Multi-Technique Fusion for Shape-Based Image Retrieval

by

Akrem Saad El-ghazal

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Akrem Saad El-ghazal

Abstract

Content-based image retrieval (CBIR) is still in its early stages, although several attempts have been made to solve or minimize challenges associated with it. CBIR techniques use such visual contents as color, texture, and shape to represent and index images. Of these, shapes contain richer information than color or texture. However, retrieval based on shape contents remains more difficult than that based on color or texture due to the diversity of shapes and the natural occurrence of shape transformations such as deformation, scaling and orientation.

This thesis presents an approach for fusing several shape-based image retrieval techniques for the purpose of achieving reliable and accurate retrieval performance. An extensive investigation of notable existing shape descriptors is reported. Two new shape descriptors have been proposed as means to overcome limitations of current shape descriptors. The first descriptor is based on a novel shape signature that includes corner information in order to enhance the performance of shape retrieval techniques that use Fourier descriptors. The second descriptor is based on the curvature of the shape contour. This invariant descriptor takes an unconventional view of the curvature-scale-space map of a contour by treating it as a 2-D binary image. The descriptor is then derived from the 2-D Fourier transform of the 2-D binary image. This technique allows the descriptor to capture the detailed dynamics of the curvature of the shape and enhances the efficiency of the shape-matching process.

Several experiments have been conducted in order to compare the proposed descriptors with several notable descriptors. The new descriptors not only speed up the online matching process, but also lead to improved retrieval accuracy.

The complexity and variety of the content of real images make it impossible for a particular choice of descriptor to be effective for all types of images. Therefore, a data-fusion formulation based on a team consensus approach is proposed as a means of achieving high accuracy performance. In this approach a select set of retrieval techniques form a team. Members of the team exchange information so as to complement each other's assessment of a database image candidate as a match to query images. Several experiments have been conducted based on the MPEG-7 contour-shape databases; the results demonstrate that the performance of the proposed fusion scheme is superior to that achieved by any technique individually.

Acknowledgements

First and foremost, I thank Allah (God) for giving me the strength and ability to pursue this work to its conclusion.

During this research, I have worked with many people who have contributed in a variety of ways to my research. It is a pleasure to convey my gratitude to all of them in my humble acknowledgment.

In the first place, I would like to record my profound gratitude to Dr. Otman Basir for his supervision, advice, and guidance from the very early stages of this research as well as for his gift of the extraordinary experiences I enjoyed during the work. Above all and most especially, I would like to thank him for being a friend as well as a supervisor and for providing me with unfailing encouragement and support in many ways. His true scientific intuition has made him a constant source of ideas and passion about science, which have extraordinarily inspired and enriched my development as a student and a researcher. He is always enthusiastic to meet me to discuss my ideas, to proofread and mark up my work; always challenged me with questions that helped me think through any difficulties. I am more indebted to him than he knows.

I gratefully thank the members of my thesis defense committee for their constructive comments about this thesis. I am thankful that in the midst of all their activities, they agreed to be members of the examination committee.

Special thanks to Dr. Majid Ahmadi from the Department of Electrical and Computer Engineering, University of Windsor for accepting to be my external examiner and for devoting his valuable time to read my thesis and to provide me with feedback on my work.

I am deeply indebted to a number of faculty members who were of great assistance during my undergraduate and graduate studies in the Electrical and Electronic

Dedication

I dedicate this thesis to my family, especially

my parents,

my wife and my children

Contents

List of Tables

List of Figures

Abbreviations and Notations

1-FD	One-dimensional Fourier Descriptor
AF	Angular Function
AR	Aspect ratio
ARC	Angular Radial Coordinates
ARTDs	Angular Radial Transform Descriptors
BE	Bending Energy
C	Circularity
CBFD	Curvature-Based Fourier Descriptor
CBIR	Content-Based Image retrieval
CC	Complex Coordinate
CD	Centroid Distance
CLD	Chord-Length Distance
CSS	Curvature Space Scale
CSSD	Curvature Scale Space Descriptor
CX	Convexity
DFT	Discrete Fourier Transform
DP	Dynamic Programming
E	Eccentricity
ECP	Elite Candidates Partition
EX	Extent
FD	Fourier Descriptor
FDs	Fourier Descriptors
F_i	Feature Extraction Scheme
FMDs	Fourier-Mellin descriptors
FMT	Fourier-Mellin transform
FPD	Farthest Point Distance
GD	Grid Descriptor
GFDs	Generic Fourier Descriptors
IHM	Invariant Hu Moment
IRA	Image Retrieval Agents
IRM	Invariant Regular Moment

IZM	Invariant Zernike Moment
M	Number of Agents
MER	Minimum Enclosing Rectangle
MTAR	Multi-Triangular Area Representation
PC	Polar Coordinates
PCP	Potential Candidates Partition
PZMDs	Pseudo-Zernike Moment Descriptors
QBIC	Query By Image Example Content
RD	Radial Distance
R_{ii}	Marginal Ranking
R_{ij}	Conditional Ranking
S	Solidity
SGD	Simple Global Descriptors
S_i	Similarity Measure
TAR	Triangular Area Representation
TCA	Triangular Centroid Area
TMDs	Tchebichef Moment Descriptors
ZMDs	Zernike Moment Descriptors
ZMs	Zernike Moments

Chapter One

Introduction

Recent years have seen an enormous increase in the number of images captured by digital cameras. The ease and convenience of capturing digital images and transmitting them between digital cameras and image databases is a contributing factor to the immense growth of image databases. These databases cover a wide range of applications, including military, environmental, astronomy, transportation, aviation, medical and multimedia. The storage format of the image data is relatively standardized; however, the effective retrieval of images from such databases remains a significant challenge.

This chapter provides an overview of research being conducted in the field of image retrieval; it highlights some of the outstanding issues in shape-based image retrieval and outlines the organization of this thesis.

1. 1 An Overview of Image Retrieval

Automatic image-retrieval techniques are required for handling massive amounts of stored and exchanged image information. As depicted in Figure 1.1, images are typically retrieved from a database based on either textual information or content information. Early retrieval techniques relied on the textual annotation of the images: images were first annotated with text, then searched for based on their textual tags. However, text-based techniques have many limitations due to their reliance on manual annotation – a tedious and error-prone process, especially for large data sets. Furthermore, the rich content typically found in images and the subjectivity of human perception make using words to describe images a difficult, if not impossible, task.

```
                    ┌─────────────────────┐
                    │   Image Retrieval   │
                    └─────────────────────┘
                               │
            ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─┼──────────────────────┐
            │                  │                       │
            ↓                  ↓                       ↓
     ┌ ─ ─ ─ ─ ─ ─ ─ ┐              ┌─────────────────────┐
     │   Text-Based  │              │   Content-Based     │
     └ ─ ─ ─ ─ ─ ─ ─ ┘              └─────────────────────┘
```
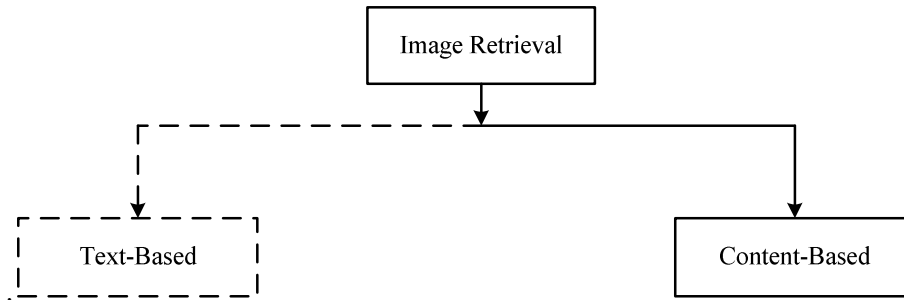
Figure 1.1 Classification of image-retrieval methods.

To overcome these difficulties, Content-Based Image Retrieval (CBIR) has been proposed [1]. This approach to image retrieval relies on the visual content of the images rather than on textual annotations and therefore has the potential to respond to more specific user queries. In 1992, the National Science Foundation of the United States organized a workshop on visual information management systems in order to identify new directions in image database management systems. At this conference, Kato [1] introduced the term Content-Based Image Retrieval (CBIR) to describe automatic retrieval of images from a database. He emphasized the use of color and shape as the criteria for the automatic image retrieval process. Since then, the term CBIR has been adopted to describe an image-retrieving process that is used for large collections of images and that is based on features that can be automatically extracted from the images themselves.

Figure 1.2 depicts a typical CBIR system. The visual contents of the images are extracted and described by feature vectors that form a feature database. During the retrieval stage, the users provide the retrieval system with a sample image (query image) or a sketched figure. The system then changes this query into feature vectors. In the matching stage, the system calculates the similarities between the feature vectors of the query sample or sketch and those of the images in the database, and then performs the retrieval accordingly. Recent retrieval systems have incorporated relevance feedback

from the users to modify the retrieval process so that it generates more perceptually and semantically meaningful results.
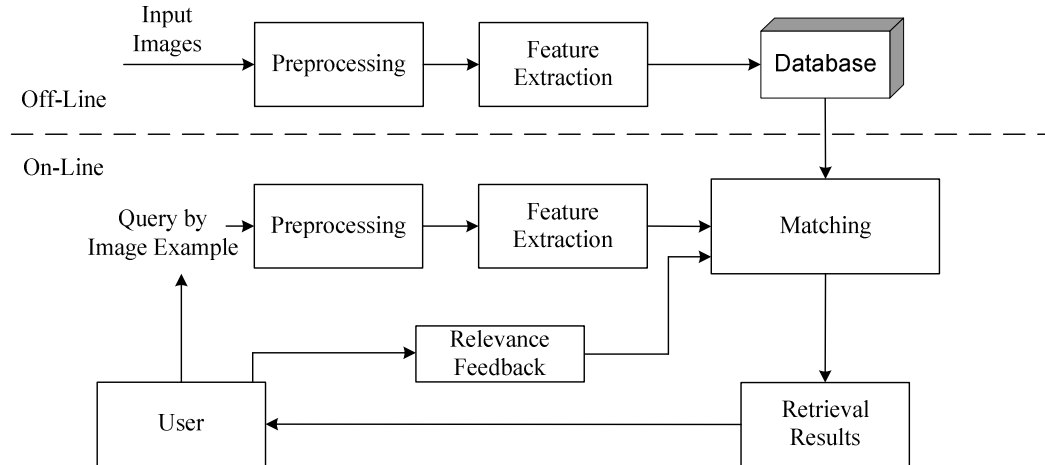
Figure 1.2 Typical content-based image retrieval (CBIR) system.

Despite the many challenges inherent in current CBIR technology, several CBIR commercial systems and CBIR research prototypes are now available. One of the best-known commercial CBIR systems is the QBIC (Query By Image Content) developed by IBM [2-4]. It offers retrieval by any combination of color, texture, or shape, as well as by text keywords. Virage is another popular commercial CBIR system used by AltaVista [5]. Other experimental systems have been designed by academic institutes in order to test the performance of new techniques. The most prominent of these systems are Photobook [6], developed by the Massachusetts Institute of Technology (MIT), and VisualSEEk [7], developed at Columbia University. A systematic and comprehensive survey of CBIR systems has been conducted by Veltkamp and Tanase, who evaluated and compared 39 CBIR systems[8].

CBIR techniques use visual contents such as color, texture, and shape to represent and index images, as shown in Figure 1.3. Color and texture have been explored more

3

thoroughly than shape. Because shape is a more intrinsic property of objects than color or texture, and given the considerable evidence that natural objects are recognized based primarily on their shape [9, 10], the increasing interest in using the shape features of objects for CBIR is not surprising. The focus of this research is therefore on shape-based image retrieval.
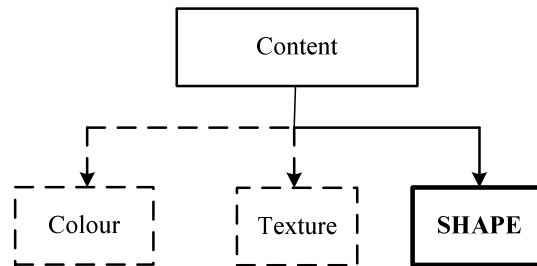
Figure 1.3 Classification of content-based image retrieval.

Since humans can often recognize the characteristics of objects using only their shapes, it is conceivable to expect shape-based techniques to be intuitive as a tool for retrieving images. A user survey of the cognition aspects of image retrieval also supports this conclusion; it indicates that users are more interested in retrieval by shape than by color or texture [11]. However, retrieval by shape is still considered to be a more intrinsically difficult task compared to retrieval based on other visual features [9]. In addition, the problem of shape retrieval becomes more complex when extracted objects are corrupted by occlusion or noise as a result of the image segmentation process.

During the last decade, significant progress has been made in both theoretical and practical research on shape-based image retrieval systems [12-14]. There are two main types of approaches to shape representation: region-based and contour-based. The region-based approach is a general method which can be applied to generic shapes. However, humans are thought to distinguish shapes mainly by their contour features [15].

Moreover, in many applications, the shape's contour is the only feature of interest, whereas the internal content of the shape is not as important. This research therefore focuses on images that can be classified based on their contours. This focus does not imply that region-based techniques are not employed in the investigation of the problem, because region-based techniques are generic and can thus be utilized in retrieving shapes based on their region or contour.

1. 2 Motivation

Although a number of studies are already being conducted with respect to shape-based image retrieval, many challenging problems continue to attract researchers:

- Techniques that have a robust structure for capturing local information tend to fail to capture global information, and vice versa.
- Previous shape-based techniques exhibit low accuracy when they attempt to classify shapes within the same class that have different contour signatures. Samples of these images are depicted in Figure 1.4.
- Techniques that capture local characteristics utilize complex matching processes that make them impractical for retrieving shapes from huge databases.

Class 1   Class 2
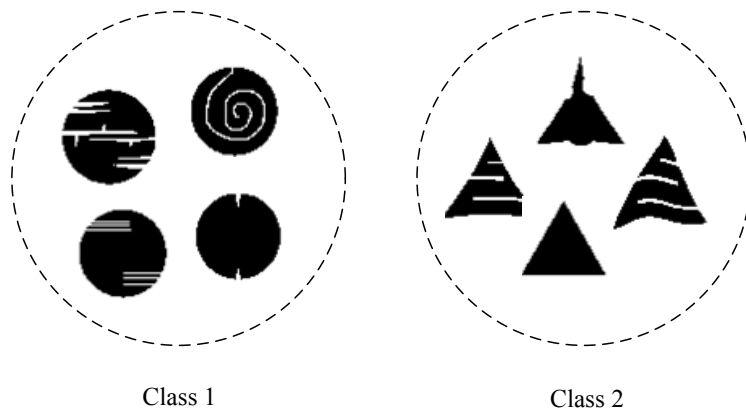
Figure 1.4 Two classes, each having shapes with different contours.

1. 3 Research Objectives

The following are the main objectives of this thesis:

- To investigate and analyze various shape-based image-retrieval techniques and compare their performance as tools for shape-based image retrieval.

- To improve the performance of shape-based image retrieval by devising new techniques that use a simple matching process to retrieve images.

- To develop a multi-technique fusion approach that uses *simple techniques* to achieve enhanced performance compared to that of existing techniques.

1. 4 Outline of the Thesis

This thesis is composed of six chapters:

Chapter 1 gives a brief background about image retrieval, describes the motivation and objectives of this research.

Chapter 2 provides a literature review of the boundary and region shape descriptors. An investigation of these descriptors is presented, and their advantages and disadvantages are clarified.

Chapter 3 introduces a simple yet powerful shape signature, named Farthest Point Distance (FPD), for Fourier Descriptor (FD). Several existing signatures are also reviewed and compared experimentally with the proposed signature.

Chapter 4 describes a novel Curvature-Based Fourier Descriptor (CBFD). The proposed descriptor is implemented and then compared with notable existing descriptors.

In **Chapter 5,** a data-fusion formulation is presented as a potential solution for addressing some of the drawbacks of traditional CBIR techniques. This new formulation utilizes the proposed descriptors in order to achieve high accuracy in shape retrieval.

Chapter 6 contains conclusions derived from the research and presents suggestions for future work.

Chapter Two

Literature Review of Region- and Boundary-Based Shape Descriptors

2.1 Introduction

Deriving shape descriptions is an important task in content-based image retrieval. In practice, a shape descriptor represents a 2-D shape as a vector in a fixed dimensional space. In other words, it is a mapping that converts the shape space into a vector space and satisfies the requirement that two similar shapes will also have close-to-identical shape descriptors. A shape-description should be unaffected by translation, rotation, and scaling. It is not necessary that the original shape be reconstructed using the extracted descriptors, but the descriptors for different shapes should also be different enough that they can be used to discriminate between shapes.

During the last decade, significant progress has been made through research into both the theoretical and practical aspects of shape description, and the literature reports a variety of techniques that have the goal of describing objects based on their shapes. The two main approaches to deriving shape descriptors are region-based and boundary-based (also known as the contour-based approach). In the region-based approach, all the pixels within a shape's region are adopted in order to obtain the shape descriptors. On the other hand, the boundary-based descriptor approach uses only the boundary of an object to extract its shape descriptors.

Each of the region and boundary approaches can be further divided into global and structural techniques [15]. Global techniques use the whole shape to derive descriptors

while structural techniques decompose the shape into sub-regions called primitives. Figure 2.1 shows the classification of shape-based descriptors.
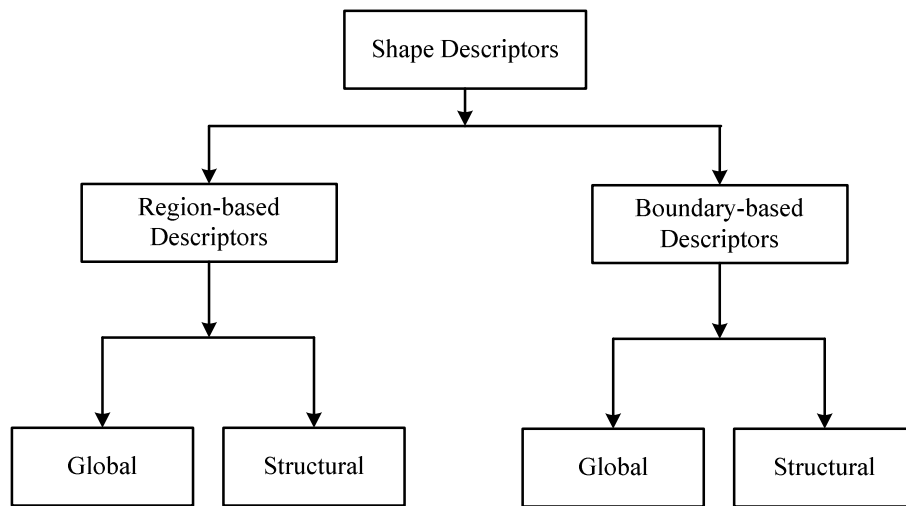
Figure 2.1 Classification of shape-based descriptors.

The effectiveness and efficiency of shape descriptors are the most challenging technical issues for a content-based image-retrieval system. Effectiveness has traditionally been the major concern with respect to shape descriptors. However, in CBIR applications, efficiency is equally important because of the demand for online retrieval. In the development of the MPEG-7 standard[16], six requirements have been set for measuring a shape descriptor: good retrieval accuracy, compact features, general application, low computational complexity, robust retrieval performance, and coarse-to-fine hierarchical representation [17].

This chapter gives a comprehensive review of shape-based descriptors and explains the ones commonly used for image retrieval. Promising descriptors for shape retrieval are indentified according to the six MPEG-7 requirements.

2.2 An Overview of Region-Based Shape Descriptors

Global region-based techniques often use invariant moments to describe shapes [18, 19]. The pioneering work in invariant moments was presented by Hu in 1962 [20]. He employed the theory of algebraic invariants to derive a set of moment invariants that are invariant under rotation, translation, and scaling transformations. This work opened the door to further research, and a number of improvements and generalizations of invariant moments have since been developed. Reddi [21] defined 2-D rotational moments in polar coordinates. Rotational moments have the property of being invariant under image rotation. Another set of invariant moments based on geometrical moments, also known as regular moments, were derived by Bamieh and De Figueriredo [22]. Abu-Mostafa and Psaltis [23] introduced complex geometrical moments and used them to derive a set of moment invariants. Flusser [24] used complex moments and reported a simple scheme for the derivation of the Hu moment. These moment invariants based on geometrical moments have been used for the recognition of images in many applications [25-31].

One of the drawbacks of the above invariants is that a non-orthogonal basis is used to derive them; hence, they suffer from a high degree of information redundancy.

To overcome the shortcomings of geometrical moments, Teague [32] introduced the notion of using an orthogonal basis to derive the invariants. He used continuous orthogonal polynomials, such as Legendre and Zernike polynomials, to derive Legendre and Zernike moments (ZMs). Zernike moments have an attractive rotation property, as a result this type of invariant moment has been used for many applications, such as character recognition [33], face recognition [34], and shape retrieval [35].

Teh and Chin [27] evaluated the performance of several moments with respect to issues such as representation ability, noise sensitivity, and information redundancy. In terms of overall performance, Teh and Chin showed that Zernike moments outperform

geometrical and Legendre moments. Khotanzad [36] explored the rotation invariance of Zernike moments and applied them in image recognition. In addition, he introduced an image reconstruction method using ZMs. Belkasim and *et al.* [33] compared different moment invariants and proposed a new scaling scheme for ZM invariants. Whoi and Kim [37] studied Zernike moments and showed that Zernike moments are very effective in retrieving an image from a large image database. Zhang and Lu [38] compared the performance of Zernike moments against the performance of several shape-retrieval techniques and showed that Zernike moments provide the most accurate retrieval of shapes from a standard database.

The Angular Radial Transform (ART) descriptor [39] is another region-based shape descriptor that also belongs to the category of shape descriptors based on moments, as defined by the MPEG-7 working group. The ART descriptors are based on cosine functions. It has been reported that Zernike moments outperform angular radial transform descriptors in retrieving shapes that can be classified based on their contours, but that ART descriptors perform better in the case of complex shapes, i.e., shapes that consist of multiple contours, which can be classified only by region-based techniques [40]. Not only do angular radial transform and Zernike descriptors both possess the property of rotation invariance but both can also be normalized with respect to changes in the size of the shape using a simple and direct procedure proposed by Bin and Jia-xiong [41].

Recently, discrete orthogonal moments have been introduced as shape descriptors [42-45]. This type of moment was first introduced by Mukundan *et al.* [42]. They adopted discrete orthogonal Tchebichef polynomials to derive a set of orthogonal moments. Mukundan *et al.* [42] claimed that Tchebichef moments may overcome the shortcomings of continuous moments because the implementation of Tchebichef moments does not involve any numerical approximation, since the basis set is orthogonal in the discrete

domain of the image coordinate space. To derive rotation invariants from Tchebichef moments, Mukundan *et al.* [46] recently proposed radial Tchebichef moments.

Other orthogonal-discrete moments that have been proposed are the Krawtchouk [43], Racah [44], and Dual Hahn [45] moments. The main drawback of these moments is that they are not invariant with respect to rotation, which makes them unsuitable for pattern-recognition and image-retrieval applications.

Ping and *et al.* [47] suggested the use of the orthogonal radial Jacobi polynomials and the angular Fourier complex exponential factor for deriving a set of moments called Jacobi-Fourier moments. Moreover, they showed that almost all orthogonal moments that consist of a radial polynomial and the angular Fourier exponential factor can be derived from the Jacobi–Fourier Moments in terms of different parameter values.

Sheng and Duvernoy [48, 49] employed the Mellin transform and the angular Fourier complex componential factor to drive invariant Fourier-Mellin descriptors. The problem associated with Fourier-Mellin descriptors is that the kernel function of the Mellin transform is not orthogonal; hence, these invariants suffer from a high degree of information redundancy. To overcome this problem, Sheng and Shen [50] applied the Gram-Schmidt orthogonalization process to orthogonalize the kernel function of the Mellin transform and used the resultant orthogonal polynomial to derive a set of orthogonal Fourier-Mellin moments. Their experiments demonstrated that, for small images, the description produced by orthogonal Fourier-Mellin moments is better than that produced by Zernike moments in terms of image-reconstruction errors and signal-to-noise ratio.

The Fourier transform is a powerful tool for shape analysis and has many applications [27, 51-55]. Most Fourier shape-based analysis techniques are based on shape contours and the 1-D Fourier transform. Recently, Zhang and Lu [56] used the 2-D

Fourier transform to derive a set of invariant descriptors, called generic Fourier descriptors (GFDs). In this method, the 2-D shape is transformed to polar coordinates, and the 2-D Fourier transform is then applied in order to drive the shape descriptors. Generic Fourier descriptors capture the spectral features of a shape in both radial and circular directions. The experiments conducted by Zhang and Lu showed that the GFD outperforms the ZM in retrieving shapes based on the content of a region. Li and Lee [57] proposed a technique which utilizes the magnitude of the Fourier transform as well as phase coefficients in order to derive a set of invariant features.

All of the above techniques use a mathematical transformation to derive shape features. However, there are other techniques that do not use a mathematical transformation in order to capture shape features. The most notable are the shape matrix and grid descriptors. In the shape matrix technique, the binary value of the shape is sampled at the points where circles that are concentrated at the centre of the shape intersect with radial lines drawn from the centre of the shape [58]. Then a matrix is formed so that the columns and rows correspond to the circles and radial lines, respectively. Taza and Suen [59] used a weighting matrix to give more weight to the pixels close to the boundaries of the image. The matching process using a shape matrix is expensive because the dimension of the feature vector is very large.

With grid descriptors, the given shape is overlaid by a grid space of a fixed size. A value of 1 is assigned to cells if they are covered by a certain percent of the shape, and a value of 0 is assigned to each of the other cells [60]. The rotation normalization is based on the major axis, which is sensitive to noise and can be unreliable even in the absence of noise effects. Moreover, online retrieval of the gird descriptors involves extremely complex computations due to the high degree of dimensionality of the feature vectors.

Several structural region-based representations have been proposed [61-66]. Generally, these techniques use the region of the shape skeleton or convex hull in order to derive shape descriptors. The idea behind obtaining the skeleton of the region of a shape is to preserve the topological information of the shape and to eliminate redundant information. In other words, the skeletonization extracts a region-based shape feature that represents the general form of an object. The shape's skeleton can be obtained by several techniques such as thinning algorithms, mathematical morphologic-based algorithms, and distance map-based algorithms [67]. After skeletonization, the skeleton is decomposed into parts and represented as a graph. Matching shapes then becomes a graph-matching problem. The difficulty with skeleton-based techniques is that a small amount of noise or a variation in the boundary often generates redundant skeleton branches that may seriously disturb the topology of the skeleton's graph [68].

Like the skeleton, the convex hull is a powerful concept that can be used for the analysis and description of shapes. The extraction of convex hulls can be achieved by finding significant convex deficiencies along the boundary of the shape. The shape can then be represented by a string of concavities. The resultant convex deficiencies can be further decomposed by obtaining the convex hulls and deficiencies of these convex deficiencies. This process can be recursively repeated until all of the derived convex deficiencies are convex. The shape is then represented as a concavity tree.

2.3 Region-Based Shape-Retrieval Descriptors

The following subsections further discuss the characteristics of region-based descriptors commonly used for shape retrieval and identify their advantages and disadvantages.

2.3.1 Simple Global Descriptors (SGDs)

A specific region of a shape can be described using scalar measures based on its simple geometrical characteristics. The simple descriptors of a shape can often discriminate only between shapes with large dissimilarities; therefore, they are usually used in the retrieval process as filters to eliminate false hits or are combined with other techniques to differentiate shapes. The advantage of these descriptors is that they have a physical meaning, but they are not suitable as standalone shape descriptors. A brief description of commonly used simple descriptors is provided in the following sections.

A. Eccentricity (E)

Eccentricity (E) has been widely used as a shape feature [2, 69]. It illustrates the way in which the region points are scattered around the centre of the image region. E is defined as the ratio of the major axis of the region to the minor axis. It is calculated using central moments such that [70]:

$$E = \frac{\mu_{20} + \mu_{02} + \sqrt{\mu_{20}^2 + \mu_{02}^2 - 2\mu_{02}\mu_{20} + 4\mu_{11}^2}}{\mu_{20} + \mu_{02} - \sqrt{\mu_{20}^2 + \mu_{02}^2 - 2\mu_{02}\mu_{20} + 4\mu_{11}^2}} \tag{2.1}$$

where μ_{pq} are the central moments.

Figure 2.2 displays various shapes and their eccentricity.

E=1   E=1   E=0.5
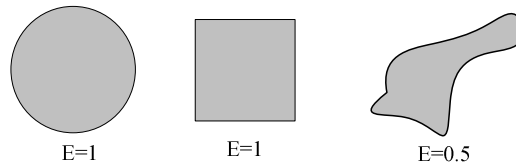
Figure 2.2 Examples of object eccentricity.

15

B. Solidity (S)

Solidity is computed as the ratio between the object area A_0 and the area of the corresponding convex hull [71, 72]. Solidity is computed as follows:

$$S = \frac{A_o}{Convex\ Area} \tag{2.2}$$

Figure 2.3 exhibits various objects and their solidity.

S=1   S=1   S=1   S=0.7549
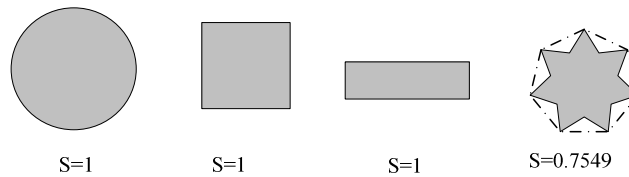
Figure 2.3 Examples of object convexity and solidity.

C. Extent (Rectangularity)(EX)

Extent is a measure that reflects the rectangularity of a shape [73], in other words, how much of a shape fills its minimum enclosing rectangle (MER). Extent is defined as

$$EX = \frac{A_o}{A_{MER}} \tag{2.3}$$

where A_{MER} is the area of the object's MER.

Figure 2.4 shows various objects and their extent. EX has a value of 1 for a rectangular shape.

EX=0.789   EX=1   EX=1   EX=0.57
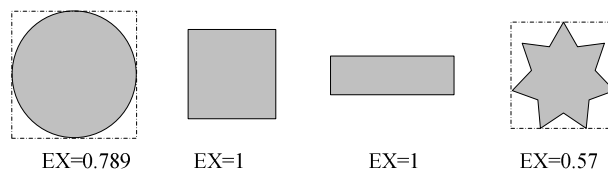
Figure 2.4 Examples of object extent.

D. Circularity (Compactness)

Circularity is a measure of roundness and is defined as follows [74]:

$$C = \frac{4\pi * A_o}{(P_o)^2}$$
(2.4)

where A_o is the object area, and P_o is the object perimeter.

It is sometimes used in the form *(1-C)* or in an inverse formula $C = \frac{(P_o)^2}{4\pi * A_o}$ [40, 75].

Figure 2.5 shows various shapes and their circularity.

C=1   C=.91   C=0.785   C=.6
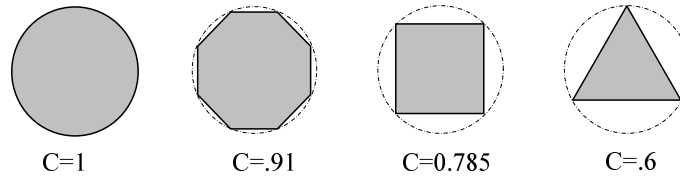
Figure 2.5 Examples of object circularity.

2.3.2 Moments

Moments can be classified as orthogonal or non-orthogonal according to the basis used to derive them. This section defines the commonly used moment-based descriptors and summarizes their properties.

2.3.2.1 Non-Orthogonal Moments

A. Regular Moment Descriptors (RMD)

Regular moments are by far the most popular type of moment [27]. Two-dimensional (2-D) regular moments of the order (p + q) of an (N x M) digital image f(x, y) are defined as follows [33, 76]:

$$RM_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q . f(x, y)$$
(2.5)

The first order moment is used to locate the centroid of the image, where

17

$$x_c = \frac{RM_{10}}{RM_{00}}, y_c = \frac{RM_{01}}{RM_{00}} \tag{2.6}$$

Central moments are constructed by subtracting the centroid from all the coordinates. These moments are invariant under translation and are represented as

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - x_c)^p (y - y_c)^q \cdot f(x, y) \tag{2.7}$$

Moments can also be invariant to both translation and scale if the central moments are divided by the zeroth moment using the following formula [77]:

$$\eta_{pq} = \frac{\mu_{pq}}{(\mu_{00})^{(p+q+2)/2}} \tag{2.8}$$

Although normalized central moments (η_{pq}) are invariant under translation and change in size, they are not invariant under rotation. To make η_{pq} rotation invariant, the shape angle is estimated using second-order moments [33]:

$$\alpha = \frac{1}{2} \, tan^{-1}(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}) \tag{2.9}$$

The regular moments of $f(x,y)$, after rotation by an angle α about the origin, are defined by [27, 33]

$$RM_{pq}^{\alpha} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x \cos(\alpha) + y\sin(\alpha))^p (y \cos(\alpha) - x\sin(\alpha))^q \cdot f(x, y) \tag{2.10}$$

The invariant regular moments (IRM) can be expressed in terms of the original normalized central moments of $f(x, y)$ such that [33]

$$IRM_{pq} = \sum_{r=0}^{p} \sum_{s=0}^{q} \binom{p}{r} \binom{q}{s} (-1)^{q-s} \cdot (\cos(\alpha))^{p-r+s} \cdot (\sin(\alpha))^{q+r-s} \eta_{p+q-r-s.r+s} \tag{2.11}$$

The main steps in extracting the invariant regular moment descriptors (RMD) are shown in Figure 2.6.

18

Input Image
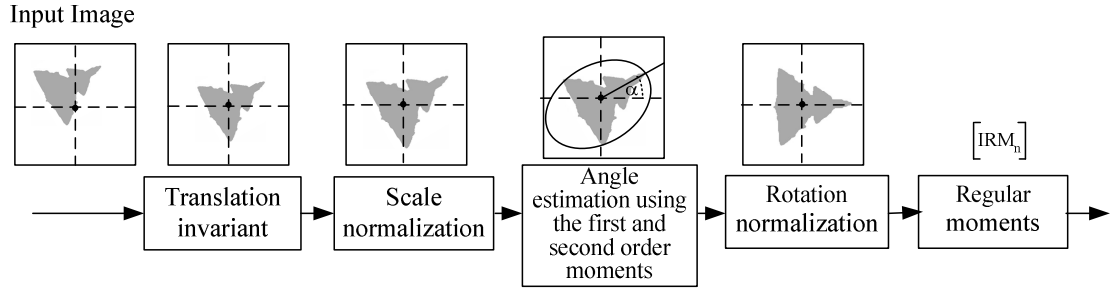
Figure 2.6 The basic steps in the generation of regular moment descriptors.

Here, the method used to produce rotation invariance is called the principal axis. It is based on the observation that moments are computed relative to a unique set of principal axes of distribution and are therefore invariant to the orientation of the distribution. The main difficulty with the RMD lies in estimating the angle α. The principal axes cannot be uniquely defined, particularly when the image is of circular or an n-fold rotational symmetry.

