

AN EFFECTIVE SHAPE DESCRIPTOR FOR CONTENT BASED IMAGE RETRIEVAL

by

YUPENG LI

B. Sc., Central South University,
China 1991

A thesis
presented to Ryerson University
in partial fulfillment of the
requirement for the degree of
Master of Applied Science
in the Program of
Electrical and Computer Engineering

Toronto, Ontario, Canada, 2006

© Yupeng Li, 2006

UMI Number: EC53601

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.



UMI Microform EC53601

Copyright 2009 by ProQuest LLC

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

(Yupeng Li)

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

(Yupeng Li)

Borrower's Page

Ryerson University requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

[illegible]

ABSTRACT

AN EFFECTIVE SHAPE DESCRIPTOR FOR CONTENT BASED IMAGE RETRIEVAL

©Yupeng Li 2006

Master of Applied Science
Department of Electrical and Computer Engineering
Ryerson University

In this work, we present a modified version of the Generic Fourier Descriptor (GFD) that operates on edge information within natural images from the COREL image database for the purpose of shape-based image retrieval. By incorporating an edge-texture characterization (ETC) measure, we reduced the complexity inherent in oversensitive edge maps typical of most gradient-based detectors that otherwise tend to contaminate the shape feature description. We find that the proposed techniques not only improve overall retrieval in terms of shape, but more importantly, provide a more accurate similarity ranking measure of retrieved results, demonstrating the need for greater consideration for dominant internal and external shape details. A feature database combined by color moments, color histograms, Gabor wavelet and shape features is applied in our image retrieval system. Relevance feedback has also been considered, bridging the gap between the high level concepts and the low level visual features. The experimental results indicate that dynamically updating weights associated with feature components by users' feedback greatly improves retrieval performance.

Acknowledgments

First of all, I would like to take this opportunity to thank my supervisor Dr. Ling Guan for his encouragement, patience, and inspirational supervising throughout the course of my graduate study. Without his guidance and support I have received, it would not have been possible to finish this work.

I would also like to thank the Department of Electrical and Computer Engineering at Ryerson University for providing research infrastructure and the Ryerson Multimedia Research Laboratory (RML). My thanks are due to Natural Science and Engineering Research Council of Canada (NSERC) and Ryerson University for providing me financial support throughout my research work.

I would like to acknowledge many professors and graduate students at Ryerson University for their instructions and suggestions to my work. Special thanks are due to my colleagues at Ryerson Multimedia Research Laboratory for their help and friendship.

Finally, I would like to say thanks to my parents, sisters, daughters, and especially, my wife for their immeasurable support and understanding. I want them to know that this thesis is the working result of the whole family, rather than mine individually.

Contents

1	Introduction	1
1.1	General Background	1
1.2	Text-Based Image Retrieval	1
1.3	Content-Based Image Retrieval (CBIR)	3
1.4	Relevance Feedback in CBIR	4
1.5	Outline of Thesis	5
2	Related Work	6
2.1	Shape Description Techniques	6
2.1.1	Contour-Based Shape Description Techniques	8
2.1.2	Region-Based Shape Description Techniques	10
2.1.3	Edge Detection	13
2.2	Content-Based Image Retrieval System	15
2.2.1	Current CBIR Systems	15
2.2.2	Relevance Feedback in CBIR	16
2.3	Summary	18
3	CBIR Based-on Shape Feature for Natural Images	19
3.1	Edge-Texture Characterization (ETC) Measure	21
3.1.1	The Edge-Texture Characterization (ETC) Algorithm	22
3.1.2	The Extraction of Texture Regions	23
3.1.3	Experimental Results	23
3.2	Generic Fourier Descriptor (GFD)	31
3.2.1	Polar Fourier Transform	33
3.2.2	Derivation and Implementation of Generic FD	36
3.2.3	Modified Edge-Based GFD	37
3.3	Experimental Results	38
3.4	Summary	46
4	Dynamic Weight Selection for CBIR using Relevance Feedback	48
4.1	Feature Extraction	49
4.1.1	Color Features	50
4.1.2	Texture Features	52

4.2	Normalization	55
4.3	Weight Updating	57
4.4	Query Modification	58
4.5	Experimental Results	60
4.6	Summary	64
5	Conclusions	71
5.1	Summary of Thesis	71
5.2	Contributions	72
5.3	Future Research Extensions	73
	Bibliography	75
A	List of Publications	84

List of Figures

1.1	Image description using different information sources	2
1.2	Google and Lycos search engines	3
2.1	Classification of shape description techniques	7
2.2	Lena's edge maps using different edge detection approaches	15
3.1	The frame structure of our proposed image retrieval	20
3.2	The preprocessing procedure of Lena image	25
3.3	The preprocessing procedure of Flower image	26
3.4	ETC measurement of Lena image	27
3.5	ETC measurement of Flower image	28
3.6	Texture removing procedure of Lena edge map	29
3.7	Texture removing procedure of Flower edge map	30
3.8	(a)Original shape image in the polar space (b)Polar image plotted in the Cartesian space	35
3.9	One sample image of each class in the database	38
3.10	Column chart of retrieval performance	40
3.11	Flag query - CLASS 9	42
3.12	Aircraft query - CLASS 10	43
3.13	Cat query - CLASS 3	44
3.14	Fruit query - CLASS 1	45
4.1	The frame for weight selection using relevance feedback	49
4.2	Contours of the designed Gabor filter dictionary	55
4.3	The graph of weight vectors based on two different query images	58
4.4	Query modification model	59
4.5	The 50 query images from Corel database	61
4.6	Retrieval performance	63
4.7	The retrieval performance based on 'Rock' query (including shape features) .	65
4.8	The retrieval performance based on 'Rock' query (excluding shape features) .	66
4.9	The retrieval performance based on 'Plane' query (including shape features)	67
4.10	The retrieval performance based on 'Plane' query (excluding shape features)	68
4.11	The retrieval performance based on 'Sunrise' query (including shape features)	69
4.12	The retrieval performance based on 'Sunrise' query (excluding shape features)	70

List of Tables

2.1	Edge detection representations based on three different edge detectors	14
3.1	Retrieval performance from three different sets of shape features	39
4.1	Extracting algorithm of three types of feature descriptors	60
4.2	Average retrieval rate (in percentage) for 300 queries in 10,000 image database	62

Chapter 1

Introduction

1.1 General Background

WITH the rapid development of computer hardware and software technology, multimedia data processing has been applied in many sectors of society. Multimedia is the coherent integration of multiple media sources, e.g. text, graphics, images, audio and video. The use of images in human work and life is very common. It is widely applied in distance education, military, commerce, digital museum and medicine, etc. Digital photography, cheap storage (CDs and DVDs) and high-capacity public networks (Internet) have led to an expedited increase in digital image applications. However, the huge amount of images stored around the world makes the utilization of images from existing databases more difficult than ever. It is necessary to find effective tools for indexing and managing digital images. As a result, image retrieval has become a very active research area since 1970s. It mainly includes two retrieval models, one being text-based retrieval and the other content-based retrieval. Figure 1.1 is an illustrative representation of image description based on different information sources.

1.2 Text-Based Image Retrieval

Traditionally, images are stored in a database and indexed using textual information. Text-based image retrieval is an extension of the modern Information Retrieval (IR). The advan-

