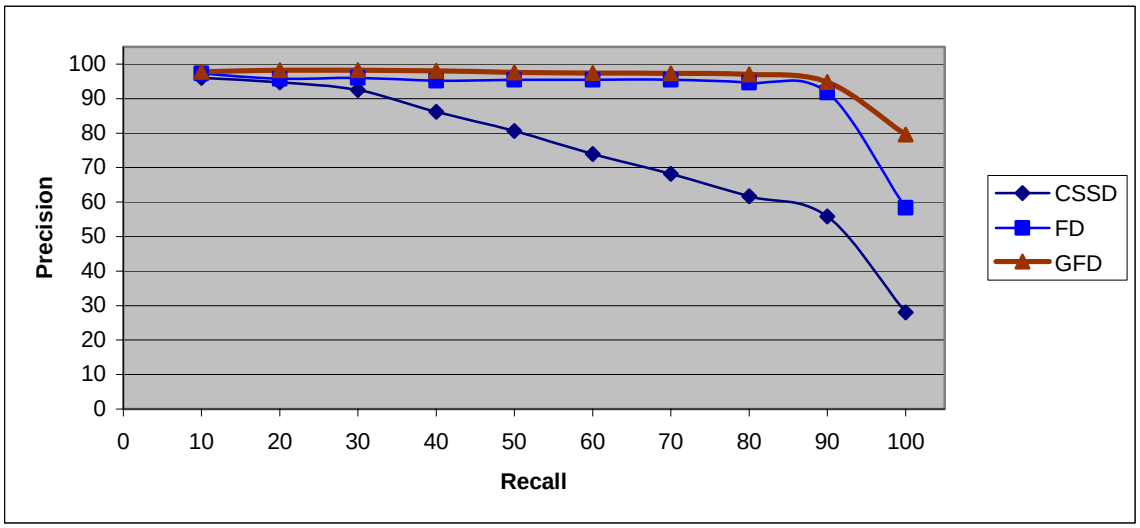
(a) Average precision-recall of 420 retrievals using GFD, FD and CSSD on Set A1 of MPEG-7 contour shape database CE-1.



(b) Average precision-recall of 420 retrievals using GFD, FD and CSSD on Set A2 of MPEG-7 contour shape database CE-1.



(c) Average precision-recall of 1400 retrievals using GFD, FD and CSSD on Set B of MPEG-7 contour shape database CE-1.



(d) Average precision-recall of 200 retrievals using GFD, FD and CSSD on Set C of MPEG-7 contour shape database CE-1.

Figure 6.7. Retrieval effectiveness of GFD, FD and CSSD on MPEG-7 contour shape database

6.6.2.2 Results Analysis

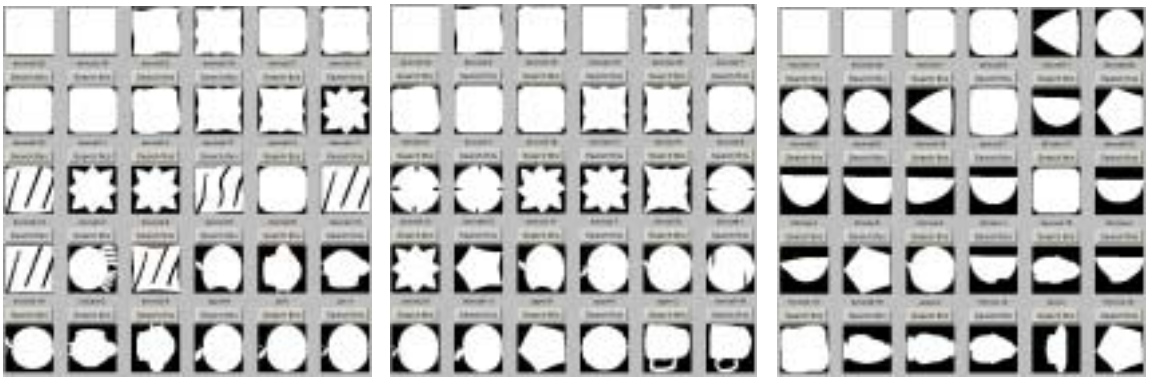
It can be seen from Figure 6.7 that the proposed GFD outperforms 1-D FD and CSSD on all the sets of MPEG-7 contour shape database. The performance of CSSD is significantly lower than GFD and 1-D FD due to its complex normalization and matching. GFD has 100% correct retrieval of rotated shapes. It has almost 100% correct retrieval of scaled shapes and non-rigid affine distorted shapes. The advantage of GFD over contour-based shape descriptors is obvious in situations where severe protrusions and indentations occur. For example, in Figure 6.8(a), GFD not only retrieves those distorted square shapes but also retrieves those squares with severe indentations. For the contour shape descriptors, they can only retrieve square shapes without any indentations. In Figure 6.8(b), GFD retrieves most of the ray fish shapes in the first screen, however, the contour shape descriptors are easily trapped by those protrusions of the shapes. Contour shape descriptors treat any shapes with hook-like parts as similar shapes to the ray fish shape which has a hook-like tail. Similar to (b), in Figure 6.8(c), contour shape descriptors are easily distracted by the complex arms of the flies. They are confused with shapes with similar protrusions, or similar number of protrusions. GFD is able to concentrate on the main body of the fly shapes, successfully retrieve most of the fly shapes in the first screen.

When a shape is scaled, its boundary can be substantially changed. The contour shape descriptors can fail completely when large scaling occurs (Figure 6.8(f)). However, GFD is not affected by large scaling.

GFD is also more robust to severe deformation of shape than the contour shape descriptors. The fish bream-120 is a severely distorted shape, however, GFD correctly retrieve its similar shapes (Figure 6.8(e)). 1-D FD only works better than GFD in situations where the protrusions and indentations constitute the main body of the shape. The fork shape in Figure 6.8(d) consists only of protrusion parts. 1-D FD has very high performance on this shape.

