Chapter 3

Performance analysis of various local and global descriptors on image retrieval system

3.1. Introduction

An effective image descriptor robust to subject change, partial occluded, distorted, rotated, and noise affected images is essential for pattern recognition and computer vision applications. Therefore, in this chapter, we study and analyze the performance of prominent local and global shape descriptors. A shape descriptor can be dense or discrete. A dense descriptor computes features on all the pixels of the image, while discrete one computes features on a subset of pixels. Thus, we designate dense descriptors as global shape descriptors and discrete descriptors as local shape descriptors. Various global shape descriptors include moment invariants (MI) [47], generic Fourier descriptors (GFD) [61], angular radial transform (ART) [60], wavelet moments (WM) [58], Zernike moments descriptor (ZMD*) [96], etc. These descriptors compute features by considering the entire image as a whole and represent the global aspects of an image. The discrete descriptors include Fourier descriptors (FD) [35], Weber’s local descriptors (WLD) [44], local binary patterns (LBP) [45, 97], local ternary patterns (LTP) [46], contour points distribution histograms (CPDH) [43], etc. These descriptors do not consider entire image as a whole. In fact, only a part of the shape such as boundary or image masks of various dimensions usually 3×3 neighborhood are used for extracting the features. These descriptors provide local characteristics of the shape. An extensive analysis of MPEG-7 shape descriptors has been presented by Zhang and Lu [98], where effectiveness of ZMD and FD is confirmed through experimental results. In [99] two local descriptors FD and curvature scale space, and two global descriptors MI and ART are evaluated against each other. However, comparative analysis of global descriptors MI, GFD, ART, WM, and ZMD and local descriptors FD, WLD, LBP, LTP, and CPDH has not been reported yet for their performance on image retrieval system. It has been investigated in the literature that the performance of WLD, LBP, and LTP is good for texture and face images but not evaluated for distorted, partial occluded, rotated, noisy, and binary images. Therefore, in this chapter, we analyze all mentioned local and global descriptors in terms of six principles set by MPEG-7, which are described in Chapter 1.

*ZMD and ZMs refer to the same descriptor and are used interchangeably in the thesis.
Apart from that, we propose improvement in the retrieval rate of global descriptors ART, GFD, and ZMD by incorporating their respective phase coefficients along with their magnitudes. The phase coefficients carry important information to represent global characteristics of images and in some situations, the important features of images are preserved only if the phase is retained. In addition, the phase coefficients are more robust to non-uniform image intensity fluctuations [55,100]. Using the phase information (together with the magnitude) in the comparison process seems a natural way to improve the similarity measure in terms of robustness against geometric deformation or noise predominantly. As per our knowledge, for image retrieval the inclusion of phase coefficients of ART, GFD, and WM has not been reported yet. Our primary observations include tremendous improvement in retrieving noise-affected images by inclusion of phase coefficients together with magnitude for ART and GFD, which is demonstrated in the experimental section. The major contributions of this chapter are as follows:

- Performance evaluation of the prominent local and global descriptors.
- Improving retrieval performance of global descriptors by incorporating phase coefficients.
- Performance evaluation of various descriptors on various standard databases by keeping in view the principles set by MPEG-7.

### 3.2. Local descriptors

#### 3.2.1. Fourier descriptor (FD)

FD is obtained by applying Fourier transform on a shape signature. The shape signature is a one-dimensional function which is derived from the shape boundary coordinates. Various shape signatures have been used to derive FD such as complex coordinates, polar coordinates, chord length distance, angular function, radial distance or centroid distance, triangular centroid area, triangular area representation, angular radial coordinates, etc. It has been demonstrated by Zhang and Lu [35] that FD derived from the centroid distance function is more effective than FD derived from any other shape signatures.

In the first step of the derivation of FD, the boundary coordinates \((x(u), y(u)); u = 0, 1, 2, ..., N - 1\), are obtained, where \(N\) represents total number of
boundary points. The radial distance between the boundary points \((x(u), y(u))\) and the centroid \((x_c, y_c)\) of the shape is represented as:

\[
r(u) = \sqrt{(x(u) - x_c)^2 + (y(u) - y_c)^2}
\]

The concept of using the centroid \((x_c, y_c)\) is to render the signature invariant to translation. The centroid is computed as follows:

\[
x_c = \frac{1}{N} \sum_{u=0}^{N-1} x(u), \quad y_c = \frac{1}{N} \sum_{u=0}^{N-1} y(u)
\]

The discrete Fourier transform of \(r(u)\) is given as:

\[
a_n = \frac{1}{N} \sum_{u=0}^{N-1} r(u) \exp\left(\frac{-j2\pi u}{N}\right), \quad n = 0, 1, ..., N - 1.
\]

The coefficients \(a_n\) are called the Fourier descriptors of the shape. FD is invariant to rotation, scale, and translation. The rotation invariance of FD is achieved by considering only the magnitude values of the descriptor and ignoring the phase information. Scale invariance is achieved by dividing the magnitude of the first half of descriptor by the DC component, i.e., \(a_0\) or \(FD_0\).

\[
FD_n = \frac{|a_n|}{|a_0|}, \quad n = 0, 1, ..., N/2
\]

\(FD_0\) is generally the largest coefficient and it represents the average energy of the signature, consequently, the normalized descriptor reside in the range \([0,1]\). Only half of the FD coefficients are used to index the shape because centroid function is a real value function and only half of the FD coefficients are distinct [34].