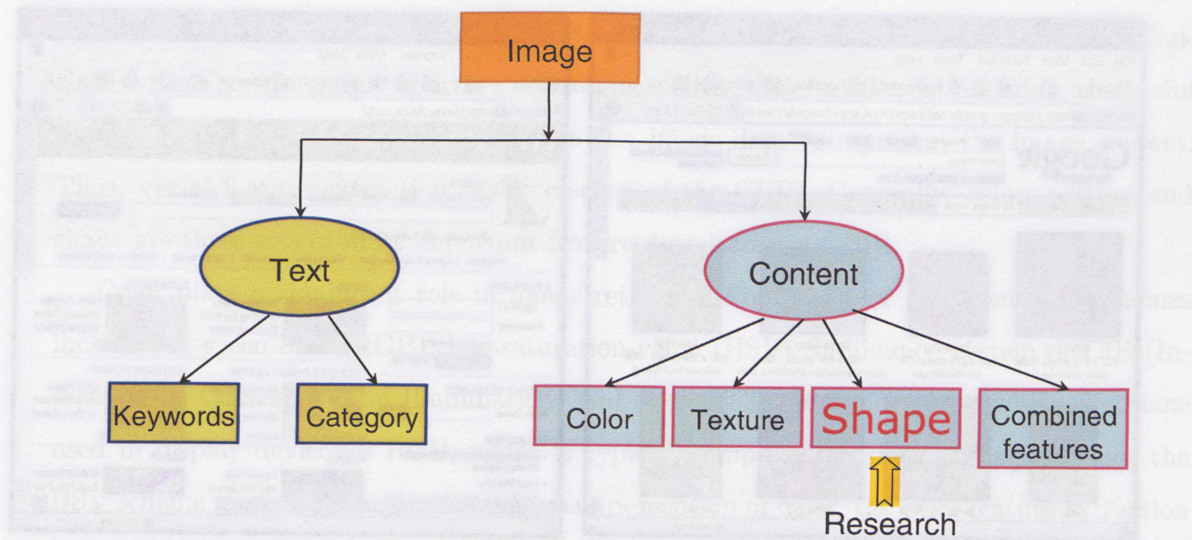


Figure 1.1: Image description using different information sources

One of the advantages of textual indexing of images is that it can provide the user with key word searching, catalogue browsing and even with query interface such as Structural Query Language (SQL), as shown in Figure 1.1. Text-based image search is often used to find information on the World Wide Web (WWW). Lycos [1], Google [2] and Altavista [3] are very popular web search engines. Users use Google search engine to search images by text queries and Lycos search engine can provide multimedia information including video, music and images. Similarly, Altavista search engine supports options of image, audio and video. Figure 1.2(a) and (b) show the Graphic User Interfaces of Google and Lycos search engines respectively. In spite of the popularity of text-based image retrieval on the website, it apparently has numerous limitations:

(a) When the image database is very large, it is almost impossible to manually annotate all the images.

(b) It is very difficult to find suitable textual words for the description of visual features of image.

(c) Manual image indexing is highly subjective, extremely laborious and only expresses the personal opinion of the user.

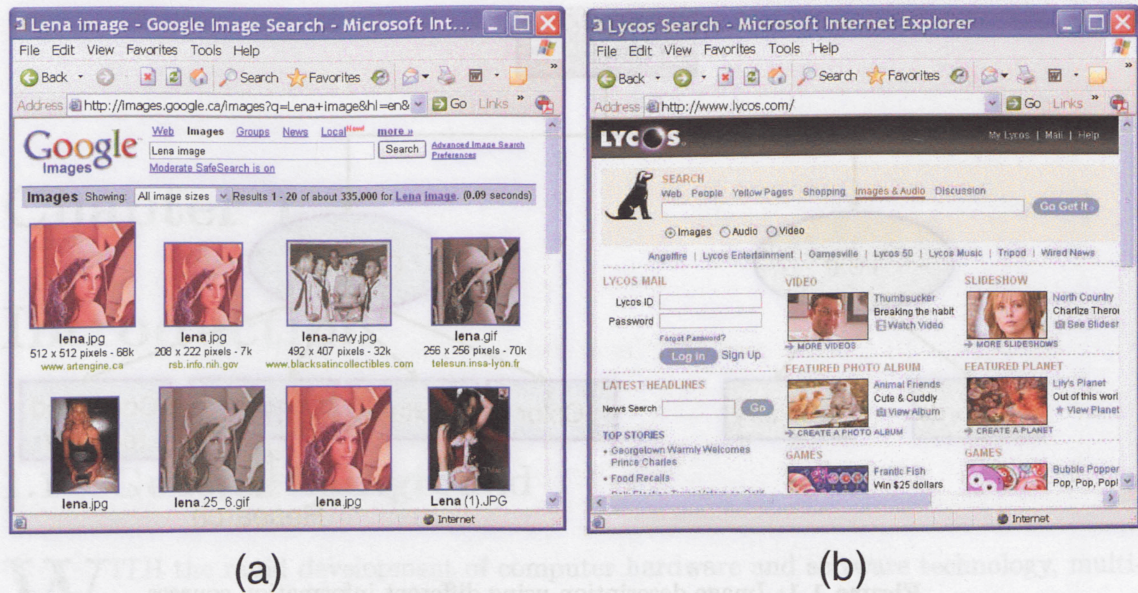


Figure 1.2: Google and Lycos search engines

(d) The users search the database by providing a textual description of the query which is, sometimes, very difficult to annotate.

Because of the limitations of text-based image retrieval, researchers have attempted to find alternative ways for indexing and retrieving images, which focus on the image indexing techniques that have the ability of retrieving images based on their content and can be automated. The technologies are now generally referred to as Content-Based Image Retrieval (CBIR).

1.3 Content-Based Image Retrieval (CBIR)

The objective of CBIR is to find all the visual documents in the image libraries similar to the query, based solely on the content information. A typical example from telemedicine area is that a medical doctor in a remote area is studying a CT scanning image of a patient, which shows a tumor like object. The doctor wants to search for similar cases from a medical database in order to make a inferred diagnostic decision. Visual feature extraction, indexing

and similarity measures are the major basis in content-based image retrieval. Because digitized images consist purely of arrays of pixel intensities with no inherent meaning, the useful information should be extracted from the raw image data to represent the image content. Thus, visual feature extraction is the essence of the CBIR. Generally, color, texture and shape are three sets of most important feature descriptors in CBIR.

Color plays a significant role in image retrieval. Popular color representation schemes include red-green-blue (RGB), hue-saturation-value (HSV), luminance system of CIE (International Commission on Illumination) and others [63]. The most acknowledged scheme used in display devices is RGB, which is typically employed in digital images. Since the HSV scheme more accurately reflects human perception of color, the color feature extraction usually employs this format.

Texture is a visual pattern where there are a large number of visible elements densely and evenly arranged. A texture element is a uniform intensity region of simple shape which is repeated. Texture has already been applied in pattern recognition and computer vision.

Shape-based image retrieval is one of the hardest problems in general image retrieval. This is mainly due to the difficulty of finding an appropriate shape descriptor. One of our main research objectives in this thesis is to focus on image indexing and retrieval using shape features, finding an effective shape descriptor in natural image retrieval (Figure 1.1).

1.4 Relevance Feedback in CBIR

The focus of the early attempts in CBIR was to develop completely automated, open loop systems without any user feedback. However, there is a significant semantic gap between the high level concepts understood by the human perception and the low level features used by computer for image representation. Different human beings based on their subjective opinions may describe the same content differently, especially for those images including more objects. For example, we have an image with a plane being in the centre of the scene. One person may be concerned with the plane's purpose (passenger, fighter or private

plane), and another may consider the plane's status (flying in the sky or stopping at the airport). Hence, the user's feedback is crucial in content-based image retrieval. The process of collecting feedback information from the users by labeling the partial retrieval results is called *Relevance Feedback* (RF). Researchers have developed different RF algorithms for CBIR to improve the retrieval performance. Some of the key approaches will be reviewed in Chapter 2.

1.5 Outline of Thesis

The remainder of this thesis is organized as follows:

Chapter 2: Related Work Previous works in the field of content-based image retrieval are reviewed, including CBIR based on different shape descriptors, image retrieval systems and relevance feedback. For the preparation of our proposed shape descriptor, the detection and extraction of edge information in images are also discussed.

Chapter 3: CBIR based-on shape feature for natural images We present the algorithm of generic Fourier descriptor (GFD) in details. The algorithm of edge-texture characterization (ETC) is applied to remove textured parts from the edge map of an image extracted by the Canny filter. A new shape descriptor: the modified generic Fourier descriptor (MGFD) is proposed for natural image retrieval.

Chapter 4: Dynamic weight selection for CBIR using relevance feedback The extraction of color and texture features in the literature are first reviewed. A feature database consisting of color (color histograms and color moments), texture (Gabor wavelet transform) and shape (modified generic Fourier descriptor) is described. The normalization approaches for the feature database are also discussed. Dynamic weight updating by user's feedback and modified query give the best performance in CBIR.

Chapter 5: Conclusions We discuss the results and summarize the contributions of this thesis. Recommendations for future research are also provided.

Chapter 2

Related Work

CONTENT Based Image Retrieval (CBIR) is used for retrieving semantically relevant images from an image database based on automatically derived image features [4]. The main goal of CBIR is to improve the efficiency of image indexing and retrieval, thereby reducing human interaction in the indexing process.

One of the main tasks for CBIR systems is feature extraction of every image based on its pixel values and similarity comparison. These features (color, texture and shape) become the image representation for measuring the similarity between query and other images in the database. The main objective of this thesis is to develop a new effective shape descriptor, in order to better understand shape-based image retrieval system, some current shape descriptors are reviewed in section 2.1. Edge detection and extraction of images are also covered in this section. Nevertheless, shape features combined with other features will ultimately be applied to CBIR systems, thus, some existent CBIR systems are discussed in section 2.2, including relevance feedback in CBIR. Color and texture descriptors will be described briefly in Chapter 4. We summarize this Chapter in section 2.3.