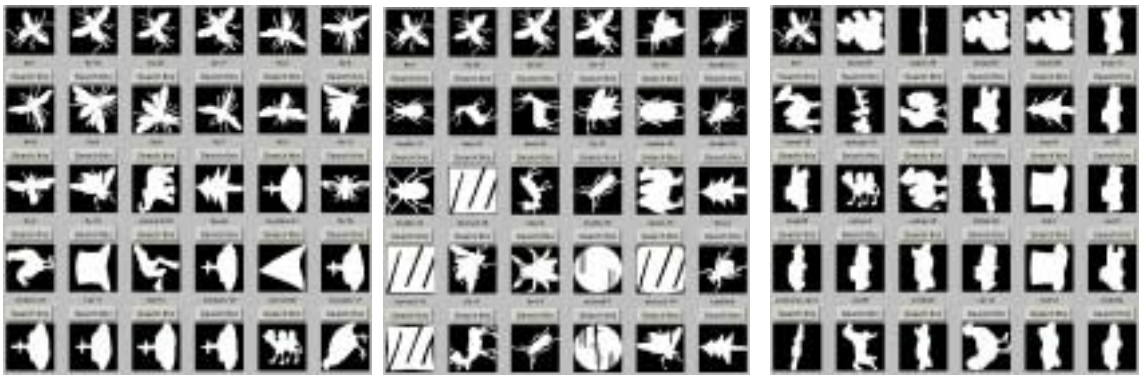
To summarize, due to using only boundary information, 1-D FD and CSSD are more likely affected by various types of shape variations such as scaling, protrusions, indentations and deformations. Due to using all the information within a shape region, GFD is more robust to shape variations than 1-D FD and CSSD.

Although the extraction of GFD requires more computation than the extraction of contour FD, the computation of online matching using GFD is about as efficient as contour FD. Because both GFD and contour FD use city block distance measure for online matching. For image retrieval application, low computation of online matching is essential although low computation of offline feature extraction is also desirable.



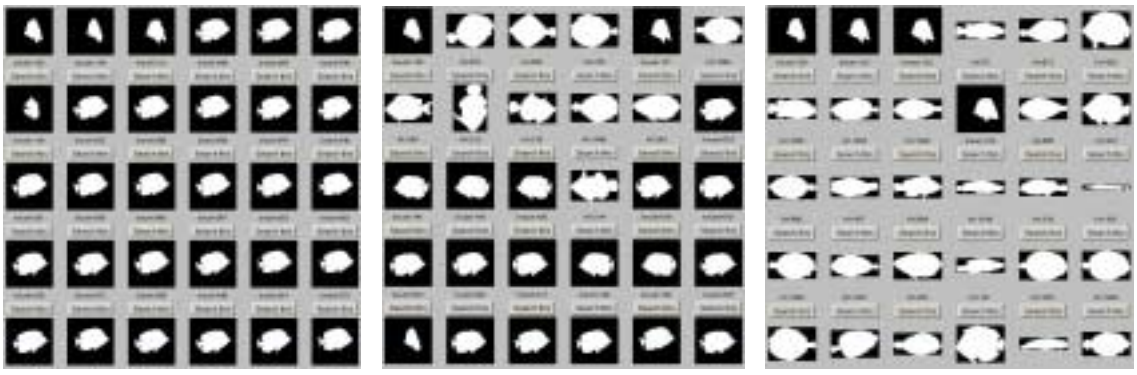
(a) Retrieval of query device3-20 on Set B using (left) GFD; (middle) 1-D FD; (right) CSSD.

(b) Retrieval of query ray-1 on Set B using (left) GFD; (middle) 1-D FD; (right) CSSD.

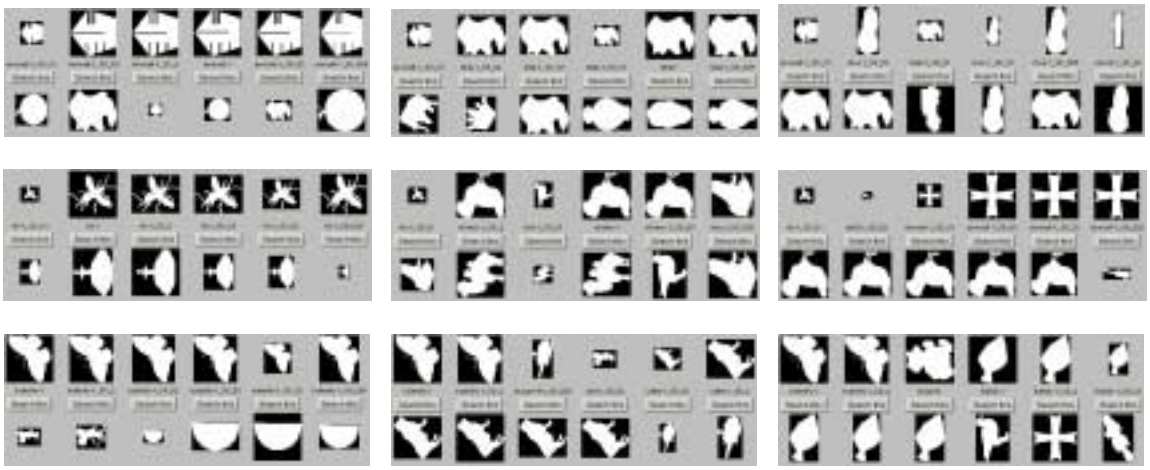


(c) Retrieval of query fly-1 on Set B using (left) GFD; (middle) 1-D FD; (right) CSSD.

(d) Retrieval of query fork-1 on Set B using (left) GFD; (middle) 1-D FD; (right) CSSD.



(e) Retrieval of query bream-120 on Set C using (left) GFD; (middle) 1-D FD; (right) CSSD.



(f) Retrievals of device6-1_SD_01, fly-1_SD_01 and butterfly-1 on Set A1 using (left) GFD; (middle) 1-D FD; (right) CSSD.

Figure 6.8. Example retrievals using GFD, 1-D FD and CSSD on different sets of CE-1.