### 3.2.2. Weber’s local descriptor (WLD)

Weber’s local descriptor is based on Weber’s law, which states that the change of a stimulus (such as sound, lighting) that will be just noticeable is a constant ratio of the
original stimulus. When the change is smaller than this constant ratio of the original stimulus, a human being would recognize it as background noise rather than a valid signal. WLD is found to be effective for texture image retrieval and human face detection. Basically, WLD histograms consist of two components: differential excitation and orientation. The differential excitation component is a function of the ratio between two terms: one term is the relative intensity difference of a current pixel against its neighbors, other term is the intensity of the current pixel. The orientation component is the gradient orientation of the current pixel. Consequently, two components are used to construct a concatenated WLD histogram. The computation of the two components differential excitation and orientation is described as follows:

### 3.2.2.1. Differential excitation

The intensity difference between a current pixel and its neighbors are determined in order to find the salient variations within an image. The differential excitation \( \xi(x_c) \) of a current pixel \( x_c \) as shown in Fig. 3.1 is computed as:

\[
\xi(x_c) = \arctan \left( \sum_{i=0}^{N} \left( \frac{x_i - x_c}{x_c} \right) \right)
\]  

(3.5)

\( \xi(x_c) \) may take a minus value if the neighbor intensities are smaller than that of the current pixel. It simulates the case of the surroundings being darker than the current pixel. If \( \xi(x_c) \) is positive, it represents the case when surroundings are lighter than the current pixel. The range of \( \xi(x_c) \) is \([-\pi/2, +\pi/2]\) and is divided into \( M \) intervals \( l_m (m=0,1,\ldots,M-1) \). Thus, for each interval \( l_m \), we have \( l_m= [\eta_{m,l}, \eta_{m,u}] \). The lower bound \( \eta_{m,l} = (m/M - 1/2)\pi \) and the upper bound \( \eta_{m,u} = ((m+1)/M - 1/2)\pi \) for example \( l_0 = [-\pi/2, -\pi/3] \).

| \( x_1 \) | \( x_2 \) | \( x_3 \) |
| \( x_8 \) | \( x_5 \) | \( x_4 \) |
| \( x_7 \) | \( x_6 \) | \( x_5 \) |

**Fig. 3.1 8×8 neighborhood**
3.2.2.2. Orientation

The orientation component of WLD is the gradient orientation, which is computed as:

$$\theta(x_c) = \arctan \left( \frac{x_6 - x_2}{x_8 - x_4} \right)$$

(3.6)

or

$$\theta(x_c) = \arctan \left( \frac{v_{ij}^{11}}{v_{ij}^{10}} \right)$$

(3.7)

where $v_{ij}^{11} = x_6 - x_2, v_{ij}^{10} = x_8 - x_4$ and $\theta$ is further quantized into $T$ dominant orientations. Before quantization, the following mapping is performed: $f : \theta \rightarrow \hat{\theta}$.

$$\hat{\theta} = \arctan 2(v_{ij}^{11}, v_{ij}^{10}) + \pi$$

(3.8)

$$\arctan 2(v_{ij}^{11}, v_{ij}^{10}) = \begin{cases} 
\theta, & v_{ij}^{11} > 0 \text{ and } v_{ij}^{10} > 0 \\
\pi + \theta, & v_{ij}^{11} > 0 \text{ and } v_{ij}^{10} < 0 \\
\theta - \pi, & v_{ij}^{11} < 0 \text{ and } v_{ij}^{10} < 0 \\
\theta, & v_{ij}^{11} < 0 \text{ and } v_{ij}^{10} > 0 
\end{cases}$$

(3.9)

where $\theta \in [-\pi/2, \pi/2]$ and $\hat{\theta} \in [0, 2\pi]$. The quantization function is given as:

$$\Phi_t = \frac{2t}{T} \pi, t = \text{mod} \left( \frac{\hat{\theta}}{2\pi/T} + \frac{1}{2}, T \right)$$

(3.10)

e.g. for $T = 8$, the dominant orientations are $\Phi_t = \frac{t\pi}{T}, (t = 0, 1, \ldots, T - 1)$. Thus, the computed orientations are divided into $T$ bins. The complete procedure for building WLD histogram can be found in [44]. In our experiments, we use $M = 6$ and $T = 8$ bins for WLD histogram.

3.2.3. Local binary pattern (LBP)

The original LBP operator, introduced by Ojala et al. [45] is a powerful means of texture description. The operator labels the pixels of an image by thresholding the neighborhood of each pixel with the centre value and considering the result as a binary
number. Then the histogram of the labels can be used as a texture description.