2.1 Shape Description Techniques

Shape is an important visual feature. It is widely applied in object recognition, matching, registration and analysis areas, it is one of the primitive features for image content description. The aim of shape description is to uniquely characterize the shape using a shape

descriptor. However, shape content description is a difficult task, because it is not easy to define perceptual shape features and make similarity measures between shapes. Moreover, shape is often corrupted with noise, defection, arbitrary distortion and occlusion. As a result, the search for shape representation and description techniques becomes a challenging task. In the following, shape description techniques are reviewed in details.

There are many methods to classify shape description techniques in the literature [5, 6, 63]. The most common and general classification is based on the use of shape boundary points as opposed to shape interior points. The resulting classes are known as *boundary* and *global* [7]. In our work, we adopt a hierarchical classification approach. The variety of shape description techniques are first classified into *contour-based* method and *region-based* method based on shape features extracted from contour only or the whole shape region. For each class, the different algorithms are further identified between *structural* and *global* based on the shape represented as a whole or sub-parts. The whole classification of main shape representation techniques is shown in Figure 2.1.

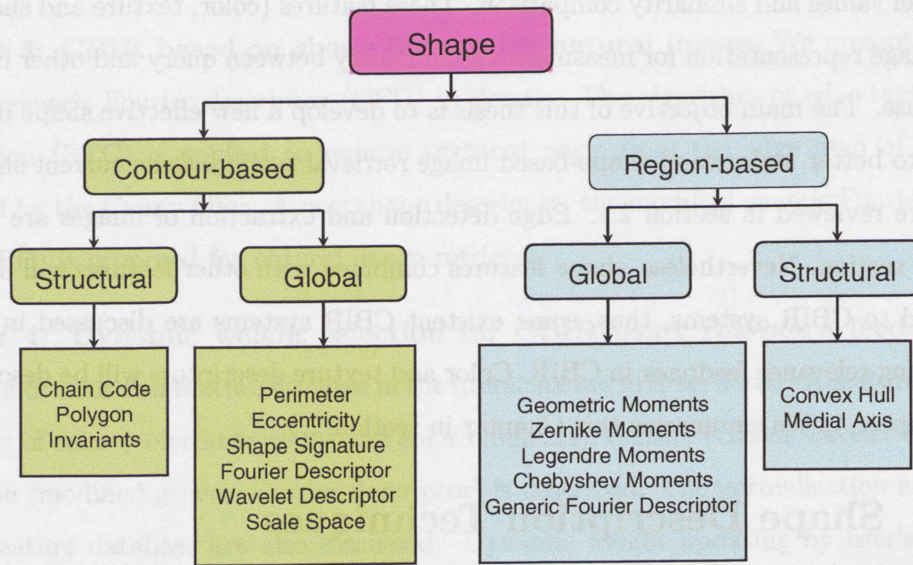


Figure 2.1: Classification of shape description techniques

2.1.1 Contour-Based Shape Description Techniques

Contour shape representation techniques only focus on shape boundary information. There are commonly two typical approaches for contour shape modeling: global and structural. Global methods do not separate shape into sub-parts, the shape is described by a feature vector derived from the integral boundary. Structural methods divide the shape boundary into segments (primitives) using a particular criterion.

Structural Methods

Chain codes represent an object by a sequence of unit-size line segments with a given orientation. Based on a selected starting point, a chain code can be produced by 4-, 8- or N-directional chain code [8]. Due to high dimension of chain code and being sensitive to boundary noise and variations, it is often used as an input to higher level analysis, not used directly for shape description.

Polygon approach approximates shape as polygon [9], and the shape is described as string of line segments which are then organized into tree data structure. The application of it for general shapes is impractical.

Shape invariants can also be viewed as a structural approach, because they also represent shape based on boundary primitives. Common invariants include geometric invariants [10], algebraic invariants [11] and differential invariants. Geometric and algebraic invariants are appropriate for shape boundaries represented by straight lines or algebraic curves. Differential invariants can be used for the boundaries which may not be represented by lines or algebraic curves.

The advantage of structural approach is its capability of dealing with occlusion problem in the scene and allowing partial matching. However, it is based on the cost of several drawbacks. The main drawbacks of structural approach are: (1) generation of primitives and features; Due to no formal definition for an object or shape, the number of primitives required for each shape is unknown. (2) Computation complexity; (3) Failing to catch global shape features which are equally important for the shape description.

Global Methods

Common simple global descriptors (*area, circularity, eccentricity and bending energy*) [12] can usually only discriminate with large dissimilarities, thus, they are not suitable for use as shape descriptors alone and are only combined with other shape descriptors to discriminate shapes. Other simple global shape descriptors proposed by Peura and Iivarinen [13] include *convexity, ratio of principle axis, circular variance and elliptic variance*.

Shape signatures, including complex coordinates, polar coordinates, tangent angle, curvature and chord-length, etc., represent shape by a one-dimensional function derived from shape boundary points. They usually achieve translation and scale invariance by normalization. Also, shape signatures can be quantized into a signature histogram, which is rotation invariant [11]. Shape signatures are sensitive to noise. Small changes in the boundary can cause large error in the matching. Thus, they should not be used directly for shape retrieval.

The problem of noise sensitivity and boundary variations in many spatial domain shape description methods stimulates the use of scale space analysis [14, 15, 16]. The *scale space* representation of a shape is achieved by tracking the position of inflection points in shape boundary filtered by low-pass Gaussian filter with variable widths. The inflection points in the representation are expected to be significant object characteristics. The result is usually an interval tree, consisting of inflection points. The matching is actually a point to point matching for two scale space images. It is not practical for complex shapes in the database, because complex shape can result in a very high interval tree.

The above shape descriptors are analyzed only in the spatial domain. Spectral descriptors, including Fourier descriptors (FD) and wavelet descriptors (WD), can overcome the problem of noise sensitivity and boundary variations by analyzing shape in the spectral domain. FD is one of the most popular shape descriptors used in literature [17, 18, 19, 20, 21, 22, 23, 25, 26, 27]. Traditional FD algorithms are only applied on closed and non-occluded curves, Lin et al. proposed the description of partial shape using FD [19], Arbter et al. developed the affine-invariant FD into the affine shape description [22]. In order to describe disjointed or articulated contour shapes, Rauber proposed a Universidade Nova de Lisboa

(UNL) FD, which is acquired by applying 2-D Fourier transform on the UNL transformed shape image [24]. The modified Fourier descriptor was proposed by Rui et al. [26], it is both robust to noise and invariant to geometric transformation. The FD can capture both coarse (global) and fine (local) features on shape boundary. However, for the local shape features, it can not detect the locations of local features. Eichmann et al. [31] developed short-time Fourier descriptor (SFD) to locate local boundary features.

In [32], Zhang and Lu compared the retrieval performance based on shape using FD and SFD. The results showed that FD outperformed SFD in the retrieval performance. The reason is that SFD may capture the local boundary features more accurately than FD, and local boundary distortions will not affect the whole set of SFD features. However, it can not detect global boundary features which are also important for shape representation. Furthermore, the lower frequency FD features employed in shape description are robust to local distortion, because significant effects from local distortions mainly concentrate on the very higher frequency FD features.

In recent years, shape description using wavelet descriptors (WD) has proposed by some researchers [28, 29, 30]. Comparing WD with FD, WD is of multi-resolution in the spatial domain, however, the increase of spatial resolution sacrifices frequency resolution. Moreover, the complicated matching scheme of wavelet description makes it impractical for online shape retrieval.

In a word, the advantages of FD over many other shape descriptors are: (i) it involves low computation by using the fast Fourier transform (FFT). (ii) The resulting descriptor is compact and any matching is very simple. (iii) Only a few lower coefficients of the Fourier transform are necessary to describe shape information. (iv) FD also overcomes noise sensitivity and captures both global and local features.

2.1.2 Region-Based Shape Description Techniques

Contour-based shape description is successfully used in many applications such as object recognition, character recognition and shape coding, etc. Contour-based shape descriptors

are useful if a shape contour is available. However, the description of shape from non-rigid and complex shapes becomes very difficult. Therefore, contour-based shape descriptors have limited applications. Alternatively, complex shape can be described using region-based shape description methods. Region-based methods can capture both boundary and interior information, we usually do not need to find shape contour.