B. Hu Moment Descriptors (HMDs)

Hu was the first to derive a set of seven moments that are invariant under the actions of translation, scaling, and rotation [20]. These are known as Hu moments and are given by the following equations:

19

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + 3(\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + 3\eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]$$
$$\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \qquad (2.12)$$

$$\phi_6 = (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{\,})$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]$$
$$\quad - (\eta_{30} + 3\eta_{12})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

where η_{pq} is obtained using Equation (2.8)

In 2000, Flusser [24] used complex moments and reported a simple scheme for the derivation of the Hu moment. The complex moments of order (p+q) are defined as [24]:

$$C_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+iy)^p (x-iy)^q f(x,y)dxdy \qquad (2.13)$$

where $i = \sqrt{-1}$

For digital images, each complex moment is expressed in terms of geometric moments η_{pq} as [24]

$$C_{pq} = \sum_{r=0}^{p} \sum_{s=0}^{q} \binom{p}{r}\binom{q}{s}(-1)^{q-s}(i)^{p+q-r-s}\,\eta_{r+s,\,p+q-r-s} \qquad (2.14)$$

The seven Hu moments are calculated using Equation (2.14), as follows [24]:

$$\Phi_1 = C_{11}$$
$$\Phi_2 = C_{20}C_{02}$$
$$\Phi_3 = C_{30}C_{03}$$
$$\Phi_4 = C_{21}C_{12} \qquad (2.15)$$
$$\Phi_5 = \mathrm{Re}(C_{30}C_{12}^3)$$
$$\Phi_6 = \mathrm{Re}(C_{20}C_{12}^2)$$
$$\Phi_7 = \mathrm{Im}(C_{30}C_{12}^3)$$

Flusser also showed that the Hu moments are dependent and that

$$\phi_3 = \frac{\phi_5^2 * \phi_7^2}{\phi_4^3}$$

(2.16)

In recognition of 2-D objects, Flusser suggested the following moments with the use of second and third order, or fourth order, respectively [24]:

- Using the second and third order:

-

$$\begin{aligned}
\Psi_1 &= C_{11} = \Phi_1 \\
\Psi_2 &= C_{21}C_{12} = \Phi_4 \\
\Psi_3 &= \mathrm{Re}(C_{20}C_{12}^2) = \Phi_6 \\
\Psi_4 &= \mathrm{Im}(C_{20}C_{12}^2) \\
\Psi_5 &= \mathrm{Re}(C_{30}C_{12}^3) = \Phi_5 \\
\Psi_6 &= \mathrm{Im}(C_{30}C_{12}^3) = \Phi_7
\end{aligned}$$

(2.17)

- Using the fourth order:

-

$$\begin{aligned}
\Psi_7 &= C_{22} \\
\Psi_8 &= \mathrm{Re}(C_{31}C_{12}^2) \\
\Psi_9 &= \mathrm{Im}(C_{31}C_{12}^2) \\
\Psi_{10} &= \mathrm{Re}(C_{40}C_{12}^4) \\
\Psi_{11} &= \mathrm{Im}(C_{40}C_{12}^4)
\end{aligned}$$

(2.18)

The advantage of using invariant Flusser moment descriptors (FMD) and HMD is that they are rotation invariant and provide a very compact representation of the shape. A few lower-order central moments have a physical meaning, such as mean, variance, and skew. However, the physical meaning of the other moment invariants is not known. Moreover, the basis used to derive the moments is not orthogonal, so these moments suffer from a high degree of information redundancy.

2.3.2.2 Orthogonal Moments

A. Zernike Moment Descriptors (ZMDs)

To overcome the shortcoming of geometrical moment-based descriptors with respect to information redundancy, Teague [32] adopted Zernike orthogonal polynomials to derive a set of invariant moments, called Zernike moments (ZMs). Zernike moments are defined as the projections of *f(x,y)* on complex Zernike polynomials which form a complete orthogonal set over only the interior of the unit circle; that is, $x^2 + y^2 \leq 1$. The function of complex Zernike moments with an order *p* and repetition *q* in polar coordinates is defined as

$$ZM_{pq} = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 f(r,\theta).V_{pq}^*(r,\theta).r\,dr\,d\theta \tag{2.19}$$

$V_{pq}(r,\theta)$ is the Zernike polynomial, defined as

$$V_{pq}(r,\theta) = R_{pq}(r).e^{iq\theta}, \quad i = \sqrt{-1} \tag{2.20}$$

where *p* is a positive integer or zero and *q* is a positive integer subject to the constraints that $p - |q|$ is even and $q \leq p$, *r* is the length of the vector from the origin to the *(x, y)* pixel; i.e., $r = \sqrt{x^2 - y^2}$ θ is the angle between the vector *r* and the *x* axis in a counterclockwise direction, the symbol (*) denotes the complex conjugate, and $R_{pq}(r)$ is a real-valued Zernike-radial polynomial defined as follows:

$$R_{pq}(r) = \sum_{k=0}^{(p-|q|)/2} (-1)^k. \frac{(p-k)!}{k!\left(\frac{p+|q|}{2}-k\right)!\left(\frac{p-|q|}{2}-s\right)!}.r^{p-2k} \tag{2.21}$$

The ZM expression can be mapped from the polar coordinates to the *xy*-plane by changing the variables in the double integral form of Equation (2.19). Then A_{nl} can be written in the *xy*-plane as follows:

$$ZM_{pq} = \frac{p+1}{\pi} \int \int_{x^2+y^2\leq1} f(x,y).V_{pq}^*(r,\theta)\,dx\,dy \qquad (2.22)$$

Here, $= rcos(\theta)$, $y = rsin(\theta)$, and the symbol (*) denotes the complex conjugate. For a digital image, the integrals are replaced by summations to obtain

$$ZM_{pq} = \frac{p+1}{\pi} \sum_x \sum_y f(x,y).V_{pq}^*(r,\theta)$$

$$= \frac{p+1}{\pi} \sum_r \sum_\theta f(rcos(\theta),rsin(\theta)).V_{pq}^*(r,\theta) \quad r \leq 1 \qquad (2.23)$$

The features defined by the Zernike moments are derived by adopting the rotational properties of these moments [32, 36]. If the rotated image through an angle α is denoted by f^α then the relationship between the original and the rotated images in the same polar coordinates is given by

$$f^\alpha(r,\theta) = f(r,\theta+\alpha) \qquad (2.24)$$

Using the polar coordinates, the ZMs for the original image are

$$ZM_{pq} = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 f(r,\theta).R_{pq}(r).e^{-iq\theta}.rdrd\theta \qquad (2.25)$$

The ZMs of the rotated image in the same coordinates are

$$ZM_{pq}^\alpha = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 f(r,\theta+\alpha).R_{pq}(r).e^{-iq\theta}.rdrd\theta \qquad (2.26)$$

A change in the variable $\theta_1 = \theta + \alpha$ produces the following:

$$ZM_{pq}^\alpha = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 f(r,\theta_1).R_{pq}(r).e^{-iq(\theta_1-\alpha)}.rdrd\theta_1$$

$$= \left(\frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 f(r,\theta_1).R_{pq}(r).e^{-iq\theta_1}.rdrd\theta_1\right)e^{iq\alpha}$$

$$ZM_{pq}^\alpha = ZM_{pq}.e^{iq\alpha} \qquad (2.27)$$

23

Equation (2.27) demonstrates that ZMs have a simple rotational transformation property: each ZM requires only a phase shift under rotation. This simple property suggests that the magnitudes of the ZMs of a rotated image function remain identical to those before rotation.

The magnitudes of the ZMs are rotation invariant but are not scale or translation invariant. Translation invariance is achieved by using the centralized moments to move the origin to the centre of the image. Scale invariance is achieved by enlarging or reducing each object so that its zeroth-order moment M_{00} is equal to a predetermined value β [32, 34]. The regular moments of $f(x, y)$ after a scale change by factor a are defined by

$$RM_{pq}^{a} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^{p} \, y^{p} \, f\left(\frac{x}{a}, \frac{y}{a}\right) dx dy \qquad (2.28)$$

which is expressed in terms of the original moments of $f(x, y)$ by

$$RM_{pq}^{a} = a^{(p+q+2)} \, RM_{pq} \qquad (2.29)$$

Since the goal is to achieve $RM_{00}^{a} = \beta$, then $a = \sqrt{\beta/RM_{00}}$, and $f(x,y)$ can be normalized with respect to scaling and translation by transforming $f(x,y)$ into $g(x,\text{y})$, as follows:

$$g(x, y) = f\left(\frac{x}{a}+x_c, \frac{y}{a}+y_c\right) \qquad (2.30)$$

If a point $\left(\frac{x}{a}+x_c, \frac{y}{a}+y_c\right)$ does not correspond to a grid location, then the function value associated with the point is interpolated from the values given by the four nearest grid locations [36]. Belkasim *et al.* [78] adopted Cartesian coordinates to explicitly make the Zernike moments directly invariant to scale, translation, and rotation without the need to use regular moments.

The ZMDs normalized by using the previous methods do not reflect the features of the original shape very well because scaling a shape before the extraction of a feature leads to loss of information about the shape, especially when the shape is shrunk. Bin and

Jia-xiong [41] proposed a normalization scheme that can be used after the extraction of the image feature. Normalization is achieved by dividing the ZMs of an image by the zeroth-order moment. The comparison study conducted by Bin and Jia-xiong [41] indicated that the normalization that occurs after the extraction of the Zernike features yields a better invariant than normalization that occurs before feature extraction. The main steps in extracting Zernike Moment Descriptors are shown in Figure 2.7.
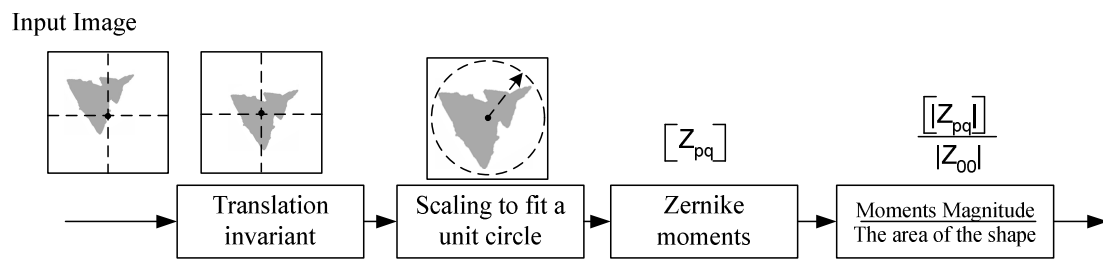
Figure 2.7 The basic steps in the generation of Zernike moment descriptors.

Many researchers have adopted Zernike moments for character recognition [33-35]. Zernike moment invariants have been shown to outperform several other shape descriptors and to be highly effective in terms of image representation [37, 38]. Using a database of about 6,000 images, Whoi and Yong [37] conducted several experiments and showed that the Zernike moment is a very effective shape descriptor in terms of exact matching under various transformations and the similarity-based retrieval.

B. Pseudo-Zernike Moment Descriptors (PZMDs)

Pseudo-Zernike moments are derived based on the basis set of pseudo-Zernike polynomials, which are similar to Zernike polynomials but have a different real-value radial polynomial $\tilde{R}_{pq}(r)$, which is defined as follows [79]:

25

$$\tilde{R}_{pq}(r) = \sum_{k=0}^{p-q}(-1)^k \cdot \frac{(2p+1-k)!}{k!\,(p-|q|-k)!\,(p+|q|+1-k)!} \cdot r^{p-k} \qquad (2.31)$$

where $p \geq 0$, and q is a positive or negative integer subject to $|p| \leq q$ only. Pseudo-Zernike moments provide more features than ZMs because PZMs are not subject to the condition that (p-q) is even. Complex PZM descriptors of order p with repetition q for a continuous function image can be derived in polar coordinates as follows:

$$PZM_{pq} = \frac{p+1}{\pi}\int_{0}^{\pi}\int_{0}^{1} f(r,\theta) \cdot \tilde{R}_{pq}(r) \cdot e^{-iq\theta} \cdot r\,dr\,d\theta \qquad (2.32)$$

For a digital image, the integrals are replaced by summations to obtain

$$PZM_{pq} = \frac{p+1}{\pi}\sum_{r}\sum_{\theta} f(r\cos\theta, r\sin\theta) \cdot \tilde{R}_{pq}(r) \cdot e^{-iq\theta} \quad , \quad r \leq 1 \qquad (2.33)$$

The same procedure can be utilized with respect to the previous section to prove that the magnitudes of the complex PZM descriptors of a rotated image function remain identical to those in place before rotation. Scale invariance of PZMs can be achieved by normalizing an image by a predetermined value using geometrical moments. Chong *et al.* [79] described a mathematical framework for deriving a new set of scale invariants of pseudo-Zernike moments based on pseudo-Zernike polynomials. However, this approach is dependent on the order of the pseudo-Zernike polynomial: as the order becomes higher, the magnitude of the invariant moments significantly diminishes, and the dynamic range between invariant moments becomes very large. The large dynamic range leads to a feature vector dominated by the larger-magnitude invariants.

C. Angular Radial Transform Descriptors (ARTDs)

For region-based shape descriptions, the MPEG-7 working group has selected angular radial transform descriptors [39]. Angular radial transform (ART) descriptors are similar to ZM descriptors. The main difference is that ART descriptors are based on

cosine functions rather than on Zernike polynomials. The ART basis functions are separable along the angular and radial directions and are defined on the unit circle as follows [39]:

$$V_{pq}(x,y) = V_{pq}(r,\theta) = \frac{1}{2\pi} R_p(r)e^{iq\theta} \tag{2.34}$$

$$R_p(r) = \begin{cases} 1 & p = 0 \\ 2\cos(p\pi r) & p \neq 0 \end{cases} \tag{2.35}$$

The complex ART descriptors of order p with repetition q for a continuous function $f(x,y)$ are defined in polar coordinates as

$$ART_{pq} = \frac{1}{2\pi}\int_0^\pi \int_0^1 f(r,\theta).R_p(r).e^{-iq\theta}\, rdrd\theta \tag{2.36}$$

For a digital image, the integrals are replaced by summations to obtain

$$ART_{pq} = \frac{1}{2\pi}\sum_r \sum_\theta f(r\cos\theta, r\sin\theta).R_p(r).e^{-iq\theta} \quad , \quad r \leq 1 \tag{2.37}$$

Proving that the magnitude of the complex ART is rotation invariant is straightforward. Scale invariance can be achieved by dividing the magnitude of the ART descriptors by the magnitude of the first descriptor.

D. Tchebichef Moment Descriptors (TMDs)

Tchebichef moments were first introduced by Mukundan *et al.* [42]. They adopted the discrete orthogonal Tchebichef polynomials in order to derive a set of orthogonal moments. For a given positive integer N (the image size), the Tchebichef polynomial is given by the following recurrence relation [42]:

$$t_p(x) = \frac{(2p-1)t_1(x)t_{p-1}(x) - (p-1)(1 - \frac{(p-1)^2}{N^2}t_{n-2}(x)}{p} \tag{2.38}$$

with the initial conditions

$$t_0(x) = 1 \tag{2.39}$$

$$t_1(x) = \frac{(2x + 1 - N)}{N}, \tag{2.40}$$

where $p = 0, 1, \ldots N - 1$

The Tchebichef moment of order (p+q) of an (NxM) image intensity function is defined as [42]

$$T_{pq} = \frac{1}{\rho(p, N)\rho(q, M)} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} t_p(x)q(y) f(x, y) \tag{2.41}$$

where $p, = 0, 1, \ldots, N - 1$ and q=0,1,…,M. The Tchebichef polynomial satisfies the property of orthogonality with

$$\rho(p, N) = \frac{N \left(1 - \frac{1}{N^2}\right)\left(1 - \frac{2^2}{N^2}\right)\cdots\left(1 - \frac{p^2}{N^2}\right)}{2p + 1} \tag{2.42}$$

The moments defined in Equation (2.41) are not rotation invariant. Mukundan *et al.* [46] recently proposed radial Tchebichef moments in order to derive rotation invariant features. Radial Tchebichef moments of order *p* and repetition *q* are defined as [46]

$$S_{pq} = \frac{1}{2\pi\rho(p, m)} \sum_{r=0}^{m-1} \sum_{\theta=0}^{2\pi} t_p(r)e^{-jq\theta} f(r, \theta) \tag{2.43}$$

where m denotes the number of samples in the radial direction. Since θ is a real quantity measured in radians, Equation (2.43) can be defined as:

$$S_{pq} = \frac{1}{n \, \rho(p, m)} \sum_{r=0}^{m-1} \sum_{\theta=0}^{n-1} t_p(r)e^{-jq\theta} f(r, \theta) \tag{2.44}$$

The magnitudes of the S_{pq} are rotation invariant but are not scale or translation invariant. Translation invariance is achieved by moving the origin to the centre of the image using centralized moments. Scale invariance is achieved by enlarging or reducing each object

so that its zeroth-order moment M_{00} is equal to a predetermined value β [32, 34]. The main steps in extracting the Tchebichef moment descriptors are depicted in Figure 2.8.

The key advantage of the radial Tchebichef moments is that they are rotation invariant and based on a discrete-orthogonal polynomial which is needed to ensure information redundancy in a moment set.

Input Image

| Translation invariant | Scale normalization | Radial Tchebichef moments | Moments magnitude |
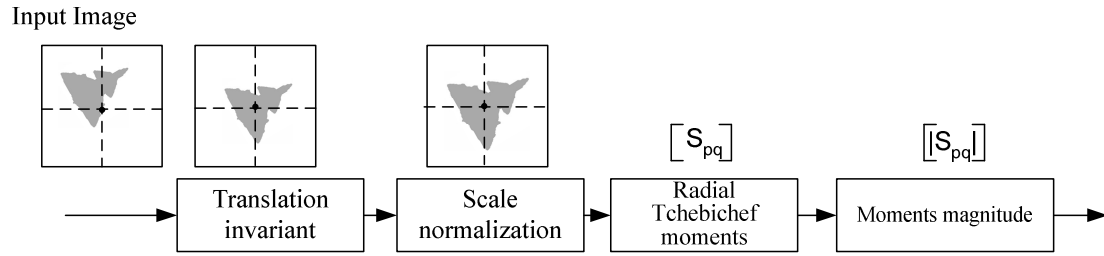
Figure 2.8 The basic steps in the generation of Tchebichef moment descriptors.

New Scale Invariants of Radial Tchebichef Moments

To achieve scale invariance of radial Tchebichef moments, researchers resort to resizing the original shape to a predetermined size. This traditional method of scaling is time-consuming and leads to loss of shape characteristics. Therefore, moments derived using the traditional normalization scheme may differ from the true moments of the original shape. This section describes a simple yet powerful scheme for deriving a new set of scale invariants of radial Tchebichef moments.

When the size of a binary shape is scaled, then the area (A) and the maximum radial distance of the shape (m) are scaled according to the scaling factor used, as shown in Figure 2.9. Therefore, it is reasonable to normalize radial Tchebichef moments using the area and the maximum radial distance of the shape.

29

	Original shape		Scaled version
		m_1	m_2
	(a)	Cartesian coordinates	(b)
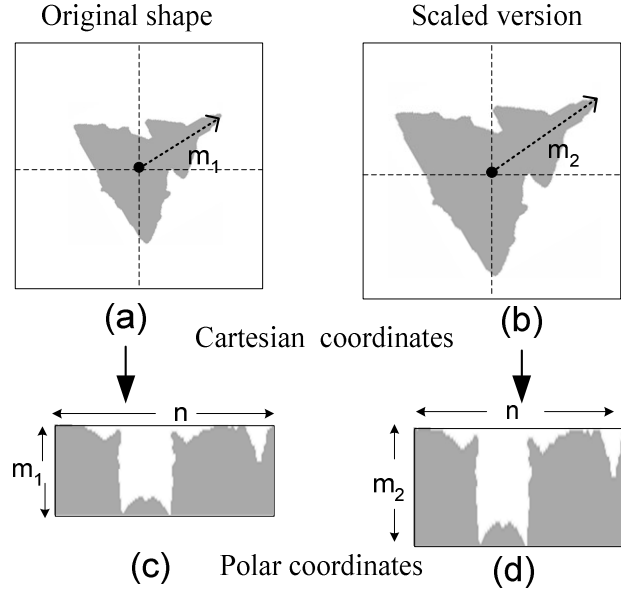
Figure 2.9 An image and its scaled version with Cartesian and polar coordinates.

The zero-order moment ($p=q=0$) of radial Tchebichef moments is obtained using Equation (2.44), as follows:

$$S_{00} = \frac{1}{n\,\rho(0,m)} \sum_{r=0}^{m-1} \sum_{\theta=0}^{n-1} t_0(r)\, e^{-j0.\theta}\, f(r,\theta) \qquad (2.45)$$

Using Equations (2.39) and (2.42), it is straightforward to prove that

$$S_{00} = \frac{1}{n\,m} \sum_{r=0}^{m-1} \sum_{\theta=0}^{n-1} f(r,\theta) = \frac{A}{n\,m} \qquad (2.46)$$

A: the area of the object.

m: the maximum radial distance.

n: the maximum resolution used in the angular direction, which is equal to 360.

From Equation (2.46), if the magnitudes of the radial Tchebichef moments are divided by $S_{00,}$ then all the moments will be scaled according to the area and the maximum radial distance, which are the two parameters affected by scaling a shape.

30

Therefore, the proposed scale invariants for the radial Tchebichef moments are as follows:

$$\acute{S}_{pq} = \frac{|S_{pq}|}{S_{00}} \qquad (2.47)$$

Table 2.1 and Table 2.2 show a binary image with different scaling factors and the corresponding magnitudes of the radial Tchebichef moments normalized using the proposed and the traditional scaling schemes, respectively. The σ and μ denote the standard deviation and mean of the normalized descriptors, respectively. The deviation of the normalized descriptors, indicated by the percentage of σ/μ, clearly shows that the proposed scaling scheme outperforms the traditional one. It can be seen from Table 2.1 and Table 2.2, that the average of the deviation for the proposed scaling scheme (12.91%) is less than that of the traditional scaling scheme (48.18%).

Table 2.1 The magnitude of the normalized Tchebicef moments using the proposed scaling scheme.

The magnitude of the normalized Tchebicef Moments	$\alpha=1$	$\alpha=0.1$	$\alpha=0.2$	$\alpha=0.3$	$\alpha=2$	$STD(\sigma)$	Mean(μ)	$\frac{\sigma}{\mu}$ (%)
$\acute{S}_{01}$	0.0692	0.0673	0.0633	0.0704	0.0511	0.0078	0.0643	12.19
$\acute{S}_{10}$	0.6890	0.7029	0.7029	0.6939	0.7272	0.0147	0.7032	2.09
$\acute{S}_{11}$	0.0796	0.0779	0.0761	0.0809	0.0489	0.0134	0 .0727	18.46
$\acute{S}_{20}$	0.2289	0.2251	0.2212	0.2300	0.1787	0.0216	0.2168	9.95
$\acute{S}_{02}$	0.1230	0.1262	0.1263	0.1455	0.1363	0.0093	0.1315	7.08
$\acute{S}_{12}$	0.0975	0.0981	0.0953	0.0970	0.1011	0.0021	0.0978	2.17
$\acute{S}_{21}$	0.0125	0.0142	0.0143	0.0072	0.0417	0.0136	0.0180	75.48
$\acute{S}_{22}$	0.1510	0.1502	0.1500	0.1487	0.1446	0.0025	0 .1489	1.71
							Average	**12.91**

Table 2. 2 The magnitude of the normalized Tchebicef moments using the traditional scaling scheme.

The magnitude of the normalized Tchebicef Moments	α=1	α=0.1	α=0.2	α=0.3	α=2	σ	Mean (μ)	$\frac{\sigma}{\mu}$ (%)
$\acute{S}_{01}$	0.4853	0.8711	0.8345	0.9008	0.4718	0.2151	0.7127	30.18
$\acute{S}_{10}$	4.8114	3.2391	3.2677	4.1768	4.8210	0.7840	4.0632	19.3
$\acute{S}_{11}$	0.5603	0.3754	0.3754	0.1558	0.5517	0.1654	0.4037	40.98
$\acute{S}_{20}$	1.5677	0.4431	0.4211	0.9856	1.5637	0.5669	0.9962	56.90
$\acute{S}_{02}$	0.8606	1.2015	1.2303	0.3924	0.8659	0.3391	0.9101	37.25
$\acute{S}_{12}$	0.6489	0.1436	0.1525	0.1521	0.6487	0.2736	0.3492	78.35
$\acute{S}_{21}$	0.0893	0.6052	0.5768	0.8869	0.0888	0.3506	0.4494	78.01
$\acute{S}_{22}$	1.0588	0.4154	0.3815	0.8934	1.0603	0.3388	0.7619	44.47
							Average	**48.18**

2.3.3 Generic Fourier Descriptors (GFDs)

A one-dimensional Fourier transform has been used successfully in image analysis to derive shape descriptors from shape contours. Recently, Zhang and Lu [56] adopted the 2-D Fourier transform to derive a set of invariant descriptors called generic Fourier descriptors (GFD). In this method, the 2-D shape is transformed to the polar coordinates, and the 2-D Fourier transform is then applied in order to derive shape descriptors. The rotation of the original shape corresponds to a shift in the normal Cartesian coordinate. This property makes it possible to derive rotation invariant features by applying the 2-D Fourier transform on the polar image and then taking the magnitude while ignoring the phase. The 2-D Fourier transform of a polar image is give by:

$$GFD(\omega_1, \omega_2) = \sum_{r=0}^{m-1} \sum_{\theta=0}^{n-1} f(r,\theta)\, e^{-j(\frac{\omega_1 r}{m} + \frac{\omega_2 \theta}{n})} \tag{2.48}$$

The main steps in the extraction of generic Fourier descriptors are shown in Figure 2.10.
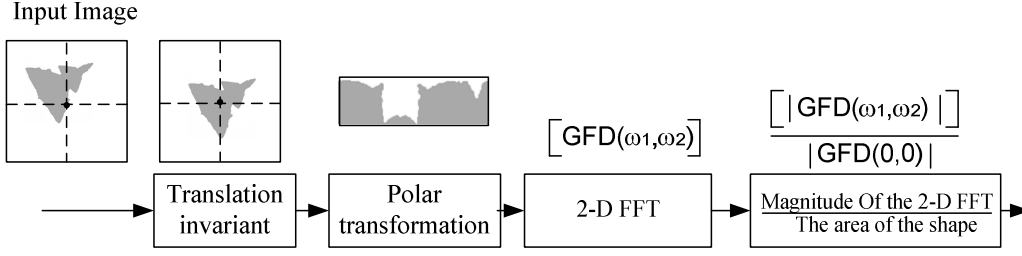
Figure 2.10 The basic steps in the generation of generic Fourier descriptors.

2.3.4 Fourier-Mellin Descriptors (FMDs)

The Fourier-Mellin transform (FMT) takes advantage of the properties of the Fourier and Mellin transforms in order to define a new set of image invariants called Fourier-Mellin descriptors (FMDs). The Fourier-Mellin transform for $f(r, \theta)$ is defined as follows [48, 49]:

$$FM_{sl} = \frac{1}{2\pi} \int_{0}^{\infty} \int_{-\pi}^{\pi} r^{s-1}.f(r,\theta).e^{-jl\theta} dr d\theta \tag{2.49}$$

This transformation shows a radial-Mellin transform with parameter *s* and an explicit circular-Fourier transform with parameter *l* [48]. The rotation invariant can be achieved by taking the magnitude of the transform. Translation invariance is conventionally achieved by placing the origin of the coordinates at the image centroid. With a given radial order *s* and the circular-harmonic frequency *l = 0, 1, . . . , N* , the scale normalization of Fourier-Mellin descriptors is defined as [49]

$$\Gamma_{sl} = \frac{|M_{sl}|^2}{\sum_{l=0}^{N}|M_{sl}|^2} \tag{2.50}$$

In the case of an imaginary order, if a nonlinear mapping $r = e^{\eta}$ is used, then the Fourier-Mellin transform can be obtained using the 2-D Fourier transform, as follows:

$$FM_{sl} = \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^\pi f(e^\eta, \theta). e^{-j(\eta s + l\theta)} d\eta d\theta \qquad (2.51)$$

In other words, the Mellin transform can be realized by logarithmically scaling the coordinates of the image and the 2-D Fourier transform of the resultant scaled function.

In Equation (2.51), it is worth noting that the GFD is similar to the FMD but rather than log-polar coordinates, the GFD uses traditional polar coordinates.

Ghorbel proposed the Analytical Fourier-Mellin Transform (AFMT) in order to construct a complete invariant to rotation and scaling. The AFMD computes the standard FMT of $r^\sigma f(r, \theta)$ rather than $f(r, \theta)$ where σ is a fixed and strictly positive real number [80]. The AFMT is defined as

$$AFM_{sl} = \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^\pi r^{\sigma - js - 1}. f(r, \theta). e^{-jl\theta} dr d\theta \qquad (2.52)$$

Let $f^{\alpha,a}$ denote the rotation and size change of an object f through angle α with a scale factor a (i.e., $f^{\alpha,a}(r, \theta) = f(ar, \theta + \alpha)$. The AFMT of $f^{\alpha,a}$ is given as:

$$AFM_{sl}^{\alpha a} = \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^\pi r^{\sigma - js - 1}. f(ar, \theta + \alpha). e^{-jl\theta} dr d\theta \qquad (2.53)$$

and a simple change in the variables r and θ gives the following relations [81]:

$$AFM_{sl}^{a\lambda} = a^{\sigma + js}. e^{j\alpha l}. AFM_{sl} \qquad (2.54)$$

Derrode and Ghorbel [81] adopted the AFM_{00} and the AFM_{10} as normalization parameters to compensate for the term $a^{\sigma + js}. e^{j\alpha l}$. They defined the invariant AFMT descriptors with respect to rotation and scaling as [81]:

$$I_{sl} = (AFM_{00})^{(-\sigma + js)/\sigma}. e^{-js \, arg(AFM_{01})}. AFM_{sl} \qquad (2.55)$$

The application of the Fourier-Mellin transform has been studied in many image-processing and pattern-recognition applications [55, 82-84]. Raman and Desai [54] used neural networks to classify invariant patterns obtained by the use of FMT. Their

experiments were based on a small set of 50 binary image shapes and showed a high recognition rate obtained even for partially occluded images. Derrode *et al.* [84] used Fourier-Mellin descriptors to retrieve gray-level images from large databases. Their experiments produced a high-quality numerical invariance and good retrieval results for a butterfly database and the well-known Columbia database.

The disadvantage of the FMDs is that the kernel function set $r^{\sigma-js-1}$ in the radial-Mellin transform is not orthogonal, which means that extracted FMDs suffer from a high degree of information redundancy. Moreover, the invariant descriptors (Equation (2.55)) defined by Derrode and Ghorbel are based on a phase angel which is very sensitive to noise.

2.3.5 Grid Descriptors (GD)

The grid-based method has attracted interest because of its simplicity with respect to representations [60]. In this technique, the given shape is overlaid by a grid space of a fixed size, as shown in Figure 2.11. A value of 1 is assigned to cells if they are at least 15% covered , and a zero is assigned to each of the other cells [60]. Then the 1s and 0s are scanned from left to right and top to bottom to obtain a binary sequence for the shape. For the shape shown in Figure 2.11, the binary sequence is 0000111100000011111-11000111 111111110000000111111. It is evident that the smaller the cell size, the more accurate the representation of the shape, but the greater the storage and computation requirements. It is evident that the representation is translation invariant, but it is not invariant to scale and rotation. Rotation and scale normalization are thus necessary when GDs are used to compare two shapes.
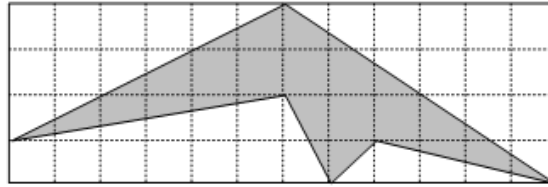
Figure 2.11 Example of grid representation.

To achieve scale normalization, all shapes are scaled proportionally so that the major axes have the same fixed length. For rotation invariance, the given shape is rotated so that its major axis is in parallel with the x-axis. As a result of the 180° rotation, two possibilities still remain for the shape placement: one of the farthest points can be either on the left or on the right. Two different binary sequences are therefore needed in order to represent these two orientations. The main steps in extracting grid descriptors are shown in Figure 2.12.
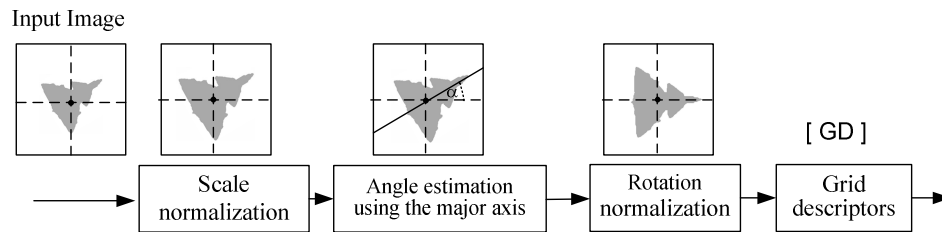
Figure 2.12 The basic steps in the generation of grid descriptors.

The key problem associated with GDs is the problematic major axis normalization. The major axis is sensitive to noise and can be unreliable even in the absence of noise effects. Moreover, online retrieval usually involves extremely complex computations due to the high degree of dimensionality of the feature vectors.

2.4 An Overview of Boundary-Based Shape Descriptors

The previous sections have provided a review of several descriptors that are based on the internal content of the shape. However, in many applications, the internal content of the shape is not as important as its boundary. Boundary-based techniques tend to be

36

more efficient for handling shapes that are describable according to their contours [40]. The literature has reported on a variety of techniques for describing objects based on the boundaries of their shape. Two main techniques are used to derive shape descriptors from shape boundaries: global and structural [15]. Global techniques use the entire contour from which to derive descriptors, while structural techniques are based on breaking down the contours into boundary segments, called primitives [15].

Global boundary techniques use the entire boundary of a shape to compute a feature vector. The descriptors are thus often compact, and the matching of shapes is a straightforward process, which can be achieved using a simple metric distance, such as Euclidean distance or city block distance. Many global boundary descriptors are based on shape signatures, which are often used to reduce a 2-D shape boundary to 1-D functions, a process that facilitates the derivation of invariant shape features. Several shape signatures have been proposed, including complex coordinates, radial distance, chord-length distance, triangular centroid area, angular function, polar coordinates, angular radial coordinates, and the farthest point distance. A shape signature can be used directly in order to distinguish different shapes. However, to make the signatures rotation invariant, shift matching must be employed as the means of finding the best match between two signatures. Using shape signatures for matching has a high cost due to the complex normalization of rotation invariance [85]. Moreover, shape signatures are very sensitive to noise, and any small change in the boundary leads to a large error in the matching [85]. Therefore, the direct use of shape signatures is impractical.