Mathematically, LBP operator takes the following form:

\[
LBP(x_c, y_c) = \sum_{n=0}^{7} 2^n s(n_i - i_c),
\]

(3.11)

where

\[
s(n_i - i_c) = \begin{cases} 
1, & n_i \geq i_c \\
0, & \text{otherwise}
\end{cases}
\]

(3.12)

where \( n \) runs over the 8 neighbors of the central pixel \( c \) and \( i_c \) and \( i_n \) are the grey level values at pixels \( c \) and \( n \), respectively. An illustration of LBP coding process is given in Fig. 3.2.

Thus, after obtaining the binary codes corresponding to each pixel, its decimal value is computed which varies from 0-255. The binary code contains 8 bits, consequently the dimension of the histogram is 256 bins corresponding to each binary pattern. This dimensionality can be reduced by using the uniform patterns as illustrated by Moore and Bowden [97]. Thus, uniform patterns contain at most two bitwise transitions from 0 to 1 or vice versa for circular binary string. The uniform patterns contain in total \((p-1)p+2\) binary patterns. Therefore, by using uniform patterns for a neighbor where \( p=8 \), reduces the histogram dimensionality from 256 to 59 bins (58 for uniform patterns and 1 for non uniform pattern). However, the performance of LBP exhibiting histograms of 256 bins or 59 bins is almost similar [97].

3.2.4. Local ternary pattern (LTP)

LBP has been proven to be highly discriminative feature for texture classification. However, LBP thresholds exactly at the value of the central pixel \( c \). Therefore, they tend to be sensitive to noise, particularly in non uniform regions. In LTP, LBP is extended to 3 valued code in which gray level in a zone of width \( \pm t \) around \( i_c \) are quantized to zero,
ones above this are quantized to +1 and ones below it are quantized to -1, i.e., the function \( s(i_n - i_c) \) given by Eq. (3.12) is replaced with a 3 valued function [46]:

\[
s'(i_n - i_c, t) = \begin{cases} 
+1, & i_n \geq i_c + t \\
0, & |i_n - i_c| < t, \\
-1, & i_n \leq i_c - t 
\end{cases}
\]  
(3.13)

and the binary LBP code is replaced with ternary LTP code. Here \( t \) is a user specified threshold and we set \( t = 5 \) in our experiments. An illustration of LTP encoding is given in Fig. 3.3.

The generated positive and negative codes of LTP can be split into negative and positive halves as represented in Fig. 3.4. For creating negative half, ones are substituted at the places of -1 in ternary pattern and rests of the codes are set to 0. Likewise, for creating positive half, ones are substituted at the places of 1 in ternary pattern and rests of the codes are set to 0. Subsequently, by treating these as two separate channels of LBP descriptor, separate histograms and similarity values are computed.
3.2.5. Contour point distribution histogram (CPDH)

CPDH is based on the distribution of points on object contour under polar coordinates. In this algorithm, the object boundary is detected using standard Canny operator to describe its shape. The resultant points on the contour can be presented as $P = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}, (x_i, y_i) \in R^2$, where $n$ represents the total number of points on contour. After the extraction of points on object boundary the centroid of the boundary image is computed. The maximum value of distance between the centroid and the boundary points on the contour is chosen as the radius of the minimum circumscribed circle. The region of the minimum circumscribed circle is divided into several bins using some concentric circles and equal interval angles. Each bin of the minimum circumscribed circle can be described as triplet $H_i = (\rho_i, \theta_i, n_i)$, where $\rho_i$ denotes the radius of the concentric circle, $\theta_i$ denotes the angle space and $n_i$ denotes the number of points located in bin $r_i$. The summarized algorithm for the construction of CPDH is given as:

Step 1: Input a binary shape image.

Step 2: Extract object contour points with Canny operator.

Step 3: Sample out $N$ points on the contour with $x$ and $y$ coordinates:

$$P = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}, (x_i, y_i) \in R^2.$$  

Step 4: Compute the centroid of the shape $(x_c, y_c)$.

Step 5: Set the centroid as the origin and translate $P$ into polar coordinates, $P = \{(\rho_1, \theta_1), (\rho_2, \theta_2), \ldots, (\rho_n, \theta_n)\}, (\rho_i, \theta_i) \in R^2$, where $\rho_i = \sqrt{(x_i-x_c)^2 + (y_i-y_c)^2}$ is the distance between the points $(x_i, y_i)$ and $(x_c, y_c)$, $\theta_i = \arctan((y_i - y_c)/(x_i - x_c))$ is the angle between $\rho_i$ and $x$-axis.

Step 6: Get the minimum circumscribed circle $C$ with the centre $(x_c, y_c)$ and radius $\rho_{\text{max}}$, where $\rho_{\text{max}} = \max\{\rho_i\}, i = 1, 2, \ldots, n$.

Step 7: Partition the area of $C$ into $u \times v$ bins with $u$ bins for $\rho_{\text{max}}$ and $v$ bins for $\theta$. 
Step 8: Construct the CPDH of the shape image by counting the number of points which are located in every bin.

Step 9: Output CPDH.