Structural Methods

Similar to the contour structural methods, region-based structural methods divide shape regions into parts used for shape representation and description. The two main region structural methods are *convex hull* and *medial axis* [33]. The extraction of convex hull is a single process which finds significant convex deficiencies along the boundary. The shape can be represented by a string of concavities. The shape description can be obtained by a recursive process to result in a concavity tree: the convex hull of an object is first extracted with its convex deficiencies, then the convex hulls and deficiencies of the convex deficiencies are found, and repeat until all the derived convex deficiencies are convex.

Like the convex hull, region *skeletons* can also be applied for shape representation and description. A skeleton can be defined as a connected set of medial lines along the limbs of a figure [33]. The skeleton methods are represented by the medial axis transform (MAT). Finding a polygonal approximation of the shape contour has been suggested as a way of overcoming the problem of medial axis being sensitive to boundary noise and variation. Region structural methods suffer from similar problems to contour structural approaches, whilst both types of region structural methods here also need to know shape boundary information. Thus, they are seldom employed as shape descriptors in shape retrieval.

Global Methods

Global methods regard shape as a whole, the resulting shape description is a numeric feature vector. Hu first proposed the seven normalized geometric moments invariant to translation, rotation and scaling for two-dimensional pattern recognition applications [34]. Geometric moment descriptors (GMD) have been used in many applications [35, 36]. The advantage of

using GMD is its compact shape representation and lower computation. However, we do not know the physical meaning of the extracted moment invariants except the physical meaning of the lower order central moments (mean, variance and skew), in addition, it is difficult to obtain higher order moment invariants.

Teague [37] introduced two different continuous orthogonal moments (*Zernike* and *Legendre* moments). *Zernike* moments have better performance than *Legendre* in shape description due to their better feature representation capability and low noise sensitivity [38]. The idea of *Zernike* moments is to expand a signal into series of orthogonal bases. Because the computation of *Zernike* moment descriptors (ZMD) does not need to know boundary information, it is suitable for complex shape representation. In addition, *Zernike* moment invariants can be determined to arbitrary order, this overcomes the shortcoming of geometric moments, in which high order moment invariants are difficult to establish. ZMD has been adopted by MPEG-7 as region-based shape descriptor.

Grid descriptor (GD) is proposed by Lu and Sajjanhar [39]. A grid of a certain number of cells is overlaid on a shape. The grid is scanned from left to right and top to bottom, forming a bitmap. The cells covered by the shape are assigned 1 and those not covered are assigned 0. The shape can be represented as a binary feature vector. Chakrabarti et al. [40] improve the grid descriptor by using an adaptive resolution (AR) representation. GD performs its simple shape representation and intuition, and it is also agreed with shape coding method in MPEG-4. However, the computation is expensive.

Mukundan et al. [41] has used discrete orthogonal moments to eliminate the problems associated with continuous orthogonal moments. *Chebyshev moments* (CM) based on the discrete orthogonal Chebyshev polynomial. Their experimental results show CM is superior to geometric, *Zernike* and *Legendre* moments in image reconstruction capability.

Recently, Zhang and Lu [42, 43] have proposed a new region-based shape descriptor: generic Fourier descriptor (GFD). The GFD is extracted from spectral domain by applying 2-D Fourier transform on polar raster sampled shape image (detailed introduction in Chapter 3). Comparing the performance in image retrieval between GFD and ZMD, the results

show that their proposed GFD outperforms ZMD. Although GFD has better performance in shape retrieval, the retrieved images are only from binary shapes of main trademarks in MPEG-7 region shape database. It is not suitable to describe natural images due to their computational complexity, we try to find some methods to extract shape characteristics and represent it using GFD. Edge detection is one of the best methods to extract edge shape information from natural images. In the following, some different methods of edge detection are reviewed.

2.1.3 Edge Detection

Edge detection is fundamental in image processing. It is a primary tool in image segmentation, image retrieval based on shape, pattern recognition and scene analysis. Therefore, the accuracy of the edge detection is the key point which people want to research, this topic has attracted many researchers and many achievements have been made.

Generally, edges in images are the areas which have strong intensity contrast, a jump in intensity from one pixel to the next [63]. Using edge detection reduces the amount of the data and filters out useless information, while keeping the important boundary properties in an image. The group of classical edge detectors includes the *Sobel*, *Prewitt* and *Canny* filters [47].

Recently, a few researchers have proposed some new methods to detect the edges of the images. Setarehdan and Soraghan [44] proposed a new fuzzy multiscale edge detection (FMED) algorithm, employing wavelet transform to produce the multiscale representation of the signal. Comparing the performance between FMED and Sobel edge detector, the results show FMED has produced more accurate and robust estimates of the edge locations. Uchiyama et al. [45] have developed the new parameters of Hopfield Neural Network (HNN) for edge detection of image. From their simulation results, the edge detection method using HNN can detect the edges of the blur images more accurately than the conventional method. In [46], an effective edge detection method based on scale correlation in wavelet transforms is proposed by Bao and Zhang. In order to filter noise, the production of two adjacent

Option	Descriptions & Uses
Sobel	Detects horizontal or vertical edges. All pixels not on the detected edges are changed to black.
Canny	Detects horizontal or vertical edges. Low error rate and well localized edge points are achieved. The detection result has a better performance.
Phase Congruency	Detects variant edge types (step, line, between step and line) in images, the local energy model postulates that features are perceived at points of maximum phase congruency in an image. The more accurate and localized edge information is detected.

Table 2.1: Edge detection representations based on three different edge detectors

wavelet subbands is defined as a correlation function to amplify edges. This approach not only achieves better edge detection, but also improves the dislocation of neighboring edges. Kovesi develops a new edge detector: phase congruency detector [48]. He reasons that edges are not just steps, but rather, a wide variety of edge types is contained within an image, many of which are somewhere between a step and line. Congruency of phase at any angle produces a clearly perceived feature. This detector can correctly detect and localize the edge information of images. The descriptions of three typical edge detectors (Sobel, Canny and Phase Congruency) are represented in Table 2.1.

In order to intuitively compare the detection results among them, an example of edge detection of Lena image using three edge detectors above is shown in Figure 2.2. The Sobel detector in Figure 2.2(a) detects a rough edge map, in which much useful edge information is lost. The Canny filter in Figure 2.2(b) gives a better detection result. It extracts almost all edge information of Lena image. However, many edge features have twin responses, such as the profiles of her face, nose and hat, etc. Moreover, her shoulder and the peak of her hat have the broken lines due to the intensity values being similar to the background. Phase congruency in Figure 2.2(c) has the best detection result: the single responses of edge features, clear and continuous edge boundaries. Although phase congruency is the best edge detector among them, it may be suitable to use in image segmentation and other areas, it

could be impractical to use in image retrieval because of its computation complexity. Based on the accuracy and simple computation of edge extraction, Canny filter is more suitable for the extraction of edge information in our image retrieval system.

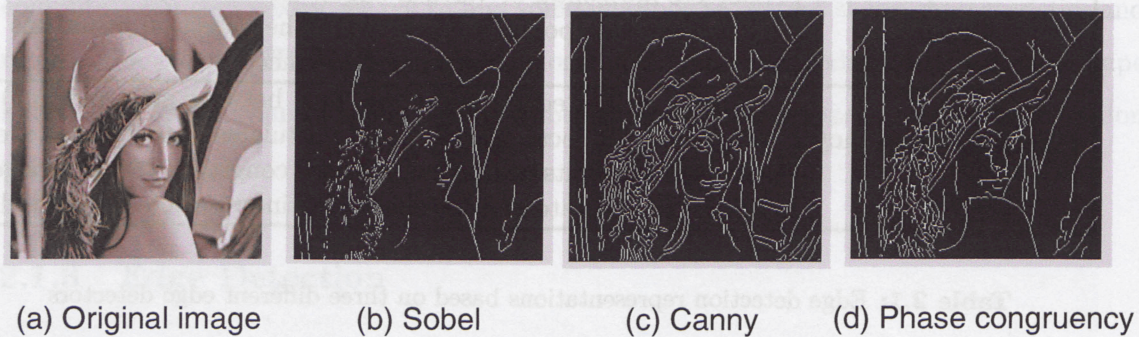


Figure 2.2: Lena's edge maps using different edge detection approaches

2.2 Content-Based Image Retrieval System

Since the term *content-based image retrieval* (CBIR) was first used to describe automatic retrieval of images from a database by color and shape features [49], CBIR has been a very active research area. One or more options of random browsing, search by example, search by sketch and search by text are supported in most image retrieval systems. Here, some of current CBIR systems described in the literature will be reviewed.