As a solution to the above problems, the Fourier transform is used to derive shape descriptors from a signature. Descriptors that are derived in this manner are often called Fourier Descriptors (FDs), which are frequently based on the application of the Fourier transform, to 1-D signature functions. In many applications, Fourier descriptors have been

proven to perform better than other boundary-based techniques [52, 86-89]. Many shape signatures have been used in Fourier descriptor techniques. However, the complex coordinate is the one most frequently described in the literature. Recent work shows that, in shape-based image retrieval, the radial distance signature outperforms the complex coordinate and other signatures [85]. To increase the ability of Fourier-based techniques to capture local features, Eichmann *et al.*, [90] have used the Short Fourier Transform (SFT). The SFT is not suitable for image retrieval because the matching process using the SFT is computationally more expensive than ones that utilize traditional FDs. Invariance to affine transformation allows considerable robustness in the case of rotating shapes in all three dimensions. Arbter *et al.* [91, 92] used a complex mathematical analysis and proposed a set of normalized descriptors which are invariant under any affine transformation. Oirrak *et al.* [93] also used one-dimensional Fourier series coefficients to derive affine invariant descriptors. Zhang and Lu [85] have shown that although the affine Fourier descriptor [91] was proposed as a means of targeting affined shape distortion, it does not perform well on the standard affine invariance retrieval set of the MPEG-7 database because affine Fourier descriptors are designed to work on polygonal shapes under affine transformation and not for non-rigid shapes [85]. Rauber and Garçao [94, 95] described a 2-D form descriptor based on a normalized polar transform, called a UNL (Universidad Nova de Lisboa) transform. The boundary of the shape is mapped to normalized polar coordinates, and then the 2-D Fourier transform is applied in order to derive the shape descriptors. Rauber and Garçao reported that if part of an object is missing, the UNL transform is abruptly changed; the derived descriptors are thus sensitive to occlusion [95].

Most Fourier-based techniques utilize the magnitude of the Fourier transform and ignore phase information in order to achieve rotation invariance as well as to make the

descriptors independent from the starting point. However, Bartolini *et al.* [96] described a technique in which phase information is exploited. Recently, El-ghazal *et al.*[97, 98] described Curvature-Based Fourier Descriptors (CBFD) for shape retrieval. The invariant descriptors of the CBFD technique are derived from the 2-D Fourier transform of the curvature-scale image obtained from the contours of the image [97].

In general, the use of Fourier descriptors is a promising boundary-based approach for shape-based image retrieval because FDs are based on the well-known Fourier theory, making them easy to compute and simple to normalize and interpret. In addition, their computational efficiency and compactness are well suited for online image retrieval.

Moment-based techniques represent a generic approach, i.e., one that can be used to derive shape descriptions even if the shape is represented by its boundary. Therefore, all moment techniques can be used to derive shape descriptors from the boundary of a shape. When moment-based techniques are used, the shape is represented by its boundary, and only the pixels that correspond to the boundary points are used to derive moment descriptors. This method reduces the amount of data storage space and computation time required [99].

Boundary moments can be derived from a shape signature by obtaining the central moments of the signature [100]. Gonzalez and Woods [67] used this principle to develop another method of deriving moment descriptors from a shape signature. In their method, the amplitude of the shape signature is treated as a random variable, and an amplitude histogram is formed, from which the moments are then derived. Another technique that uses shape signatures to derive shape descriptors is the autoregressive modeling technique [101-104]. The autoregressive model is applied to the shape signature function, and the autoregressive model coefficients are then taken as shape descriptors. Sekita and *et al.* [104] conducted several experiments with the goal of comparing techniques based on the

39

autoregressive model and Fourier-based techniques. Their experimental results clearly show that for shape recognition, Fourier-based techniques outperform techniques based on the autoregressive model. The disadvantage of the autoregressive model technique is its lack of a straightforward way to determine the correct model order. Moreover, in the case of a complex shape, a small order is not sufficient for a description [15].

Multi-scale representation is a powerful method of representation that allows the handling of boundary characteristics at different scales. Asada and Brady [105, 106] introduced the concept of *curvature primal sketch*, which is a multi-resolution description of a shape based on curvature. Their *curvature primal sketch* is acquired from both a Gaussian filter and second derivatives of a Gaussian filter. They extracted what they called primitive events, such as corners, smooth joints, cranks, ends, and bumps. Mokhtarian and Mackworth [107] proposed the use of the curvature scale space (CSS) representation for extracting shape descriptors. The representation is computed by convolving a path-based parametric representation of the curve with a Gaussian filter as the standard deviation of the Gaussian filter varies from a small to a large value, and extracting the curvature zero-crossing points of the resulting curves [107]. The peaks of the CSS representation are used as descriptors. Recent research has focused on the *curvature scale space* (CSS) method since it has been selected for use in the contour-based shape descriptions of MPEG-7 standards. Many researchers have modified or compared with the CSS method [108-114]. El Rube and *et al.* [114, 115] proposed the Multi-Triangular Area Representation (MTAR) signature, which is computed by calculating the area formed by three points on the boundary of the shape. Rather than using a Gaussian filter, El Rube *et al.* achieved a multi-scale representation by using a wavelength transform. CSS and the MTAR technique are not rotation invariant unless a circular shift of the peaks is used in the matching stage. Poel *et al.* [116] employed a

correlation coefficient as a measure of similarity between two different CSS representations. Since the whole CSS representation is used in the matching process, the dimension of the feature vector becomes very high and thus impractical for indexing and retrieval. El-ghazal *et al*. [97, 98] recently proposed the use of a 2-D Fourier transform to capture dynamic changes in the curvature of the contour of the shape and to simplify the matching process.

Wavelet analysis is used as a powerful tool in several applications, including shape analysis and recognition [87, 117, 118]. Kunttu *et al.* [87] introduced multi-scale Fourier descriptors for shape-based image retrieval. These descriptors are presented in multiple scales by adopting the wavelet and the Fourier transform, which improves the shape-retrieval accuracy of traditional Fourier descriptors. The complicated matching scheme of wavelet-based shape descriptors makes them impractical for online shape retrieval [119].

Dynamic Programming (DP) is another recently adopted method of achieving a high accuracy rate using shape boundaries [96, 120-123]. Although DP-based techniques generally offer better performance than other techniques that do not use DP, they have the disadvantage of being computationally expensive, making them impractical for large databases.

Several structure-based descriptors have been proposed for describing shape contours. In general structure-based descriptors are based on polygonal approximation and boundary decomposition [124-127]. One of the most well-known structural-boundary representations is the chain code, also known as the freeman code [128, 129]. The chain code is defined as the direction of the object's contour from a starting point. This representation is typically based on 4- or 8-connectivity, and the resulting chain code is a sequence of numbers. To make the chain code invariant with respect to rotation, the

difference of the chain code is used rather than the code itself. Then the difference is treated as a circular sequence, and the starting point is redefined so that the resulting sequence of numbers forms an integer of minimum magnitude, called the shape number. For object recognition, Iivarinen and Visa [130] proposed a chain code histogram that shows the probability of each direction, which reduces the dimension of the chain code representation. Bribiesca [125] proposed a new chain code, termed the vertex chain code (VCC). The vertex chain code indicates the number of cell vertices that are in touch with the bounding contour of the shape in that element's position. The advantage of the vertex chain code over a traditional chain code is that the shape number of the vertex chain code can be obtained directly by rotating the digits of the chain until the number is the minimum. Salem *et al.*[131] conducted several experiments in order to compare classical and vertex chain codes. Their results demonstrate that the vertex chain code recognizes shapes better than the classical chain code. Mehtre et al. [132] compared the retrieval ability of chain codes, Fourier descriptors, and moment-based descriptors. The worst performance was produced by chain codes. The disadvantages of using chain code representation is that it is very sensitive to boundary noise and variations and that it has a high dimension feature vector.

Groskey and Mehrotra [124] proposed a technique based on polygon decomposition. With this technique, the shape is broken into segments by polygon approximation. The polygon vertices are then used as primitives, and four features are extracted at each primitive. The similarity between any two features is the editing distance of the two feature strings. Berretti *et al.* [126] also proposed local shape descriptors based on partitioning the boundary of the shape into primitives. The zero-crossing points of the curvature from the Gaussian smoothed boundary are used to obtain primitives called tokens. Each token is described by two descriptors: its maximum

curvature and its orientation. An M-tree is used to index the tokens into the feature database. Two distance measures are defined which model the token similarity and shape similarity. The fundamental shortcoming of these descriptors is that they include curve orientation, which is not rotation invariant. Attalla and Siy [127] described a multi-resolution polygonal shape descriptor that is invariant to scale, rotation, and translation. To derive the shape descriptors, the contour of the shape is divided into equally spaced segments. An elastic comparison of the shape descriptors is then employed to measure the similarity of the features.

2.5 Boundary-Based Shape-Retrieval Descriptors

Many techniques have been proposed for describing images using the boundaries of the image. The following sections provide a description of commonly used boundary-based shape-retrieval descriptors and their advantages and disadvantages.

2.5.1 Simple Global Descriptors (SGDs)

The boundary of a shape can be described using scalar measures that are based on its simple geometrical characteristics. Brief descriptions of commonly used simple boundary descriptors follow.

A. Convexity (CX)

A convex hull is the minimal covering of an object. A shape convexity can be defined as the ratio of perimeters of the convex hull to that of the original contour (P_0). The convexity is represented as [133]:

$$CX = \frac{Convex\ Permiter}{P_o} \tag{2.56}$$

Examples for object convexity are provided in Figure 2.13.
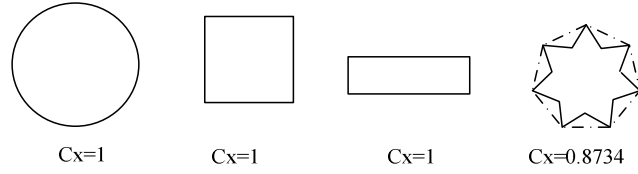
Figure 2.13 Examples of object convexities and solidities.

B. Bending Energy (BE)

Bending energy is the amount of energy required to transform a given closed contour into a circle with the same perimeter as the original contour.

$$BE = \frac{1}{N} \sum_{n=0}^{N-1} k(n)^2 \qquad (2.57)$$

N is the number of points of the contour, and k is the contour curvature, which can be calculated using Equation (2.60).

C. Aspect Ratio (AR)

The aspect ratio is defined as the shape's height divided by its width and is expressed as

$$AR = \frac{Height\ of\ the\ shape}{Width\ of\ the\ shape} \qquad (2.58)$$

2.5.2 Fourier Descriptor (FD)

The idea of a FD is to use the Fourier transformed boundary as a shape feature [52, 86, 134]. The Fourier transform is usually applied to shape signatures. Typically, a shape signature z(u) is any 1-D function that represents 2-D areas or boundaries. It usually describes a unique shape and captures the perceptual features of that shape.

Applying the Fourier Transform to Shape Signatures

The discrete Fourier transform (DFT) of a signature $z(u)$ is given as follows [67]:

$$a_n = \frac{1}{N} \sum_{u=0}^{N-1} z(u).e^{-j2\pi nu/N} \qquad , n = 0,1....N-1 \qquad (2.59)$$

44

The descriptors of a signature should be invariant for rotation, scale, and translation. Translation invariance can be achieved using the centroid of the shape. The rotation invariance of FDs is achieved by ignoring the phase information and by taking only the magnitude values of the FDs. For real-value signatures the scale invariant is achieved by dividing the magnitude of the Fourier descriptor by the DC components [135].

For complex-value signatures the DC component depends only on the position of the shape, so it cannot be used to describe shapes. For the scale normalization, the magnitudes of the other descriptors are divided by the second descriptor [52]. Figure 2.14 shows the main steps of generation of Fourier descriptors.
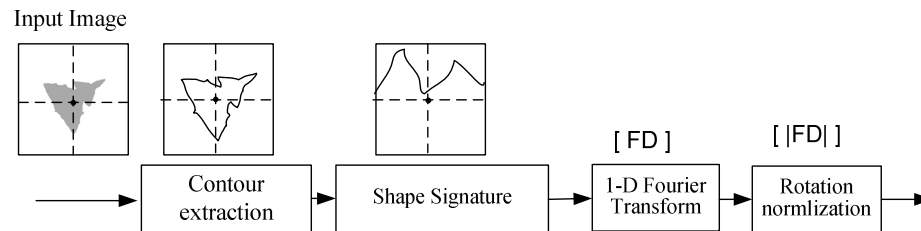
Figure 2.14 The basic steps in the generation of Fourier descriptors.

FDs are based on the well-known Fourier theory. The advantages of FDs compared to other shape descriptors is that they are easy to compute and simple to normalize [85]. The disadvantage of FDs is that each coefficient is calculated from each boundary point and therefore FDs perform poorly in the presence of occlusion.

2.5.3 Curvature Scale Space Descriptors (CSSDs)

The Curvature Scale Space (CSS) is based on the multi-scale curvature information of the inflection points of the contour as it evolves. In this technique, first an image contour is parameterized using an arc-length [107-110]. This is accomplished by sampling the contour at equal intervals, and recording the 2-D coordinates of each sampled point. The curvature is then derived from the shape boundary as follows [107]:

45

$$k(u) = \frac{\dot{x}(u)\ddot{y}(u) - \ddot{x}(u)\dot{y}(u)}{(\dot{x}^2(u) + \dot{y}^2(u))^{\frac{3}{2}}} \qquad\qquad (2.60)$$

In the next step, the curvature zero-crossing points are recovered and mapped to the Curvature Scale Space (CSS) image. The image boundaries are then convolved by Gaussian function of increasing width or standard deviation, and the zero crossing points are recovered and mapped to the CSS map at different values of the standard deviation. This process continues, and new zero crossing points are recovered and mapped until no curvature zero-crossing points are found. The result of the mapping is usually an interval tree, called a CSS map that consists of inflection points. The peaks that are more than 1/6 of the highest peak of the **CSS** map are extracted and used as descriptors to index the shape [40]. Figure 2.15 depicts an image and the peaks of its CSS representation.

A CSS descriptor is translation invariant. Scale invariance can be achieved by normalizing all the shapes into a fixed number of boundary points [102]. Rotation causes a circular shifting on the u axis, as shown in Figure 2.15.

A CSS descriptor is not robust in a global sense. Consequently, additional global descriptors such as eccentricity, aspect ratio, and circularity are used with CSS descriptors. CSS descriptors can fail to distinguish a shallow concavity from a deep concavity on the boundary of a shape. Abbasi and *et al.* proposed an enhanced CSS descriptor in order to distinguish a shallow concavity from a deep one [105].
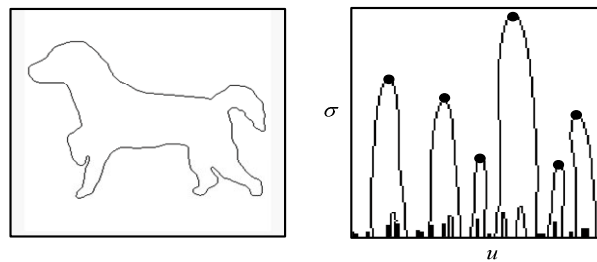
Figure 2.15 A CSS representation and its peaks.

2.6 Summary

A review of the literature related to shape-based image retrieval indicates that the number of types of shape descriptors for image retrieval has been rapidly increasing; however, each technique has a number of shortcomings.

The main advantage of structural techniques is their ability to handle the occlusion problem. However, this benefit comes at the expense of many drawbacks that make them unsuitable for shape-based image retrieval applications. First, shape partitioning leads to a complex matching process, which is the most important step in shape retrieval because it is an online process. Second, the structural approach does not capture the global features of shapes. Finally, there are no clear rules for obtaining the number of primitives required for each shape.

On the other hand, global techniques have many advantages which make them more effective and efficient for shape-retrieval applications. Global techniques are often based on mathematical transformations, which simplify the extraction of invariant features. Moreover, an inverse transformation can be used to determine the size of the feature's vector. When a global approach is used, features are extracted directly from the object without the use of shape decomposition. Global techniques can therefore be used in general applications. The matching process used in global techniques can often be accomplished through a simple procedure such as the Euclidean distance, which makes global techniques more efficient than structural ones.

For the above reasons and because of the disadvantages associated with structural techniques, global techniques are more effective and efficient than structural techniques for shape-retrieval applications.

The main drawback of Hu moments, Flusser moments, regular moments, complex moments, and Fourier-Mellin descriptors is that the basis used to derive them is not orthogonal; hence, these invariants suffer from a high degree of information redundancy. The main advantage of orthogonal moments is that they are based on orthogonal functions; hence, they have minimal information redundancy. Moreover, the selection of the required maximum order can be determined from the results of the image reconstruction.

The key advantage of ZMDs, PZMDs, ARTDs, TMDs, and FMDs is that they do not have a rotation problem. The direct scale normalization of the PZMD proposed by Chong *et al.* [79] is dependent on the order of the pseudo-Zernike polynomial: as the order becomes higher, the magnitude of the invariant moments is significantly diminished, and the dynamic range between invariant moments becomes very large. On the other hand, the direct scale normalization of ZMDs, ARTDs, and TMDs is not dependent on the order of the moments, and hence, the dynamic range between invariant moments is small.

The key problem associated with GDs is their problematic major axis normalization. The major axis is sensitive to noise and can be unreliable even in the absence of noise effects. Moreover, online retrieval usually involves a high level of computational complexity due to the high degree of dimensionality of the feature vectors.

The GFD is one of the most promising descriptors because the GFD is based on the well-known Fourier theory, which simplifies its implementation and the interpretation of the results. In addition, GFDs are well suited for online image retrieval because of their compactness and computational efficiency.

In the author's opinion, extensive experimentation should be conducted in order to investigate the performance of ZMDs, ARTDs, TMDs, and GFDs for shape-retrieval applications.

Fourier descriptors derived from different signatures can have significantly different effects on the results of the retrieval. Chapter 3 presents an investigation of a commonly used shape signature, which is compared with the proposed shape signature.

The advantage of a CSS descriptor is that it is compact. However, a CSS descriptor is able to detect only concavity and cannot represent the convexity of a shape. It therefore cannot distinguish between simple shapes that consist only of convexities, such as circles, triangles, or squares. Moreover, a CSS descriptor is not rotation invariant unless a circle shift is applied during the matching stage; hence, the online matching process with a CSS involves many schemes of circular shifts in order to align the peaks. Chapter 4 discusses CSS descriptors in greater depth and compares them with the proposed curvature-based Fourier descriptors.

Chapter Three

Farthest Point Distance: A New Shape Signature for Fourier Descriptors

3.1 Introduction

Shape description is an important task in content-based image retrieval (CBIR). The literature describes a variety of techniques that have been developed to represent objects based on their shapes. As explained in Chapter 2, each of these techniques has advantages and disadvantages. A Fourier descriptor (FD) is one of these techniques: a simple yet powerful technique that offers attractive properties such as rotational, scale, and translational invariance. Shape signatures, which constitute an essential component of Fourier descriptors, reduce 2-D shapes to 1-D functions and, hence, facilitate the process of deriving invariant shape features using the Fourier transform. A good number of shape signatures have been reported in the literature. These shape signatures lack important shape information, such as corners, in their representations. This information plays a major role in distinguishing between shapes.

This chapter presents the farthest point distance (FPD) signature, a novel shape signature that includes corner information in order to enhance the performance of shape-retrieval techniques that are based on Fourier descriptors. The signature is calculated at each point on the contour of a shape. This signature yields distances calculated between the different shape corners, and captures points within the shape on which human visual attention is focused in order to classify shapes.

In order to reach a comprehensive conclusion about the merit of the proposed signature, the signature was compared with eight popular signatures. Furthermore, the performance of the proposed signature was evaluated against the standard boundary-based descriptor:

the Curvature Scale Space (CSS) and the commonly used region-based descriptor: the Zernike Moments (ZMs).

3.2 Shape Signatures

A shape signature $z(u)$ is a 1-D function that represents 2-D areas or boundaries, usually describing a unique shape and capturing the perceptual features of the shape. Shape signatures are either real or complex. Many signatures have been proposed in the literature [17, 85, 86]. The complex coordinates (CC), the radial distance (RD), and the triangular centroid area (TCA) are notable signatures available for deriving FDs. Fourier descriptors derived from different signatures can produce significantly different retrieval results [85]. Brief descriptions of the most commonly used shape signatures are presented in the following sections.

3.2.1 Radial Distance (RD)

The Radial Distance (RD) represents the distance between the boundary points $(x(u),y(u))$ and the centroid (x_c,y_c) of the shape [85, 89]:

$$r(u) = \sqrt{((x(u) - x_c)^2 + (y(u) - y_c)^2)}$$
(3.1)

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

$$x_c = \frac{1}{N}\sum_{u=0}^{N-1} x(u) \quad , y_c = \frac{1}{N}\sum_{u=0}^{N-1} y(u)$$
(3. 2)

Other common names for this signature are the centroid distance and the radius vector. The basic concept of the radial distance signature is graphically depicted in Figure 3.1.b.

3.2.2 Chord-Length Distance (CLD)

The chord-length distance (CLD) is derived from a shape boundary without the use of a reference point. It is the distance between a and another boundary point b such that ab is

perpendicular to the tangent vector at *a*, as shown in Figure 3.1.c. In the case of two candidates, the one whose chord is within the shape is the one chosen [85].

3.2.3 Angular Function (AF)

The Angular Function (AF) $\varphi(u)$ represents changes in the direction of the boundary of a shape. These changes are important to the human visual system and can be used as a shape signature. The angular function $\varphi(u)$ at different points on the boundary of a shape is defined as [86]:

$$\varphi(u) = \arctan \left(\frac{y(u) - y(u - w)}{(x(u) - x(u - w)} \right) \tag{3.3}$$

where *w* is a step of selected length. The basic concept of the angular function signature is shown graphically in Figure 3.1.d.

3.2.4 Triangular Centroid Area (TCA)

The triangular centroid area (TCA), formed by two boundary points $(x_1(u), y_1(u)), (x_2(u), y_2(u))$ and the centre of the object, is changed along with changes in the points on the boundary. This area is denoted as a shape signature (Figure 3.1.e) and can be calculated as follows [85]:

$$TCA\ (u) = \frac{1}{2} \left| x_1(u)\, y_2(u) - x_2(u)\, y_1(u) \right| \tag{3.4}$$

3.2.5 Triangular Area Representation (TAR)

The triangular area representation (TAR) signature is computed by calculating the area formed by three points on the boundary of the shape [114]. The TAR signature is different from the TCA signature, which calculates the area formed by two boundary points and the centre of the object. In the TAR signature, the area of the triangle that has two sides of equal length *s* and is formed by three points P(u-s), P(u), and P(u+s) is calculated as follows[114]:

$$TAR(u,s) = \frac{1}{2}[-p_x(u)p_y(u-s) + p_x(u+s)p_y(u-s) + p_x(u-s)p_y(u)$$
$$- p_x(u+s)p_y(u) - p_x(u-s)p_y(u+s) + p_x(u)p_y(u+s)]$$

(3. 5)

When contours are traced in a counter-clockwise direction, convex, concave, and straight lines have positive, negative, and zero areas, respectively [114]. Figure 3.1.f depicts these three different types of area for a TAR signature.

3.2.6 Complex Coordinates (CCs)

The complex coordinate (CC) signature is formed by treating each coordinate pair ((x(u), y(u)) , u=0,1,2…..,N) of pixels on the xy-plane of a particular shape as a complex number, as follows [67, 89]:

$$CC(u) = (x(u) - x_c) + j(y(u) - y_c)$$

(3. 6)

The complex coordinate is translation invariant because of the subtraction of the centroid from the boundary coordinates of the shape. This signature is also frequently called a position function. The basic concept of the complex coordinates signature is shown graphically in Figure 3.1.g.

3.2.7 Polar Coordinates (PCs)

The polar coordinate (PC) signature is formed by combining the radial distance signature RD (u) and the polar angle $\theta(u)$ signature, as shown in Figure 3.1.h. The result of the combination is another complex-value signature, called the polar coordinate (PC) signature [17]:

$$PC(u) = RD(u) + j\theta(u)$$

(3. 7)

3.2.8 Angular Radial Coordinates (ARCs)

The angular radial coordinate (ARC) signature is similar to the polar coordinate signature; however, the angular function $\varphi(u)$ is used rather than the polar angle [17]:

$$ARC(u) = RD(u) + j\varphi(u) \qquad \text{(3. 8)}$$

Figure 3.1.i depicts the basic concept behind the generation of this signature.

(a) The Original Shape   (b) Radial Distance *RD(u)*   (c) Chord Length Distance ,*CLD(u)*

(d) Angular Function, *AF(u)*   (e) Triangular Centroid Area, *TCA(u)*   (f) Triangular Area Representation, *TAR(u)*

(g) Complex Coordinates, *CC(u)*   (h) Polar Coordinates, *PC(u)*   (i) Angular Radial Coordinates, *ARC(u)*
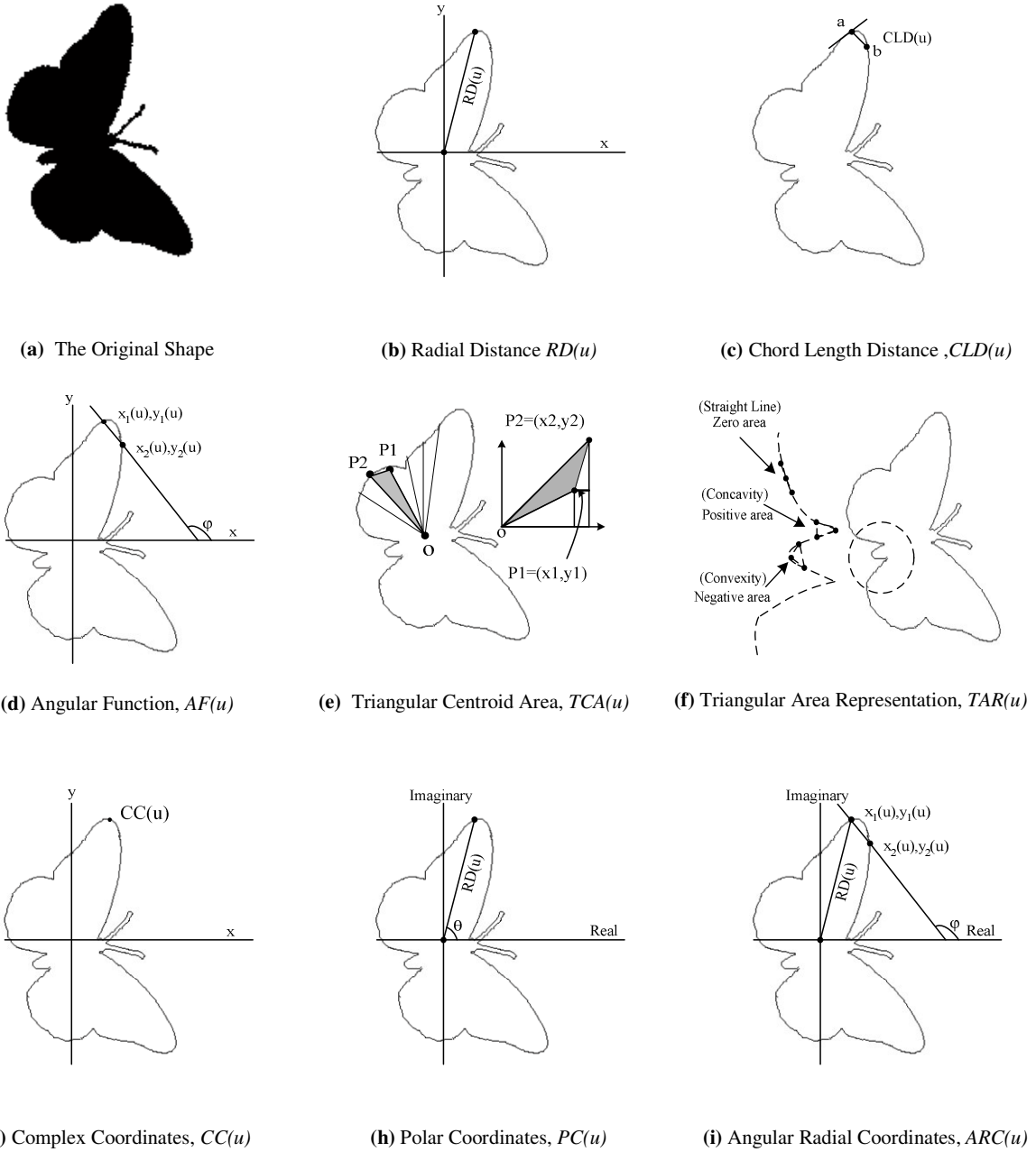
Figure 3.1 The basic concepts of eight commonly used signatures.

3.3 The Farthest Point Distance (FPD) Signature

This section presents the farthest point distance (FPD) signature, a novel technique that exploits the differential properties of shapes, such as corner points and transition details.

The FPD was developed to overcome some of the drawbacks of existing techniques, such that they ignore the distances between corners. The value of the signature at a given point *a* is defined as the distance between *a* and the point farthest from it, say *b*. The signature is calculated by adding the Euclidean distance between point *a* and the centroid *c* to that between the centroid *c* and the farthest point *b*. Assuming that shape coordinates *(x(u),y(u)),u=0,1,…N-1* of the boundary have been extracted, the FPD signature at boundary point *(x(u),y(u))* can be calculated as follows [136]:

$$FPD(u) = \sqrt{([x(u)-x_c]^2 + [y(u)-y_c]^2)} + \sqrt{([x_{fp}(u)-x_c]^2 + [y_{fp}(u)-y_c]^2)} \qquad (3.9)$$

where *(x_{fp}(u), y_{fp}(u))* is the farthest point from *(x(u),y(u))*, and *(x_c, y_c)* is the centroid of the shape.

Figure 3.3 shows how the distance from point *a* to its farthest point *b* is calculated. This signature captures distances between corners. Transition points and corners are elements of focus to the human visual system, and corner points play a major role in most shape-matching techniques. Figure 3.3 provides examples of FPD signatures for three randomly selected classes. It can be seen that shapes within the same class have similar FPD signatures.
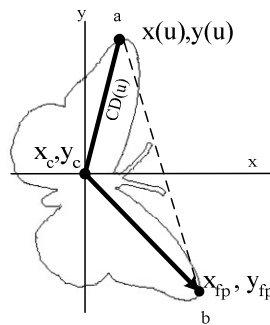
Figure 3.2 The basic concept of the farthest point distance (FPD) signature
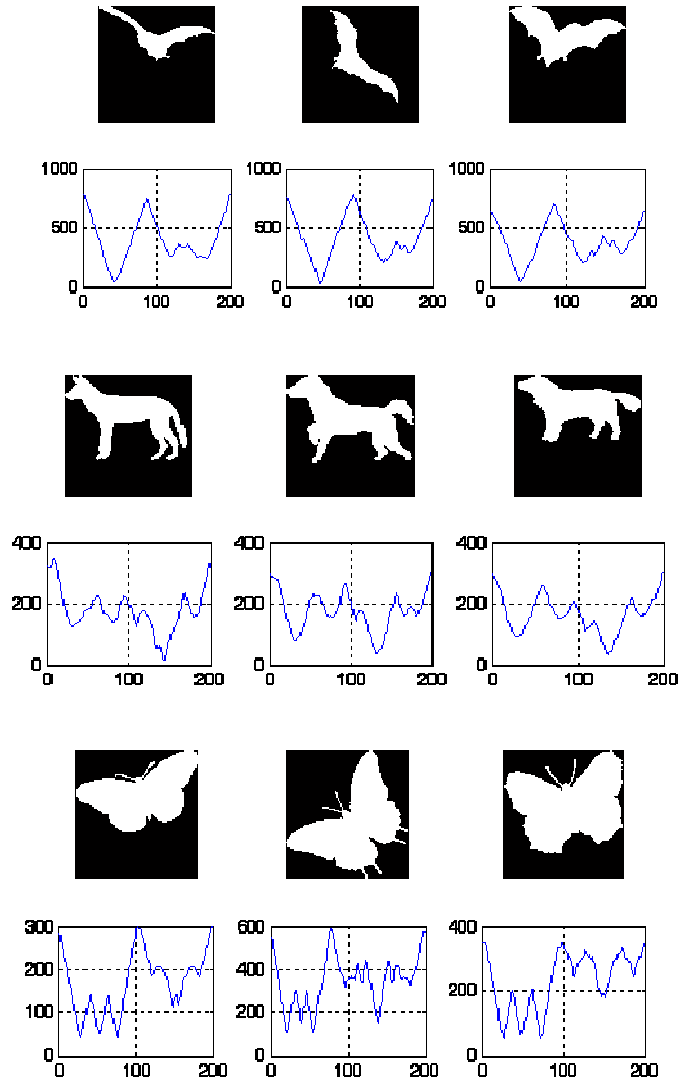
Figure 3.3 The FPD signatures for shapes from three randomly selected classes.

3.4 Generation of Fourier Descriptors

The direct use of shape signatures in a spatial domain leads to a high matching cost as a result of the complex normalization of the rotation invariance [96][123] . Therefore, the discrete Fourier transform (DFT) is used to simplify the matching stage in the retrieval process and to reduce the noise sensitivity of the signature used. The Fourier descriptor (FD) is a powerful tool for shape analysis that has many applications [51] [83, 84, 92]

[137] [138]. The idea of the FD is to use the Fourier transformed boundary as a shape feature [52, 86, 134]. The discrete Fourier transform of an arbitrary signature $z(u)$ is defined as follows [67]:

$$a_n = \frac{1}{N} \sum_{u=0}^{N-1} z(u) \, e^{-j2\pi nu/N}$$

(3. 10)

where $n = 0,1,2,\dots N-1$

The coefficients a_n ($n=0,1,..N-1$) are called the Fourier descriptors (FDs) of the shape, and are denoted by FD_n. Fourier descriptors are invariant to rotation, scale, and translation. Their rotation invariance can be established by taking into consideration the magnitude values of the descriptors and ignoring the phase information. Their scale invariance for real-valued signatures can be achieved by dividing the magnitude of the first half of the descriptors by the FD_0 [135]:

$$F = [\frac{|FD_1|}{|FD_0|}, \frac{|FD_2|}{|FD_0|},\dots, \frac{|FD_{N/2}|}{|FD_0|}]$$

(3. 11)

The reason for choosing the FD_0 as a factor in scale normalization is that it represents the average energy of the signature. Moreover, the FD_0 is often the largest coefficient, and consequently the range of the values of the normalized descriptors should be [0 1] [85].

For complex-value signatures, the DC component depends only on the position of the shape, so it cannot be used to describe shapes. For scale normalization, the magnitude of the other descriptors are divided by the FD_1 descriptor, as follows [52]:

$$F = [\frac{|FD_2|}{|FD_1|}, \frac{|FD_3|}{|FD_1|},\dots, \frac{|FD_{N-1}|}{|FD_1|}]$$

(3. 12)

The FPD signature is translation invariant because it is obtained with respect to the centroid of the shape. Proving that all the other signatures are translation invariant is a

straightforward process. A variety of signatures are systematically reviewed and their invariant properties are summarized in [139].