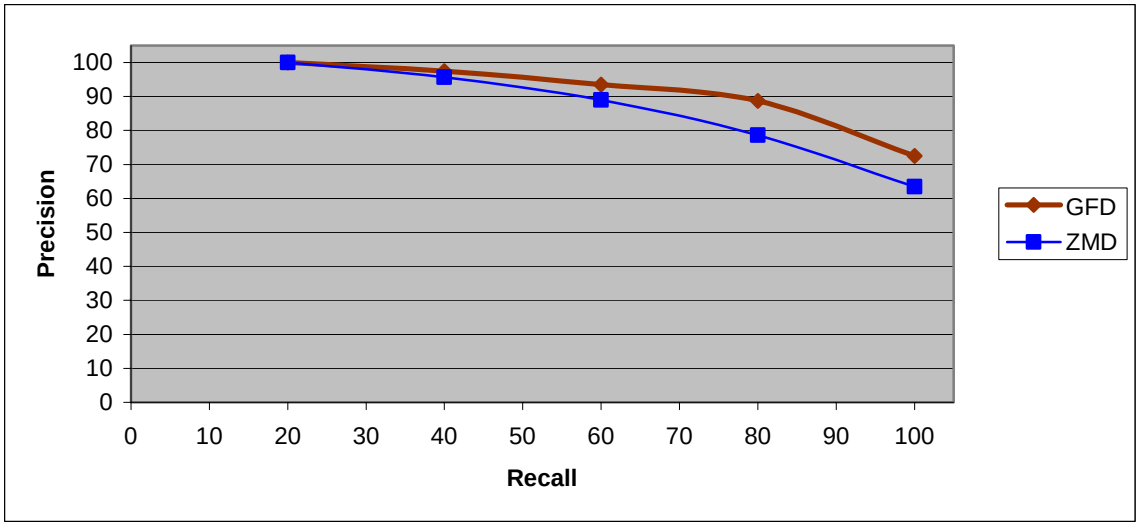
6.6.3 Comparison of GFD and Region-based Shape descriptor

In this section, GFD is compared with ZMD which has been adopted as region-based shape descriptor in MPEG-7. Both the retrieval effectiveness and efficiency are compared.

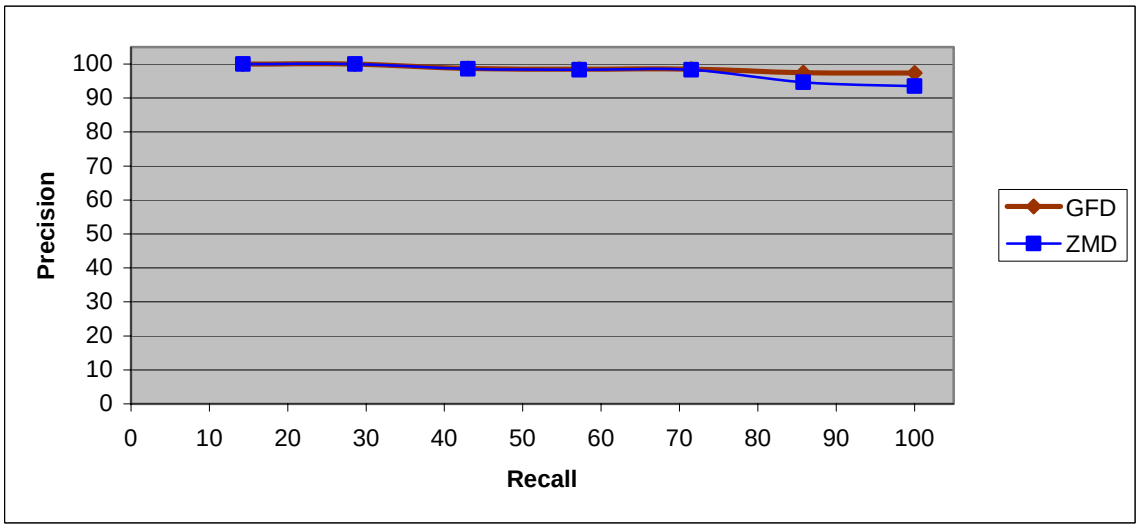
6.6.3.1 Test Setup and Retrieval Results

The comparison is conducted on MPEG-7 region-based shape database (CE-2). MPEG-7 region shape database CE-2 consists of 3621 shapes of mainly trademarks. MPEG-7 region shape database CE-2 has been introduced in Chapter 5. It has been organized by MPEG-7 into 6 subsets A1, A2, A3, A4, B and the whole database to test shape descriptor's behavior under scaling, rotation, scaling/rotation, perspective transformation, subjective test and general distortion.

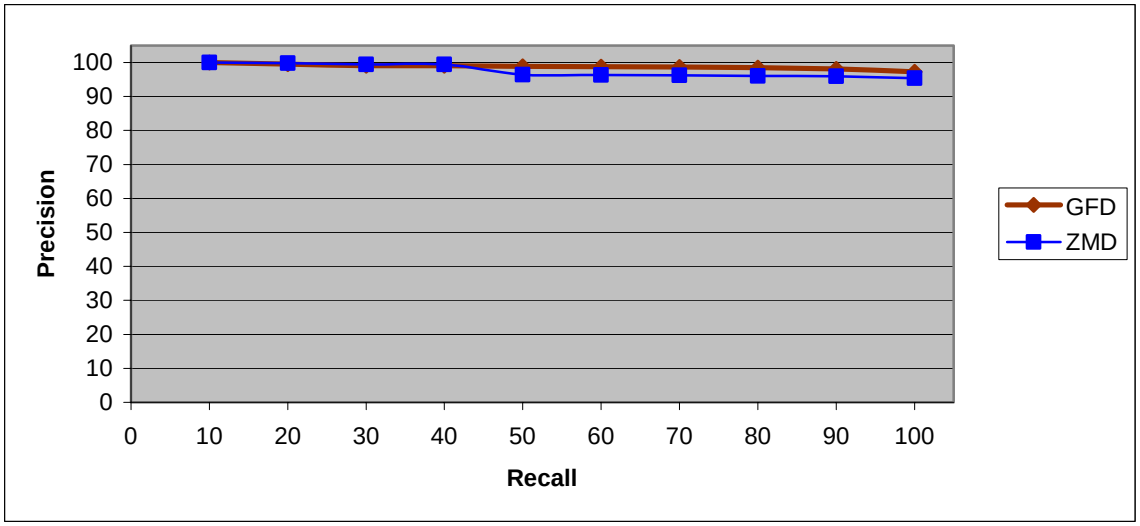
100 shapes in A1, 140 shapes in A2, 330 shape in A3 and A4, 682 shapes in B and 651 shapes in CE-2 are used as queries in each test dataset respectively. The precision-recall is used for evaluation of retrieval effectiveness. The average precision-recall of retrieval using the two shape descriptors on each set are shown in Figure 6.9(a)-(f).