2.2.1 Current CBIR Systems

The well-known and representative commercial CBIR system is the *QBIC* system developed by IBM [50]. It can support image retrieval by any combination of color, texture or shape as well as text keyword. Image queries can be selected from example images or by sketching a desired shape on the screen. QBIC is one of a family of systems that use high-dimensional indexing techniques to realize a fast search. The available demos are at <http://www.qbic.almaden.ibm.com/>.

Photobook is the representative research CBIR system developed at MIT Media Lab [51]. Similar to QBIC, images in the database are described by color, texture, shape and other appropriate features. The system can offer retrieval of textures, shapes and human faces. The other two research systems developed at Columbia University are the visual feature search engine: *VisualSEEk* [52] and World-Wide-Web-oriented text/image search engine: *WebSEEk* [53]. *VisualSEEk* supports searching by image region, color, shape, spatial locations and keyword. Users can make image queries by specifying areas of defined shape and color at absolute or relative locations within the image. *WebSEEk* consists of three main modules including an image/video collecting module, a subject classification and indexing module and a searching, browsing and retrieval module. The queries are selected based on both text and visual content. The corresponding demos are at <http://www.ee.columbia.edu/sfchang/demos.html>.

MARS (multimedia analysis and retrieval system) was developed at University of Illinois by Huang et al. [54]. The system represents each object in an image by a variety of features, and applies different similarity measures to compare query and stored objects. User's feedback is used to adjust weights and improve retrieval performance. *MARS* proposes the relevance feedback architecture in image retrieval. In the following, we will review the applications of relevance feedback in CBIR.

2.2.2 Relevance Feedback in CBIR

Recently, relevance feedback has been a powerful tool for interactive content-based image retrieval. It is a process of automatically adjusting an existing query using information from user's feedback based on the relevance of previously retrieved images, bridging the gap between the low-level features and the high-level concepts. Relevance feedback in CBIR systems has attracted much attention within the research community.

Rui et al. use the relevance feedback to dynamically update weights associated with the component features [66]. This method is applied successfully in *MARS* system [55]. Tian et al. [56] proposed a novel approach by providing both positive and negative feedback for Support Vector Machines (SVM) learning instead of *MARS* [55] only using user's positive

feedback to update weight vector. This approach releases the user from manually providing preference weights. Su et al. [57] proposed approach is based on a Bayesian classifier, positive and negative feedback examples have the different usages, positive examples for estimating a Gaussian distribution that represents the desired images for a given query, negative examples for updating the ranking of the retrieved images. The Principal Component Analysis (PCA) technique is also used to reduce the dimensionality of feature spaces.

A graphic-theoretic model for incremental relevance feedback is proposed by Zhuang et al. [58], they first introduce a *two-layered graph model* to represent the correlations between images. Based on this model, which is enriched with semantic correlations between images selected from user's feedbacks by a learning strategy, they propose the *link analysis* approach for image retrieval and relevance feedback. The Kernel-based approach is another popular choice for implementing relevance feedback in CBIR system, however, most kernel approaches have regarded the input as a long flat vector. This may increase the chances of "polluting" the feature elements that uniquely represent the selected images. Alternately, Chung and Fung propose a two layer kernel configuration with an objective and improve the retrieval performance [59].

Muneesawang and Guan have adopted a non-linear radial basis function (RBF) method which can better characterize the behavior of human users in an interactive retrieval session and implement an adaptive metric which progressively models the notion of image similarity through continual feedback from the users. This proposed approach is applied on an image database compressed by the wavelet transform and vector quantization coders [60]. Yap and Wu develop a fuzzy relevance feedback to integrate the user's fuzzy interpretation of visual content into the notion of relevance feedback. A fuzzy radial basis function network (FRBFN) is constructed based on the user's feedback, and the underlying parameters and network structure are also optimized using a gradient-descent training strategy [61].

2.3 Summary

In this Chapter, we have first provided a substantial review on some existing shape representation techniques. Generally, there are two kinds of shape description techniques: contour-based and region-based. Contour-based methods usually involve less computation than region-based counterparts. However, contour shape descriptors are more easily affected by noise and variation than region-based shape descriptors because they use less shape information than region-based methods. Region-based methods are usually more robust and application independent. However, they usually need more computational power and more storage than the contour-based descriptors. In addition, edge detection algorithms and CBIR systems are also reviewed. The Canny filter is a better edge detector due to its accuracy and simple computation of edge extraction. Relevance feedback is popularly applied in CBIR systems for improving retrieval performance.

Apparently, every shape technique has its advantages and disadvantages. MPEG-7 has a set of principles to evaluate the suitability of shape technique: good retrieval accuracy, compact features, general application, low computation complexity, robust retrieval performance and hierarchical coarse to fine representation. The Generic Fourier Descriptor (GFD) satisfies these six principles and has demonstrated better performance [43]. Thus, in Chapter 3, we will introduce GFD algorithm and our proposed shape description method in detail.

Chapter 3

CBIR Based-on Shape Feature for Natural Images

IN Content Based Image Retrieval (CBIR), shape information is widely considered to play a key role in the characterization of scenes, however, it is difficult to measure its property, this is why most researchers seldom extract shape features and apply them in natural image retrieval. Most research has been restricted to collections of simplified shape contour images, such as those of binary logos and trademarks. In such collections, scenes often involve only a single object with a well defined shape, wherein a single class is often comprised of a set of images with only minor variations to the dominant shape, that whilst altered, remain well defined.

In collections involving natural scenes however, such as those found in the well known COREL database, scenes are generally much more complex, involving many combinations of objects, of a variety of shapes and sizes that may or may not be embedded in equally complex backgrounds. Shape information often becomes contaminated by the mixture of content in a scene, rendering shape based retrieval results, relatively poor. In this Chapter, we propose an enhanced region-based technique to better deal with CBIR applications in the domain of natural image databases. The structure of our proposed approach is shown in Figure 3.1.

In this Chapter, we will focus on image indexing, detailing our proposed methods for shape description. In general, there are two components in CBIR: image indexing and

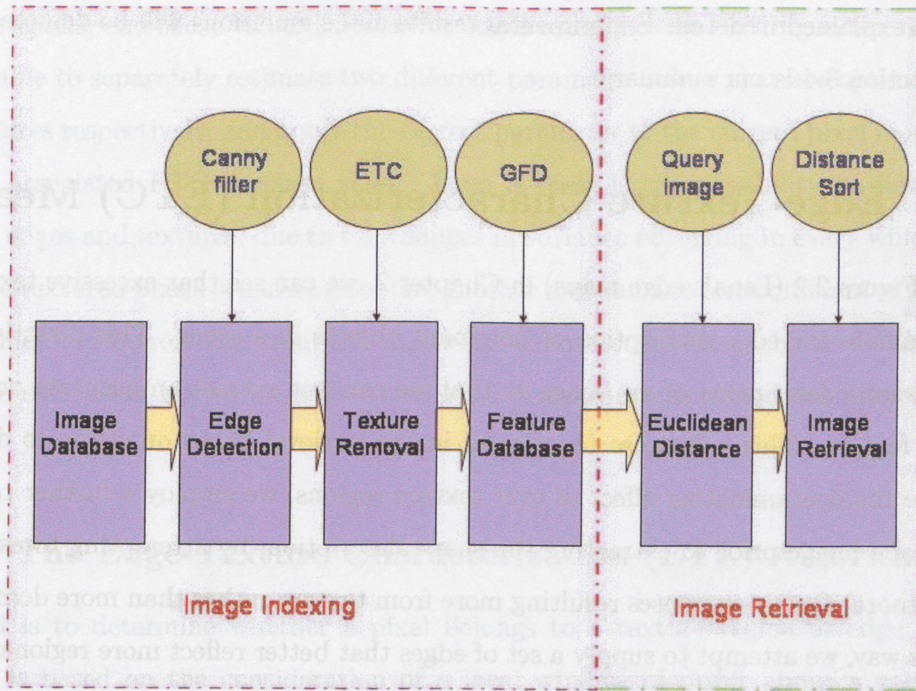


Figure 3.1: The frame structure of our proposed image retrieval

retrieval depicted in Figure 3.1. The block on the far left is the original image database. At first we extract the edge map of every image in the database using Canny edge detection filter (discussed in Chapter 2). Next, we eliminate the contaminating textured regions from the edge map of every image using an Edge-Texture Characterization (ETC) [62] approach and form the edge map database with texture removed. Extracted shape features using Generic Fourier Descriptor (GFD) [43] form the final index that will be used for shape based image retrieval. The image retrieval part of the system, as with many traditional approaches, involves recalling indexed images that are closest to a given query. We will evaluate by the performance of such indexing, considering different query images selected randomly from the image database.