The measure of the similarity between two shapes indexed with M normalized Fourier descriptors is the Euclidean distance D between the normalized Fourier descriptors of the query image F^q and the normalized Fourier descriptors of an image from the database F^d :

$$D(F^q, F^d) = \sqrt{\sum_{i=1}^{M} (f_i^q - f_i^d)^2} \qquad (3.13)$$

3.5 Comparative Study

To obtain a comprehensive evaluation of the proposed signature, it was compared with eight signatures commonly used to derive Fourier descriptors. Moreover, the proposed signature was compared with the two notable techniques: the Curvature Scale Space (CSS) and the Zernike Moments (ZM). These two techniques are used in our comparison because in MPEG-7 standard, the curvature scale space (CSS) has been adopted as the contour-based shape descriptor and in recent work it has been shown that the Zernike moment descriptors (ZMDs) outperform many region-based descriptors [38, 40]. Moreover, the proposed and the Zernike moment descriptors use the same simple matching distance measure in order to retrieve shapes.

3.5.1 MPEG-7 Databases

Due to the lack of a standard database, the evaluation of shape signatures is not an easy task. Researchers in this field tend to develop their own databases, which are often limited in size or application scope or both. The MPEG-7 developers have set up a database of a reasonable size and generality [40]. It consists of three main sets: set A, set B, and set C. Set A consists of two subsets, A_1 and A_2, which are used to test invariance to scaling and rotation, respectively. Subset A_1 includes 420 shapes: 70 primary shapes and 5 shapes

derived from each primary shape with scale factors ranging from 0.1 to 2. Subset A_2 includes 420 shapes: 70 primary shapes and 5 subset shapes generated by rotating the primary shape with angles ranging from 9° to 150°. Sample shapes from set A of the MPEG-7 database are shown in Figure 3.4.

Set B consists of 1400 images that are classified into 70 classes, each class having 20 images. Set B is used to test for similarity-based retrieval performance, and to test the shape descriptors for robustness to various arbitrary shape distortions that include rotation, scaling, arbitrary skew, stretching, defection, and indentation. Samples of shapes from set B of the MPEG-7 database are shown in Figure 3.5.

Set C consists of 200 affine transformed bream fish images and 1100 images of marine fish that are unclassified. The 200 bream fish images are frames extracted from a short video clip of a bream fish swimming. This set is used to test shape descriptors for robustness to non-rigid object distortions. The first frame of the video is usually used as a query, and the number of bream shapes in the top 200 retrieved shapes is counted. However, for the experiment conducted in the current research, the 200 beam fish were designated as queries in order to obtain a comprehensive comparison. Samples of the images from this database are shown in Figure 3.6.

In this study, a noise database (set D) is created in order to test the performance of the proposed technique in the presence of noise. This set consists of 420 shapes: 70 primary shapes and 5 shapes derived from each primary shape by adding random Gaussian noise to the boundary of the primary shape. The signal-to-noise ratios for the distorted shapes are 40, 35, 30, 25, and 20 dB. Sample shapes from the proposed set D database are shown in Figure 3.7.

9º 36º 45º 90º 150º

The rotated versions of the primary shape

A primary shape

0.1 0.2 0.25 0.3 2.0

The scaled versions of the primary shape
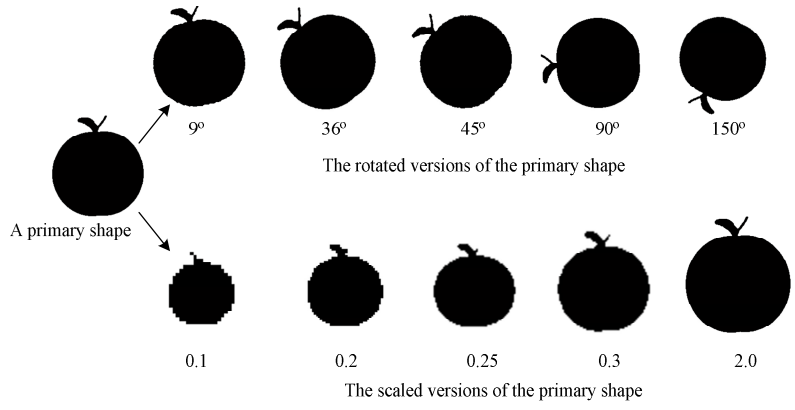
Figure 3.4 A primary shape from set A of the MPEG-7 database and its rotated and scaled versions.
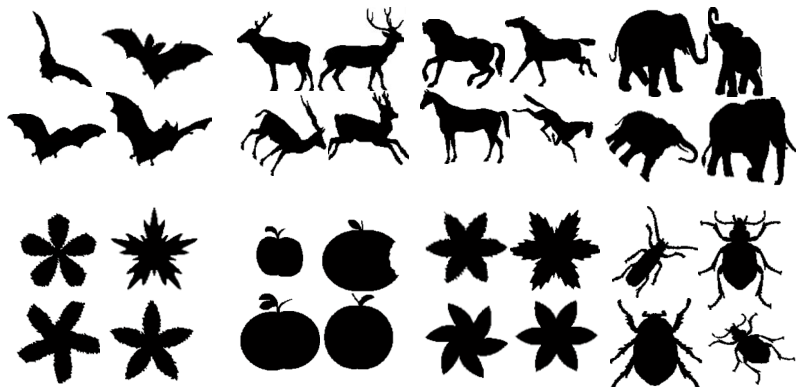
Figure 3.5 Samples of shapes from set B of the MPEG-7 database.

Bream-0 Bream-10 Bream-112 Bream-118 Bream-119 Bream-124 Bream-200
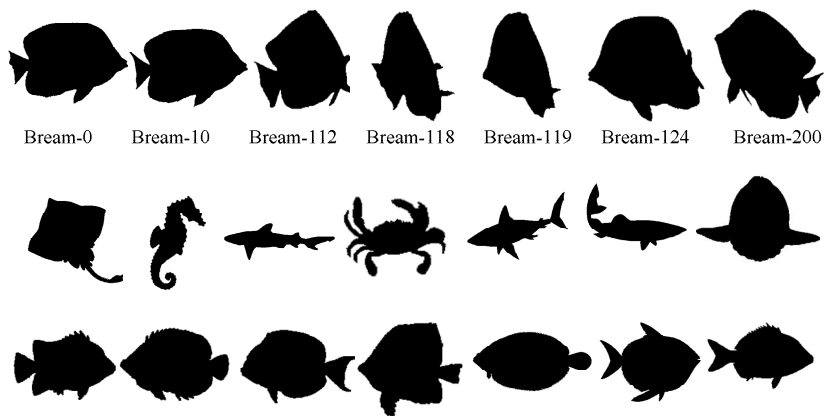
Figure 3.6 Samples of shapes from set C of the MPEG-7 database.

60

| The primary shape | 40 dB | 35 dB | 30 dB | 25 dB | 20 dB |

The distorted versions of the primary shapes
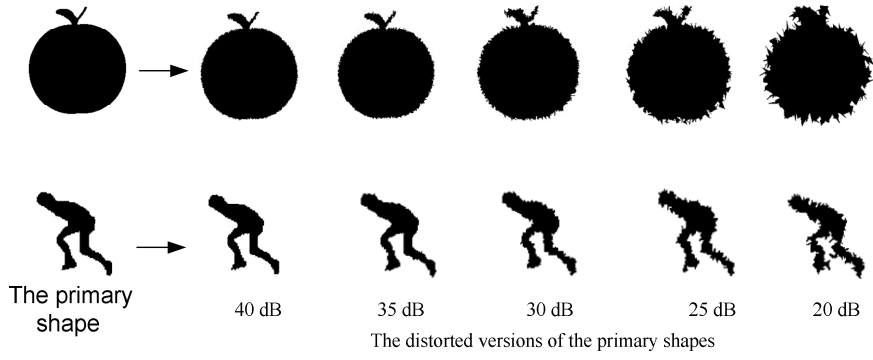
Figure 3.7 Samples from set D: the new database of noisy shapes.

3.5.2 The Performance Measure

To evaluate the performance of the techniques with respect to image retrieval, a performance measure is required. Precision and recall measures are the most commonly used and are deemed appropriate for measuring performance with respect to retrieval from classified datasets. If A is the number of relevant retrieved shapes, B the total number of retrieved shapes, and C the number of relevant shapes in the whole database used, then precision and recall are defined as A/B and A/C, respectively. Precision measures the retrieval accuracy, whereas recall measures the capability to retrieve relevant items from the database [85]. The precision value of a specific recall is the average of the precision values of all the database shapes for that recall.

3.5.3 Performance of the Farthest Point Distance Signature

3.5.3.1 Performance of the FPD Signature Compared with That of Other Signatures

The proposed farthest point distance signature was compared with the most popular and best-performing signatures [85], including the radial distance (RD), the triangular centroid area (TCA), the triangular area representation (TAR), the complex coordinate (CC), the chord-length distance (CLD) [85], the angular function (AF), the polar coordinate (PC) [17], and the angular radial coordinate (ARC) signatures [17]. Set B was

selected for evaluating the performance of the proposed signature against other signatures since it includes all possible situations with respect to shape distortion and variability.

To evaluate the performance of the proposed signature, experiments were conducted using set B of the MPEG-7 database. All shapes in the experiments were resampled to have 128 points. Selecting a small number of points will affect the retrieval performance for the proposed method as well as the other techniques. On the other hand, if the selected number of points is too large, the retrieval process will require more processing time and more storage space. The number of sampling points was carefully selected to be 128 because that number is the smallest rate that can be used without introducing sampling distortion. For consistency and fair comparison, the number of descriptors was limited to 63 for all methods.

Table 3.1 shows the average precision for low and high recalls for the FPD and other signatures using set B. From Table 3.1, it can be seen that the performance of the proposed FPD signature's is the highest and that the performance of the AF signature is the lowest. The RD and FPD signatures show comparable results in the case of low recall; however, in the case of high recall, the FPD performs better than the RD. This improvement is due to the tendency of the FPD to capture the farthest corners. The high performance of the FPD enables the retrieval of both complex and simple shapes.

The AF, PC, and ARC signatures do not perform as well as the FPD, RD, and ARC because $\varphi(u)$ and $\theta(u)$, which are utilized by the AF, PC, and ARC signatures, are very sensitive to changes in shape boundaries. The CC and TAR signatures capture the local information about the shape boundary, while the FPD, RD, and TCA signatures capture both local and global information. The advantage of the TAC signature over the FPD and RD signatures is its robustness to affine transform; nonetheless, the FPD and RD signatures outperform the TAC signature.

Table 3.1 The average precision for low and high recall for the FPD and other signatures using set B

Signature	Low Recall	High Recall
	The average precision for recall rates ≤ 50%	The average precision for recall rates > 50%
The proposed signature (FPD)	75.82 %	42.13 %
RD[89]	75.69 %	41.77 %
TCA[85]	73.40 %	38.50 %
CC[67]	64.76 %	22.59 %
PC[17]	64.40 %	35.12 %
ARC[17]	58.93 %	26.83 %
TAR[114]	58.70%	23.54%
CLD[85]	57.80 %	24.00 %
AF[86]	57.39%	27.88%

3.5.3.2 Comparison of the Performance of the FPD Signatures with That of the CSS and ZM Techniques

In another set of experiments, the proposed farthest point distance signature was combined with four simple global descriptors (SGDs): solidity (S), circularity (C), eccentricity (E), and aspect ratio (AR) [40]. The simple global descriptors enhance the ability of the proposed signature to capture global shape information. The distance (D_{FPD}), obtained by Equation (3. 13), of the Fourier descriptors obtained from the FPD signature is added directly to the average distance (D_{SGD}) of the simple global descriptors. The total distance between a query image q and an image from the database d is expressed follows:

$$D(q,d) = D_{FPD}(q,d) + D_{SGD}(q,d) \qquad (3.14)$$

where

$$D_{SGD}(q,d) = \frac{D_S(q,d) + D_C(q,d) + D_E(q,d) + D_{AR}(q,d)}{4} \qquad (3.15)$$

$D_S(q,d) = \dfrac{\left|S^q - S^d\right|}{\max(S^d)}$ is the solidity distance, $D_C(q,d) = \dfrac{\left|C^q - C^d\right|}{\max(C^d)}$ is the circularity distance, $D_E(q,d) = \dfrac{\left|E^q - E^d\right|}{\max(E^d)}$ is the eccentricity distance, and $D_{AR}(q,d) = \dfrac{\left|AR^q - AR^d\right|}{\max(AR^d)}$ is the aspect ratio distance.

The combined signatures are then compared with the Zernike moments [36] and the curvature scale space. In this case, the CSS is combined with the four simple global descriptors in order to maintain consistency [112]. Since the proposed signature and the CSS and ZM techniques utilize different criteria for normalizing shapes with respect to scale and rotation, all sets of the MPEG-7 database were used in order to obtain a comprehensive comparison.

The number of features for the CSS technique is not constant, while the number of features for the proposed method and the Zernike moments technique can be specified in advance. However, selecting a small number of features will affect retrieval performance for the proposed method and the Zernike moment technique. On the other hand, if the selected number of features is too large, the system will require more processing time and more storage space. Since the proposed signature is a real-value signature and all image contours in the database have been re-sampled to have 128 points, the maximum number of Fourier descriptors derived from the proposed signature is 63.

Zhang and Lu [85] found that 10 Fourier descriptors are sufficient for describing a shape. Thus, in the first experiment, the performance of the first 14 features of the

proposed technique (10 FDs plus 4 simple global descriptors) was compared with the performance of the Zernike moment technique with 14 features, which corresponds to Zernike moments up to the sixth order. Table 3.2 shows the number of features for Zernike moments at each order and the accumulated number of features up to each order. The recall precision curves of retrieval from set B obtained by the proposed and Zernike moment techniques for 14 features are shown in Figure 3.8. It is obvious from Figure 3.8 that the performance of the Zernike moment technique is much lower than that of the proposed technique. The performance of the Zernike moment technique can be improved by using a higher order of Zernike moments.

To select a number of features that provides a good compromise between retrieval performance and the dimension of the features, the average retrieval rates of the first top 20 shapes retrieved from set B of the MPEG-7 database were calculated at different numbers of features for the proposed and the Zernike moment techniques. Figure 3.9, which shows these average retrieval rates, reveals that the average retrieval rates of the proposed technique are higher than those of the Zernike moment technique for the same number of features. Moreover, it is interesting that the performance of the proposed technique does not significantly improve after the first 15 features, which include the four simple global descriptors. Figure 3.9 also reveals that no significant improvement in the performance of the ZM technique results after 28 features. To balance the number of features and the performance of the Zernike moment technique, the first 28 features, which correspond to the ninth order of Zernike moments, were used in the next experiments.
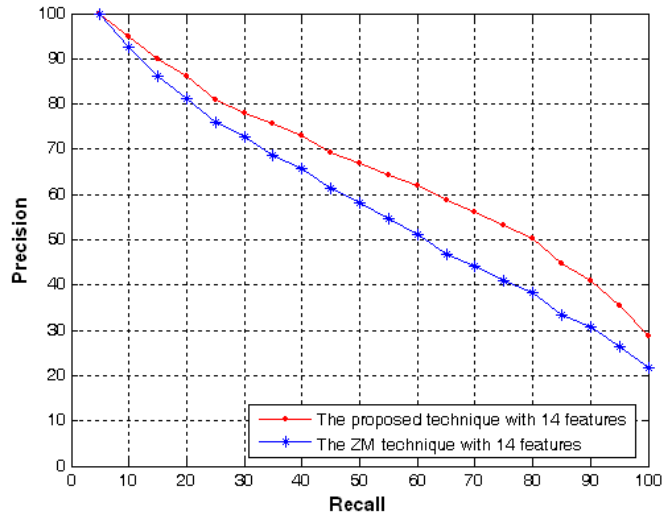
Figure 3.8 Precision-recall curves for the proposed and ZM techniques with 14 features.
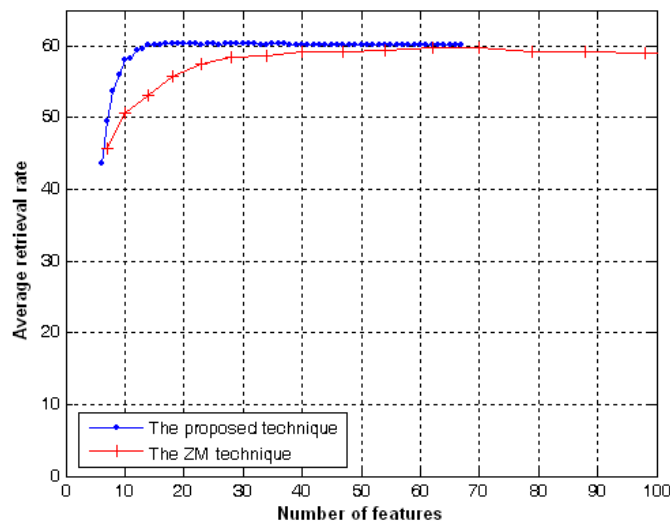
Figure 3.9 The average retrieval rates for different numbers of features for the proposed and the ZM techniques.

Table 3.2 The number of features for Zernike moments at each order and the accumulated number of features for each order.

Order (n)	Zernike moments of order n with repetition m (A_{nm})	The number of moments in each order n	The accumulated number of features up to each order
0	$A_{0,0}$	1	Not counted because it is used for scale invariance
1	$A_{1,1}$	1	Not counted because it is used for translation invariance
2	$A_{2,0}$, $A_{2,2}$	2	2
3	$A_{3,1}$, $A_{3,3}$	2	4
4	$A_{4,0}$, $A_{4,2}$, $A_{4,4}$	3	7
5	$A_{5,1}$, $A_{5,3}$, $A_{5,5}$	3	10
6	$A_{6,0}$, $A_{6,2}$, $A_{6,4}$, $A_{6,6}$	4	14
7	$A_{7,1}$, $A_{7,3}$, $A_{7,5}$, $A_{7,7}$	4	18
8	$A_{8,0}$, $A_{8,2}$, $A_{8,4}$, $A_{8,6}$, $A_{8,8}$	5	23
9	$A_{9,1}$, $A_{9,3}$, $A_{9,5}$, $A_{9,7}$, $A_{9,9}$	5	28

The recall and precision curves using the four sets from the MPEG-7 database and the created noisy database (set D) were plotted as shown in Figure 3.10 to 3.14, and the average precision rates for low and high recall are shown in Table 3.3 to 3.7. In addition, three different screen shots for three queries from set B of the MPEG-7 database are provided in Figure 3.15 in which the top left shape of each screen shot is the query shape and the "x" symbol is used to indicate an irrelevant shape.

The proposed signature outperforms the CSS technique, as can be determined from Figure 3.10 to 3.14. Figure 3.10 shows that the ZM technique and the proposed signature yield comparable results in the case of low recall; however, the proposed signature outperforms the ZM technique in the case of high recall with only 15 features. From Table 3.1 and Table 3.3, it can be seen that the performance of the Fourier descriptors has been improved by adding only four simple global descriptors. These simple global descriptors enhance the ability of the proposed signature to capture global shape information.

The performance of the techniques under scale invariance was tested using the database from subset A_1, for which the CSS technique has the lowest accuracies for both high and low recall, as shown in Figure 3.11, and the ZM technique has the highest accuracy for high recall. The FPD technique produces a result comparable to that of the ZM technique for low recall. The ZM technique gives the best result in the scale invariance test because scale normalization is implied when the image is limited to the unit circle.

As shown in Figure 3.12, the combined FPD signature and ZM techniques produce almost perfect results in the rotation invariance test, whereas the CSS produces the worst results. The CSS technique has compact features; however, its matching algorithm is very complex and fails to distinguish between objects within the same class but with different rotations.

In the case of the database of images with non-rigid object distortions (set C), all three techniques have high accuracies for low recall, as shown in Figure 3.13 and Table 3.6. However, in the case of high recall, the CSS technique has the lowest accuracy, while the FPD technique shows a result similar to that of the ZM technique.

In the case of noisy shapes (set D), it is clear that the performance of each technique is decreased when the signal-to-noise ratio is reduced, as shown in Figure 3.14. The proposed technique produces the best performance in all cases because the proposed technique uses a small number of Fourier descriptors (the first 11 descriptors) that correspond to low frequencies and ignores the higher order of Fourier descriptors that are more sensitive to noise.

In Figure 3.15, it is worth noting, not only that the proposed signature has better accuracy, but also that all irrelevant retrieved shapes have been ranked in the last row of each of the three examples.
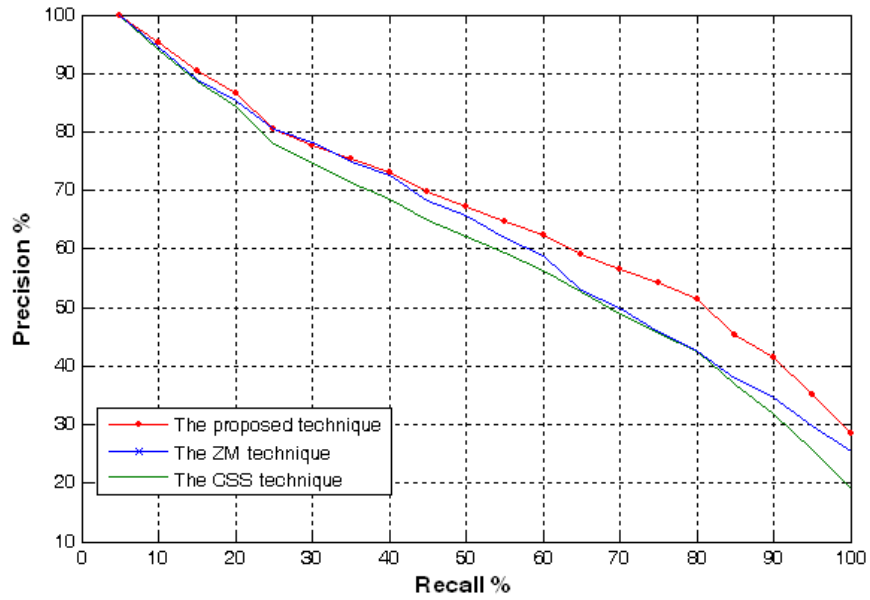
Figure 3.10 Precision-recall curves for the FPD+SGD, ZM, and CSS+SGD techniques using set B.

Table 3.3 The average precision rates of low and high recall for the proposed, ZM, and CSS techniques using set B

Method	Low Recall	High Recall
	The average precision for recall rates ≤ 50%	The average precision for recall rates > 50%
Proposed technique	81.16 %	49.15 %
ZM technique	80.88 %	43.94 %
CSS technique	78.62 %	41.81 %
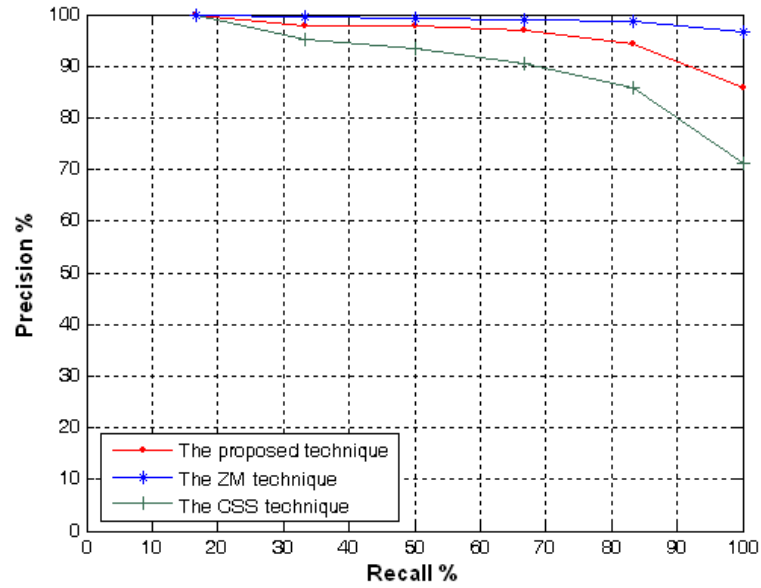
Figure 3.11 Precision-recall curves for the FPD+SGD, ZM , and CSS+SGD techniques
using subset A_1.

Table 3.4 The average precision rates of low and high recall for the proposed, ZM, and
CSS techniques using subset A_1

Method	Low Recall	High Recall
	The average precision for recall rates $\leq 50\%$	The average precision for recall rates $> 50\%$
The proposed technique	98.60 %	93.23 %
The ZM technique	99.69%	98.18 %
The CSS technique	96.18 %	82.52 %
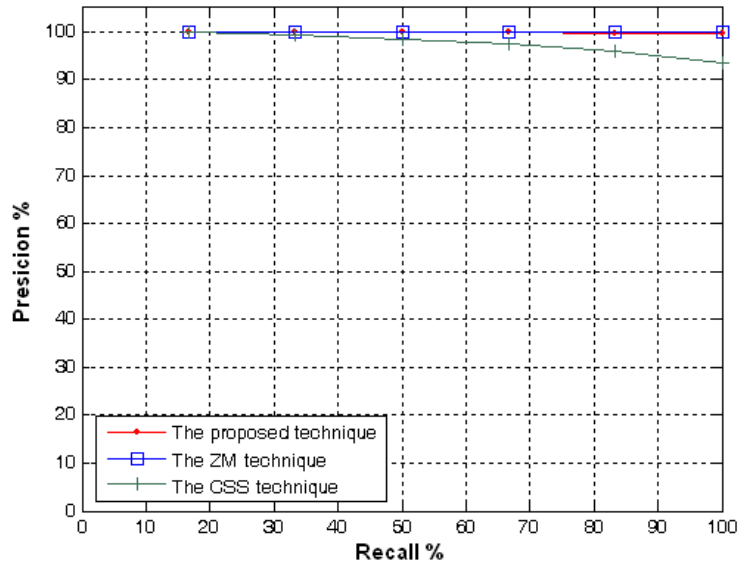
Figure 3.12 Precision-recall curves for the FPD+SGD, ZM, and CSS+SGD techniques using subset A_2.

Table 3.5 The average precision rates of low and high recall for the proposed, ZM, and CSS techniques using subset A_2

Method	Low Recall The average precision for recall rates ≤ 50%	High Recall The average precision for recall rates >50%
The proposed technique	100.00 %	99.89 %
The ZM technique	100.00 %	100.00 %
The CSS technique	99.25 %	95.55 %

71
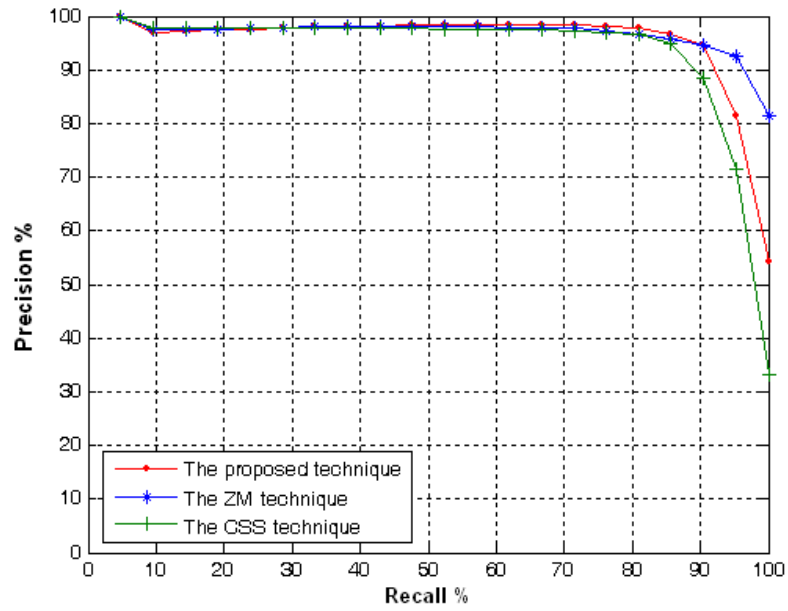
Figure 3.13 Precision-recall curves for the FPD+SGD, ZM, and CSS+SGD techniques using set C.

Table 3.6 The average precision rates of low and high recall for the proposed, ZM, and CSS techniques using set C.

Method	Low Recall	High Recall
	The average precision for recall rates $\leq 50\%$	The average precision for recall rates $> 50\%$
The proposed technique	97.96 %	92.23%
The ZM technique	98.02 %	95.25 %
The CSS technique	97.95 %	88.1 %
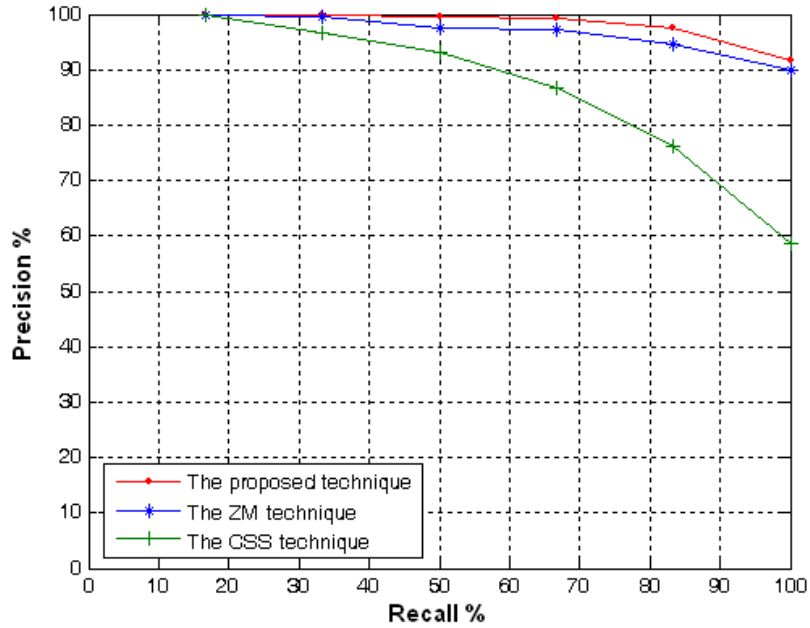
Figure 3.14 Precision-recall curves for the FPD+SGD, ZM, and CSS+SGD techniques using set D

Table 3.7 The average precision rates of low and high recall for the proposed, ZM, and CSS techniques using set D.

Method	Low Recall	High Recall
	The average precision for recall rates ≤ 50%	The average precision for recall rates > 50%
The proposed technique	99.85 %	96.08 %
The ZM technique	99.09 %	93.91 %
The CSS technique	96.65 %	73.77 %

To compare the computational efficiency of the three techniques, the processing time in the matching stage for set B of the MPEG-7 database was computed using the same processor and software for all three techniques. Matlab (version 7.0), running on a Pentium IV CPU 2.6 GHZ PC with a memory of 1.5 GB, was used as a testing platform. Table 3. 8 shows the number of features and the average processing time (the test was repeated 40 times) for each query in the matching stage for the proposed, ZM, and CSS techniques. Data in Table 3. 8 reveal that the proposed technique has the lowest feature

vector size and gives the least processing time, whereas the CSS was found to have the highest time. Because many factors must be considered in order to align two peaks of the CSS features, the time required by the CSS is high compared with that used by the proposed and ZM techniques.

Table 3. 8 The average time required for each query in the matching stage using set B of the MPEG-7 database

Method	Number of features	Average time required for each query
The Proposed technique	15	0.001404 (s)
The ZM technique	28	0.0017645 (s)
The CSS technique	Depends on the number of shape concavities (the average is 20)	2.1640 (s)
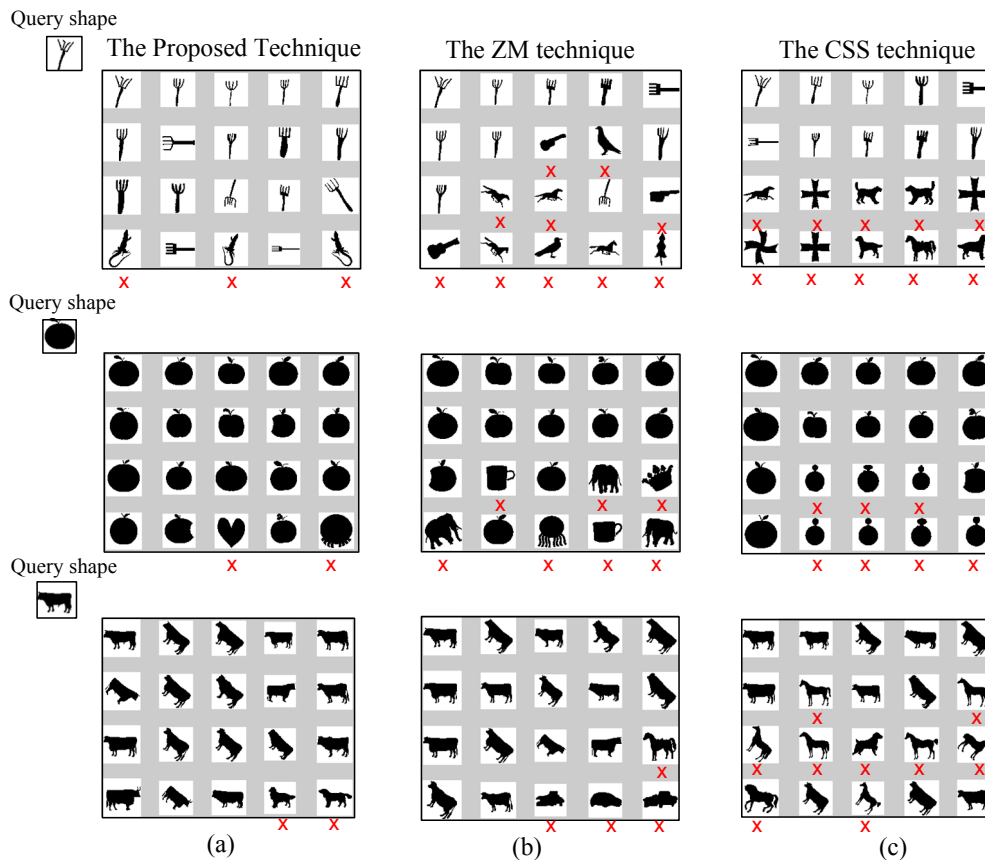
Figure 3.15 Retrieval of fork, apple, and cow shapes from set B of the MPEG-7 database: (a) Proposed technique (b) ZM technique (c) CSS technique

74

3.6 Summary

This chapter has presented a new shape signature for Fourier descriptors. The proposed signature has been evaluated by comparing it with several commonly used shape signatures. The performance of the proposed signature has been examined based on several experiments using standard databases.

The experimental results demonstrate that the proposed signature and the radial distance (RD) signature yield comparable results; however, the proposed signature (FPD) performs better in the case of high recall. This improvement is due to the fact that the FPD signature tends to capture corner information for each object, which are the extreme shape points on which visual attention naturally focuses.