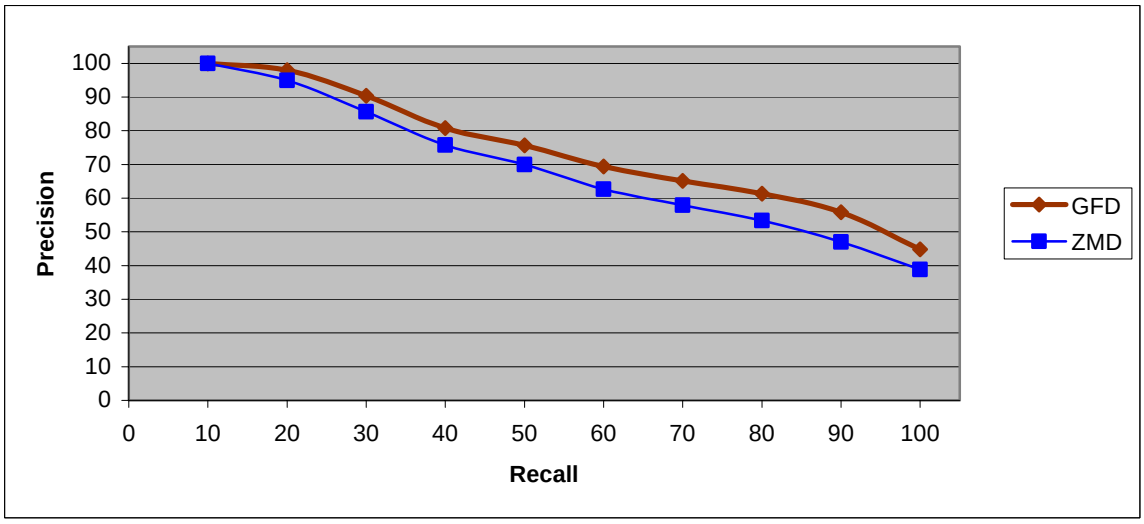
(a) Average precision-recall of 100 retrievals using GFD, ZMD on Set A1 of MPEG-7 region shape database CE-2.



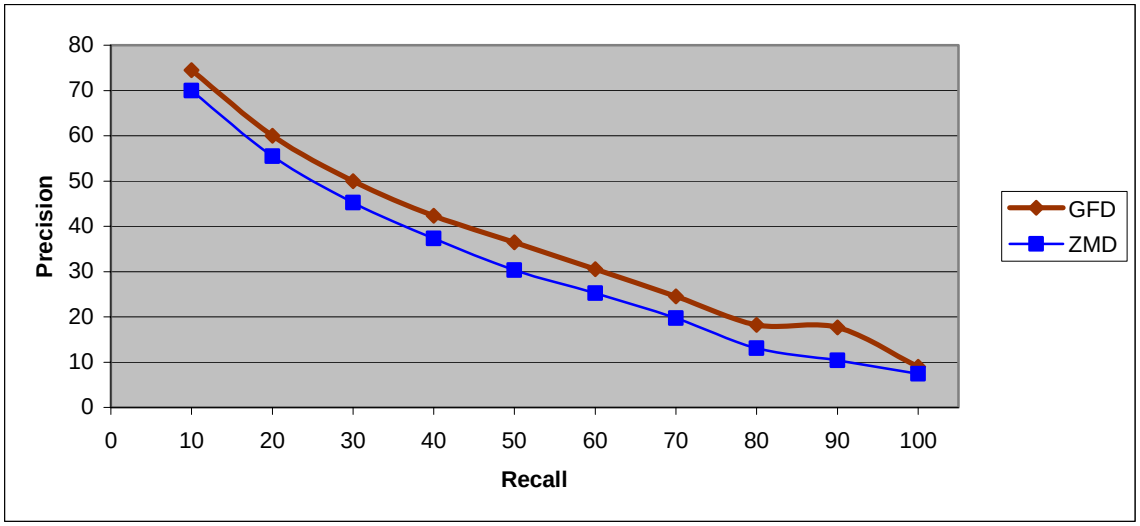
(b) Average precision-recall of 140 retrievals using GFD, ZMD on Set A2 of MPEG-7 region shape database CE-2.



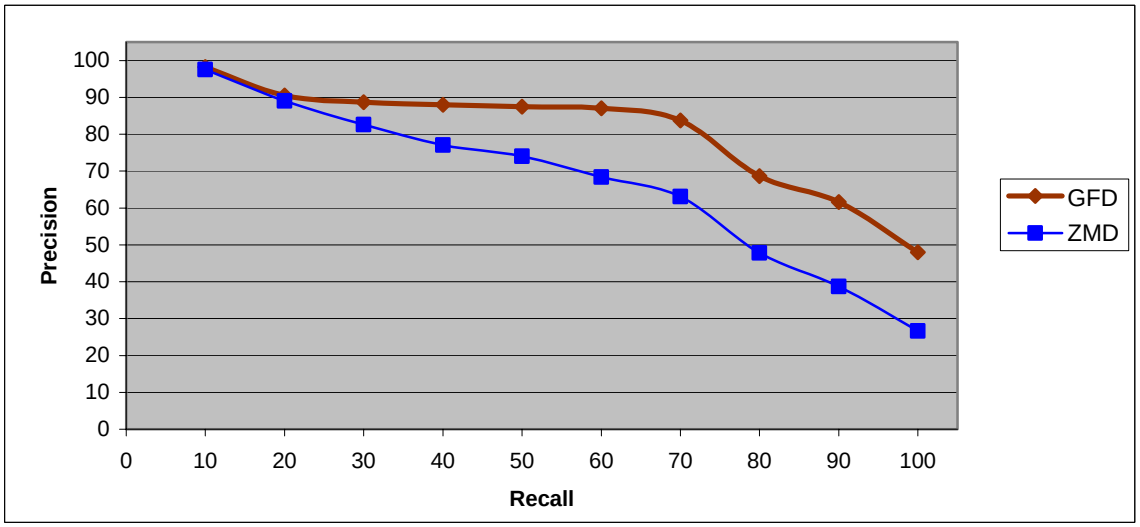
(c) Average precision-recall of 330 retrievals using GFD, ZMD on Set A3 of MPEG-7 region shape database CE-2.