The rest of this Chapter is organized as following. In Section 3.1, we will introduce the edge-texture characterization (ETC) algorithm, including the ETC method, the extraction of texture regions and experimental results. In Section 3.2, the generic Fourier descriptor

will be explained in detail. Experimental results and simulations will be depicted in Section 3.3. Section 3.4 is our summary.

3.1 Edge-Texture Characterization (ETC) Measure

From Figure 2.2 (Lena's edge maps) in Chapter 2, we can see that excessive texture regions contaminate the edge description of dominant objects in a scene. This directly effects the exact shape description of an image so that we can not extract an accurate or meaningful shape feature. This will cause the rate of image retrieval based on shape to decrease. To remove the contaminating effect of over texture regions, we employ a further refinement in the shape image prior to extracting the shape description, by attempting to establish (and thus ignore) Canny responses resulting more from texture rather than more dominant edges. In this way, we attempt to supply a set of edges that better reflect more regional boundaries within an image rather than every intensity variation. The edge-texture characterization (ETC) approach introduced in [62] provides a fuzzy discrimination between edge and texture regions and is adopted in this work. Generally, pixel value arrangements corresponding to textures and edges will exhibit different values for this measurement.

The principal of ETC is founded in examining the changes in variance occurring in a windowed local region when it is blurred by an average filter. In smooth regions the variance does not really change much, however the response in textured versus edge images is quite marked. Thus it becomes easy to distinguish between smooth feature types and those edge or texture. However, due to the possibility that both edges and textures may exhibit similar levels of image activities in terms of gray level variations around their neighborhoods, it is usually difficult to distinguish these two feature types. In order to separate edge and texture features well, we introduce some noise into the original image. Because of the different noise masking capabilities of these two feature types, it is usually not desired to apply similar values of regularization parameters to both of them. The optimal parameter values to texture regions usually result in a noisy appearance for the edge due to its surrounding

smooth regions, and those values suitable for edges usually cause blurring in texture regions. We are able to separately estimate two different parameter values that are optimal to edges and textures respectively, and apply the correct parameter to the current pixel in accordance with its associated ETC measure value. Even if there is the possibility of distinguishing between edges and textures, due to the changes in variance occurring in every windowed size filter, the textured pixels separate from one another and can not form different small texture areas, which can be removed from the edge map. To overcome the problem, we also develop a texture extraction algorithm based on the ETC measure using morphological operations [63].

3.1.1 The Edge-Texture Characterization (ETC) Algorithm

The goal is to determine whether a pixel belongs to a texture region or edge, the ETC measure is based on the consideration of a local window of pixels about a subject pixel in the image, the ratio $k = \bar{\sigma}/\bar{\sigma}'$ between the standard deviation of original $\bar{\sigma}$ versus that of blurred intensities $\bar{\sigma}'$, yields a measure of deviation k due to the underlying nature of the image content in that region. A simple range of values captured by this measure can be attributed to a texture region, thus we can establish a regional mask over the texture regions so that they may be later ignored in the shape descriptor calculation. The equations of $\bar{\sigma}$ and $\bar{\sigma}'$ are defined as:

$$\bar{\sigma}^2 = \frac{1}{|N|} \sum_{(i,j) \in N} (x_{i,j} - \bar{x})^2 \quad (3.1)$$

$$\bar{\sigma}'^2 = \frac{1}{|N|} \sum_{(i,j) \in N} (x'_{i,j} - \bar{x}')^2 \quad (3.2)$$

In the equations 3.1 and 3.2, N denotes a neighborhood set around the current pixel, $x_{i,j}$ denotes the gray level value of pixel (i, j) in the set, and $x'_{i,j}$ is the corresponding smoothed gray level value under $K \times K$ averaging. \bar{x} and \bar{x}' are the mean of the gray level values of the original and smoothed variables in the neighborhood set. In general, most of the estimated k values are restricted to the interval $[0, K]$.

3.1.2 The Extraction of Texture Regions

Having distinguished smooth, shape and texture features using the ETC approach, the next step is to extract the texture regions from the edge map and eliminate their contamination of the overall shape information. Even with the ETC as an indicator of textured pixels, there are still occasional pixels that are confused and thus leave holes or ‘pitting’ in the region mask. In order to fix this problem, morphological operations are a very useful tool for extracting image components, they are often applied for image pre- and postprocessing.

The two fundamental morphological operations are dilation and erosion, our algorithm of extracting texture regions are based on these operations. Dilation operations are used to ‘grow’ or ‘thicken’ objects in a binary images, the manner and extent of this thickening is controlled by a structuring element. ‘Shrinking’ or ‘thinning’ objects in a binary image is achieved through erosion operations. As with dilation operations, a structuring element can control the manner and extent of shrinking. The structuring element thus plays an important role in morphological operations. Various combinations of dilation and erosion are often used in practical image processing. Generally an image will go through a series of dilations and/or erosions using the same or different structuring elements. Our detail morphological operations will be depicted in the following experimental section.

3.1.3 Experimental Results

The ETC calculation, as with many texture based measure, is performed in gray scale mode, thus, before calculating the ETC values, some preprocessing is necessary for natural color images. Figure 3.2(a), Lena, is an original color image. We first change it into a 2-D gray level image, Figure 3.2(b), and form the noisy image, Figure 3.2(c). After adding Gaussian White noise with mean 0.005, we finally smooth it with a 5X5 averaging filter as shown in Figure 3.2(d). Figure 3.3 (Flower image) shows another preprocessing procedure.

In order to demonstrate the capability of the ETC algorithm to distinguish between smooth, edge and texture regions, and especially for the edge and texture areas which are not easily separated, we extract the ETC-maps, which are the descriptions of the ETC values

as gray level images as shown in Figure 3.4 and 3.5. In our experiments, we select a 5×5 averaging filter as a windowed local region, so the estimated k values are restricted in $[0, 5]$. Those image pixels corresponding to the calculated k values were found empirically, to fall roughly within 3 continuous ranges $R1 = [0, 1.2]$, $R2 = [1.2, 1.8]$ and $R3 = [1.8, 5]$. The pixels demonstrating the k values within the ranges $R1$, $R2$ and $R3$ are approximately related to smooth pixels, edge pixels and texture pixels respectively. This can be verified by visually displaying the binary maps in Figure 3.4 and 3.5.

Figure 3.4 shows the ETC-map and the feature segmentation binary maps for the Lena image. From the ETC-map in Figure 3.4(a), we can see that the feather texture area of the hat in the image exhibits the brightest response in the map, corresponding to the highest values of k . This supports our previous affirmation that texture pixels should exhibit higher values of k . It is also found that the edges correspond to middle values of k , and the smooth areas display darkest in the map.

In order to show the results of estimating the smooth, edge and texture regions using different the ranges of k values, we construct binary maps in Figures 3.4(b)-(d), where the white pixels are those with their k values corresponding to the ranges $R1$, $R2$ and $R3$ in Figure 3.4(b)-(d), respectively. Those pixels with their k values within the interval $R1$ correspond to the smooth regions in Figure 3.4(b), Figures 3.4(c) and (d) show the edge pixels and texture pixels with their k values within interval $R2$ and $R3$ respectively.

Figure 3.5 shows the same performance for the image Flower. From the ETC-map in Figure 3.5(a), it can be seen that the brightest part is the center of the flower, which is the texture area. Figures 3.5(b)-(d) indicate the smooth, edge and texture regions respectively.