Furthermore, the performance of the descriptors derived from the FPD signature was compared with that of two commonly used techniques: the Curvature Scale Space and the Zernike moment techniques. The results show that the proposed descriptors outperform the Curvature Scale Space technique with respect to both high and low recall for all sets of the MPEG-7 database. The results also show that the proposed signature performs better than the Zernike moment technique for low and high recall in the most challenging database (set B), while maintaining comparable results for both low and high recall in the other sets. Moreover, the feature size (15 descriptors) used by the proposed descriptors is almost half the size of the feature size (28 descriptors) used by the Zernike moment technique. This low feature size renders the proposed descriptors computationally more efficient for large databases. The proposed FPD technique satisfies the six principles set by MPEG-7: good retrieval accuracy, compact features, general application, low level of computation complexity, robust retrieval performance, and hierarchical coarse-to-fine representation. The main advantages of the Fourier descriptors derived from the proposed signature are their compactness and the simplicity of their matching process. However,

these descriptors are not capable of capturing the local characteristics of a shape. Consequently, another shape descriptor that is able to capture the local shape characteristics is proposed in Chapter 4.

Chapter Four

A Curvature-Based Fourier Descriptor for Shape Retrieval

4.1 Introduction

Fourier descriptors are considered to be promising descriptors as they are based on a sound theoretical foundation and also have the advantages of computational efficiency and attractive invariance properties. This chapter proposes a new curvature-based Fourier descriptor (CBFD) for shape retrieval. The proposed descriptor takes an unconventional view of the curvature-scale-space representation of the contour of a shape as it treats it as a 2-D binary image, hence referred to as a curvature-scale image, or CSI. The invariant descriptor is derived from the 2-D Fourier transform of the curvature-scale image. This method allows the descriptor to capture the detailed dynamics of the curvature of the shape and to enhance the efficiency of the shape-matching process. Experiments have been conducted using the widely known MPEG-7 databases in conjunction with a created noisy database in order to compare the performance of the proposed descriptor with that of six commonly used shape-retrieval descriptors: curvature-scale-space descriptors (CSSDs), angular radial transform descriptors (ARTDs), Zernike moment descriptors (ZMDs), radial Tchebichef moment descriptors (RTMDs), generic Fourier descriptors (GFDs), and 1-D Fourier descriptors (1-FDs).

4.2 Related work

The curvature-space-scale (CSS) technique is based on multi-scale curvature information obtained from the inflection points of the contour as it evolves. To obtain the curvature-scale-space map, an image contour is first parameterized using an arc length [107]. This step is accomplished by sampling the contour at equal intervals and recording

the 2-D coordinates of each sampled point $(x(s), y(s))$, $s=0,1,..L$. The curvature is then

derived from the shape boundary as follows [40]:

$$k(s,\sigma) = \frac{\dot{X}(s,\sigma)\ddot{Y}(s,\sigma) - \ddot{X}(s,\sigma)\dot{Y}(s,\sigma)}{(\dot{X}(s,\sigma)^2 + \dot{Y}(s,\sigma)^2)^{3/2}}$$ (4.1)

where $\dot{X}(s,\sigma) = x(s)*\dot{g}(s,\sigma),\ \ddot{X}(s,\sigma) = x(s)*\ddot{g}(s,\sigma)$

, $\dot{Y}(s,\sigma) = y(s)*\dot{g}(s,\sigma),\ \ddot{Y}(s,\sigma) = y(s)*\ddot{g}(s,\sigma)$, "*" is the convolution operator, and

$g(s,\sigma)$ is a Gaussian function of standard deviation σ, while $\dot{g}(s,\sigma)$ and $\ddot{g}(s,\sigma)$ are the

first and second derivatives of $g(s,\sigma)$, respectively.

After each convolution, the zero-crossing points of the curvature are recovered and mapped to the curvature-scale space. This process continues, and new zero-crossing points are recovered and mapped until no curvature zero-crossing points are found. The result of the mapping is usually an interval tree, called a CSS map, consisting of inflection points. The CSS technique is based on the extraction of only peaks that are more than 1/6 of the highest peak of the CSS map. These peaks are used as descriptors for indexing the shape. Figure 4.1 depicts a shape, a sample of the contours of the shape during the evolution process, the resulting CSS map, and the peaks of the CSS map.
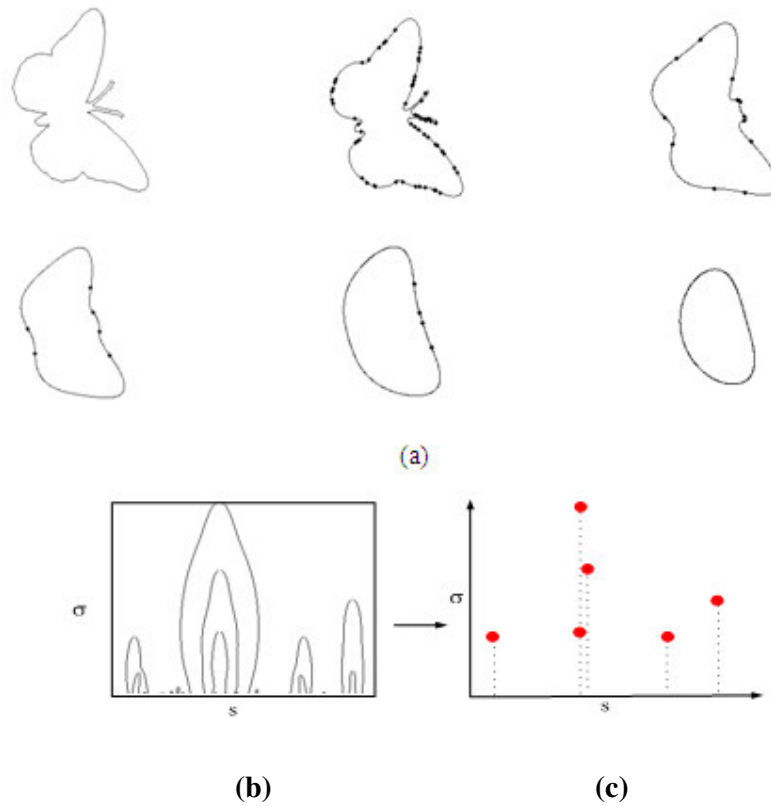
Figure 4.1 (a) A shape and samples of its contour during the evolution process, showing the zero-crossings (b) The resulting CSS map (c) The peaks of the CSS map

The CSS descriptor is translation invariant. Scale invariance can be achieved by normalizing all the shapes into a fixed number of boundary points [110]. Rotation causes a circular shifting on the *s* axis, as shown in Figure 4.2. To achieve rotation invariance, the highest peak is shifted to the origin of the CSS map. The similarities of the two shapes A and B are then measured by summing the differences between the peak values of all the matched peaks (peaks within a threshold value are considered to be matched), and the peak values of the unmatched peaks. To increase the accuracy of the results, other schemes of circular shifting matching are used. For example, rather than shifting the primary peak of A to match the primary peak of B, the primary peak of A can be shifted to match the secondary peak of B, or the secondary peak of A can be shifted to match the primary peak of B. The mirror shape has a CSS that differs from the original one, and

matching must include the mirrored shape as well. The details of the implementation of the CSS descriptor are described in [40].
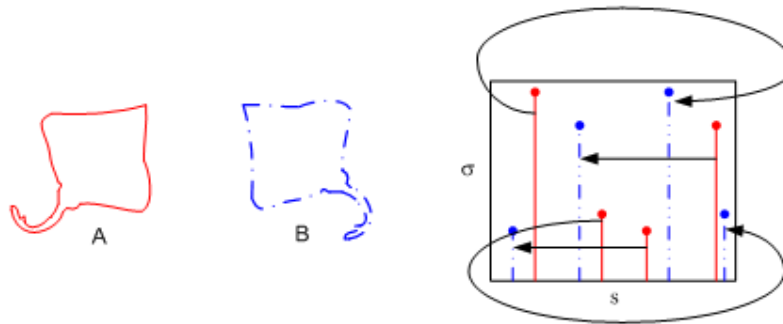
Figure 4.2 A shape before and after rotation and the peaks of the CSS map before and after rotation.

Although the CSS technique has many advantages, including its ability to capture local shape features and its robustness with respect to noise, it also has drawbacks. It uses only the peaks of the CSS map and ignores other important features, such as the dynamic changes in curvature during the process of extracting the zero-crossing points. This kind of feature might be a key factor in distinguishing between shapes. Moreover, the CSS descriptor is not rotation invariant unless a circular shift is applied in the matching stage, and hence, the online matching process of the CSS involves many schemes of circular shift in order to align the peaks [56, 140]. Another disadvantage of the CSS technique is that the number of peaks varies by shape, and these peaks are often mismatched and can be ordered quite differently. This fact adds to the complexity of the matching stage and increases the processing time required. These difficulties affect the overall performance of the CSS descriptors with respect to shape retrieval. The following section proposes a new curvature-based Fourier descriptor that will overcome the drawbacks of the CSS descriptors.

4.3 A Novel Curvature-Based Fourier Descriptor (CBFD)

This section proposes a new Fourier descriptor based on mapping the curvature-scale space of the contour of a shape into the 2-D Fourier domain in order to overcome the disadvantages of the well-known CSS descriptors, as described in the previous section.

The 2-D Fourier transform can be applied to any binary image $f(x,y)$ in the normal Cartesian coordinate system in order to extract Fourier descriptors, as follows:

$$F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y).\exp[-j2\pi(ux/M + vy/N)] \qquad (4.2)$$

The resulting descriptor, however, is not rotation invariant even though the phase of the Fourier transform is ignored. In other words, the 2-D Fourier transform produces different descriptors if a 2-D shape $f(x,y)$ is rotated by θ in the normal Cartesian space:

$$f(x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta) \leftrightarrow F(u\cos\theta + v\sin\theta, -u\sin\theta + v\cos\theta) \qquad (4.3)$$

Figure 4.3 shows the Fourier transform of a shape and its rotated version. According to Equation (4.3), the magnitudes of the Fourier transform for the original shape differs from that for its rotated version.

It is difficult to obtain rotation invariance of the features when the images are represented in the normal Cartesian coordinate system [56]. On the other hand, the 2-D Fourier transform has many properties that make it a useful tool for image processing and analysis. One of these is the fact that a shift in the normal Cartesian coordinate system affects only the phase of the Fourier transform:

$$f(x - x_0, y - y_0) \leftrightarrow F(u,v) \cdot \exp[-j2\pi(ux_0 + vy_0)] \qquad (4.4)$$

(a)

(b)

(c)

(d)
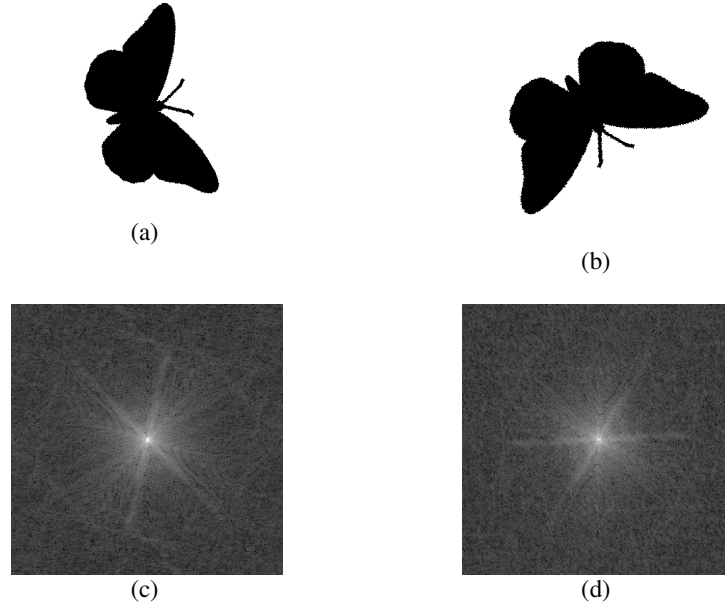
Figure 4.3 The magnitude of the Fourier transform for a shape and its rotated version.

Practical application of the 2-D Fourier transform requires a representation of the image that affects the phase of the 2-D Fourier transform when the original shape is rotated. To apply the 2-D Fourier transform, the procedure explained in Section 4.2 is used, and the contour of the original shape is represented by a binary image that has boundaries similar to those of the CSS map. This image will be referred to as a curvature-scale image (CSI), which can be obtained as follows:

$$CSI(s,\sigma) = \begin{cases} 0 & \text{if } k(s,\sigma) < 0 \\ 1 & \text{if } k(s,\sigma) \geq 0 \end{cases} \tag{4. 5}$$

where $k(s,\sigma)$ is the curvature of the contour given by Equation (4.1).

Figure 4.4.a shows the CSI of the shape presented in Figure 4.3.a. It can be seen that the boundaries of the CSI are exactly the same as those of the CSS map shown in Figure 4.1.b. Figure 4.4.b shows the CSI of a rotated version of the shape from Figure 4.1.a. It can be seen in Figure 4.4 that the effect of the rotation in the original shape corresponds to that of a circular shift of the CSI image. This property makes it possible to derive

rotation invariant descriptors using the 2-D Fourier transform and the property given in Equation (4. 4). To apply the 2-D Fourier transform, the CSI (s,σ) is treated as a two-dimensional rectangular binary image in the normal Cartesian coordinate system. Therefore, if the 2-D Fourier transform is applied to the CSI (s,σ), the effect of the phase on the resulting descriptor can be eliminated by taking only the magnitude and ignoring the phase of the 2-D Fourier transform of the CSI (s,σ). If the CSI (s,σ) is constructed from the contour of a binary image, then the 2-D Fourier transform of the CSI (s,σ) is defined as follows:

$$FCSI(u,v) = \sum_{t=0}^{m-1}\sum_{\sigma=0}^{n-1} CSI(s,\sigma).\exp[-j2\pi(us/m+v\sigma/n)] \qquad (4.\ 6)$$

where m is the maximum number of contour points and n is the maximum number of iterations needed in order to construct the CSI image using different standard deviations σ.
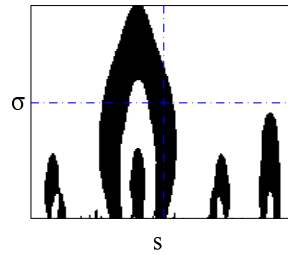
(a) CSI before rotation
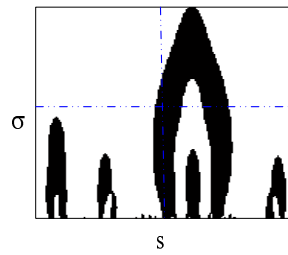
(b) CSI after rotation
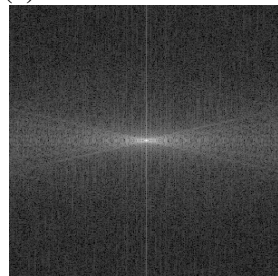
(c) Magnitude of the Fourier transform before rotation
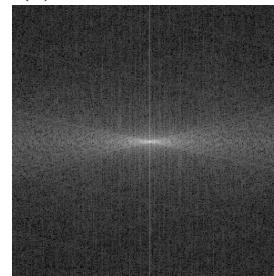
(d) Magnitude of the Fourier transform after rotation

Figure 4.4 The CSI images for both the image shown in Figure 4.3.a and its rotated version, and the magnitude of the Fourier transform for each.

The Fourier descriptors derived from FCSI (u,v) coefficients have to be invariant with respect to translation, rotation, and scaling. Given that the CSI (s,σ) image is translation invariant, it is obvious that the FCSI (u,v) coefficients also possess translation invariance. Rotation invariance of the FCSI (u,v) coefficients is achieved by ignoring the phase of the FCSI (u,v) coefficients. To achieve scale invariance, all FCSI (u,v) coefficients are divided by the magnitude of the first coefficient [56]. The reason for choosing the first coefficient as a factor in scale normalization is that it represents the average energy of the CSI image. Moreover, $|FCSI(0,0)|$ is often the largest coefficient, and consequently, the normalized descriptors are within [0 1] [56]. The proposed translation, scaling, and rotation invariance descriptors are as follows:

$$\text{FD} = \frac{|FCSI(0,1)|}{|FCSI(0,0)|}, \frac{|FCSI(0,2)|}{|FCSI(0,0)|}, ..., \frac{|FCSI(0,n)|}{|FCSI(0,0)|}, ..., \frac{|FCSI(m,0)|}{|FCSI(0,0)|}, ..., \frac{|FCSI(m,n)|}{|FCSI(0,0)|} \tag{4. 7}$$

The first descriptor is ignored because it is used for scale normalization. For efficient retrieval, only a few descriptors are used to index the shape. The similarity measure between two shapes indexed with N normalized Fourier descriptors is the Euclidian distance D_{FCSI} between the proposed normalized Fourier descriptors of the query image F^q and the normalized Fourier descriptors of an image from the database F^d, as follows:

$$D_{FCSI}(F^q, F^d) = \sqrt{\sum_{i=1}^{N}(f_i^q - f_i^d)^2} \tag{4. 8}$$

The proposed descriptor has many advantages over the CSS descriptor. First, the proposed descriptor uses all of the information from the curvature-scale image while the CSS technique uses only the peaks of the CSS map. Second, the number of features of the proposed descriptor can be determined in advance, and hence, the number of features for

different shapes can be equal. With the CSS technique, on the other hand, the number of features is not equal for different shapes. Third, the proposed descriptor possesses rotation invariance, so there is no need for the circular shifts used in the matching stage of the CSS technique, which translates into significant computational savings. With all of these advantages, the proposed descriptor clearly outperforms the CSS descriptor.

4.4 Combined Curvature Scale Image

The proposed descriptors cannot represent a convex shape because the CSI is constructed for a concave shape only. To overcome this drawback, the procedure for mapping a convex shape to a concave shape proposed in [116] is applied:

$$D_{x(s),y(s)} = \sqrt{(x_c - x(s))^2 + (y_c - y(s))^2} \qquad (4.9)$$

$$x_m(s) = (x(s) - x_c) \cdot \frac{2.R - D_{x(s),y(s)}}{D_{x(s),y(s)}} + x_c \qquad (4.10)$$

$$y_m(s) = (y(s) - y_c) \cdot \frac{2.R - D_{x(s),y(s)}}{D_{x(s),y(s)}} + y_c \qquad (4.11)$$

where

R: the circle that encloses the shape

(x_c, y_c): the centre of the circle

$(x_m(s), y_m(s))$: the coordinates of the mapped shape

The CSIs of the mapped shape and the original shape are combined into one image, referred to as the combined CSI image. The upper side of the combined CSI image represents the shape's concavities, and the lower side represents the shape's convexities. The proposed features are derived from the combined CSI image using the procedure explained in Section 4.3.

85

Figure 4.5 shows a convex shape (a square) and its mapped version along with the CSIs for both cases. It is clear that the combined CSI image makes it possible to distinguish between convex shapes such as a triangle and a square. Furthermore, it is noteworthy that for any convex shape, the upper side of the combined CSI image is represented by a blank image that has the same size as the lower side. The CSI for the shape shown in Figure 4.1.a, its CSI for the mapped version, and the combined version are shown in Figure 4.6.

(a) Original shape

(b) Mapped shape

(c) CSI image for the original shape

(d) CSI image for the mapped shape

(e) Combined CSI image
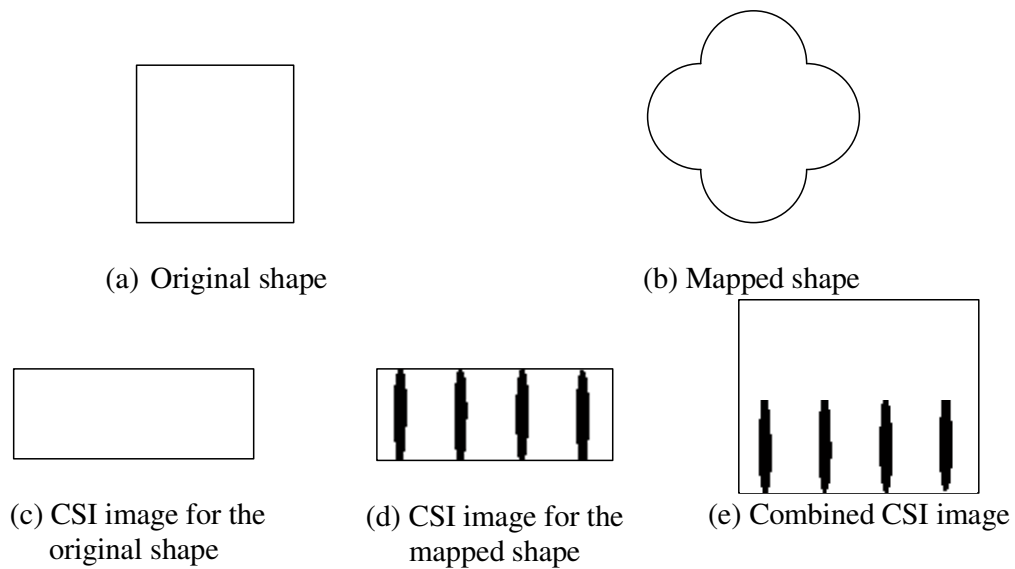
Figure 4.5 A square shape, its transformed version, the CSI for the square shape, the CSI for the mapped version, and the combined CSI image.

(a) CSI Image for the shape shown in Figure 4.1a

(b) CSI Image for the mapped version of the shape shown in Figure 4.1.a.

(c) Combined CSI Image
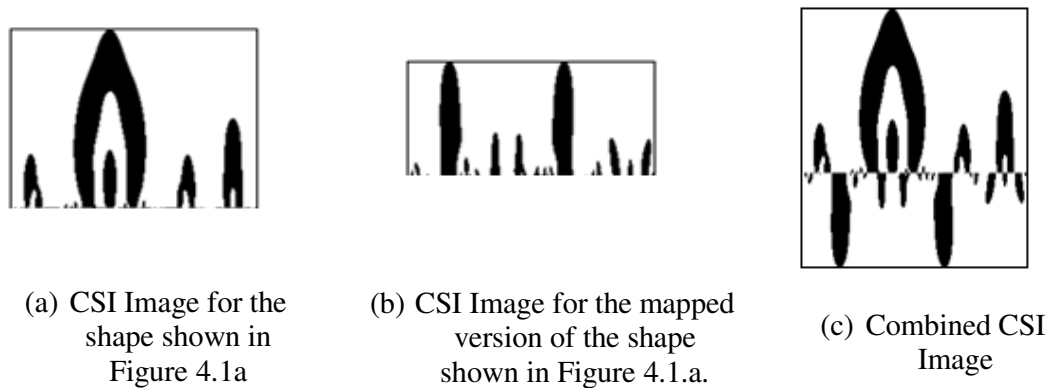
Figure 4.6 CSI for the shape shown in Figure 4.1.a, CSI for the transformed shape, and the combined CSI image.

4.5 Comparative Study

The theoretical properties of the proposed invariant descriptors are confirmed by the experimental results presented in this section. To obtain comprehensive results, the proposed descriptors were compared with six commonly used descriptors, the first two of which are the curvature scale space (CSS) and the radial angular transform descriptor (RATD) techniques. These techniques were used for the comparison because in the MPEG-7 standard, the curvature scale space (CSS) and radial angular transform descriptors have been adopted as the contour-based shape descriptor and the region-based shape descriptor, respectively [40]. The other comparative descriptors are the 1-D descriptor, Zernike moment descriptor (ZMD), radial Tchebichef moment descriptor (RTMD), 1-D Fourier descriptor (1-FD), and generic Fourier descriptor (GFD). These descriptors were adopted for the comparison because it has recently been shown that they are efficient and effective techniques for shape retrieval [13, 38, 56, 85, 89, 141]. Furthermore, both 1-FD and GFD techniques are based on a Fourier transform as are the newly proposed descriptors. In the case of the 1-FD, the centroid signature was utilized because recent work has proven that the centroid distance signature outperforms other signatures for shape retrieval [85].

4.5.1 Comparison of the Proposed Descriptors and Other Descriptors

The proposed descriptors (CBFDs) and the CSS descriptors are not robust in a global sense. Consequently, the proposed descriptors and the CSS descriptors were combined with four simple global descriptors (SGDs): eccentricity (E), aspect ratio (AR), circularity (C), and solidity (S) [40]. These simple global descriptors enhance the ability of the proposed descriptors to capture global shape information. The distance of the proposed descriptors(D_{CBFD}), obtained by Equation (4. 8) is added directly to the average distance

of the simple global descriptors (D_{SGD}) obtained by Equation (3. 15). The total distance between a query shape q and a shape from the database d is expressed as follows:

$$D(q,d) = D_{CBFD}(q,d) + D_{SGD}(q,d) \qquad (4.\ 12)$$

The number of features for the CSS descriptors is not constant, while the number of features for the proposed descriptors can be specified in advance. However, selecting a small number of features negatively affects the retrieval performance of the proposed descriptors. On the other hand, if the selected number of features is too large, the system requires more processing time and more storage space. Since the proposed descriptors derived from the 2-D Fourier transform there are two different parameters (frequencies) (u, v) that should be selected. To select theses parameters the average retrieval rates of the first top 20 shapes retrieved from set B of the MPEG-7 database were calculated at different frequencies and the result is shown in Figure 4.7. The experiment revealed that the 52 descriptors (u=4, v=13 according to Equation (4. 6)) of the 2-D Fourier transform of the CSI result in a performance superior to that of other descriptors. The first descriptor is not included because it is used for scale normalization. Thus, the total number of descriptors used for the proposed technique is 55 (51 Fourier descriptors plus 4 simple global descriptors). For consistency and fair comparison, the number of descriptors is limited to 55 for all other descriptors.
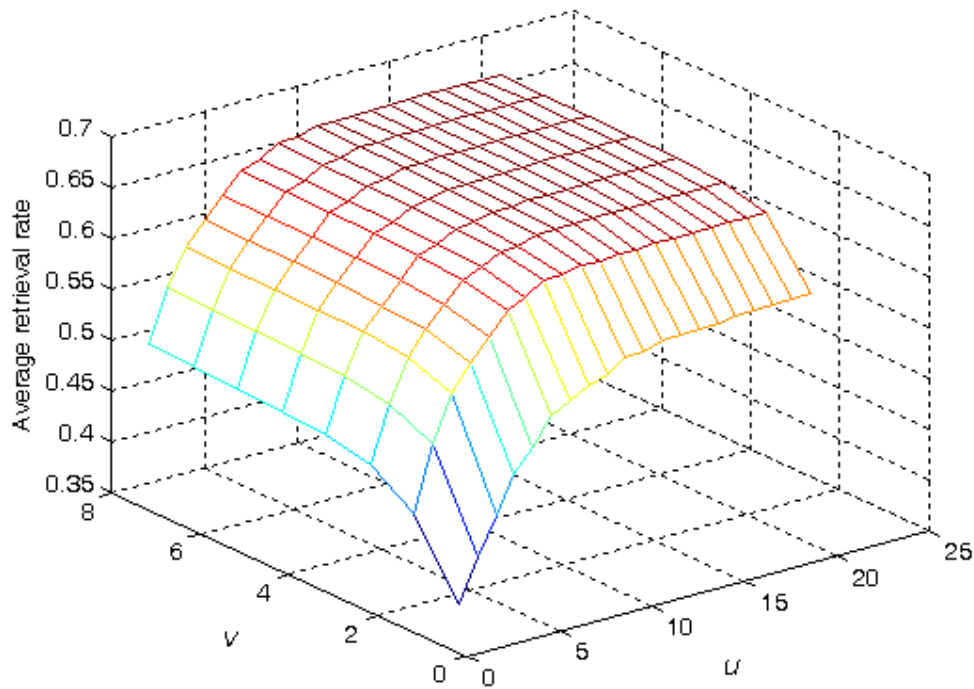
Figure 4.7 The average retrieval rate for set B at different frequencies: *u* and *v*.

The recall and precision curves using the four sets of the MPEG-7 database and the created noisy database are plotted in Figure 4.8 to Figure 4.12, and the average precision rates for low and high recall are shown in Table 4.1 to Table 4. 6. The proposed descriptors outperform the CSSD, as can be seen in Figure 4.8 to Figure 4.12 and they produce results comparable to those of the CSSD in the case of the database of images with non-rigid object distortion. The reason for the improvement is that the proposed descriptor uses all of the information from the curvature scale image (CSI) while the CSS descriptor uses only the peaks of the CSS map. Figure 4.8 shows that the proposed descriptors outperform all other descriptors with respect to both high and low recall. The CSSD, GFD, and RTMD yield comparable results in cases of low and high recall; however, their results are not as accurate as those of the ZMD and the ARTD. The ARTD, ZMD, RTMD, GFD, and 1-FD techniques often misclassify shapes that are globally similar but have different local characteristics. The proposed descriptors, on the

other hand, have a strong structure for capturing both local and global features using the curvature of the contour of the shape and the simple global descriptors, respectively.

The performance of the proposed descriptors under scale invariance was tested using the database from subset A_1, as shown in Figure 4.9. The RTMDs give the highest accuracy for both low and high recall. In the case of low recall, the proposed descriptors give results comparable to those of the ARTD, ZMD, and GFD. The GFD and the 1-FD give comparable results, which are considered very low in the case of high recall compared to the results from the other descriptors. In this test, region-based descriptors perform better than boundary-based descriptors because scaling a shape changes the local characteristics of the shape and keeps the global ones, which are the ones captured by region-based descriptors.

In the case of the rotation invariance test, as shown in Figure 4.10, all descriptors give perfect results except the CSSD and 1-FD. The CSS technique has compact features; however, its matching algorithm is very complex and fails to distinguish between objects within the same class but having different rotations.

In the case of the database of images with non-rigid object distortion (set C), as shown in Figure 4.11, all descriptors produce high accuracy for low recall. However, in the case of high recall, the results of the CBFD are not as accurate as those of the ARTD, ZMD, RTMD, GFD, and 1-FD because set C has many relevant shapes that are not visually close to the query shapes. Furthermore, it contains irrelevant shapes that are visually closer to the query shapes than are the relevant shapes. For example, the first three shapes in the last row of Figure 3.6 are more similar to the bream-0 fish than to the bream-118 or bream-119 fish shown in the first row in the same figure. This observation has also been mentioned by Latecki *et al.* [142].

In the case of noisy shapes, it is clear that the performance of each technique decreases when the signal-to-noise ratio is reduced, as shown in Figure 4.12. The proposed descriptors produce the best performance in all cases. It is noteworthy that with the proposed technique the boundary noise is filtered because of the smoothing process during the construction of the CSI map. Moreover, the noise is filtered because the higher order of the Fourier descriptors that are more sensitive to noise are ignored. These two filtering processes are the reason the proposed descriptors outperform the others.
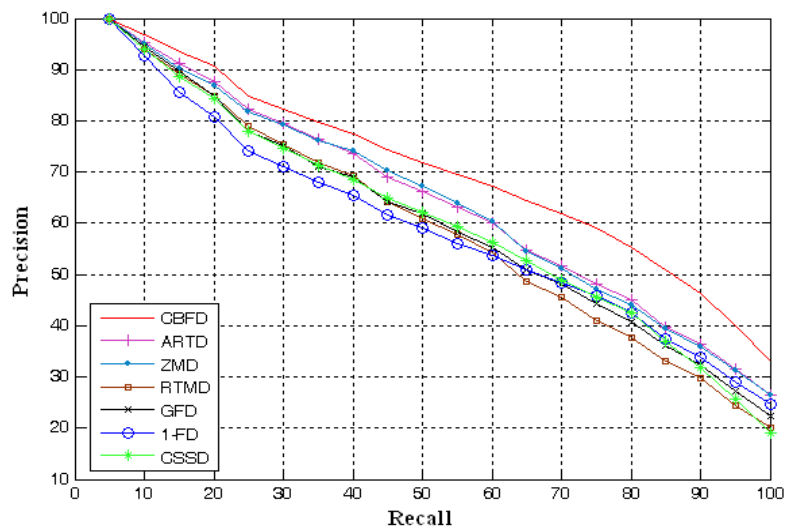
Figure 4.8 Precision-recall curves for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using set B.

Table 4.1 The average precision rates of low and high recall for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using set B.

Method	Low Recall The average precision for recall rates ≤ 50%	High Recall The average precision for recall rates > 50%
The proposed descriptors (CBFD)	85.15 %	54.77 %
ARTD	82.10 %	45.69 %
ZMD	82.11 %	45.42 %
RTMD	78.87 %	39.23 %
GFD	78.80 %	41.56 %
1-FD	75.83 %	42.25 %
CSSD	78.61 %	41.81 %
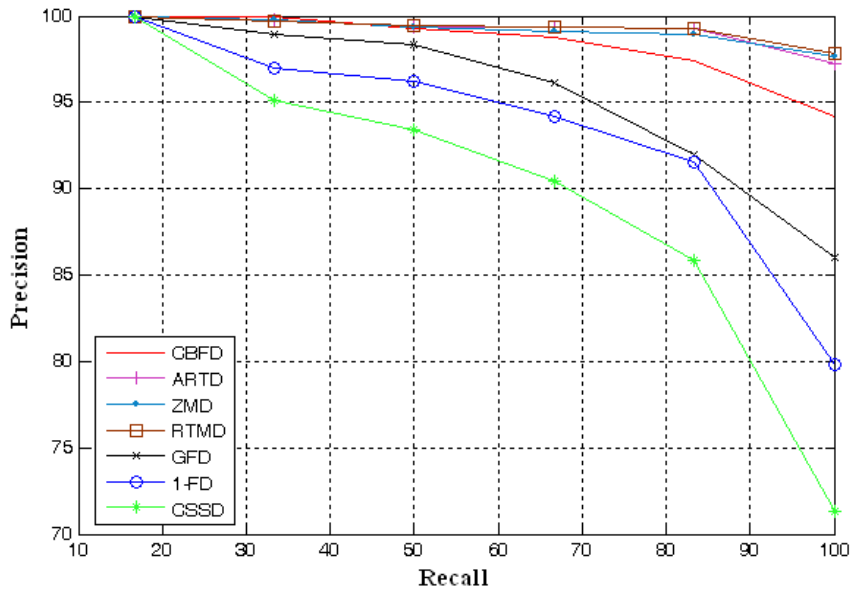
Figure 4.9 Precision-recall curves for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using set A1.

Table 4.2 The average precision rates of low and high recall for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using A1.

Method	Low Recall	High Recall
	The average precision for recall rates $\leq$ 50%	The average precision for recall rates > 50%
The proposed descriptors (CBFD)	99.74 %	96.78 %
ARTD	99.71 %	98.62 %
ZMD	99.71 %	98.59 %
RTMD	99.77 %	99.38 %
GFD	99.11 %	91.39 %
1-FD	97.76 %	88.50 %
CSSD	96.18 %	82.52 %
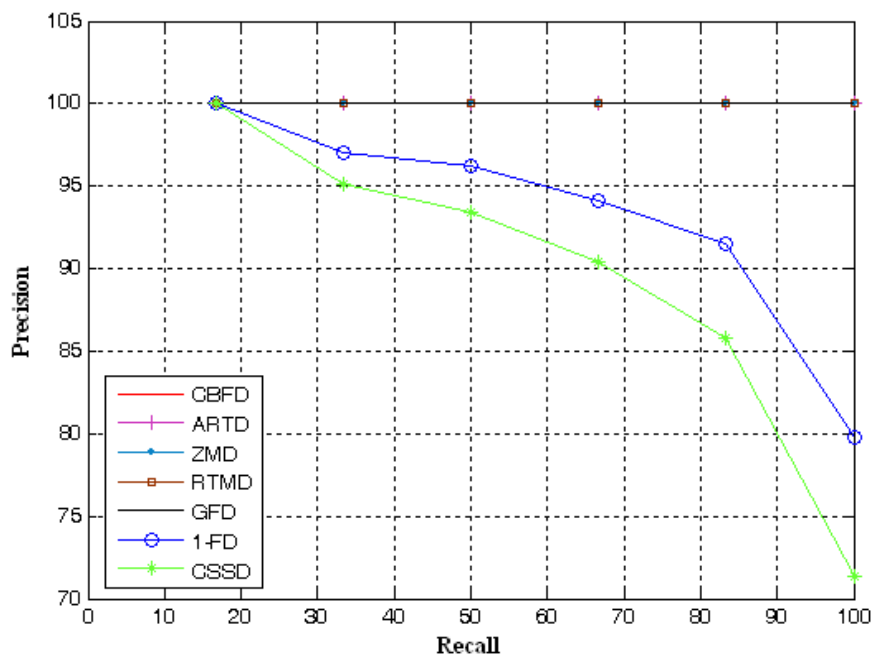
Figure 4.10 Precision-recall curves for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using set A2.