(d) Average precision-recall of 330 retrievals using GFD, ZMD on Set A4 of MPEG-7 region shape database CE-2.



(e) Average precision-recall of 682 retrievals using GFD, ZMD on Set B of MPEG-7 region shape database CE-2.



(f) Average precision-recall of 651 retrievals using GFD, ZMD on MPEG-7 region shape database CE-2.

Figure 6.9. Retrieval effectiveness of GFD and ZMD on each dataset of MPEG-7 region shape database.

6.6.3.2 Results Analysis

It can be seen from Figure 6.9 that GFD performs slightly better than ZMD on Set A2 and A3 (overall precision after full recall is about 2% different). Both GFD and ZMD have very high performance on these two sets. However, the difference between GFD and ZMD on Set A1, A4 and Set B is substantial (difference of overall precision after full recall on each set is over 5%) and the difference between GFD and ZMD on the whole database is significant (difference of overall precision after full recall 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 situations because it is only able to examine shape in circular direction. However, GFD can successfully deal with this type of situations by examining shape more carefully on the radial directions (Figure 6.10(a)(b), Figure 6.12(a)).
- Perspective deformations can also result in scaling effect, as a result, shape spatial distribution can be significantly changed (Figure 6.11(b) and Figure 6.12(b)(c)). Parts of shape can be lost due to the transform (Figure 6.11(a)). GFD can cope with this type of situations by examining shape features in radial directions.
- Due to the capturing of shape features in both radial and circular directions, the retrieved shapes are more perceptually acceptable. For example, in Figure 6.13, both GFD and ZMD retrieve all the similar shapes to the query. However, GFD not only retrieve those similar shapes, but also retrieve perceptually relevant shapes such as the members in group 1002_p144. Examples from Set B (Figure 6.14) and Set A4 (Figure 6.11(c)) also support that retrievals using GFD are more perceptually acceptable than those using ZMD.
- GFD is more robust than ZMD when the size of the shape database is increased. This is indicated in the retrieval performance on the whole database (Figure 6.9(f)).

The comparative low retrieval performance of GFD on Set B is due to that the grouping within the set is too rough, as can been seen from the example shapes in Figure 5.8. The comparative low retrieval performance of GFD on Set A4 is due to the concentric circular sampling during the feature extraction. This intrinsic problem will be considered in Section 6.7 to increase GFD robustness to affine or perspective deformed shapes which are expected common in nature.

6.6.3.3 Computation Efficiency

The computation complexity 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 polar Fourier transform of (6.7) is simpler than the Zernike moments of (5.8). PFT avoids the complex computation of Zernike polynomials. On average, the feature extraction time for each shape is 693*ms* for GFD and 1194*ms* for ZMD. 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 of GFD features used to index the shape is the same as the number of ZMD features used to index the shape. On average, it takes only 33*ms* for matching two sets of GFDs using city block distance.

(a) Retrieval of query 368_SD_033 using GFD (left) and using ZMD (right).

(b) Retrieval of query 702_SD_033 using GFD (left) and using ZMD (right).

Figure 6.10. Example retrievals on Set A1 of CE-2. There are 5 members in each query group.

(a) Retrieval of query 533_p687_pa_3 using GFD(left) and using ZMD (right)

(b) Retrieval of query 1001_p144_pa_5 using GFD(left) and using ZMD (right)

(c) Retrieval of query 1605 using GFD(left) and using ZMD (right)

Figure 6.11. Example retrievals on Set A4 of CE-2. There are 11 members in each query group.

(a) Retrieval of query 1006 using GFD (left) and using ZMD (right)

(b) Retrieval of query 1006_p144_pa_7 using GFD (left) and using ZMD (right)

(c) Retrieval of query 1004_p144_pa_3 using GFD (left) and using ZMD (right)

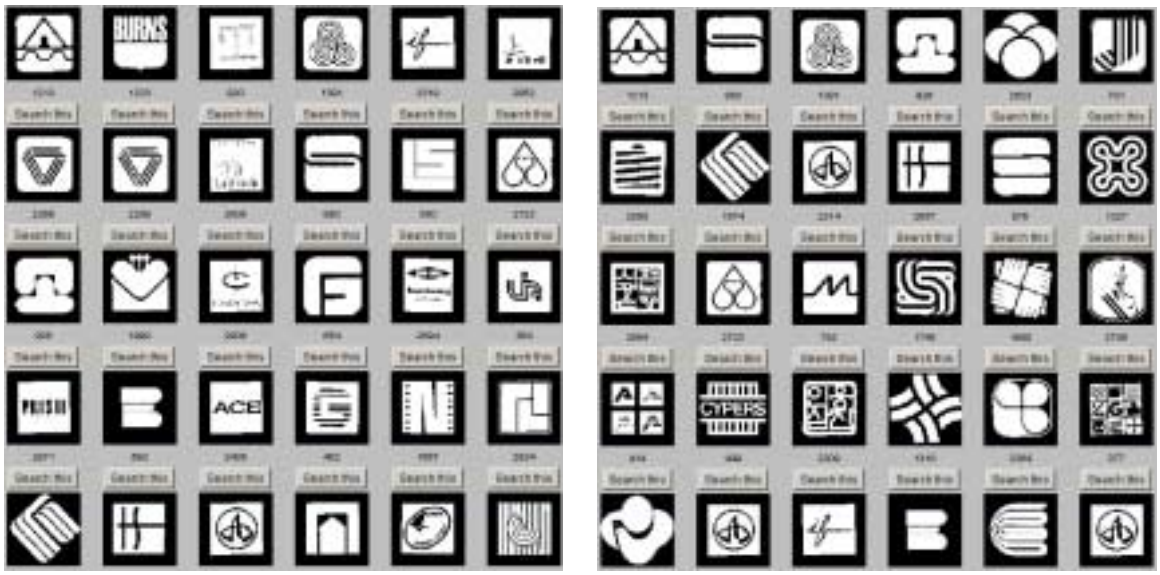
Figure 6.12. Example retrievals on CE-2. There are 21 members in each group.