Invariance to scale and translation are the intrinsic properties of CPDH and the invariance to rotation is partially resolved during the matching process by using circular shift operation. The complete details of the similarity matching process can be found in [43].

3.3. Global descriptors

3.3.1. Moment invariants (MI)

Geometric moment invariants (MI) are one of the earliest and widely used moment based descriptors. The \((p+q)th\) order of an intensity or gradient image \(f(r, \theta)\) is defined as follows:

\[
m_{pq} = \int_0^1 \int_0^\pi r^p \theta^q f(r, \theta) \, r \, dr \, d\theta, \quad p = q = 0, 1, 2, ...
\] (3.14)

Based on the geometric moments, Hu [47] derived a set of moment invariants from the non linear combination of geometric moments to achieve affine invariance. For the second and third order moments, we have the following seven orthogonal invariants:

\[
\phi_1 = m_{20} + m_{02} \\
\phi_2 = (m_{20} - m_{02})^2 + 4m_{11}^2 \\
\phi_3 = (m_{20} - 3m_{12})^2 + (3m_{21} - m_{03})^2 \\
\phi_4 = (m_{30} - m_{21})^2 + (m_{12} - m_{03})^2 \\
\phi_5 = (m_{20} - 3m_{12}) + (m_{30} + m_{12}) \left[(m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2\right] \\
\quad + (3m_{21} + m_{03})(m_{21} + m_{03}) \left[3(m_{30} + m_{12})^2 - (m_{21} - m_{03})^2\right] \\
\phi_6 = (m_{20} - m_{02})[(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2] + 4m_{11}(m_{30} - m_{12})(m_{21} - m_{03}) \\
\phi_7 = (3m_{21} - m_{03})(m_{30} + m_{12})[(m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2] \\
\quad + (3m_{12} - m_{30})(m_{21} + m_{03})[3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2]
\]
where $\phi_i$ is the skew orthogonal moment. This set of moments is invariant to translation, scale, mirroring (with a negative sign) and rotation.

### 3.3.2. Generic Fourier descriptor (GFD)

GFD is based on Fourier transform (FT). Shape analysis using FT is supported by well developed Fourier theory. However, it is not desirable to acquire shape features using FT directly, because the acquired features are not compact and rotation invariant. Therefore, a modified polar FT (PFT) is derived by treating the polar image in polar space as a normal two dimensional rectangular image in Cartesian space. If 2DFT is applied on a rectangular image, the PFT has the similar form to the conventional 2D discrete FT in Cartesian space. Consequently, the modified PFT is defined as:

$$ PF(\rho, \phi) = \sum_{r} \sum_{i} f(r, \theta_i) \exp \left[-j2\pi \left(\frac{r}{R} \rho + \frac{2\pi}{T} \phi\right)\right], \quad (3.16) $$

where $0 \leq r < R$ and $\theta_i = i(2\pi/T)(0 \leq i < T); 0 \leq \rho < R, 0 \leq \phi < T$. $R$ and $T$ are the radial and angular frequencies. The acquired polar Fourier coefficients are translation invariant because the maximum radius of the bounding circle is computed from the centroid of the shape. Rotation and scale invariance are acquired by the following normalization [61]:

$$ GFD = \left[ \begin{array}{c} \frac{PF(\rho, \phi)}{PF(0,0)} \end{array} \right], \quad 0 \leq \rho < R, 0 \leq \phi < T \quad (3.17) $$

The similarity between two shapes is measured by Euclidean distance between the two feature vectors of the shapes. In our experiments, we use 36 GFD features based on $R = 4$ radial and $T = 9$ angular frequencies as suggested by Ref. [101].

### 3.3.3. Angular radial transform (ART)

ART is a transform based image descriptor, which is adopted by MPEG-7 as a region based descriptor [60]. ART provides an efficient and compact solution to pixel
spatial distribution within an object region. It is well suited to describe both connected and disconnected regions of the shape. Mathematically, ART coefficients are defined as [60]:

\[ F_{pq} = \frac{1}{2\pi} \int_{0}^{\pi} \int_{0}^{2\pi} f(r, \theta) V_{pq}^* (r, \theta) r \, dr \, d\theta, \]

(3.18)

where \( f(r, \theta) \) is an image intensity function in polar co-ordinates, and \( V_{pq}^* (r, \theta) \) is the kernel function of ART, which is the complex conjugate of \( V_{pq} (r, \theta) \) defined by

\[ V_{pq} (r, \theta) = R_p (r) e^{i \theta}, \]

(3.19)

\[ R_p (r) = \begin{cases} 1 & p = 0 \\ 2 \cos (\pi pr) & p > 0 \end{cases}, \]

(3.20)

where \( p = 0, 1, 2, \ldots \), and \( q = 0, 1, 2, \ldots \), are the order and repetition of ART, respectively. As per MPEG-7 standard, ART is expressed by 140 bits, consisting of 35 coefficients having 4 bits each. Thus, in our solution, we use ART coefficients with order \( p < 3 \) and repetition \( q < 12 \). Rotation invariance is achieved by using the magnitude of the coefficients. For scale invariance, the coefficients of ART are divided by the magnitude of the first coefficient of order \( (p = 0, q = 0) \), i.e., by \( |F_{00}| \). The translation invariance is achieved by considering the centre of mass \( (x_c, y_c) \) as the centre of the image, which is derived by the geometric moments as follows:

\[ x_c = \frac{m_{10}}{m_{00}}, \quad y_c = \frac{m_{01}}{m_{00}}, \]

(3.21)

where \( m_{10}, m_{01} \) and \( m_{00} \) are the geometric moments of \( f(r, \theta) \), derived using Eq. (3.14).