Table 4.3 The average precision rates of low and high recall for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using set A2.

Method	Low Recall	High Recall
	The average precision for recall rates $\leq 50\%$	The average precision for recall rates $> 50\%$
The proposed descriptors (CBFD)	100.00%	100.00%
ARTD	100.00%	100.00%
ZMD	100.00%	100.00%
RTMD	100.00%	100.00%
GFD	100.00%	100.00%
1-FD	100.00%	98.35%
CSSD	99.25%	95.55 %
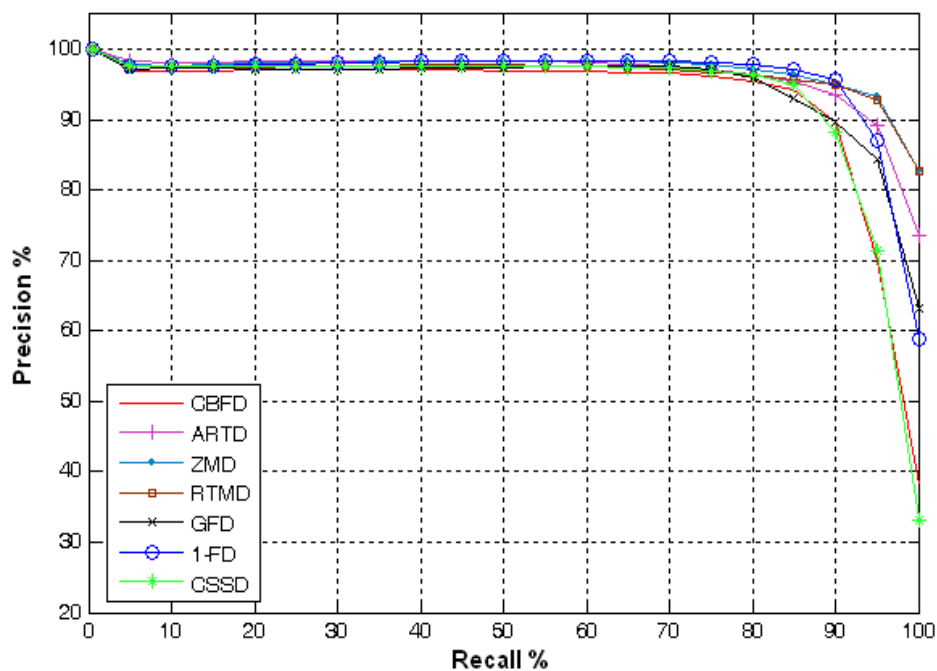
Figure 4.11 Precision-recall curves for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using set C.

Table 4.4 The average precision rates of low and high recall for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using set C.

Method	Low Recall	High Recall
	The average precision for recall rates $\leq$ 50%	The average precision for recall rates $>$ 50%
The proposed descriptors (CBFD)	97.38 %	88.03 %
ARTD	98.48 %	94.16 %
ZMD	98.32 %	95.85 %
RTMD	97.79 %	95.22 %
GFD	97.54 %	91.98 %
1-FD	98.15 %	93.34 %
CSSD	97.95 %	88.10 %
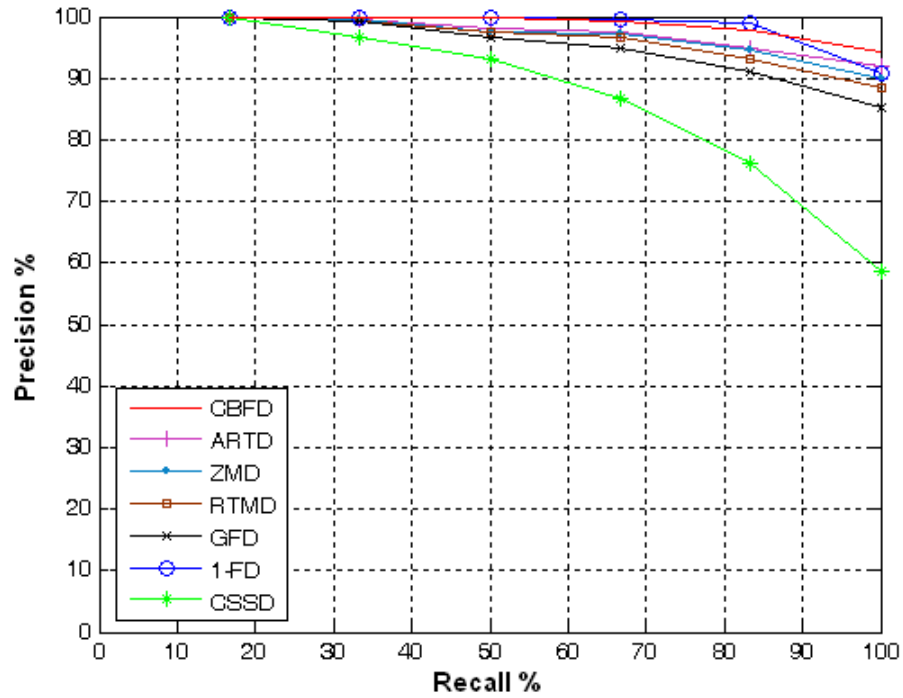
Figure 4.12 Precision-recall curves for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using set D.

Table 4.5 The average precision rates of low and high recall for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD using set D.

	Low Recall	High Recall
Method	The average precision for recall rates $\leq$ 50%	The average precision for recall rates $>$ 50%
The proposed descriptors (CBFD)	99.96 %	97.19 %
ARTD	99.75 %	94.75 %
ZMD	99.24 %	94.92 %
RTMD	98.99 %	92.83 %
GFD	98.66 %	90.38 %
1-FD	99.91 %	96.5 %
CSS D	96.65 %	73.77 %

4.5.2 Overall Performance Using the Mpeg-7 Database

The overall average performance of the proposed descriptors was compared with that of the other descriptors based on the following weighted average:

95

$$Weighted \quad Average = \frac{420}{2860}*A_1 + \frac{420}{2860}*A_2 + \frac{1400}{2860}*B + \frac{200}{2860}*C + \frac{420}{2860}*D \quad (4.\,13)$$

The weights are calculated based on the number of query images used to obtain the recall-precision curve for each set. Table 4. 6 shows the overall average performance of the CBFD descriptors compared to that of the ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD. Based on Table 4. 6, it is clear that, compared with other descriptors, the proposed descriptors provide the best performance with respect to both low and high recall.

Table 4. 6 The overall average precision rates of low and high recall for the CBFD, ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD.

Method	Low Recall	High Recall
	The average precision for recall rates $\leq 50\%$	The average precision for recall rates $> 50\%$
The proposed descriptors (CBFD)	92.50%	76.14%
ARTD	91.05 %	72.03 %
ZMD	90.97 %	72.04 %
RTMD	89.32 %	68.77 %
GFD	89.12 %	68.16 %
1-FD	87.70 %	68.82 %
CSSD	88.22 %	63.61 %

4.5.3 Complexity Analysis

To compare the computational efficiency of the five descriptors, the processing time in the matching stage for set B of the MPEG-7 database was computed using the same processor and software for all descriptors. Matlab (version 7.0), running on a Pentium IV CPU 2.6 GHZ PC with 1.5 GB of memory was used as a testing platform. The test was repeated 40 times, and

Table 4.7 shows the average processing time for each query in the matching stage for the proposed descriptors, as well as for the ARTD, ZMD, RTMD, GFD, 1-FD, and CSSD. The data displayed in

Table 4.7 reveal that the CSSDs require the highest time whereas the other descriptors have the same processing time because they all use the same matching procedure with the same number of features. The CSSD time is high compared to that of the proposed and the other descriptors because many factors must be considered in order for two peaks of the CSSD features to be aligned.

Table 4.7 The average time required for each query in the matching stage using set B of the MPEG-7 database.

Descriptor	The average time required for each query
The proposed descriptors (CBFD), ARTD, ZMD, RTMD, GFD, and 1-FD	0.0559 (s)
CSSD	2.16 (s)

4.6 Summary

This chapter has presented new curvature-based Fourier descriptors for shape retrieval. The proposed descriptors have been evaluated against six commonly used descriptors. Several experiments based on the MPEG-7 database were conducted in order to confirm the theoretical properties of the proposed descriptors.

The experimental results demonstrate that the proposed descriptors (CBFDs) outperform the CSS and 1-D Fourier descriptors for three sets from the MPEG-7 database, and they produce results comparable to those of the CSSD in the case of the database of images with non-rigid object distortion. The CBFD descriptors also outperform six notable descriptors in one of the most challenging databases (set B) for both high and low recall. The superior performance results from the ability of the proposed descriptors to capture both the local and global properties of an object, while the ARTD, ZMD, RTMD, and the GFD tend to capture only the global properties. The

experimental results also show that the overall average performance of the proposed descriptors is better than that of all the other descriptors.

The proposed descriptors overcome several disadvantages of the well-known CSS descriptors. The overall performance of the proposed descriptors is 12.53% better than that of the CSSD for high recall. As well, the proposed descriptors use a simple matching procedure that makes them efficient for retrieving images from huge databases.

The complexity and variety of the content of real images make it impossible for a particular choice of descriptors to be effective for all types of images. Therefore, in the Chapter 5 a data- fusion formulation based on a team consensus approach is proposed as a means of achieving high performance accuracy.

Chapter Five

Consensus-Based Fusion Algorithm for Shape-Based Image Retrieval

5.1 Introduction

The enormous variation in the content of images makes it impossible for a particular choice of descriptor to be effective for all types of images. It is therefore reasonable to approach the problem by combining a group of descriptors. This chapter proposes a novel co-ranking scheme that exploits the complementary benefits obtained using several shape-based image retrieval techniques and integrating their assessments based on a pair-wise co-ranking process. A consensus-based fusion algorithm is also proposed to integrate several shape-based image retrieval techniques so as to enhance the performance of the image retrieval process. This algorithm causes several techniques to work as a team: they exchange their ranking information based on the pair-wise co-ranking scheme to reach a consensus that will improve their final ranking decisions. Since the effectiveness of a descriptor is image dependent, the author maintains that for this strategy to achieve the intended goal, the combining scheme must be dynamic as a function of the query context. Consequently, a context-based fusion algorithm is proposed for integrating several shape-based retrieval techniques. The context of the query is used to determine the most appropriate technique or combination of techniques that will produce the best performance.

5.2 Formulation of the Problem

Consider a set of M image-retrieval agents indexed by the set IRA = *{IRA₁, IRA2,… IRAₘ}*, cooperating to retrieve query image *x* from an ensemble of images denoted by Θ .

Each agent IRA$_i$ uses a feature extraction scheme F_i and a matching strategy Γ_i to determine the similarity measure S_i between query image x and all images $y \in \Theta$; that is,

$$S_i(x, y) = \Gamma_i(F_i(y), F_i(x)), \forall y \in \Theta \qquad (5.1)$$

Each agent *IRA*i establishes a ranking $R_i(y \mid F_i(y)) \in \{1,2,...N\}, \forall y \in \Theta$ such that $R_i(y \mid F_i(y)) \le R_j(z \mid F_j(z))$ implies $S_i(x, y) \ge S_j(x, z)$.

Without loss of generality, $R_i(y \mid F_i(y))$ can be viewed as an index set from 1 to N, where index 1, for example, points to the candidate image closest to query image x, and index N points to the candidate image most dissimilar to query image x. In general, index l in this set points to the image that is preceded by l-1 candidates; these candidates are viewed by *IRA$_i$* as better candidates for the query image x than the candidates ranked l. Since each agent uses a different feature extraction scheme, it is expected that the agents make different decisions for ranking the images of the set. Furthermore, since a CBIR technique is selected as an agent in this proposed approach, it must have demonstrated acceptable performance (i.e., not completely random). Consequently, it is reasonable to expect that good candidates will be clustered at the top of the ranking by all agents.

5.3 Pair-Wise Co-Ranking Scheme

The following describes a proposed information exchange scheme between the agents to make a decision about the candidate images that are closest to the query image. This process exploits the relative advantages and disadvantages of each agent.

The concept of the new pair-wise co-ranking scheme is based on the hypothesis that agent *IRA$_j$* may readjust its ranking if it is exposed to the ranking results of another agent *IRA$_i$*. The communication process between two agents is depicted in Figure 5.1. Revised ranking information is referred to as conditional ranking information in order to signify

the influence of other agents on the ranking. On the other hand, the initial ranking is referred to as marginal ranking information.
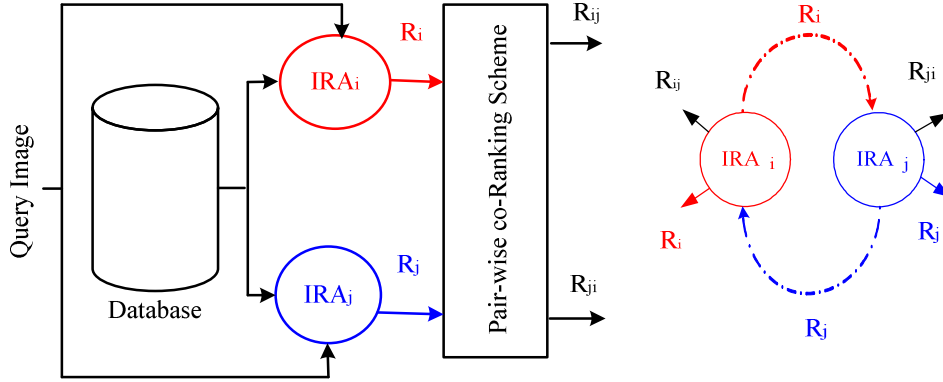
Figure 5.1 Two image-retrieval agents exchange information about the ranking.

To set up the process of the exchange of information among agents, the ranking set of each agent IRA_i is divided into two partitions: an elite candidates partition (ECP_i) and a potential candidates partition (PCP_i). It is expected that good matches to the query will be clustered in the elite partition. ECP contains the first m candidates; PCP contains the last N-m candidates. Thus, the ranking produced by agent IRA_i based only on its feature set can be viewed as a concatenation of two ranking sets: $R_i = \{R_i^{ECP}, R_i^{PCP}\}$ where R_i^{ECP} is the ranking result for $R_i \leq m$, and R_i^{PCP} is the ranking result for $R_i > m$.

Figure 5.2 shows an example of the way agent IRA_j computes its conditional elite partition ranking R_{ji}^{ECP}, based on marginal ranking information from agent IRA_i. Image z is an elite image in the ranking from agent IRA_i. Agent IRA_j uses its feature extraction scheme F_j to determine the rank of image z in its conditional elite candidates partition; that is,

$$R_{ji}^{ECP}(z \mid F_j(z), R_i(z \mid F_i(z)) \leq m), z \in ECP_i \qquad (5.\,2)$$

This formula can be read as follows: agent *j* re-ranks image *z* based on feature extraction scheme $F_j(z)$, given that image *z* has been ranked in the elite candidate partition by agent *i*, based on the feature extraction scheme $F_i(z)$.

The fact that image *z* is placed in the conditional elite partition of agent IRA_j does not necessarily imply that image *z* is also in the marginal elite partition of IRA_j. A similar procedure is followed for computing the conditional ranking with respect to potential candidates partitions, that is,

$$R_{ji}^{PCP}(z \mid F_j(z), R_i(z \mid F_i(z)) > m), z \in PCP_i \qquad (5.3)$$

The conditional ranking from agent IRA_j, based on information received from agent IRA_i is viewed as a concatenation of two ranking sets; that is,

$$R_{ji} = \{R_{ji}^{ECP}, R_{ji}^{PCP}\} \qquad (5.4)$$

It should be noted that for a given agent, the revised ranking information is used to complement its initial ranking and not to replace it. The results of the above process are summarized in Table 5.1.

$$R_i = \{R_i^{ECP}, R_i^{PCP}\}$$

$$R_{ji} = \{R_{ji}^{ECP}, R_{ji}^{PCP}\}$$

Elite Candidates Partition (ECPi)

Potential Candidates Partition (PCPi)

z

$z \in ECP_i, \ R_i(z \mid F_i(z)) \le m$

IRA_j

z

m

m

$R_{ji}^{ECP}(z \mid F_j(z)), R_i(z \mid F_i(z)) \le m), z \in ECP_i$
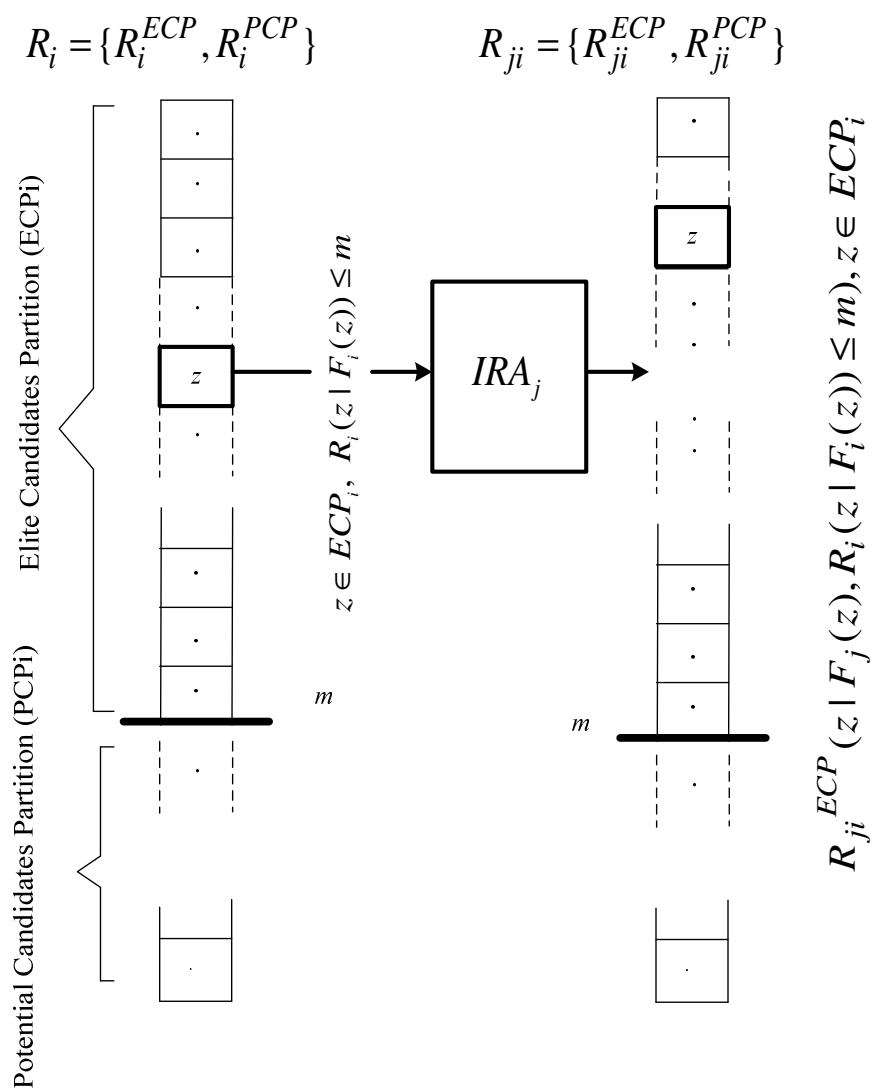
Figure 5.2 Illustration of the pair-wise co-ranking scheme.

Table 5.1 Marginal and Pair-Wise Conditional Rankings

$$R_i^{ECP}(z|F_i(z))\} \; \forall \; R_i \; \leq m$$

$$R_i^{PCP}(z|F_i(z))\} \; \forall \; R_i \; > m$$

$$R_i = \{R_i^{ECP}, R_i^{PCP}\}$$

$$R_{ij}^{ECP}(y|F_i(z),R_j(y|F_j(z)) \leq m)\} \; \forall \; R_{ij} \; < m$$

$$R_{ij}^{PCP}(y|F_i(z),R_j(y|F_j(z)) > m)\} \; \forall \; R_{ij} \; > m$$

$$R_{ij} = \{R_{ij}^{ECP}, R_{ij}^{PCP}\}$$

$$R_j^{ECP}(z|F_j(z))\} \forall R_j \; \leq m$$

$$R_j^{PCP}(z|F_j(z))\} \forall R_j \; > m$$

$$R_j = \{R_j^{ECP}, R_j^{PCP}\}$$

$$R_{ji}^{ECP}(y|F_j(z),R_i(y|F_i(z)) \leq m)\} \; \forall \; R_{ji} \; \leq m$$

$$R_{ji}^{PCP}(y|F_j(z),R_i(y|F_i(z)) > m)\} \; \forall \; R_{ji} \; > m$$

$$R_{ji} = \{R_{ji}^{ECP}, R_{ji}^{PCP}\}$$

Illustrative Example of the Pair-wise Co-ranking Scheme

The database described in Chapter 3 consists of 70 classes, each having 20 objects. The goal is to have the relevant images ranked as the top 20 positions. Figure 5.3 and Figure 5.4 display the results produced by agents IRA_1 and IRA_2, respectively.
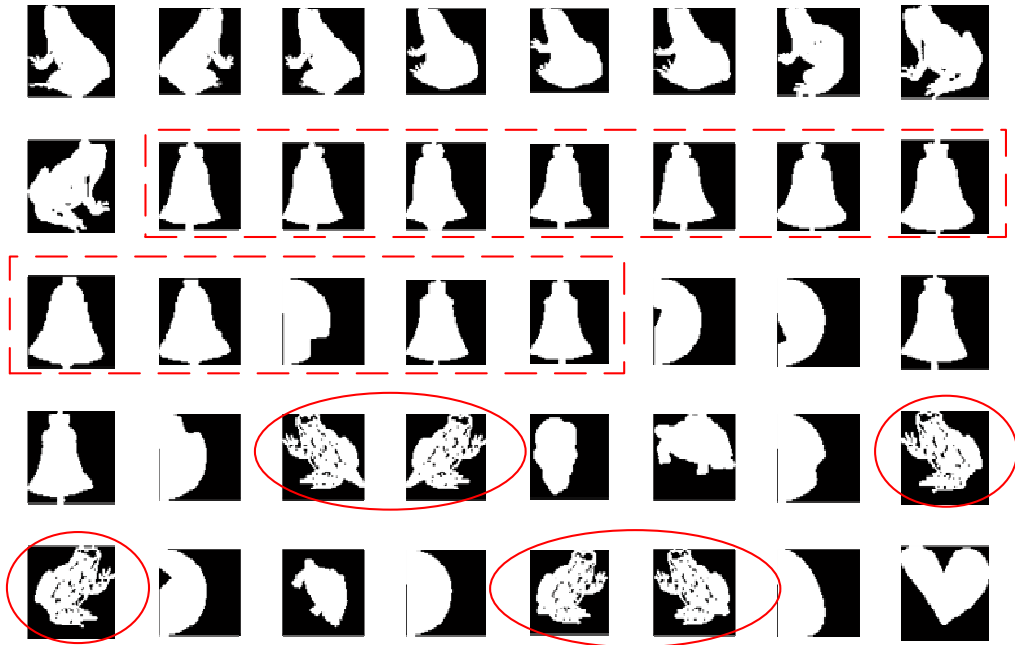
Figure 5.3 Results retrieved by $IRA_1 (R_1)$.
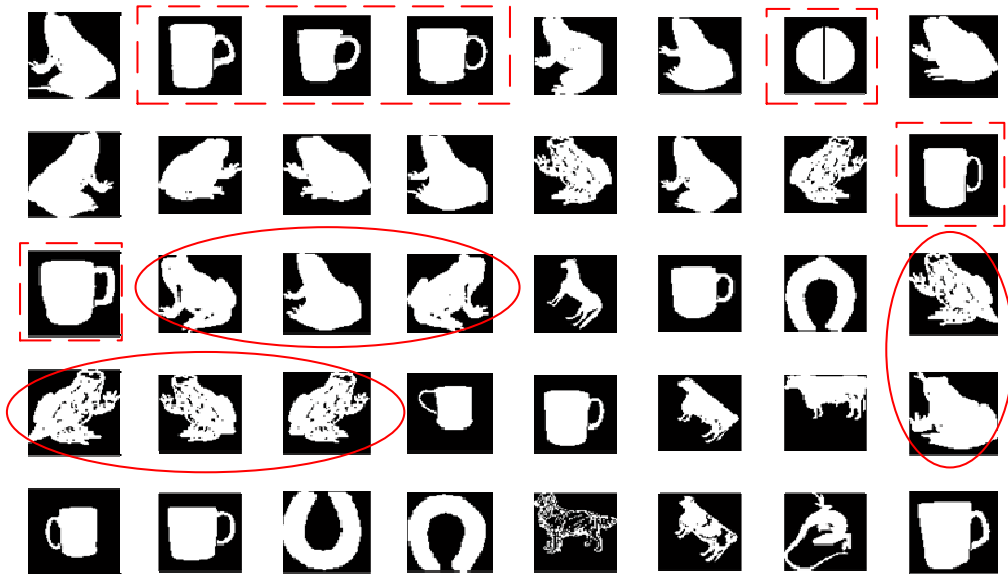
Figure 5.4 Results retrieved by $IRA_2 (R_2)$.

`
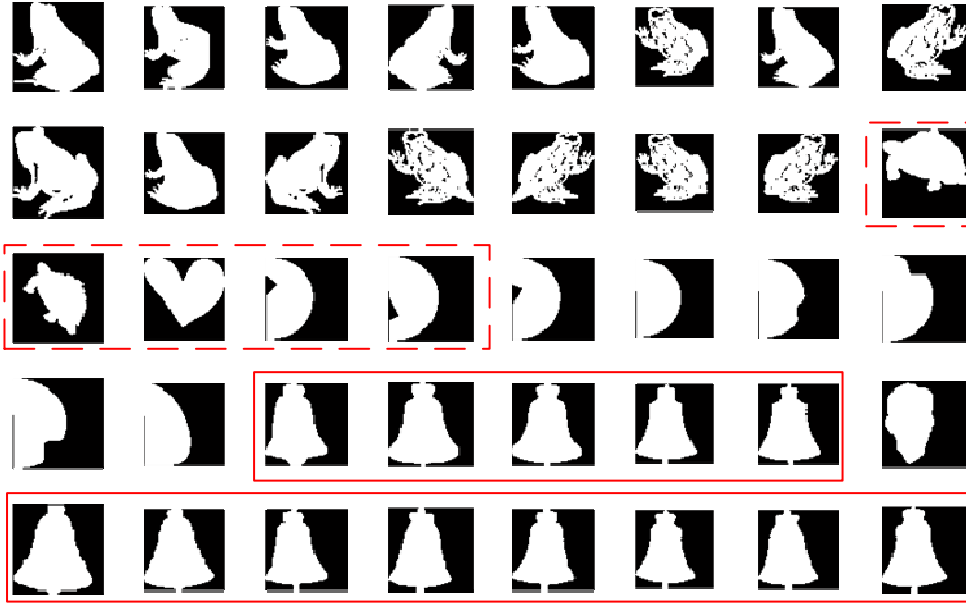
Figure 5.5 Results of R_1 revised based on R_2 (R_{21}).

In both figures, the top left shape is the query shape (frog); the retrieved shapes are arranged in descending order according to the similarity of each shape to the query shape. Figure 5.3 shows that the first agent, IRA$_1$, has managed to rank 9 of 20 shapes correctly (45%). Furthermore, most of the irrelevant shapes, indicated by the dashed frames, are objects that belong to the same class (bell). In Figure 5.4, it is evident that the irrelevant shapes from agent IRA$_2$, indicated by the dashed frames, differ from those from agent IRA$_1$. Figure 5.5 shows the conditional ranking, i.e. the revised results, from agent IRA$_1$ based on the ranking information it has received from the second agent, IRA$_2$. In Figure 5.5, the irrelevant shapes from the first agent, indicated by the dashed frames in Figure 5.3, do not appear in top-ranked positions. The positions of the relevant shapes, encircled by the ellipses in Figure 5.3, produced by the first agent, IRA$_1$, are repositioned into the top 20 positions in R21, resulting in an accuracy rate of 75% (15 out of 20). It is clear that the conditional ranking results (the revised results) R21 are much better than the marginal ranking results (the unrevised results) R1 and R2.

5.4 Consensus-Based Fusion Algorithm

The main motivation for this research is to identify image-retrieval techniques that capture different characteristics of an image and combine them in a way that enhances retrieval performance through consensus about the rankings. These techniques can be considered to be a team of agents who cooperate in order to determine the image in the database that best matches the query image. This collaboration is accomplished through an exchange of information about the ranking of candidates. Each agent uses its feature extraction scheme to compute a ranking R_{ii}. Double indexing is used in order to simplify the mathematical notations. The ranking reflects an individual technique's preference for the candidate that best matches the query image (i.e., marginal ranking). Each agent is additionally exposed to the results of the rankings by the other agents so that a conditional ranking is also computed for each candidate image, as explained above. Therefore, M retrieval agents yield M^2 rankings: M marginal rankings plus M (M-1) conditional rankings. Figure 5.6 depicts the ranking set of a three-agent system.
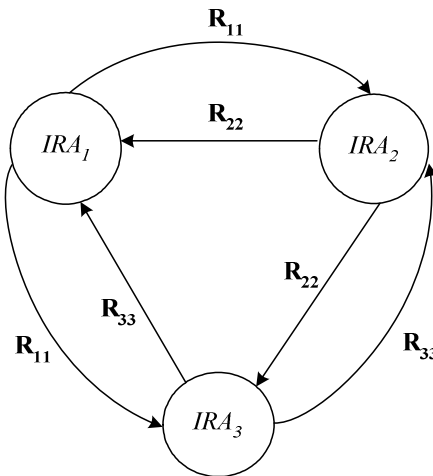
Figure 5.6 Information exchange among three image retrieval agents.

For M agents, the ranking sets can be organized in a matrix format:

$$R = \begin{bmatrix} R_{11} & R_{12}...... & R_{1M} \\ R_{21} & R_{22} & ...... & R_{2M} \\ . & & . \\ . & & . \\ . & & . \\ R_{M1} & R_{M2}...... & R_{MM} \end{bmatrix} \qquad (5.5)$$

where R_{ij} is the conditional ranking of the ith agent, given the ranking of the jth agent, and R_{ii} is the marginal ranking of the ith agent.

Figure 5.7 portrays the exchange of ranking information that yields the pair-wise conditional rankings.
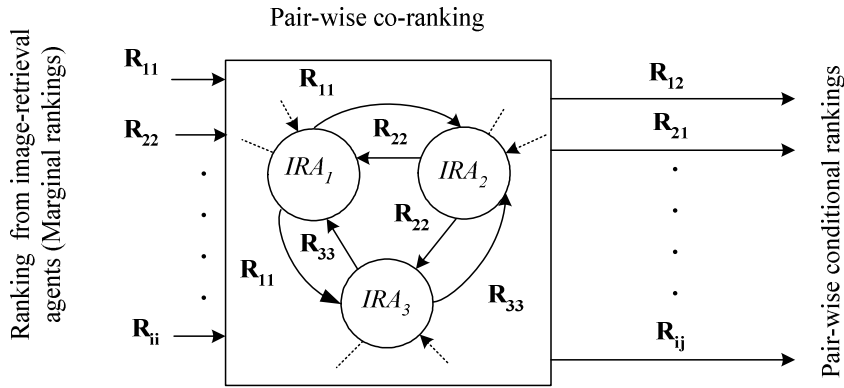
Figure 5.7 Steps in the application of the pair-wise co-ranking.

To obtain a final decision about the rankings, a pair-wise consensus algorithm is proposed. This algorithm first combines each set of the pair-wise conditional rankings that were obtained from the cooperation of each pair of techniques used, by averaging their rankings:

$$R_{cc} = (R_{ij} + R_{ji})/2 \qquad (5.6)$$

where $c=1, 2,... M (M-1)/2$, and M is the number of techniques used.

Then the pair-wise co-ranking scheme is applied to the combined conditional rankings R_{cc} to produce a new set of pair-wise conditional rankings R_{ij}^{t} (t represents the number of

108

iterations). At each iteration t, the algorithm produces another set of $(M\ (M\text{-}1)\ /2)$ combined conditional rankings R_{cc}^t. The initial rankings of the individual agents are represented by R_{ij}^0.$(i=j,\ i=1,2,…m)$.

In this manner, each technique influences the ranking information of the other techniques until all combined pair-wise conditional rankings R_{cc}^t result in consensus about the rankings. Figure 5.8 shows the main steps in the proposed algorithm.
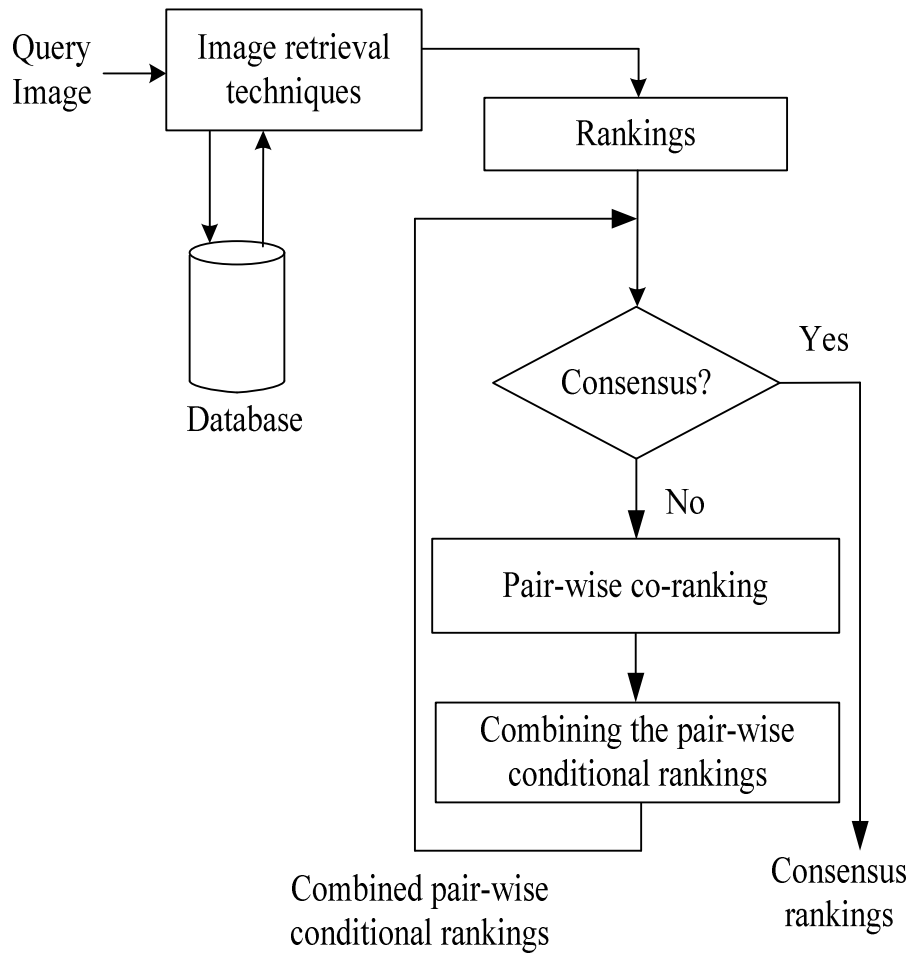
Figure 5.8 The main steps in the consensus-based algorithm.