Retrieval of query 1009 using GFD (left) and using ZMD (right)

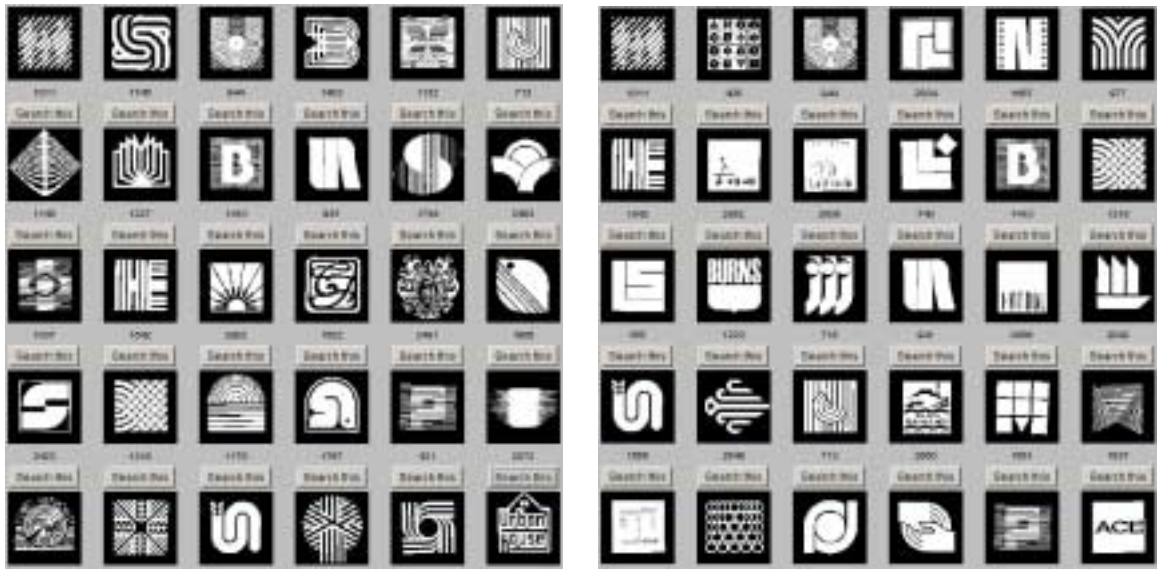
Figure 6.13. Example retrievals on Set A3 of CE-2. There are 11 members in each query group.

(a) Retrieval of 2992 using GFD (left) and using ZMD (right)

(b) Retrieval of 1180 using GFD (left) and using ZMD (right)



(c) Retrieval of 1213 using GFD (left) and using ZMD (right)



(d) Retrieval of 1011using GFD (left) and using ZMD (right)

Figure 6.14. Example retrievals on Set B of CE-2.

6.7 Enhance GFD Robustness to Affine Distortion

In the above, it can be seen that the proposed GFD has a very high retrieval performance on scaled and rotated region-based shapes. Its overall retrieval precision on full recall is 98.6% for rotated shapes (Set A2), 90.5% for scaled shapes (Set A1) and 98.8% (Set A3) for rotated and scaled shapes. However, comparing with rotation and scale invariance test, the retrieval performance on perspective transform test (overall precision is 74.1% on Set A4) and general distortions test (overall precision 80.5% on CE-2) are significantly lower.

The reason for the lower retrieval performance on Set A4 and CE-2 is because when severe skew or stretching occur, the shape region distribution within the circle the shape resides is changed significantly. For example, in Figure 6.15, perceptually, shape (a) and shape (d) are similar. However, due to severe skew, shape (a) only occupies half of the circle it resides while shape (b) fully occupies the circle it resides. When extracting GFD, the shape is scanned circularly on concentric circles. The scanning expands from center towards periphery. When the scanning reaches towards 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, GFD is enhanced to be able to retrieve severe distorted shapes.

6.7.1 Enhancing GFD

To enhance GFD, 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. An optimized *major axis algorithm* (MAA) is used. The MAA is similar to the MAA described in Section 5.4 of Chapter 5, except now the searching of boundary point pairs is started from the bounding circle of the shape. Figure 6.15(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 (consisting 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 the shape into a square of size of 128×128. Figure

6.15(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.

GFD is then extracted from the normalized shape. The acquired GFD is the enhanced GFD (EGFD).