3.3.4. Wavelet moments (WM)

Wavelet moments invariants combine two sorts of features viz. moment invariants features and wavelet features of 2D shapes. Wavelet moments are based on wavelet transform. Wavelet transform is a method for accomplishing localized analysis [102-104]. The characteristic of wavelet transform is particularly suited to extracting local features. However, by utilizing wavelet transform in terms of moments, it provides global aspects
of the image. Mathematically, wavelet moments for an image intensity function \( f(r, \theta) \) are defined as:

\[
W_{mnq} = \int_0^{2\pi} \int_0^1 f(r, \theta) e^{i\theta} \psi_{mn}(r) r \, dr,
\]

(3.22)

and

\[
\psi_{mn}(r) = 2^{m/2} \psi(2^m r - 0.5n)
\]

(3.23)

Wavelet basis function is defined as:

\[
\psi^{a,b}(r) = \frac{1}{\sqrt{a}} 2^{m/2} \psi\left(\frac{r-b}{a}\right),
\]

(3.24)

where \( a \) is a dilation parameter and \( b \) is a shifting parameter. In our experiments, we consider the cubic \( B - \text{spline} \) wavelets [105], which are optimally localized in space frequency and are close to Zernike polynomials. Therefore, the mother wavelet \( \psi(\rho) \) of the cubic \( B - \text{spline} \) in Gaussian approximation form is [105]:

\[
\psi(r) = \frac{4a^{n+1}}{\sqrt{2\pi}(n+1)} \sigma_n \cos(2\pi f_0 (2r - 1)) \times \exp\left\{-\frac{(2r-1)^2}{2\sigma_n^2(n+1)}\right\},
\]

(3.25)

where \( n = 3, a = 0.697066, f_0 = 0.409177, \) and \( \sigma_n^2 = 0.561145. \) The domain of \( m \) and \( n \) are restricted as:

\[
a = 0.5^m \quad m = 0, 1, 2, 3
\]
\[
b = 0.5n.0.5^n \quad n = 0, 1, ..., 2^{m+1}
\]

(3.26)

Using different scale index \( m \) and shift index \( n \), wavelets cover the whole radial domain. Thus, the wavelet moment invariants \( W_{mnq} \) provide features of the object \( f(r, \theta) \) at different scales.
3.3.5. Zernike moments descriptor (ZMD)

The set of orthogonal ZMs for an image intensity function \( f(r, \theta) \) with order \( p \) and repetition \( q \) are defined over a continuous unit disk \( 0 \leq r \leq 1, 0 \leq \theta < 2\pi \) [30] as:

\[
Z_{pq} = \frac{p+1}{\pi} \int_0^1 \int_0^{2\pi} f(r, \theta) V_{pq}^*(r, \theta) r dr d\theta,
\]

(3.27)

where \( V_{pq}^*(r, \theta) \) is the complex conjugate of the Zernike polynomials \( V_{pq}(r, \theta) \), defined as:

\[
V_{pq}(r, \theta) = R_{pq}(r) e^{i\theta},
\]

(3.28)

where \( p \geq 0, 0 \leq |q| \leq p, p - |q| = \text{even}, j = \sqrt{-1}, \text{and } \theta = \tan^{-1}(y/x). \)

The radial polynomials \( R_{pq}(r) \) are defined by:

\[
R_{pq}(r) = \sum_{k=0}^{(p-|q|)/2} (-1)^k \frac{(p-k)!}{k! \left( \frac{p+|q|}{2} - k \right)! \left( \frac{p-|q|}{2} - k \right)!} r^{p-2k},
\]

(3.29)

The radial polynomials satisfy the orthogonality relation:

\[
\int_0^1 R_{pq}(r) R_{p'q'}(r) r dr = \frac{1}{2(p+1)} \delta_{pp'},
\]

(3.30)

where \( \delta_{ij} \) is Kronecker delta. The set of Zernike polynomials \( V_{pq}(r, \theta) \) form a complete orthogonal set within the unit disc as:

\[
\int_0^{2\pi} \int_0^1 V_{pq}(r, \theta) V_{p'q'}^*(r, \theta) r dr d\theta = \frac{\pi}{p+1} \delta_{pp'} \delta_{qq'},
\]

(3.31)
In our experiments, we use moment order $p_{\text{max}} = 12$ as an appropriate order for feature selection. The moment order $p_{\text{max}} = 12$ is considered to be a good tradeoff between the computation complexity and image description capability [55], which provides 47 moment features. The moments $Z_{0,0}$ and $Z_{1,1}$ are excluded from the features set as $Z_{0,0}$ signifies an average gray value of image and $Z_{1,1}$ is the first order moment, which is zero if the centroid of the image falls on the centre of the disc. A small number of features do not provide satisfactory results, while the high numbers of features are prone to “overtraining” and reduce the computation efficiency. In addition, higher order ZMs suffer from numerical integration error and numerical instability.