We have applied our ETC method to the two images, Lena and Flower, which contain substantial texture information. For Figure 3.4(d) the image Lena, the extracted texture regions mainly concentrate around the feathers of her hat, which would typically be perceived as textures by our observer. Similarly, for Figure 3.5(d) the image Flower: the extracted texture regions are mainly located around the flower stamen. Our final target is to remove the texture areas which contaminate the edge shape information in the edge map. Based



(a) Original image



(b) Gray level image



(c) Noisy image



(d) Averaging image

Figure 3.2: The preprocessing procedure of Lena image



(a) Original image



(b) Gray level image

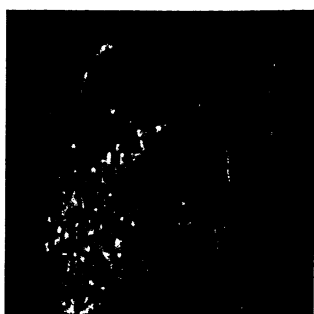


(c) Noisy image



(d) Averaging image

Figure 3.3: The preprocessing procedure of Flower image



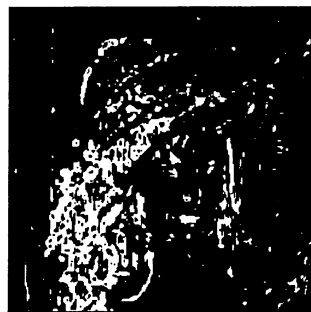
(a) ETC-map



(b) Smooth region map

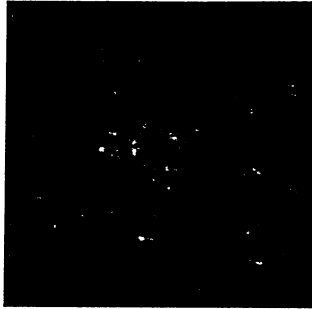


(c) Edge map



(d) Texture map

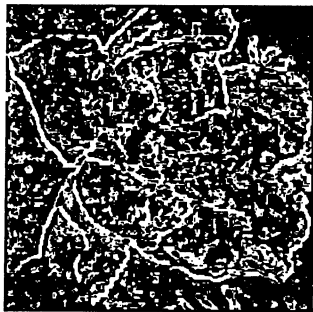
Figure 3.4: ETC measurement of Lena image



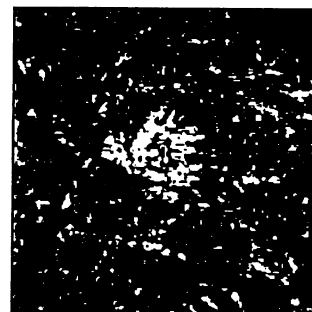
(a) ETC-map



(b) Smooth region map



(c) Edge map



(d) Texture map

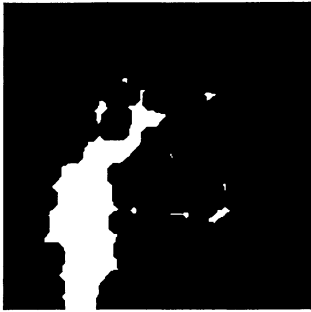
Figure 3.5: ETC measurement of Flower image



(a) Original edge map



(b) Erosion operation



(c) Closing operation(dilation and erosion)

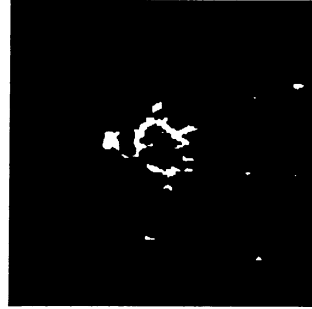


(d) Edge map(texture removed)

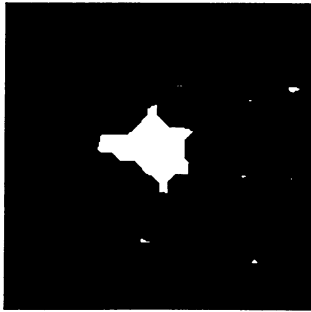
Figure 3.6: Texture removing procedure of Lena edge map



(a) Original edge map



(b) Erosion operation



(c) Closing operation(dilation and erosion)



(d) Edge map(texture removed)

Figure 3.7: Texture removing procedure of Flower edge map

on the ETC algorithm alone, we can not remove them well, because there are separated texture regions with scattered patches and the complete texture areas are not formed in Figure 3.4(d) and 3.5(d). Choosing a fixed threshold may not be ideal for all images thereby resulting in such scattered texture regions. To ensure clarity in the final edge map, we seek a thorough removal of such regions - thus employ morphological operations to eliminate noisy estimated and extract a clean and well closed region.

Figures 3.6 and 3.7 show us the procedure of removing the texture areas from the edge maps. Figures 3.6(a) and 3.7(a) are the original edge maps by using the Canny filter, Figures 3.6(d) and 3.7(d) are the edge maps after removing the texture regions. Figures 3.6(b), (c) and Figures 3.7(b), (c) are the maps which are formed by using morphological operations.

In morphological operations, the selection of the structuring elements is very crucial. We implement many kinds of structuring elements in our experiments and find that the following structuring elements with related parameters give us the best performance. We first select a horizontal 'line' with length 10 and degree 0 and a vertical 'line' with length 5 and degree 90 as the structuring elements, applying respectively the erosion operation to 'shrink' the maps in Figures 3.4(d) and 3.5(d) and get Figures 3.6(b) and 3.7(b) maps. Then, we choose the 'disk' as a second structuring element with radius 5 for the closing operation, being a dilation followed by an erosion, to 'thicken' and 'shrink' Figures 3.6(b) and 3.7(b) maps and achieve the more complete texture regions in Figure 3.6(c) and Figure 3.7(c), although there is still some sparse noise in the maps, it does not affect the texture description and removal. From Figures 3.6(d) and 3.7(d), more accurate edge shape information has been made available.

3.2 Generic Fourier Descriptor (GFD)

In Section 3.1, we applied the ETC algorithm and morphological operations on edge maps of the image database and achieved more accurate edge shape information. The next step is to find a better shape descriptor to characterize the edge shape. Generic Fourier Descriptor (GFD), proposed by Dr. D. S. Zhang [43], will be applied for the task. In order to show

why we select GFD as our shape descriptor, some shape descriptors introduced in Chapter 2 will be discussed briefly.

In the past studies, contour-based shape descriptors are usually available for representing contour shape with a continuous single boundary. The one-dimensional Fourier Descriptor (FD) has been successfully applied to many shape description applications. The good characteristics of FD, such as simple calculation, simplicity for matching, robustness to noise and compact coarse to fine representation, make it a popular shape descriptor. One of the major problems in attempting to use the contour-based standard Fourier descriptor as a feature for assessing the similarity between images based on shape, lies in its dependence on the prior knowledge of boundary information. In particular, it assumes that for each image, we have an ordered description of the points that form the connected path responsible for a particular boundary. As such, when considering the boundary of regions of interest, some form of higher level segmentation becomes necessary. Unfortunately, even if such segmentation is available, it is often the case that multiple boundaries will occur within the image (either due to internal shape content or multiple regions of interest). In previous work [64], the FD has been applied to the description of raw edge information of natural images found in the COREL database, and has met with limited success in terms of shape identification.

Region-based shape descriptors can be applied to more general applications and are more robust in describing shape with complex boundaries compared with contour-based shape descriptors, because they exploit more edge distribution to derive shape features. Spatial domain shape features, such as, Zernike moment descriptors (ZMD) have been proposed as a preferred technique over other region based techniques such as geometric moments (e.g. Hu's moments) [65]. Derived from a complex set of orthogonal polynomials over the unit disk, a more rotationally invariant description of shape information is achieved, independent of boundary information. Limitations exist in terms of computational complexity and a tendency to capture spatial moments in the radial directions rather than spectral features, thus spectral information is not captured evenly at each order resulting in the loss of significant features useful for shape description. As an alternative, a region-based 2-D polar Fourier

transform (PFT) attempts to better capture the spectral content of angular and radial information by transforming the polar description of an image into a rectangular image of radial vs. angular distribution of image intensities, upon which a standard 2-D FT may be applied.

3.2.1 Polar Fourier Transform

The Fourier transform has been widely applied for multimedia processing and applications. The advantage of analyzing images in the spectral domain rather than analyzing images in the spatial domain is that it is easy to overcome the noise problem which is general to digital images. On the other hand, the spectral features of an image are usually more compact than the features extracted from spatial domain, comparing of only a few lower frequency Fourier coefficients that can approximately characterize the shape information, this greatly reduces the computational complexity. The one-dimensional FT has been widely applied to contour shape to derive FD. And there is no related work on region based FD. The generic Fourier descriptor based on region shape will be introduced from the two-dimensional Polar FT.