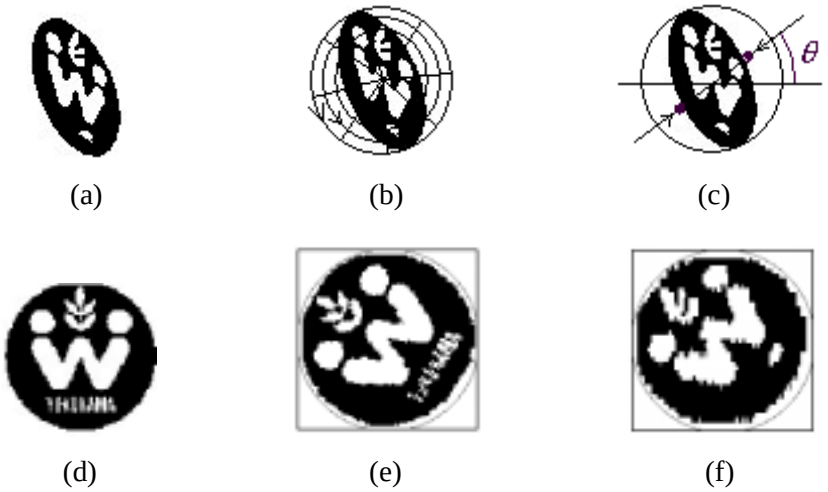
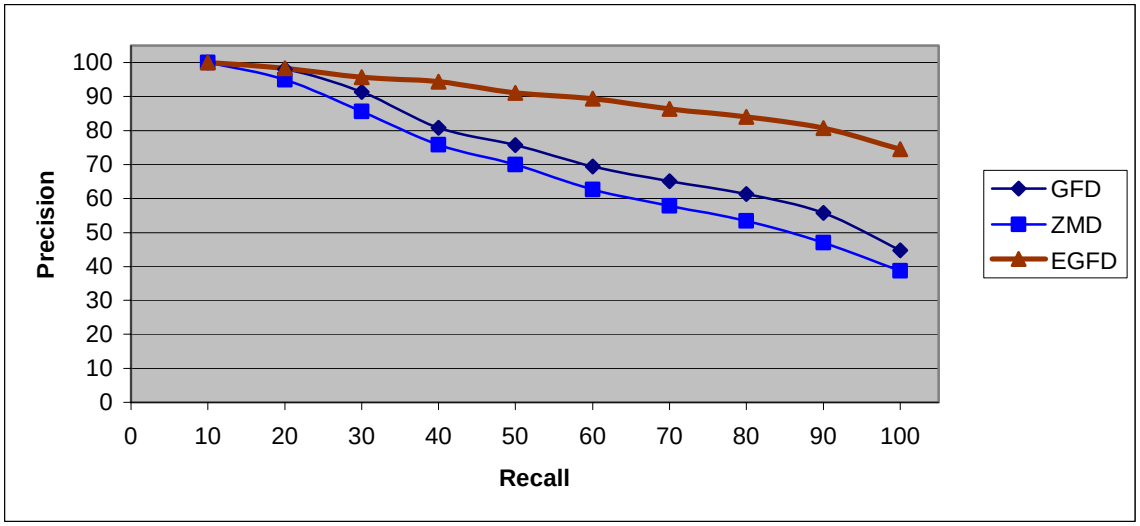


Figure 6.15. (a) A skewed shape; (b) polar scanning of image (a); (c) searching boundary points of shape (a) in one particular direction; (d) a similar shape of (a); (e) normalized shape of (d); (f) normalized shape of (a).

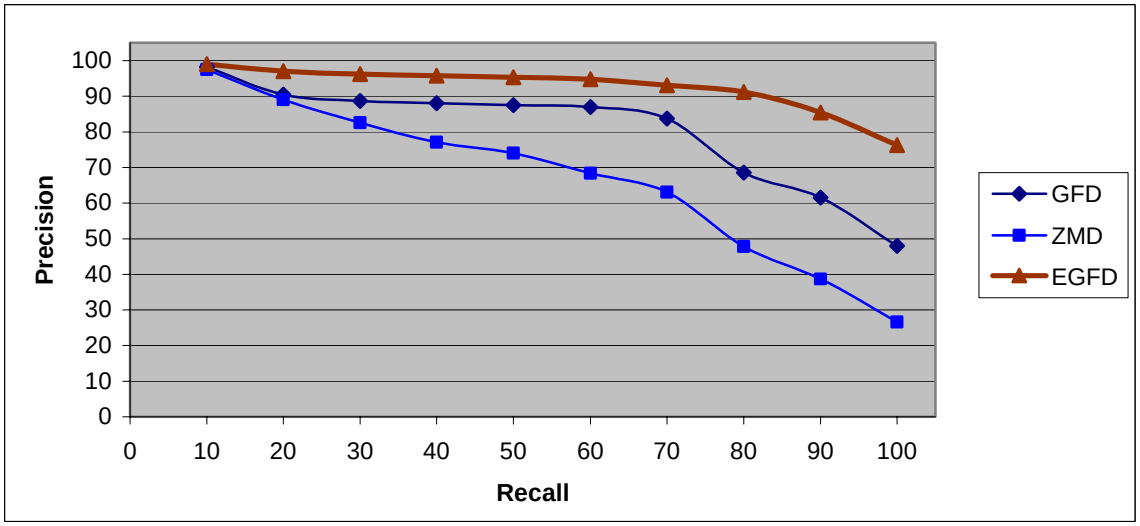
6.7.2 Retrieval Effectiveness of Enhanced GFD

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 retrieval performance measurement is still precision-recall. For each query, the precision of the retrieval at each level of the recall is obtained. The final precision of retrieval is the average precision of all the query retrievals. The average precision and recall of the retrieval using the three types of shape descriptors on Set A4 and CE-2 are shown in Figure 6.16(a)(b).



(a) Average precision-recall of 330 retrievals using GFD, ZMD and EGFD on Set A4 of MPEG-7 region shape database CE-2.



(b) Average precision-recall of 651 retrievals using GFD, ZMD and EGFD MPEG-7 region shape database CE-2.

Figure 6.16. Comparison of retrieval effectiveness of GFD, ZMD and EGFD on Set A4 and CE-2.

It can be observed from Figure 6.16 that on both datasets, EGFD outperforms GFD and ZMD significantly. Comparing with GFD, the improvement on Set A4 is 15.4%, the overall precision after full recall is increased from 74.1% to 89.5%. The improvement on CE-2 is 12%, the overall precision after full recall is increased from 80.5% to 92.5%.

Figure 6.17 shows two cases of retrievals on the two datasets using EGFD, ZMD and GFD respectively. 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 using EGFD is 890*ms* on Windows platform of a PC-III of 866MHZ, while this is 693*ms* for GFD and 1194*ms* for ZMD. The online retrieval time for all the three shape descriptors is the same, since the same number (36) of shape features are used for shape representation.