3.4. Improvement in GFD, ART, WM, and ZMD

It has been studied in the literature [60,61,96] that only the magnitude components of a descriptor are considered because they exhibit rotation invariance property. Accordingly, the phase coefficients of GFD, ART, WM, and ZMD are excluded from the features set. Nevertheless, it is observed that phase coefficients carry important information about the image. The phase coefficients are more robust to non uniform image intensity fluctuations [55,100] and they play important role in image reconstruction. In addition, if the phase information is used with magnitude, it significantly improves the retrieval rate in terms of robustness against geometric deformation or noise predominantly. Keeping this in view, recently several researchers incorporated the phase information in representing the image [53-55]. The phase coefficients of unrotated image and rotated query image are related as:

\[
\psi'_{pq} = \psi_{pq} - q\alpha 
\]

or

\[
q\alpha = \psi_{pq} - \psi'_{pq},
\]

where

\[
\psi_{pq} = \tan^{-1}\left(\frac{I_{pq}}{R_{pq}}\right),
\]

where $\psi'_{pq}$ and $\psi_{pq}$ are the phase coefficients of the rotated and unrotated image respectively, $I_{pq}$ and $R_{pq}$ represent respectively, the real and imaginary parts of a phase
coefficients for GFD, ART, WM, and ZMD. The acquired phase coefficients of the query image are compared with that of database images by using the following similarity measure:

\[
D_\psi = \sum_p \sum_q \min \left\{ \frac{|q \alpha - (2\pi - q \alpha)|}{\pi} \right\},
\]

where \( q \alpha = \psi^O - \psi^D \), \( \psi^O \) and \( \psi^D \) represent the phase coefficients of query and database images, respectively. The expression given by Eq. (3.35) provides normalized phase based distance. These distances are combined with the normalized Euclidean distance measure for magnitude components of GFD, ART, and ZMD to obtain the improved retrieval results.

### 3.5. Experimental results and performance evaluation

For image matching process, we use Euclidean distance measure for FD, MI, GFD, ART, WM, and ZMD based features. Since other methods WLD, LBP, LTP, and CPDH, are histogram based, we use histogram intersection similarity measure [86] for comparing the histograms of query and database images. In addition, for assessing the complete image retrieval performance, all the images in the database are served as query. In order to compare the performance of various local and global descriptors, we conduct five sorts of tests. In the first test, the performance is analyzed for partial occluded and distorted shapes. In the second test, the performance is evaluated for subject changed images, i.e., for a database containing high variability among images of same class. Third test scrutinizes the performance for 3D objects. Fourth test is examined for rotation invariance and the fifth test is conducted to assess the performance for robustness to noise. For all these tests, we use different databases, which are Kimia-99, MPEG-7 CE shape-1 part B, COIL-100, Rotation, and Noise (the description of all these databases is given Appendix A). The retrieval accuracy of the system is measured in terms of precision \((P)\) and recall \((R)\) rates.
Fig. 3.5 $P-R$ performance of (a) GFD (b) ART (c) WM (d) ZMD and (e) all four by using magnitude only, phase only, and combining both phase and magnitude features
3.5.1. Performance comparison of image retrieval accuracy

3.5.1.1. Retrieval analysis for phase and magnitude based global features

Here, we present the retrieval effectiveness of the phase based features against magnitude based features for GFD, ART, WM, and ZMD in terms of $P-R$ curves for MPEG-7 CE shape-1 part B database. In Fig. 3.5(a) $P-R$ curves for GFD are presented, which demonstrates that the performance of phase based coefficients and magnitude based coefficients is similar and their $P-R$ curves overlap with each other. However, by combining both phase and magnitude based global features the retrieval rate significantly improves. Similarly, it can be seen from Fig. 3.5(b) for ART that by combining both phase and magnitude based features the retrieval effectiveness improves again. From Fig. 3.5(c) and Fig. 3.5(d), it is apparent that the performance of WM and ZMD also improves considerably by amalgamating both phase and magnitude based features. While comparing GFD, ART, WM, and ZMD against each other by combining both phase and magnitude components, we observe that ZMD outperforms ART, GFD, and WM, in order as illustrated in Fig. 3.5(e). Hence, we observe that by combining both phase and magnitude based features the performance of system improves significantly. Thus, in the rest of the experiments, we compare combined phase and magnitude based features of global descriptors GFD, ART, WM, and ZMD, for all the databases rather than comparing magnitude only features.