5.4.1 Illustrative Example of the Consensus-Based Algorithm

This section introduces an example using the MPEG-7 database, which illustrates the effectiveness of the proposed algorithm. For simplicity, numbers are used to refer to all images in the database. An image that has an index of 746 is considered to be a query image. In this example, three techniques were used to retrieve images from set B of the MPEG-7 database, with the goal of finding images similar to the given query image.

The proposed algorithm was applied, and the rankings after different iterations are shown in Table 5.2 and Table 5.3. The first three columns of Table 5.2 show the rankings from the three individual techniques. Only the top 20 images retrieved by each technique are listed, and star signs are used to indicate the irrelevant retrieved images. The rankings that resulted from all the techniques reaching consensus are highlighted.

From the first three columns of Table 5.2, it can be seen that the three individual techniques reached consensus about the rankings of only two images. Moreover, many images were irrelevant, as indicated by the star signs. After the fourth iteration, most of the irrelevant images (indicated by star signs) from the individual rankings no longer appear in the top 20 positions and have been replaced with relevant images.

Table 5.2 The rankings from the three individual techniques and from the proposed algorithm after the fourth iteration

R_{11}^0	R_{22}^0	R_{33}^0	R_{11}^4	R_{22}^4	R_{33}^4
748	748	748	748	748	748
749	749	749	749	749	749
746	746	759	757	741	741
747	747	758	747	747	757
757	756	741	741	757	747
756	757	760	760	760	760
745	760	753	744	744	744
744	593^*	757	746	746	758
750	592^*	431^*	758	754	754
751	741	702	754	756	746
742	565^*	742	756	758	756
743	567^*	430^*	43^*	43^*	753
760	754	754	745	745	43^*
741	744	720	753	753	745
1312^*	561^*	752	750	759	759
1319^*	579	291^*	743	750	742
1317^*	600^*	717	742	743	743
1314^*	599^*	743	751	752	752
1313^*	745	191^*	759	742	750
678^*	576^*	57^*	752	751	191^*

Table 5.3 The rankings from the proposed algorithm after the sixth and ninth iterations

R_{11}^{6}	R_{22}^{6}	R_{33}^{6}	R_{11}^{9}	R_{22}^{9}	R_{33}^{9}
748	748	748	748	748	748
749	749	749	749	749	749
741	741	741	741	741	741
757	747	757	747	747	747
747	757	747	757	757	757
760	760	760	760	760	760
744	744	744	744	744	744
746	746	746	746	746	746
758	754	758	754	754	754
754	758	754	758	758	758
756	756	756	756	756	756
43^{*}	43^{*}	43^{*}	43^{*}	43^{*}	43^{*}
745	745	753	745	745	745
753	753	745	753	753	753
743	743	742	743	743	743
750	759	743	742	742	742
742	750	759	750	750	750
759	742	750	759	759	759
751	751	752	751	751	751
752	752	751	752	752	752
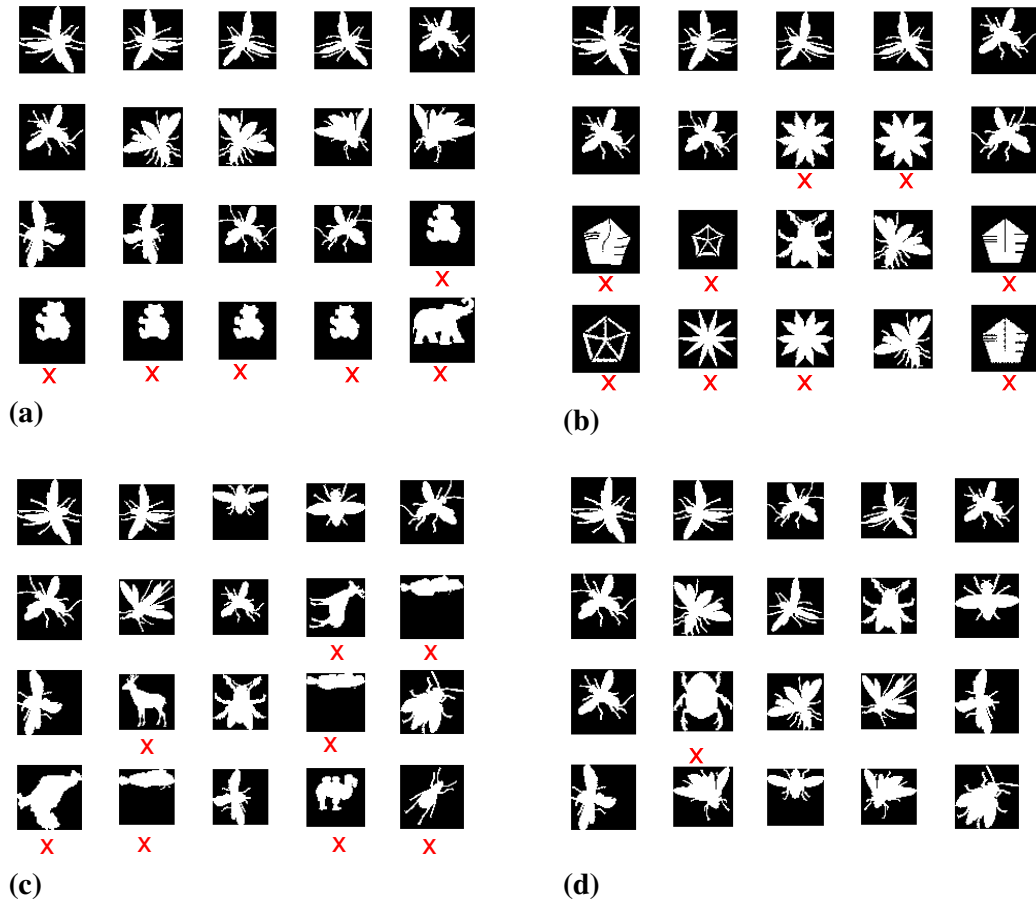
Figure 5. 9 **(a)** Retrieval results from the first technique (R_{11}^0) **(b)** Retrieval results from the second technique (R_{22}^0) **(c)** Retrieval results from the third technique (R_{33}^0) **(d)** Retrieval results from the proposed algorithm after the ninth iteration.

Table 5.3 shows the rankings of the proposed algorithm after the sixth and ninth iterations. From this table, it is clear that the number of irrelevant images has been reduced and that consensus about the ranking has been reached for 50% of the images. After the ninth iteration, a 100% consensus about all the rankings has been obtained. Moreover, not only has consensus been reached after the ninth iteration, but an accuracy rate of 95% has also been achieved.

Figure 5. 9 shows the retrieval results from the three selected techniques and from the proposed algorithm after the ninth iteration. In Figure 5. 9, the top left shape is the query

image, and the irrelevant retrieved images are marked by "x". It is clear from the figure that the proposed algorithm guides the three techniques so that together they achieve better accuracy (95%) than any individual technique obtains on its own: R_{11}^0 achieved 70% accuracy, R_{22}^0 achieved 55% accuracy, and R_{33}^0 achieved 60% accuracy.

5.4.2 Complexity of the Consensus-Based Algorithm

To estimate the computational complexity of the consensus-based algorithm the Big-Oh notation is adopted. The main steps of the consensus-based algorithm and the complexity of each step are given as follows:

- Ranking the database according to the marginal techniques $\rightarrow$ *O(N+ N log(N)).O(N)* to calculate the Euclidean distance and O(N log(N)) to sort the distance to compute a ranking.

- Applying the co-ranking scheme to obtain the conditional ranking$\rightarrow$ *O(m(N-m)).*

- Combining the conditional ranking$\rightarrow$ *O(N+N log N).*

where *N* is the number of the images in the database and *m* is the number of the Elite candidate images.

For the second step, the worst case scenario occurs when *m=N/2* and hence the complexity of the co-ranking scheme is *O (N²).*

5.5 Context-Based Fusion Algorithm

Analysis of the performance of the consensus algorithm reveals that for some queries the consensus algorithm forces the techniques to retrieve irrelevant shapes in order to reach consensus. This effect occurs because the consensus algorithm does not take into account the differences in the quality of either the marginal or the conditional ranking assessments. In some cases, the marginal or conditional rankings are better than the consensus ranking because the database is semantically classified. In other words, human

beings often classify shapes based on their global features and ignore the local features within the shapes. For example, Figure 5.10 shows two classes, each of which has four shapes. As can be seen, the four shapes that belong to the same class have completely different local characteristics, but they are semantically classified in the same class.

What is required is an algorithm that looks at the semantic features, or the context, of the shape and not just its low-level features. To achieve this goal, we propose what we called a look-ahead selector that uses the context of the query in order to influence the final ranking. The look-ahead selector uses the characteristics of the query image to determine whether the marginal, conditional, or consensus ranking is the best ranking. The current study employed the features of the three descriptors used along with simple global descriptors extracted from the query image. The next step in the algorithm is to train a neural network to select the best ranking from the marginal and consensus rankings.

Class 1   Class 2
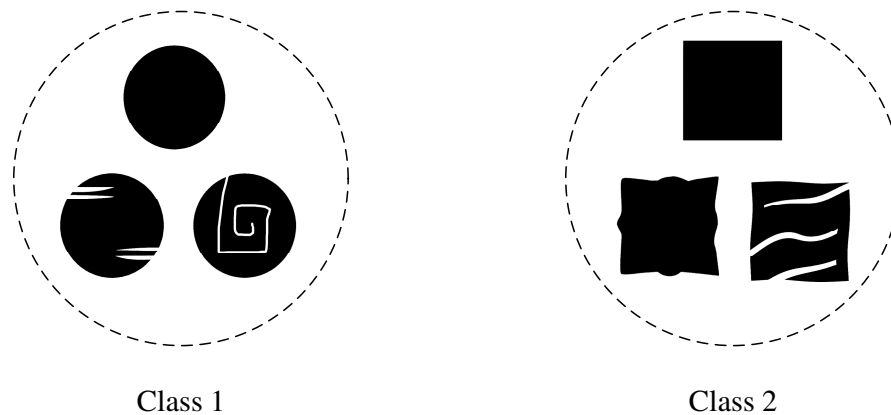
Figure 5.10 Two classes, each of which include shapes that have different local characteristics.

To illustrate the input and output of the look-ahead selector for an example with three techniques, four rankings will be produced: three marginal rankings and one final consensus ranking. Therefore, the look-ahead selector must have n inputs representing the

descriptors of the selected techniques and four input channels for the marginal and consensus rankings. A general block diagram of the proposed algorithm for the three-technique example is shown in Figure 5.11.

The look-ahead selector uses a neural network that must be trained using a set of classified shapes so that it can select the appropriate ranking from the four rankings. The input of the neural network is the descriptors of the selected techniques and the output is four elements that select the appropriate ranking. Three of the four output elements must show 0 for an undesirable ranking, with only one output element showing 1 for the candidate ranking.

The look-ahead selector improves the overall performance of the retrieval process because the quality of each ranking is considered through the preprocessing of the query image. It also reduces the time required to retrieve images because, in some cases, the look-ahead selector selects one of the marginal rankings, which saves the time required to obtain the other conditional or consensus rankings.
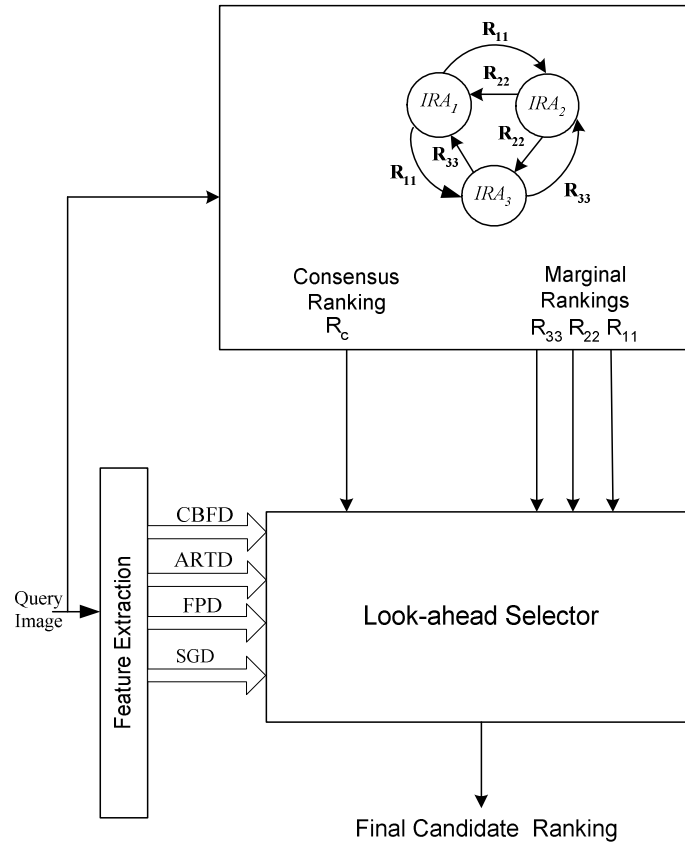
Figure 5.11 General block diagram of the context-based algorithm.

5.6 Experimental Results

The five databases described in Chapter 3 have been used to test the performance of the algorithms presented in this chapter. To implement and test the newly developed consensus algorithm, three shape descriptor techniques were selected.

The first, the angular radial transform descriptor (ARTD), is a region-based technique that provides the global characteristics of an image. This technique has been adopted by MPEG-7 as a standard region descriptor.

The second technique, the curvature-based Fourier descriptor (CBFD) [97], is a boundary-based technique used to provide the local characteristics of an image for the proposed algorithm. This technique is designed to capture local information using the curvature of the shape and a Fourier transform. As shown in Chapter 4, the CBFD

outperforms the curvature scale space descriptor (CSSD), which has been adopted by MPEG-7 as a standard boundary technique.

The third technique is the Fourier descriptor (FD) technique, based on the farthest point distance (FPD) signature. Chapter 3 describes how the FPD signature outperforms the commonly used signatures. The FD technique was selected because it is a boundary-based technique that can capture both local and global characteristics. Low frequencies provide global characteristics, and high frequencies symbolize the local characteristics of an image [85].

The set B images from MPEG-7 were used to investigate the effect of the number of elite candidates (M) and the number of iterations (t) on the consensus algorithm. Figure 5.12 shows the average precision for different values of m and t. It can be seen that after the fourth iteration (t=3) there is no significant improvement. This result was expected because at this iteration (t=3) all marginal rankings revise one another based on the pair-wise co-ranking scheme. The results also show that the best performance is obtained at m =40, which represents 2.85% of the total number of shapes in the database.
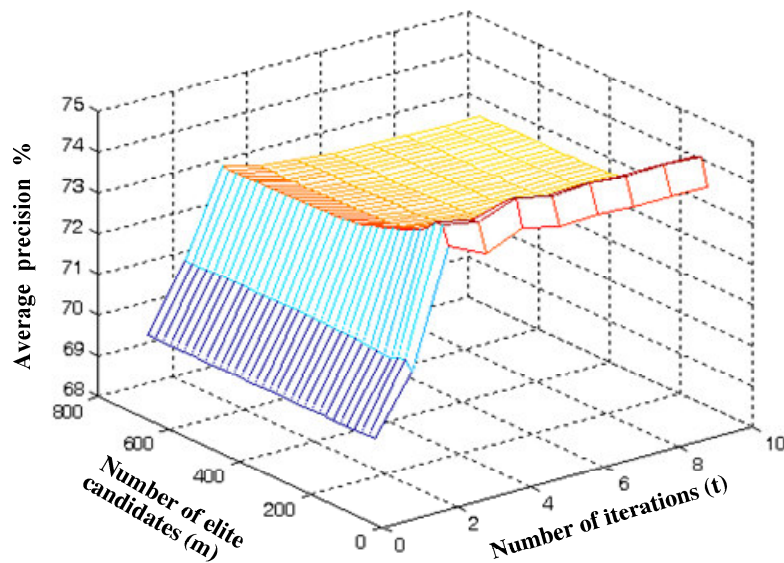
Figure 5.12 The average precision for different values of m and t.

118

The ranking of the proposed consensus-based algorithm was compared with the marginal rankings produced by the three selected techniques, and the ranking produced by averaging the rankings of the three selected techniques to pool their results (pooled marginal ranking). Presented in Figure 5.13 to Figure 5.17 are the precision-recall curves for the consensus-based algorithm as well as those for the ARTD, CBFD, and FPD techniques and those resulting from the pooling of the three techniques. The average precision rates for low and high recall are shown in Table 5.4 to Table 5.8. Table 5.9 shows the comparison of the overall average performance of the proposed descriptors with that of the other techniques based on the weighted average given in Equation (4. 13)

Figure 5.13 to Figure 5.17 indicate the effectiveness of the proposed consensus-based algorithm, whose performance is significantly superior to that of the, CBFD, FPD, and ARTD techniques and better than that produced by pooling the rankings from the three techniques. Based on Table 5.9, it is clear that, compared with other techniques, the consensus-based algorithm descriptors provide the highest performance for both low and high recall.
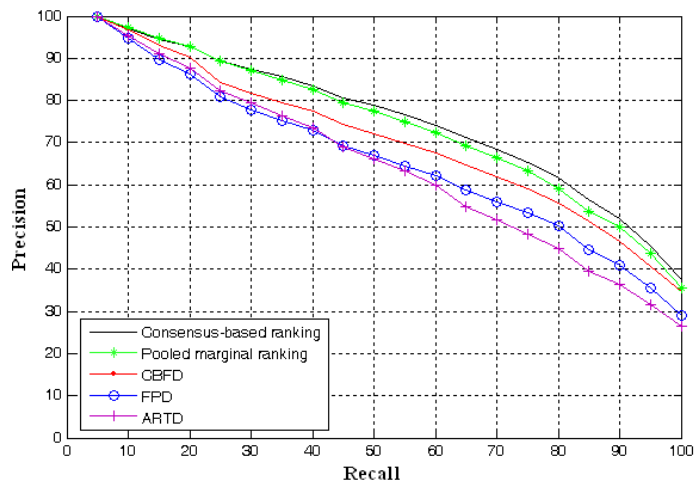
Figure 5.13 Precision-recall curves for set B results from the consensus-based algorithm; for the results from the pooled marginal ranking of the three selected techniques; and for the marginal results from the CBFD, FPD, and ARTD techniques.

119

Table 5.4 The set B average precision rates of low and high recall for the consensus-based algorithm; the pooled marginal ranking from the three selected techniques; and the marginal rankings from CBFD, FPD, and ARTD techniques

Method	Low Recall	High Recall
	Average precision for recall rates ≤ 50%	Average precision for recall rates >50%
Consensus-based ranking	88.99 %	60.89 %
Pooled marginal ranking	88.53 %	58.80 %
CBFD (Chapter 4)	85.15 %	54.77 %
FPD (Chapter 3)	81.16 %	49.15 %
ARTD	82.10 %	45.69 %
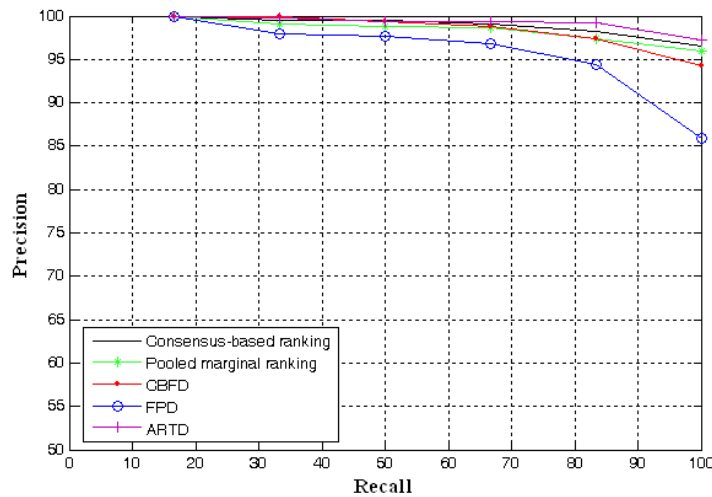
Figure 5.14 Precision-recall curves for set A1 results from the consensus-based algorithm; for the results from the pooled marginal ranking of the three selected techniques; and for the results from the CBFD, FPD, and ARTD techniques.

Table 5.5 The set A1 average precision rates of low and high recall for the consensus-based algorithm; the pooled marginal ranking from the three selected techniques; and the marginal rankings from the CBFD, FPD, and ARTD techniques

Method	Low Recall	High Recall
	Average precision for recall rates ≤ 50%	Average precision for recall rates >50%
Consensus-based ranking	99.69 %	97.97 %
Pooled marginal ranking	99.29 %	97.33 %
CBFD (Chapter 4)	99.74 %	96.78 %
FPD (Chapter 3)	98.60 %	93.23 %
ARTD	99.71 %	98.62 %
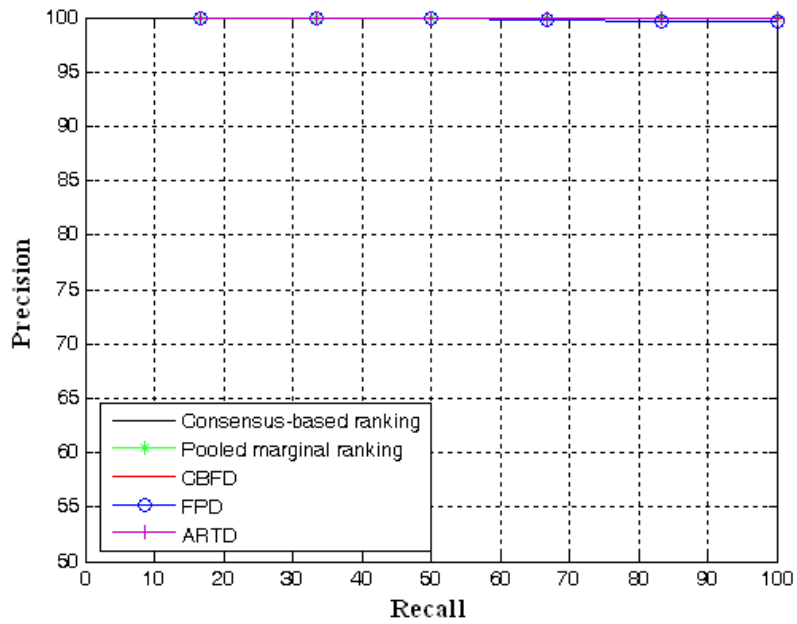
Figure 5.15 Precision-recall curves for set A2 results from the consensus-based algorithm; for the results from the pooled marginal ranking of the three selected techniques; and for the marginal results from the CBFD, FPD, and ARTD techniques.

Table 5.6 The set A2 average precision rates of low and high recall for the consensus-based algorithm; the pooled marginal ranking of the three selected techniques; and the marginal rankings from the CBFD, FPD, and ARTD techniques.

Method	Low Recall Average precision for recall rates ≤ 50%	High Recall Average precision for recall rates >50%
Consensus-based ranking	100.00 %	100.00 %
Pooled marginal ranking	100.00 %	100.00 %
CBFD (Chapter 4)	100.00 %	100.00 %
FPD (Chapter 3)	100.00 %	99.89 %
ARTD	100.00 %	100.00 %
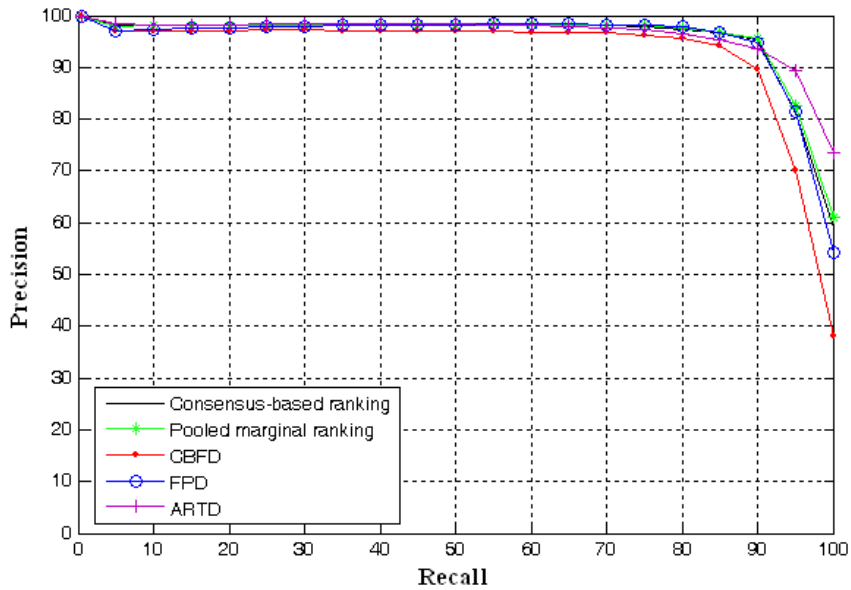
Figure 5.16 Precision-recall curves for set C results from the consensus-based algorithm; for the results from the pooled marginal ranking of the three selected techniques; and for the marginal results from the CBFD, FPD, and ARTD techniques.

Table 5.7 The set C average precision rates of low and high recall for the consensus-based algorithm; the pooled marginal ranking of the three selected techniques; and the marginal rankings from the CBFD, FPD, and ARTD techniques.

| Method	Low Recall	High Recall
	Average precision for recall rates ≤ 50%	Average precision for recall rates > 50%
Consensus-based ranking	98.42 %	92.61 %
Pooled marginal ranking	98.40 %	92.97 %
CBFD (Chapter 4)	97.38 %	88.03 %
FPD (Chapter 3)	97.96 %	92.23 %
ARTD	98.48 %	94.16 %
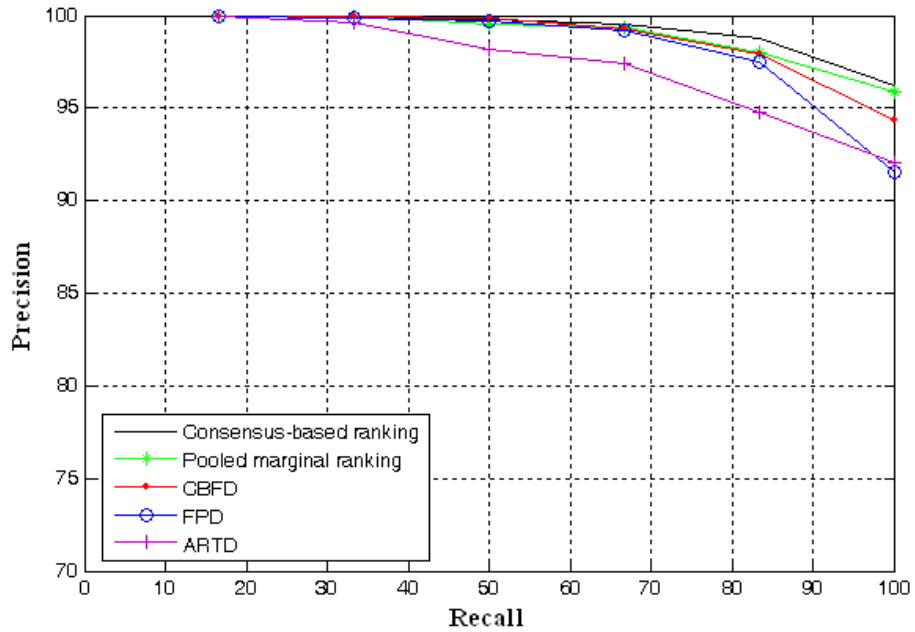
Figure 5.17 Precision-recall curves for set D results from the consensus-based algorithm; for the results from the pooled marginal ranking of the three selected techniques; and for the marginal results from the CBFD, FPD, and ARTD techniques.

Table 5.8 The set D average precision rates of low and high recall for the consensus-based algorithm; the pooled marginal ranking of the three selected techniques; and the marginal rankings from the CBFD, FPD, and ARTD techniques.

Method	Low Recall Average precision for recall rates ≤ 50%	High Recall Average precision for recall rates >50%
Consensus-based ranking	99.92 %	98.21%
Pooled marginal ranking	99.84 %	97.74 %
CBFD (Chapter 4)	99.96 %	97.19 %
FPD (Chapter 3)	99.85 %	96.08 %
ARTD	99.75 %	94.75 %

Table 5.9 The overall average precision rates of low and high recall for the consensus-based algorithm; the pooled marginal ranking of the three selected techniques; and the marginal rankings from the CBFD, FPD, and ARTD techniques

Method	Low Recall	High Recall
	Average precision for recall rates ≤ 50%	Average precision for recall rates >50%
Consensus-based ranking	94.42 %	79.78 %
Pooled marginal ranking	94.13 %	78.62 %
(CBFD) (Chapter 4)	92.50 %	76.14 %
FPD (Chapter 3)	90.41 %	72.98 %
ARTD	91.05 %	72.03 %

5.7 Analysis and Discussion

The experimental results presented in the previous section show that the consensus-based algorithm achieves a high degree of accuracy for all of the databases. However, it has been noted that for some specific classes from set B of the database, the performance of the marginal techniques is better than that of the consensus-based algorithm. Examples of these classes are shown in Figure 5.10. For these classes, the performance of the consensus-based algorithm was very low because these classes were classified based on consideration of only global characteristics, and local ones were ignored. For each of these classes, Figure 5.18 shows the accuracy of the first top 40 retrieved shapes for the consensus-based algorithm and that of the three marginal techniques. As can be seen in Figure 5.18, for class 32 (indicated by arrows), the accuracy of the consensus-based algorithm was very low compared to the accuracy of the ARTD technique.

ARDT

FPD

CBFD

Censuses-based algorithm
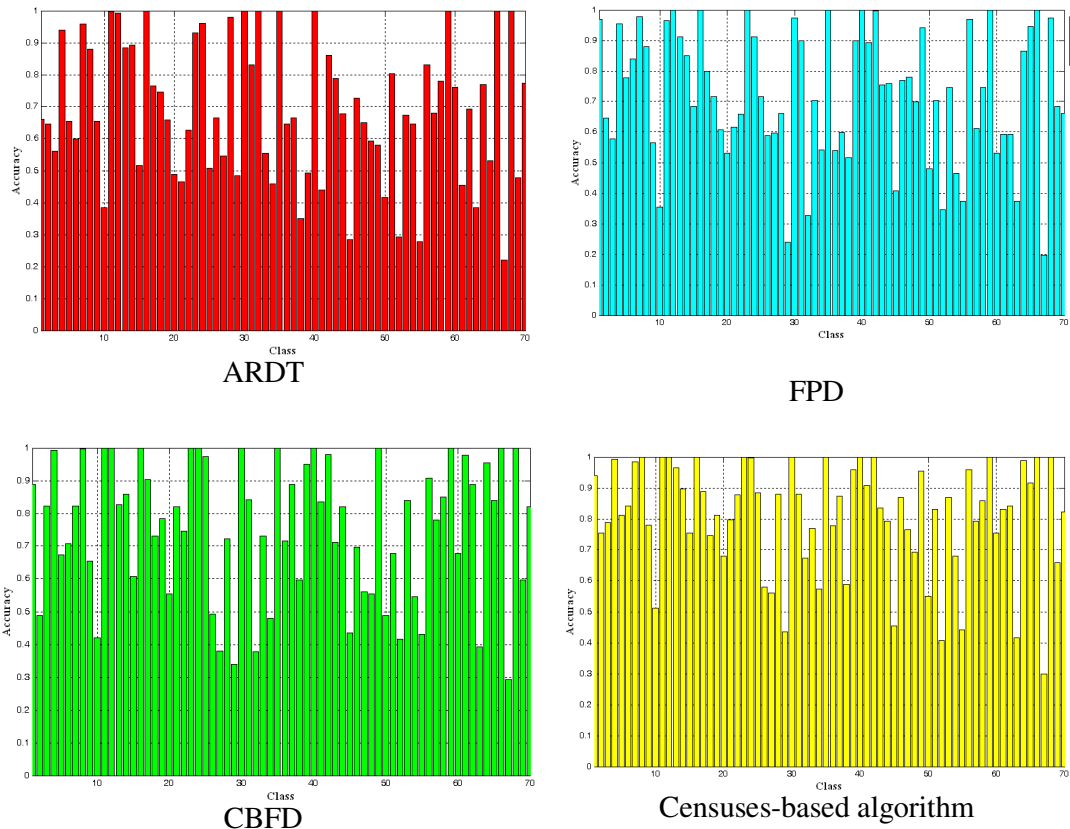
Figure 5.18 The retrieval accuracy by class for the ARTD, FPD, and CBFD techniques and for the consensus-based algorithm.

To overcome this problem, the context-based algorithm was used. In this algorithm, part of the set B database was used as a training set (the first seven shapes of each class), and the other shapes were used as a testing set. To test the concept of the context-based algorithm the Matlab neural network toolbox was used to facilitate the training and testing stages. For each class, Figure 5.19 shows the accuracy of the context-based algorithm and the three marginal techniques. As can be seen from Figure 5.19, the accuracy of the context-based algorithm is better than that of the consensus-based algorithm. For several classes, the context-based algorithm was able to solve or minimize the problem encountered with the consensus-based algorithm. The performance of the context-based algorithm is based on the number of training sets used to train the neural network and can be improved through the online learning of the neural network that

results from feedback from the users. This method will allow the retrieval system to learn through feedback from each user query.
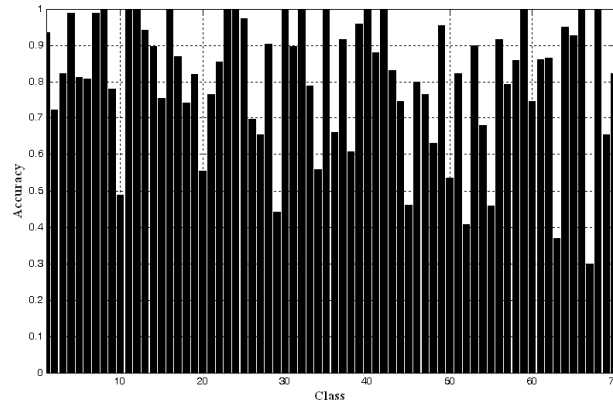
Figure 5.19 The retrieval accuracy by class for the context-based algorithm

5.8 Summary

This chapter has presented a new pair-wise co-ranking scheme based on a revision of the rankings for the retrieved shapes produced by several techniques. A consensus-based fusion algorithm is also proposed as a means of improving the retrieval performance of a group of shape-based image retrieval techniques by enabling them to achieve consensus with respect to their rankings. Five databases have been used to test the performance of the proposed algorithm, and the experimental results have been presented.