(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 6.17. 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.

6.7.3 Application of Enhanced GFD

The enhancement above addresses a very common shape variation, that is, perspective transformation. Perspective transformations are expected common in nature due to same objects being viewed from different angles and similar objects being viewed from different angles. The application of the enhancement process, however, is database/application dependent. If the database has abundant perspective shapes, this technique can be very effective in retrieving similar shapes. However, if the database does not have perspective shapes, or the user wants finer distinction between similar shapes, the enhanced process may not be desirable. For example, if the user wants to distinguish between rectangles and squares, or to distinguish between ellipses and circles, the enhanced GFD can fail, because it normalizes all the shapes into same eccentricity (=1). Therefore, in general applications, the enhancement is a useful option to the retrieval system rather than the replacement of GFD.

6.8 Discussions

In the above, the proposed GFD has been studied and evaluated in details. Comparisons of GFD with both contour-based and region-based shape descriptors are given. Results show GFD outperforms 1-D FD, CSSD and ZMD in terms of retrieval effectiveness. 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.

- Good retrieval accuracy. Overall, the GFD has high retrieval performance on most of the test datasets. For contour shapes, the overall retrieval precision on full recall is respectively 99.3% for scaled shapes, 100% for rotated shapes, 63.7% for similarity shapes and 95.8% for affined shapes. For region shapes, the overall retrieval precision on full recall is respectively 90.5% for scaled shapes, 98.6% for rotated shapes, 98.8% for rotated and scaled shapes, 74.1% (89.5% if enhanced) for perspective shapes and 80.5% (92.5% if enhanced) for generally distorted shapes.
- Compact features. The dimension of GFD is 36. Although it is higher than the dimension of 1-D FD, it is relatively low considering its online matching delay is only 33*ms*.
- General applications. GFD is designed for generic shape description, it does not assume any knowledge of shape database.
- Low computation. The computation of GFD is simple, it is simply derived by applying Fourier transform on a radial-angular sampled image. The computation of enhanced GFD

adds computation cost, however due to the use of an optimized major axis algorithm, the computation is relatively efficient.

- Robust retrieval performance. Because GFD is extracted from spectral domain, and due to an enhanced process, it is very robust to noise and severe shape variations. This is demonstrated in the retrieval performance test.
- Hierarchical coarse to fine representation. Like all spectral features, GFD approximate image in a coarse to fine manner. The lower order of GFD features capture shape coarser features, as such, they can be used to filter out very different shapes in the initial retrieval. The higher order GFD features capture shape finer features, they can be added to the lower order GFD features to create a more accurate descriptor, as such, they can be used to refine the retrieval result from the initial retrieval.

It is important to note that for a shape descriptor to be feasible and practical, two key characteristics are essential, that is, stability and clarity.

Stability means when the descriptor is used to describe different types of shapes or applied to different applications, its performance should be least affected. This requires that the derivation of descriptor should depends on as few parameters as possible. This equivalently requires that the derivation process of the descriptor be not complex. Because the more complex the derivation process, the more parameters are involved in obtaining the descriptor; consequently, the descriptor depends on more uncertain or empirical factors. Therefore, when a descriptor derived under delicately controlled conditions is applied to another application or real world applications, its performance is drastically discounted. The many structural methods and shape invariants are among the examples of these types of unstable descriptors. Specific example of instable shape descriptors are CSSD and GD.

Clarity means that a shape descriptor should have a clear perceptual or physical meaning and have a simple interpretation. This characteristic requires not only the modeling process or the feature extraction process be simple, but also requires the normalization be simple. The simple feature advantage can be failed by a complex normalization process. For example, the extraction of geometric moments is very simple and the extracted lower order moments have clear physical meaning. However, the physical meaning of the acquired features has lost due to the complex normalization process. In fact, the physical meaning of the seven geometric moment invariants is not known. This contributes to GMD's poor retrieval performance.

GFD achieves both stability and clarity, because of the following truths:

- GFD is simple. Basically, GFD is acquired by applying 2-D Fourier transform on polar raster sampled image.
- Its normalization process and matching process are also very simple. Normalization is achieved by multiplying the magnitudes of the transformed coefficients with a scalar factor. The matching is simply the city block distance between feature vectors.
- The physical meaning of GFD is clear. Its derivation is based on the well understood Fourier theory. It captures the radial and angular spectral features of the shape.
- It is stable. There is no empirical or database dependent parameters used in the derivation of the descriptor. The only factor needs for determination is the dimension of the descriptor. The dimension is predictable and has been studied. The determination of the dimension can be more easily achieved by a hierarchical manner. That is to say, to match shape roughly using a few low order GFD features and refine the matching result using higher order GFD features.

This two characteristics make it different from most of the existing shape descriptors in literature.

6.9 Summary

In this chapter, a generic Fourier descriptor (GFD) for general shape description and retrieval has been proposed. The new shape descriptor has been tested on both MPEG-7 contour shape database and MPEG-7 region shape database. Compared with existing techniques, the new shape descriptor has the following advantages.

- It improves the conventional 1-D Fourier descriptor (FD) in that: **i**) it does not assume shape contour information which may not be available; **ii**) it captures shape interior content as well as shape boundary features; **iii**) it is more robust in terms of retrieval accuracy.
- It improves Zernike moment descriptor (ZMD) in that: **i**) it captures spectral features in both radial and circular directions, the features in both radial and circular directions are coherent; **ii**) it is simpler to compute; iii) it is more robust and perceptually meaningful in terms of retrieval accuracy.

GFD satisfies the six principles set by MPEG-7. It has two more characteristics, that is, stability and clarity, making it distinguished from most of the shape descriptors in literature. The significance of the proposed GFD is that it outperforms both the curvature scale space descriptor (CSSD) and ZMD, which are adopted as shape descriptors by MPEG-7. The application of GFD to general image retrieval will be given in the next chapter.