3.5.1.2. Comparison of local and global descriptors autonomously

As mentioned earlier, the retrieval accuracy of the system is evaluated in terms of precision and recall. Hence, here we present the $P-R$ curves for local descriptors and global descriptors. In the first test, we analyze the performance for Kimia-99 database and the $P-R$ curves for local descriptors are given in Fig. 3.6(a). It is observed that the performance of WLD is poorer than other methods. The performance of LTP is superior to other methods, followed by FD, CPDH, and LBP. It is mentioned here that the performance of LTP is better than FD for higher values of recall and the performance of FD is better than LTP for lower values of recall. While evaluating the performance of global descriptors for Kimia-99 database as shown in Fig. 3.6(b), we see that MI gives the worst performance among other methods. However, ZMD (mag+phase) outperforms rest of the methods by giving quite high retrieval accuracy followed by GFD (mag+phase),
WM (mag+phase), and ART (mag+phase). The performance of WM and ART is similar and their $P-R$ curves coincide. The next test is performed for MPEG-7 database, in which large variations exist among the instances of a class. The $P-R$ curves for local descriptors are presented in Fig. 3.7(a), which demonstrates that WLD has the worst performance and the performance of LTP is slightly superior to that of LBP. FD performs well but the performance of CPDH is superior to all the descriptors. In Fig. 3.7(b), the performance of global descriptors is presented for MPEG-7 database. We see that ZMD (mag+phase) gives the best performance among others followed by ART (mag+phase), GFD (mag+phase), WM (mag+phase), and MI. The performance of local descriptors for COIL-100 database is given in Fig. 3.8(a). We see that the worst performance is given by FD and WLD performs moderately. The performance of LBP, LTP, and CPDH is good and provide almost similar results. While evaluating the performance of global descriptors for COIL-100 database as shown in Fig 3.8(b), we observe that MI gives poor performance. The performance of rest of the methods is almost similar and their $P-R$ curves overlap with each other, in which ZMD supersedes others.

The analysis of local descriptors for the rotation invariance is presented in Fig. 3.9(a). It is observed that FD is invariant to rotation and give high retrieval accuracy as compared to other methods. It is perceived that LTP is better than LBP for rotated images. The performance of WLD is moderate for rotation invariance test. However, LBP and CPDH give poor performance for rotation invariance test. While considering the $P-R$ performance of global descriptors as shown in Fig. 3.9(b), we observe that ZMD(mag+phase), WM(mag+phase), and ART(mag+phase) possess the highest retrieval accuracy for rotation invariance. The performance of MI is slightly lower than ZMD (mag+phase) and ART (mag+phase), whereas GFD (mag+phase) gives the worst performance amongst all. The next analysis is performed for examining the robustness to noise. The $P-R$ curves for the Noise database are given in Fig. 3.10(a) for local descriptors. It is seen that FD gives the worst performance for noise affected images. The performance of LTP is superior to that of LBP, WLD, and CPDH in order. The similar test is performed for evaluating the performance of global descriptors and the $P-R$ curves are given in Fig. 3.10(b). The figure demonstrates that by combining both phase and magnitude based features, tremendous improvement in the performance ZMD, GFD, and ART is observed. ZMD (mag+phase), WM(mag+phase), and ART (mag+phase) confer very high retrieval accuracy and their $P-R$ curves overlap with each other. On the other
hand, MI provides the worst performance again. The performance of ART is superior to GFD for noise test.

Fig. 3.6 Comparison of the $P - R$ performance for Kimia-99 database for (a) local descriptors (b) global descriptors

Fig. 3.7 Comparison of the $P - R$ performance for MPEG-7 database for (a) local descriptors (b) global descriptors
Fig. 3.8 Comparison of the $P - R$ performance for COIL-100 database for (a) local descriptors (b) global descriptors

Fig. 3.9 Comparison of the $P - R$ performance for Rotation database for (a) local descriptors (b) global descriptors
3.5.2. Comparison of computation time

To study the computation time of the local and global shape descriptors for extracting the features for offline and online matching, we consider MPEG-7 database, which contains 1400 images. All the images of this database are used to compute the average feature extraction time. The CPU elapsed time is given in Table 1 for both local and global descriptors. We observe that the local descriptor WLD takes the least amount of time followed by FD and LBP, which take similar amount of time. On the other hand, in global descriptors MI takes the least and WM takes the highest amount of CPU time. Therefore, we see that in local descriptors WLD and in global descriptors MI are the most efficient methods.

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>Offline (1400 images)</th>
<th>Online (1 image)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Local descriptors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD</td>
<td>29.68</td>
<td>0.021</td>
</tr>
<tr>
<td>WLD</td>
<td>19.74</td>
<td>0.006</td>
</tr>
<tr>
<td>LBP</td>
<td>29.68</td>
<td>0.021</td>
</tr>
<tr>
<td>LTP</td>
<td>32.51</td>
<td>0.023</td>
</tr>
<tr>
<td>CPDH</td>
<td>39.48</td>
<td>0.027</td>
</tr>
<tr>
<td><strong>Global descriptors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>48.63</td>
<td>0.069</td>
</tr>
<tr>
<td>GFD</td>
<td>653.46</td>
<td>0.466</td>
</tr>
<tr>
<td>ART</td>
<td>289.22</td>
<td>0.109</td>
</tr>
<tr>
<td>WM</td>
<td>4789.2</td>
<td>3.782</td>
</tr>
<tr>
<td>ZMD</td>
<td>375.3</td>
<td>0.268</td>
</tr>
</tbody>
</table>