The experiments demonstrate that the proposed algorithm yields results that are better than the marginal rankings produced from any of the three selected techniques. The performance of the proposed algorithm is also superior to that produced by pooling the marginal rankings from the three techniques. This improvement is the result of allowing the selected techniques to recursively exchange their ranking information until all of the techniques reach a consensus about their rankings. The experimental results also show, however, that the performance of the consensus-based algorithm is low for specific

classes. To overcome this difficulty with the consensus-based algorithm, a context-based algorithm has been proposed, which makes the final ranking a function of the query context.   Based on the results, the context-based algorithm is able to eliminate or minimize the problem with the consensus-based algorithm.

Chapter Six

Conclusions and Future Work

The main objective of this research is to investigate and improve the shape-based image retrieval process.

State-of-the-art shape descriptors are reviewed, several are comprehensively studied and their advantages and disadvantages identified, and the shape-description techniques most commonly used for shape retrieval are determined.

A new shape signature for Fourier descriptors, called the farthest point distance (FPD) signature, is proposed as a method of overcoming some of the disadvantages of previous shape signatures. The proposed signature calculates the distances between different corners of a shape and captures the points within the shape that humans use as a visual focus in order to classify shapes. To investigate the merit of the newly proposed signature, it is experimentally evaluated and compared to eight commonly used shape signatures. These signatures include the radial distance (RD), the triangular centroid area (TCA), the triangular area representation (TAR), the complex coordinate (CC), the chord-length distance (CLD), the angular function (AF), the polar coordinates (PC), and the angular radial coordinates (ARC) signatures. The experimental results clearly show that the proposed signature outperforms commonly used signatures because it tends to capture information about the corners of each object. This information, that is, the extreme points of a shape, also corresponds to the natural focus of human visual attention. To improve the ability of the proposed descriptors to capture the global characteristics of shapes, simple global descriptors are used with the proposed descriptors. Experimental investigations are conducted with respect to the size of the feature vector of the proposed

descriptors that is derived using the FPD signature and simple global descriptors. The results show that retrieving images from standard databases based on the shape of the image requires only 15 Fourier descriptors. These compact features are compared with the most effective and compact shape descriptors in the literature. The main advantages of the Fourier descriptors derived from the proposed signature are their compactness and the simplicity of their matching process. However, these descriptors are not capable of capturing the local characteristics of a shape. Consequently, another shape descriptor that is able to capture the local shape characteristics is also proposed. This descriptor is called a curvature-based Fourier shape descriptor (CBFD).

The curvature-based Fourier shape technique employs an unconventional view of the curvature-scale-space map of a shape contour because it treats it as a 2-D binary image. Invariant descriptors are derived from the 2-D Fourier transform of the curvature-scale image, which enables the descriptors to capture the detailed dynamics of the curvature of the shape contour and enhances the efficiency of the shape-matching process. The strength of the curvature-based shape techniques lies in the fact that it captures the local characteristics of a shape by means of descriptors that can facilitate a simple matching process. Since the curvature-based Fourier descriptors are 2-D, they have been compared with other (notable) 2-D descriptors that use a simple matching distance measure such as Euclidean distance in order to retrieve shapes. These other descriptors include angular radial transform descriptors, Zernike moment descriptors, radial Tchebichef moment descriptors, and generic Fourier descriptors. The CBFD has also been compared with standard contour-based shape descriptors (CSS) and 1-D Fourier descriptors.

The results of experiments using MPEG-7 databases in conjunction with a specially created noisy database demonstrate the effectiveness of the proposed CBFDs. The high performance of the CBFD in the presence of noise is due to their construction using two

filtering processes: the first is a smoothing process using a Gaussian filter to obtain the curvature scale image, and the second ignores high-frequency components that are sensitive to noise.

The rich content of images makes it impossible for one particular technique to be effective for all types of shapes. Consequently, a fusion algorithm has been proposed so that a group of techniques can be combined. In this algorithm, several techniques exchange their ranking information based on a pair-wise co-ranking scheme in order to produce a set of pair-wise conditional rankings and improve their ranking decisions. Then, each set of the pair-wise conditional rankings obtained through the cooperation of each pair of techniques is combined by averaging their rankings. The pair-wise co-ranking scheme is then applied recursively in order to reach a consensus about the final ranking decisions. Since the effectiveness of the descriptor used is image dependent, the combination scheme must be dynamic as a function of the query context. Consequently, a context-based fusion algorithm has been proposed so that the context of the query image can be considered in order to integrate several shape-based retrieval techniques. To test the performance of the proposed fusion algorithm, several techniques have been selected to work together, all of which use a simple matching process. Experiments based on the MPEG-7 databases and the specially created noisy database have been conducted in order to validate the performance of the proposed fusion algorithm. The results clearly show the effectiveness of the algorithm in retrieving images based on their shapes.

The main contributions of this study can be summarized as follows:

1. The merit of the most notable shape descriptors for image shape retrieval has been systematically and comprehensively investigated.

2. A shape signature has been proposed for deriving Fourier descriptors.

3. Curvature-based Fourier descriptors that are capable of capturing local characteristics have been proposed.

4. A fusion algorithm based on a team consensus approach has been proposed as a method of combining several shape-retrieval techniques and enhancing the performance of the shape-based image retrieval process.

6.1 Suggestions for Future Work

Although the present research provides solutions to shape-based image retrieval problems, extensive opportunities exist for further research. To improve the performance of shape-based image retrieval, investigations of the following areas of future work are suggested:

- The proposed curvature-based shape descriptors use only the magnitude and ignore the phase of the Fourier transform. This deficiency leads to losses in some of the shape characteristics, which might be ones that are a key factor in distinguishing between two shapes. The overall accuracy of the proposed descriptors will be improved if the phase information of the Fourier transform is also considered. An extension of the curvature-based Fourier descriptors so that they can handle 3-D shapes also deserves investigation because such an extension may make it possible to retrieve 3-D shapes.

- The proposed fusion algorithm can be used to integrate text and image content so as to develop a hybrid retrieval system that can be developed to support both query by keyword and query by image content. Furthermore the context of the keywords can be used to guide a group of techniques to reach a better final decision.

- Relevance feedback [143] can be added to the shape-based image retrieval system in order to improve the effectiveness of the retrieval. This addition would allow the retrieval system to learn through feedback from each user query. Clustering techniques can also be used to organize image collection and improve retrieval efficiency.

- Current research into the use of high-level features is still in its early stages [144]. A key issue in content-based image retrieval (CBIR) research is the exploration of ways to bridge the gap between the high-level semantics of an image and its lower-level properties that are derived from shape, color, and texture. The semantic features can be used with the low level features to improve the proposed fusion algorithm.

6.2 Publications Related to This Thesis

The following is a list of the articles published, submitted, and to be submitted

Book Chapters

1. A. El-Ghazal, O. Basir and S. Belkasim "Shape-Based Image Retrieval using Pair-wise Candidate Co-ranking", in M. Kamel and A. Campilho (Eds), *Image Analysis and Recognition,* Springer-Verlag Berlin Heidelberg: Lecture Notes in Computer Science, pp. 650-661, 2007.

Journal Papers

2. A. El-Ghazal, O. Basir and S. Belkasim "Farthest Point Distance: A New Shape Signature for Fourier Descriptors" *Signal Processing: Image Communicatio*n, Volume 24, Issue 7, pp. 572-586, 2009.

3. A. El-Ghazal, O. Basir and S. Belkasim "Curvature-Based Fourier Descriptor for shape Retrieval", To be submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

4. A. El-Ghazal, and O. Basir and S. Belkasim "Invariant shape descriptors: A survey", To be submitted to *Pattern Recognition, 2009*.

5. A. El-Ghazal, and O. Basir and S. Belkasim "Shape-based image retrieval: A team consensuses approach" To be submitted to: *Journal of Advances in Information Fusion*, 2009.

Conference Papers

6. A. El-Ghazal, O. Basir and S. Belkasim "Scale invariants of Radial Tchebichef moments", accepted in *the IEEE International Symposium on Multimedia,* San Diego, California, U.S.A, 2009.

7. A. El-Ghazal, O. Basir and S. Belkasim, (2008) "A Novel Curvature-Based Shape Fourier Descriptor", in *the 15th IEEE International Conference on Image Processing*, San Diego, California, U.S.A., pp. 953-956, October 2008.

8. A. El-Ghazal, O. Basir and S. Belkasim, "A Context-Based Fusion Algorithm for Shape Retrieval", in *the 11th International Conference on Information Fusion*, July, Cologne, Germany, pp. 1-8, July 2008.

9. A. El-Ghazal, O. Basir and S. Belkasim, "A Novel Wavelet-Based Optimal Thresholding Algorithm", *The 10th IASTED International Conference on Signal and Image Processing,* Hawaii, USA. August 2008.

10. A. El-Ghazal, O. Basir and S. Belkasim, "A New Shape Signature for Fourier Descriptors" *The 14th IEEE International Conference on Image Processing*, San Antonio, Texas, U.S.A., pp. I-161-I-164, September 2007.

11. A. El-Ghazal, O. Basir and S. Belkasim, "A Consensus-based Fusion Algorithm in Shape-based Image Retrieval", *The 10th International Conference on Information Fusion*. Quebec City, Canada, pp. 1-8, July 2007.

Bibliography

[1]   K. Hirata and T. Kato, "Query by visual example - content-based image retrieval," in *Third International Conference on Extending Database Technology*, 1992, pp. 56-71.

[2]   W. N. R. Baraber, W. Equitz, M. Flickner, e. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, "The QBIC project: Querying image by content using color, texture and shape," in *Storage and Retrieval for Image and Video Databases(SPIE)*, 1993, pp. 173-187.

[3]   M. Flickner, "Query by Image and Video Content: the QBIC System," *IEEE Transactions on Computers,* vol. 28, pp. 23-32, 1995.

[4]   "http://wwwqbic.almaden.ibm.com."

[5]   J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, Jain, and C.-f. Shu, "An Open Framework for Image Management.," in *SPIE Conference On Storage and Retrieval for Image and Video Databases* San Jose, CA, 1996, pp. 76-87.

[6]   A. Pentland, "Photobook: Tools for Content-Based Manipulation of Image Databases," *International Journal of Computer Vision,* vol. 18, pp. 233-254, 1996.

[7]   J. R. Smith and S. F. Chang., "VisualSEEK: A Fully Automated Content-based Image Query System," in *ACM Multimedia Conference* Boston, MA, 1996, pp. 87-98.

[8]   R. C. Veltkamp and M. Tanase, "A Survey of Content-Based Image Retrieval Systems. ," in *O. Marques, B. Furht, Content-Based Image and Video Retrieval, by Oge Marques and Borko Furht* Kluwer, 2002, pp. 47-101.

[9]   D. Mumford, "Mathematical theories of shape: Do they model perception?," in *SPIE Conference on Geometric Methods in Computer Vision*. vol. 1570 San Diego, California, 1991, pp. 2-10.

[10]   I. Biederman, "Recognition-by-components: a theory of human image understanding," *Psychological Review* vol. 94, pp. 115-147, 1987.

[11]   L. Schomaker, E. d. Leau, and L. Vuurpijl, "Using Pen-Based Outlines for Object-Based Annotation and Image-Based Queries," *Visual Information and Information Systems,* pp. 585-592, 1999.

[12]   Z. Wang, Z. Chi, and D. Feng, "Shape based leaf image retrieval," in *IEE Proceedings, Vision Image and Signal Processing*. vol. 150, 2003, pp. 34-43.

[13]   M. E. Celebi and Y. A. Aslandogan, "A Comparative Study of Three Moment-Based Shape Descriptors," in *International Conference on Information Technology: Coding and Computing*. vol. 1, 2005, pp. 788-793.

[14]   Q. Chenl, E. Petriul, and X. Yang, "A Comparative Study of Fourier Descriptors and Hu's Seven Moment Invariants for Image Recognition," in *Canadian Conference on Electrical and Computer Engineering (CCECE   )*, Niagara Falls, Maylmai 2004, pp. 103-106.

[15]   D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition,* vol. 37, pp. 1-19, 2004.

[16]   "The MPEG Home Page. http://www.chiariglione.org/mpeg/index.htm."

[17]   I. Kunttu and L. Lepistö, "Shape-based retrieval of industrial surface defects using angular radius Fourier descriptor," *IET Image Processing* vol. 1, pp. 231-236, 2007.

[18]   S. Loncaric, "A survey of shape analysis techniques," *Pattern Recognition,* vol. 31, pp. 983-1001, 1998.

[19]   R. B. Yadav, N. K. Nishchal, A. K. Gupta, and V. K. Rastogi, "Retrieval and classification of shape-based objects using Fourier, generic Fourier, and wavelet-Fourier descriptors technique: A comparative Study," *Optics and Lasers in Engineering,* vol. 45, pp. 695-708, 2008.

[20]   M. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Transactions on Information Theory,* vol. IT-8, pp. 179-187, 1962.

[21]   S. S. Reddi, "Radial and angular moment invariants for image identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 3, pp. 240-242, 1981.

[22]   B. Bamieh and R. De Figueiredo, "A general moment-invariants/attributed-graph method for three-dimensional object recognition from a single image," *IEEE Journal of Robotics and Automation,* vol. 2, pp. 31-41, 1986.

[23]   Y. S. Abu-Mostafa and D. Psaltis, "Recognitive aspects of moment invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 6, pp. 698-706, 1984.

[24]   J. Flusser, "On the Independence of Rotation Moment Invariants," *Pattern Recognition,* vol. 33, pp. 1405-1410, 2000.

[25]   S. A. Dudani and K. J. Breeding, "Aircraft identification by moment invariants," *IEEE Transactions on Computers,* vol. C-26, pp. 39-45, 1977.

[26]   R. G. Casey, "Moment Normalization of Handprinted Characters," *IBM J. Res. Develop,* vol. 14, pp. 548-557, 1970.

[27]   A. P. Reeves, R. J. Prkop, S. E. Andrews, and F. P. Kuhl, "Three-Dimensional Shape Analysis using Moments and Fourier Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 10, pp. 937 - 943   1988.

[28] J. C. Yang, J. W. Shin, and D. S. Park, "Fingerprint Matching Using Invariant Moment Features " in *Lecture Notes in Computer Science, Computational Intelligence and Security* 2007, pp. 1029-1038.

[29] C. Anant, E. Bruno, L. Helene, and R. Christophe, "Comparative study of global invariant descriptors for object recognition," *Journal of Electronic Imaging,* vol. 17, pp. 023015-10, 2008.

[30] D. Xua and H. Li, "Geometric moment invariants " *Pattern Recognition,* vol. 41, pp. 240-249, 2008.

[31] V. Constantin-Iulian, M. Doru, and M. Cristian, "A new version of Flusser moment set for pattern feature extraction," *WSEAS Transactions on Information Science and Applications* vol. 5, pp. 396-406, 2008.

[32] M. Teague, "Image analysis via the general theory of moments," *Journal of the Optical Society of America,* vol. 70, pp. 920-930 1980.

[33] S. O. Belkasim, M. Shridhar, and M. Ahmadi, "Pattern recognition with moment invariants: a comparative study and new results," *Pattern recognition,* vol. 24, pp. 1117-1138, 1991.

[34] R. Bailey and M. Srinath, "Orthogonal moment features for use with parametric and non-parametric classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 18, pp. 389 - 399 1996.

[35] C. Kan and M. D. Srinath, "Invariant character recognition with Zernike and orthogonal Fourier Mellin moments," *Pattern Recognition,* vol. 35, pp. 143-154, 2002.

[36] A. Khotanzad, "Invariant Image Recognition by Zernike Moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 12, pp. 489-497, 1990.

[37] W.-Y. Kim and Y.-S. Kim, "A region-based shape descriptor using Zernike moments," *Signal Processing: Image Communication,* vol. 16, pp. 95-102, 2000.

[38] D. Zhang and G. Lu, "Evaluation of MPEG-7 Shape Descriptors against other Shape Descriptors," *Multimedia Systems,* vol. 9, pp. 15-30, 2003.

[39] M. Bober, "MPEG-7 visual shape descriptors," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 11, pp. 716-719, 2001.

[40] F. Mokhtarian and M. Bober, *Curvature scale space representation: Theory application and MPEG-7 standardization*, first ed.: Kluwer Academic Publishers, 2003.

[41] Y. Bin and P. Jia-xiong, "Improvement and invariance analysis of Zernike moments using as a region-based shape descriptor," in *XV Brazilian Symposium on Computer Graphics and Image Processing* 2002, pp. 120 - 127.

[42] R. Mukundan, S. H. Ong, and P. A. Lee, "Image Analysis by Tchebichef Moments," *IEEE Transactions on Image Processing,* vol. 10, pp. 1357-1364, 2001.

[43] P. T. P. Yap and R. S. H. Ong, "Image analysis by Krawtchouk moments," *IEEE Transactions on Image Processing* vol. 12, pp. 1367-1377, 2003.

[44] H. Zhu, H. Shu, J. Liang, L. Luo, and J. L. Coatrieux, "Image analysis by discrete orthogonal Racah moments," *Signal Processing,* vol. 87, pp. 687-708, 2007.

[45] H. Zhu, H. Shu, J. Zhou, L. Luo, and J. L. Coatrieux, "Image analysis by discrete orthogonal dual Hahn moments," *Pattern Recognition Letters,* vol. 28, pp. 1688-1704, 2007.

[46] R. Mukundan, "Radial Tchebichef Invariants for Pattern Recognition," in *IEEE Tencon Conference*, 2005, pp. 2098–2103.

[47]   Z. Ping, H. Ren, J. Zou, Y. Sheng, and W. Bo, "Generic orthogonal moments: Jacobi–Fourier moments for invariant image description," *Pattern Recognition,* vol. 40, pp. 1245-1254, 2007.

[48]   Y. Sheng and J. Duvernoy, "Circular-Fourier-radial-Mellin transform descriptors for pattern recognition," *Journal of the American Society for Information Science and Technology,* vol. 3, pp. 885-888, 1986.

[49]   Y. Sheng and H. H. Arsenault, "Experiments on pattern recognition using invariant Fourier-Mellin descriptors," *Journal of the Optical Society of America A,* vol. 3, pp. 771-776, 1986.

[50]   Y. Sheng and L. Shen, "Orthogonal Fourier-Mellin moments for invariant pattern recognition," *Journal of the Optical Society of America-A,* vol. 11, pp. 1748-1757, 1994.

[51]   G. Granlund, "Fourier preprocessing for hand print character recognition," *IEEE Transactions on Computers,* vol. 21, pp. 195-201, 1972.

[52]   T. P. Wallace and P A Wintz, "An efficient three dimensional aircraft recognition algorithm using normalized Fourier descriptors," *Computer Graphics and Image Processing,* vol. 13, pp. 99-126, 1980.

[53]   R. Tello, "Fourier descriptors for computer graphics," *Systems, Man and Cybernetics, IEEE Transactions on,* vol. 25, pp. 861-865, 1995.

[54]   S. P. Raman and U. B. Desai, "2-D object recognition using Fourier Mellin transform and a MLP network," in *IEEE International Conference on Neural Networks*, 1995, pp. 2154–2156.

[55]   S. Adam, J. M. Ogier, C. Cariou, R. Mullot, J. Garde, and J. Labiche, "Character Recognition, Orientation, and Scale Estimation Thanks to the Fourier Mellin Transform," *Lecture Notes in Computer Science,* pp. 472-481, 2000.

[56]   D. Zhang and G. Lu, "Shape-based image retrieval using generic Fourier descriptor," *Signal Processing: Image Communication,* vol. 17, pp. 825–848, 2002.

[57]   S. Li and M.-C. Lee, "Effective invariant features for shape-based image retrieval," *Journal   of the American Society for Information Science and Technology,* vol. 56, pp. 729–740, 2005.

[58]   A. Goshtasby, "Description and discrimination of planar shapes using shape matrices," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 7, pp. 738-743, 1985.

[59]   A. Taza and C. Y. Suen, "Discrimination of planar shapes using shape matrices," *IEEE Transactions on Systems, Man and Cybernetics,* vol. 19, pp. 1281-1289, 1989.

[60]   G. Lu and A. Sajjanhar, "Region-based shape representation and similarity measure suitable for content based image retrieval," *Multimedia Systems* vol. 7, pp. 165–174, 1999.

[61]   K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker, "Shock Graphs and Shape Matching," *International Journal of Computer Vision,* vol. 35, pp. 13-32, 1999.

[62]   T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Shock-Based Indexing into Large Shape Databases," *Lecture Notes in Computer Science,* pp. 731-746, 2002.

[63]   L. He, C. Y. Han, X. Wang, X. Li, and W. G. Wee, "A skeleton based shape matching and recovery approach," in *International Conference on Image Processing*, 2002.

[64]   T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Recognition of Shapes by Editing Their Shock Graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 550-571, 2004.

[65]   W. B. Goh, "Strategies for shape matching using skeletons," *Computer Vision and Image Understanding,* vol. 110, pp. 326-345, 2008.

[66]   H. Zaboli, M. Rahmati, and A. Mirzaei, "Shape Recognition and Retrieval: A Structural Approach Using Velocity Function " in *The 12th International Conference on Computer Analysis of Images and Patterns*, Vienna, Austria 2007.

[67]   R. C. Gonzalez and R. E. Woods, *Digital Image Processing*: Addison-Wesley, 2002.

[68]   X. Bai, L. J. Latecki, and W. Y. Liu, "Skeleton Pruning by Contour Partitioning with Discrete Curve Evolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 449-462, 2007.

[69]   P. Eggleston, "Constraint-based feature indexing and retrieval for image database," in *SPIE*, 1992, pp. 27-39.

[70]   J.-K. Wu, M. S. Kankanhalli, K. W. J. K. Wu, L. J.-H. Lim, and H. D. Hong, *Perspectives on Content-Based Multimedia Systems*: Kluwer Academic Publishers, 2000.

[71]   J. C. Russ, *The Image Processing Handbook*, Fourth ed.: CRC Press 2002.

[72]   O. E. Badawy and M. Kamel, "Shape-Based Image Retrieval Applied To Trademark Images," *International Journal of Image and Graphics,* vol. 2, pp. 1-19, 2002.

[73]   K. R. Castleman, *Digital image processing*: Prentices Hall, 1996.

[74]   B. Furht, *Handbook of multimedia computing*: CRC Press 1998.

[75] B. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*: John Wiley and Sons, 2002.

[76] C.-H. Teh and R. Chin, "On Image Analysis by the Methods of Moments " *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 10, pp. 496-513, 1988.

[77] Y. Li, "Reforming the Theory of Invariant Moments for Pattern Recognition," *Pattern Recognition,* vol. 25, pp. 723-730, 1992.

[78] S. Belkasim, E. Hassan, and T. Obeidi, "Explicit invariance of Cartesian Zernike moments," *Pattern Recognition Letters,* vol. 28, pp. 1969-1980, 2007.

[79] C. W. Chong, P. Raveendran, and R. Mukundan, "The scale invariants of pseudo-Zernike moments," *Pattern Analysis & Applications,* vol. 6, pp. 176-184, 2003.

[80] F. Ghorbel, "A complete invariant description for gray-level images by the harmonic analysis approach," *Pattern Recognition Letters,* vol. 15, pp. 1043-1051, 1994.

[81] S. Derrode and F. Ghorbel, "Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description," *Computer Vision and Image Understanding,* vol. 83, pp. 57-78, 2001.

[82] Q. S. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transformsfor image registration and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 16, pp. 1156-1168, 1994.

[83] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing,* vol. 5, pp. 1266-1271, 1996.

[84]   S. Derrode, M. Daoudi, and F. Ghorbel, "Invariant content-based image retrieval using a complete set of Fourier-Mellin descriptors," in *IEEE International Conference on Multimedia Computing and Systems*, 1999, pp. 877-881

[85]   D. S. Zhang and G. Lu, "Study and Evaluation of Different Fourier Methods for Image Retrieval," *Image and Vision Computing,* vol. 23, pp. 33-49, 2005.

[86]   T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Transactions on Computers,* vol. 21, pp. 269-281, 1972.

[87]   I. Kunttu, L. Lepisto, J. Rauhamaa, and A. Visa, " Multiscale Fourier descriptors for defect image retrieval," *Pattern Recognition Letters,* vol. 27, pp. 123–132, 2006.

[88]   E. Persoon and K. S. Fu., "Shape discrimination using Fourier descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 8, pp. 388-397, 1986.

[89]   D. S. Z. a. G. Lu., "A Comparative Study of Curvature Scale Space and Fourier Descriptors," *Journal of Visual Communication and Image Representation,* vol. 14, pp. 41-60, 2003.

[90]   G. Eichmann, C. Lu, M. Jankowski, and R. Tolimieri, "Shape representation by Gabor expansion," in *Hybrid Image and Signal Processing II*. vol. 1297, 1990, pp. 86-94.

[91]   K. Arbter, *Affine-invariant Fourier descriptors, in From Pixels to Feature, Amsterdam, The Netherlands*. Amsterdam, The Netherlands: Elsevier Science, 1989.

[92]   K. Arbter, W. E. Snyder, H. Burkhardt, and G. Hirzinger, "Application of Affine-Invariant Fourier Descriptors to Recognition of 3-D Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 12, pp. 640-647, 1990.

[93] A. E. Oirrak, M. Daoudi, and D. Aboutajdin, "Affine Invariant Descriptors using Fourier Series," *Pattern Recognition Letters,* vol. 23, pp. 1109-1118, 2002.

[94] T. W. Rauber and S. Garcao, "Shape description by UNL Fourier features-an application to handwritten character recognition," in *11th IAPR International Conference on on Pattern Recognition*, The Hague, Netherlands, 1992, pp. 466-469.

[95] T. W. Rauber and A. S. Steiger Garcao, "2-D form descriptors based on a normalized parametric polar transform (UNL transform)'," in *IAPR Workshop on Machine Vision Applications*, Tokyo, Japan, 1992, pp. 561-566.

[96] I. Bartolini, P. Ciaccia, and M. Patella, "WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 142-147, 2005.

[97] A. El-ghazal, O. Basir, and S. Belkasim, "Curvature-based Fourier Shape Descriptor," in *15th IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 953-956.

[98] A. El-ghazal, O. Basir, and S. Belkasim, "Farthest point distance: A new shape signature for Fourier descriptors " *Signal Processing: Image Communication,* vol. 24, pp. 572-586, 2009.

[99] R. Mukundan and K. R. Ramakrishnan, *Moment Functions in Image Analysis: Theory and Applications*: World Scientific, 1998.

[100] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*: Chapman & Hall Computing, 1993.

[101] R. Kashyap and R. Chellappa, "Stochastic models for closed boundary analysis: Representation and reconstruction," *IEEE Transactions on Information Theory,* vol. 27, pp. 627-637, 1981.

[102] S. R. Dubois and F. H. Glanz, "An autoregressive model approach to two-dimensional shape classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 8, pp. 55-66, 1986.

[103] M. Das, M. J. Paulik, and N. K. Loh, "A bivariate autoregressive technique for analysis andclassification of planar shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 12, pp. 97-103, 1990.

[104] I. Sekita, T. Kurita, and N. Otsu, "Complex autoregressive model for shape recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, pp. 489-496, 1992.

[105] H. Asada and M. Brady, "The Curvature Primal Sketch," *MIT AI Memo,* 1984.

[106] H. Asada and M. Brady, "The Curvature Primal Sketch," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 8, pp. 2-14, 1986.

[107] F. Mokhtarian and A. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 8, pp. 34-43, 1986.

[108] F. Mokhtarian and A. Mackworth, "A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, pp. 789-805, 1992.

[109] F. Mokhtarian, "Silhouette-based isolated object recognition through curvature scale space," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 17, pp. 539-544, 1995.

[110] F. Mokhtarian, S. Abbasi, and J. Kittler, "Efficient and Robust Retrieval by Shape Content through Curvature Scale Space,'" in *International Workshop on Image DataBases and MultiMedia Search*, 1996, pp. 35-42.

[111] F. Mokhtarian, "Silhouette-based occluded object recognition through curvature scale space," *Machine Vision and Applications,* vol. 10, pp. 87-97, 1997.

[112] S. Abbasi, F. Mokhtarian, and J. Kittler, "Curvature Scale Space Image in Shape Similarity Retrieval," *Multimedia Systems,* vol. 7, pp. 467-476, 1999.

[113] S. Abbasi, F. Mokhtarian, and J. Kittler, "Enhancing CSS-based shape retrieval for objects with shallow concavities," *Image and Vision Computing,* vol. 18, pp. 199-211, 2000.

[114] I. El-Rube, N. Alajlan, M. Kamel, M. Ahmed, and G. Freeman, "MTAR: A Robust 2D Shape Representation," *International Journal of Image and Graphics* vol. 6, pp. 421-443, 2006.

[115] I. El-Rube, N. Alajlan, M. Kamel, M. Ahmed, and G. Freeman, "Efficient Multiscale Shape-Based Representation and Retrieval," in *International Conference on Image Analysis and Recognition* Toronto, 2005, pp. 415-422.

[116] J. v. d. Poel, C. W. D. d. Almeida, and L. V. Batista, "A new multiscale, curvature-based shape representation technique for image retrieval based on DSP techniques," in *Fifth International Conference on Hybrid Intelligent Systems* 2005, pp. 373 -378.

[117] Q. M. Tieng and W. W. Boles, "Recognition of 2D Object Contours Using the Wavelet Transform Zero-Crossing Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 910-916, 1997.

147

[118] H. S. Yang, S. U. Lee, and K. M. Lee, "Recognition of 2D Object Contours Using Starting-Point-Independent Wavelet Coefficient Matching," *Journal of Visual Communication and Image Representation,* vol. 9, pp. 171-181, 1998.

[119] D. Zhang, "Image Retrieval Based on Shape." vol. PhD: Monash University, 2002.

[120] E. G. M. Petrakis, A. Diplaros, and E. Milios, "Matching and Retrieval of Distorted and Occluded Shapes Using Dynamic Programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 1501-1516, 2002.

[121] N. Arica and F. T. Y. Vural, "BAS: a perceptual shade descriptor based on the beam angle statistics," *Pattern Recognition Letters,* vol. 24, pp. 1627-1639, 2003.

[122] L. J. Lateckia, R. Lakaempera, and D. Wolter, "Optimal partial shape similarity," *Image and Vision Computing,* vol. 23, pp. 227-236, 2005 2005.

[123] N. Alajlan, I. E. Rube, M. S. Kamel, and G. Freeman, "Shape Retrieval Using Triangle-area Representation and Dynamic Space Warping," *Pattern Recognition,* vol. 40, pp. 1911-1920, 2007.

[124] W. I. Grosky and R. Mehrotra, "Index-based object recognition in pictorial data management," *Computer Vision, Graphics, and Image Processing,* vol. 52, pp. 416-436, 1990.

[125] E. Bribiesca, "A new chain code," *Pattern Recognition,* vol. 32, pp. 235-251, 1999.

[126] S. Berretti, A. Del Bimbo, and P. Pala, "Retrieval by shape similarity with perceptual distance and effective indexing," *Multimedia, IEEE Transactions on,* vol. 2, pp. 225-239, 2000.

[127] E. Attalla and P. Siy, "Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching," *Pattern Recognition,* vol. 38, pp. 2229-2241, 2005.

[128] H. Freeman, "On the encoding of arbitrary geometry configurations," *IEEE Transactions on Electronic Computers,* vol. 10, pp. 260–268, 1961.

[129] H. Freeman, "Computer Processing of Line-Drawing Images," *ACM Computing Surveys,* vol. 6, pp. 57-97, 1974.

[130] J. Iivarinen and A. J. E. Visa, "Shape recognition of irregular objects," in *The International Society for Optical Engineering*, 1996, pp. 25-32.

[131] A.-B. M. Salem, A. A. Sewisy, and U. A. Elyan, "A Vertex Chain Code Approach for Image Recognition " *International Journal on Graphics, Vision and Image Processing,* vol. 5, 2005.

[132] B. M. Mehtre, M. S. Kankanhalli, and W. F. Lee, "Shape measures for content based image retrieval: A comparison," in *Information Processing and Management*. vol. 33, 1997, pp. 319-337.

[133] Peura and J. Iivarinen, "Efficiency of Simple Shape Descriptors," in *the 3rd Intl. Workshop on Visual Form*, Capri, Italy, 1997, pp. 28-30.

[134] E. Persoon and K. S. Fu, "Shape Discrimination Using Fourier Descriptors," *IEEE Transactions on Systems, Man and Cybernetics,* vol. 7, pp. 170-179, 1977.

[135] H. Kauppinen, T. Seppänen, and M. Pietikäinen, "An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 17, pp. 201-207 1995.

[136] A. El-Ghazal, O. Basir, and S. Belkasim, "A New Shape Signature for Fourier Descriptors," in *the 14th IEEE International Conference on Image Processing* San Antonio, Texas, USA, 2007, pp. 161-164.

[137] A. S. Aguado, M. E. Montiel, and M. S. Nixon, "Parameterising Arbitrary Shapes via Fourier Descriptors for Evidence-Gathering Extraction," *Computer Vision and Image Understanding* vol. 69, pp. 202-221, 1998.

[138] B. Pinkowski, "Multiscale Fourier Descriptors For Classifying Semivowels In Spectrograms," *Pattern Recognition,* vol. 26, pp. 1593-1602, 1993.

[139] V. V. Kindratenko, "On Using Functions to Describe the Shape," *Journal of Mathematical Imaging and Vision,* vol. 18, pp. 225-245, 2003.

[140] C. Yufeng, L. Fengxia, and H. Tianyu, "Curvature Feature Based Shape Analysis," in *Proceedings of the 4th international conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications - with Aspects of Theoretical and Methodological Issues* Shanghai, China: Springer-Verlag, 2008.

[141] S. Li, M. C. Lee, and C. M. Pun, "Complex Zernike Moments Features for Shape-Based Image Retrieval," *IEEE Transactions on Systems, Man and Cybernetics, Part A,* vol. 39, pp. 227-237, 2009.

[142] L. J. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *IEEE Conference Computer Vision and Pattern Recognition*, 2000.

[143] M. L. Kherfi and D. Ziou, "Relevance feedback for CBIR: a new approach based on probabilistic feature weighting with positive and negative examples," *IEEE Transactions on Image Processing,* vol. 15, pp. 1017-1030, 2006.

[144] L. Ying, Z. Dengsheng, and L. Guojun, "Region-based image retrieval with high-level semantics using decision tree learning," *Pattern Recognition,* vol. 41, pp. 2554-2570, 2008.