Table 3.1 Average CPU elapsed time in seconds for offline and online feature extraction using the global and local descriptors
3.6. Discussion

In the previous sections, prominent local descriptors viz. FD, WLD, LBP, LTP, and CPDH and global descriptors MI, GFD, ART, WM, and ZMD are studied in detail. The experiments are performed on several sorts of images. The performance of local descriptors and global descriptors is analyzed on image retrieval system autonomously. Among the local descriptors LTP, FD, and CPDH and among the global descriptors ZMD (mag+phase) are found to be the most effective descriptors. Here in this section, we discuss their performance in terms of principles set by MPEG-7, which is described as follows:

- Good retrieval accuracy: While comparing the retrieval accuracy of local descriptors, we see that LTP outperforms rest of the descriptors for partial occluded, distorted, 3D objects, and noise affected shapes while for rotation invariance FD supersedes others. CPDH and LBP also perform well. However, we see that the performance of LTP is always superior to that of LBP and for noisy images, LTP performs far better than LBP. While considering the performance of global descriptors, we observe that the proposed ZMD, which includes both magnitude and phase based features overpowers other global descriptors because of the incorporation of sinusoid function in their radial kernel, and they possess similar properties of spectral features. The performance of ART is comparable to that of ZMD. MI gives poor image retrieval performance except for rotation invariance test. It is observed that moment invariants derived from lower orders of moments are not sufficient enough to accurately describe the complete essence of a shape. The performance of GFD and WM is moderate for all kinds of images. However, WM performs superior to GFD for Rotation and Noise tests. Nevertheless, it is observed that by incorporating the phase coefficients along with magnitude of GFD, ART, WM, and ZMD, the performance of the system improves significantly as compared to considering only magnitude based features.

- Compact features: The number of features to be used for local descriptors FD, WLD, LBP, LTP, and CPDH are 10, 48, 256 (59 for uniform patterns), twice 256 (positive and negative) (59 for uniform patterns), and 72, respectively. We observe that FD provides compact features along with good retrieval accuracy. On the other side, for global descriptors MI, GFD, ART, WM, and ZMD, the feature dimensionality is 7, 36, 36, 136, and 47, respectively. It can be seen that MI
provides the smallest number of features and WM provides the largest one. However, while considering the retrieval accuracy, ZMD is superior in performance as compared to MI. Therefore, ZMD can be considered as a good feature extractor. Feature dimensionality of ART and GFD is the same. However, ART overpowers WM and GFD in its retrieval performance.

- General applications: The performance of local descriptors CPDH and LBP is poor for rotated and noisy images, where FD performs poor for noise affected images. Therefore, we observe that LBP is not generic enough to recognize geometric and photometric transformed images. The same is true for FD in case of photometric transformed images. The rest of the local descriptors can be applied for general image retrieval. On the other hand, by analyzing the $P-R$ performance of global descriptors, we can see that MI is not designated to be generic as it performs poor for noisy, subject changed and distorted shapes. In other methods, ZMD supersedes them and can be termed as a generic descriptor.

- Low computation complexity: In local descriptors, WLD takes the least amount of CPU time, whereas FD and LBP take the second place. Therefore, if speed is concerned then FD can be used as effective local descriptor since it also provides good retrieval effectiveness. In global descriptors, MI takes the least amount of CPU time, whereas WM takes the highest amount of CPU time. Therefore, if speed is the issue then ART can be considered as a good image descriptor. It is worth to mention here that various algorithms are available to solve the issue of computation complexity of ZMD [81], which significantly reduce the CPU elapsed time. By using them, ZMD can be considered as an effective global image descriptor because it also provides good retrieval accuracy against rest of the methods.

- Robust retrieval performance: While comparing the robustness of local descriptors, it is perceived that LTP outperforms others and FD gives the worst performance. In case of global descriptors, ZMD, WM, and ART surpass to rest of the methods and the performance of MI is drastically reduced.

- Hierarchical coarse to fine representation: All the mentioned descriptors support hierarchical coarse to fine representation, which make them suitable for refining images from coarse level to finer details.
3.7. Conclusion

In this chapter, we compare a few of the best local descriptors viz. FD, WLD, LBP, LTP and CPDH for evaluating their performance on image retrieval system. For this purpose, we have considered various sorts of images representing several aspects of image shape. An extensive set of experiments are performed and it is observed that FD, LTP, and CPDH can be recommended as good local descriptors. The prominent global descriptors viz. MI, GFD, ART, WM, and ZMD are also compared against each other in terms of their image retrieval performance. Besides, an improvement in the performance of GFD, ART, WM, and ZMD is proposed by including phase coefficients along with their magnitude components, which represent high improvement principally for noise affected images. It is also observed that ZMD(mag+phase) surpasses to rest of the global descriptors for various types of images.