The continuous and discrete 2-D Fourier Transform of a shape image $f(x, y)$ ($1 \leq x \leq M, 1 \leq y \leq N$, where M and N are the width and height of the image) are given by the following two equations:

$$F(u, v) = \int_x \int_y f(x, y) \exp[-j2\pi(ux + vy)] dx dy \quad (3.3)$$

$$F(u, v) = \sum_{x=1}^M \sum_{y=1}^N f(x, y) \exp[-j2\pi(ux/M + vy/N)] \quad (3.4)$$

where u and v are the uth and vth spatial frequency in the horizontal and the vertical directions, respectively. The 2-D FT may be applied to any shape image without assuming the knowledge of boundary information. However, it is not practical to apply the 2-D FT directly on a shape image in the Cartesian space to derive FD, because the shape features captured by the 2-D FT are not invariant to rotation. Rotation invariance of a shape is crucial, because similar shapes may be located at different positions and orientations in

an image. For example, two shapes are similar in the image plane, however, their Fourier spectral distributions (phases) in frequency plane are completely different. Therefore, shape images would be better considered in polar space using a polar Fourier transform (PFT). The PFT is rotation-invariant, making it especially suitable for accurate extraction of orientation features. Two PTF are discussed and described as follows:

In order to produce PFT, both the data $f(x, y)$ and the spectra $F(u, v)$ are converted into polar space, that is, let

$$x = r \cdot \cos \theta, y = r \cdot \sin \theta \quad (3.5)$$

$$u = \rho \cdot \cos \psi, v = \rho \cdot \sin \psi \quad (3.6)$$

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

$$dx = \cos \theta dr - r \sin \theta d\theta \quad (3.7)$$

$$dy = \sin \theta dr + r \cos \theta d\theta \quad (3.8)$$

By substituting equations 3.5-3.8 into 3.3, we obtain the polar Fourier transform (PFT1):

$$PFT_1(\rho, \psi) = \int_r \int_\theta r f(r, \theta) \exp[-j2\pi r \rho \sin(\theta + \psi)] dr d\theta \quad (3.9)$$

The discrete PFT1 is obtained as

$$PFT_1(\rho_l, \psi_m) = \sum_p \sum_i f(r_p, \theta_i) \cdot r_p \cdot \exp[-j2\pi r_p \rho_l \sin(\theta_i + \psi_m)] \quad (3.10)$$

where $r_p = p/R, \theta_i = i(2\pi/T) (1 \leq i \leq T); \rho_l = l(1 \leq l \leq R)$ and $\psi_m = m\theta_i$. R and T are the resolution of radial frequency and angular frequency respectively. The polar Fourier coefficients $PFT_1(\rho_l, \psi_m)$ are applied to derive a normalized FD for shape description.

So far we have already achieved rotation invariance for the shape features by deriving the PFT1 result from the polar Fourier transform of equation 3.3. However, there still exists one problem: because of the appearance of ψ_m within *sine* function $\sin(\theta_i + \psi_m)$ in equation 3.10, the physical meaning of it is no longer the m th angular frequency. The features extracted by the PFT1 lose physical meaning in circular direction. In order to deal with this problem, a modified polar FT (PFT2) is derived by regarding the polar image in polar space as a normal two-dimensional rectangular image in the Cartesian space. Figure 3.8 displays the rectangular polar images [43]. Figure 3.8(a) is the original polar image in polar space, Figure 3.8(b) is the rectangular polar image plotted in the Cartesian space.

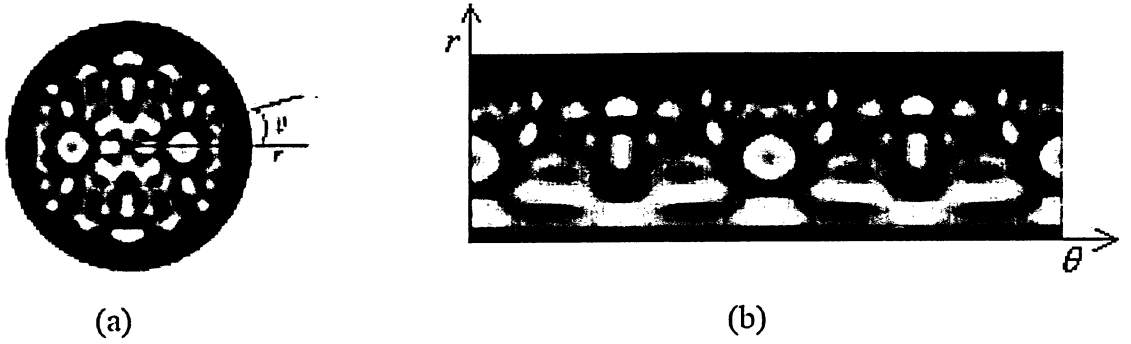


Figure 3.8: (a)Original shape image in the polar space (b)Polar image plotted in the Cartesian space

The polar image in Figure 3.8(b) is the normal rectangular image. Thus, if 2-D FT is applied on this rectangular image, the polar FT has similar form to the normal 2-D discrete FT of equation 3.4 in Cartesian space. As a result, the modified polar FT is defined as the following equation:

$$PFT_2(\rho, \phi) = \sum_r \sum_i f(r, \theta_i) \exp[j2\pi(\frac{r}{R}\rho + \frac{2\pi i}{T}\phi)] \quad (3.11)$$

where $1 \leq r \leq R$ and $\theta_i = i(2\pi/T)(1 \leq i \leq T)$; $1 \leq \rho \leq R$, $0 \leq \phi \leq T$. R and T are the radial frequency resolution and angular frequency resolution respectively. Comparing equation 3.10 and 3.11, PFT2 has a simpler expression form than PFT1. In addition, we

need not constrain the shape into a unit circle. By comparison, ZMD must be defined in a unit circle. The physical meaning of ρ and ϕ in equation 3.11 is similar to u and v in equations 3.3 and 3.4. ρ and ϕ are simply the radial frequency and the angular frequency respectively. And that the number of ρ and ϕ can be determined easily, because significant shape features are usually represented by a few low frequency components.

3.2.2 Derivation and Implementation of Generic FD

In this section, the derivation and implementation of GFD are introduced in details. Given a shape image $I = f(x, y); 1 \leq x \leq M, 1 \leq y \leq N$, we apply the PFT by first converting the shape image I from Cartesian plane to polar plane $I_p = f(r, \theta); 1 \leq r \leq R, 1 \leq \theta \leq 2\pi$. The origin of the polar plane is set to be the center of the shape, making the shape translation invariant. The center point (x_c, y_c) is given by equations 3.12 and 3.13:

$$x_c = \frac{1}{M} \sum_{x=1}^M x \quad (3.12)$$

$$y_c = \frac{1}{N} \sum_{y=1}^N y \quad (3.13)$$

and (r, θ) is given by equations 3.14 and 3.15

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2} \quad (3.14)$$

$$\theta = \arctan \frac{y - y_c}{x - x_c} \quad (3.15)$$

So PFT2 is applied on I_p . Due to the fact that the center of the shape image is the origin of the polar plane, the acquired coefficients of PFT2 are translation invariant. Rotation invariance is achieved by neglecting the phase information in the coefficients and only keeping the magnitudes of the coefficients. So far we have achieved translation and rotation invariance, we now address the issue of scale invariance. To obtain scale invariance, the first magnitude value is normalized by the area of the circle (*area*) within which the polar image

resides, and all the other magnitude values are normalized by the magnitude of the first coefficient. The translation, rotation and scale normalized PFT2 coefficients are applied as the shape descriptors. To sum up, the shape descriptor GFD derived from PFT2 is shown as follows

$$GFD = \left(\frac{|PFT_2(0,0)|}{area}, \frac{|PFT_2(0,1)|}{|PFT_2(0,0)|}, \dots, \frac{|PFT_2(0,n)|}{|PFT_2(0,0)|}, \dots, \frac{|PFT_2(m,0)|}{|PFT_2(0,0)|}, \dots, \frac{|PFT_2(m,n)|}{|PFT_2(0,0)|} \right) \quad (3.16)$$

where m is the maximum number of the radial frequencies selected and n is the maximum number of angular frequencies selected. Normally, in Fourier Transform, the first coefficient or the DC component is used as the normalization factor and is eliminated after normalization. However, this component is used in this work as an additional feature, because it reflects the average energy of the shape which is useful for shape representation. For efficient shape description, only a small number of the FD features are selected for shape representation. These selected FD shape features form a feature vector which is used to index the shape. This met with good results in simplified binary shape databases, although no real application to natural image data has been reported.

3.2.3 Modified Edge-Based GFD

In this current work, unlike in that of [43], we consider the Canny edge description of natural image queries as input to a GFD inspired operator for shape description. Direct application of [43] might see a binary image formed by thresholding from the original, such that a ‘caricature’ of the original could be utilized as the input to the GFD. The problem with this is that the computation becomes quite expensive as more pixels need to be considered in the shape image. In addition, achieving a consistent thresholding for a natural image is not trivial as it is very sensitive to contrast, etc. By using an edge description we reduce the computational load (less pixels to consider in the polar mapping). We call this proposed approach a modified GFD (MGFD).

3.3 Experimental Results

In order to test the retrieval performance of our proposed algorithm, we select three different shape descriptors: the standard Fourier Descriptor (FD), Modified Generic Fourier Descriptor (MGFD) and MGFD after removing texture part from the edge map. Our simulations were carried out using a subset of the COREL image database consisting of 1,000 natural color images (JPEG), from 10 classes that appeared to be more dominated by shape. Each class included 100 conceptually similar images. Figure 3.9 shows an example of each class in the database.

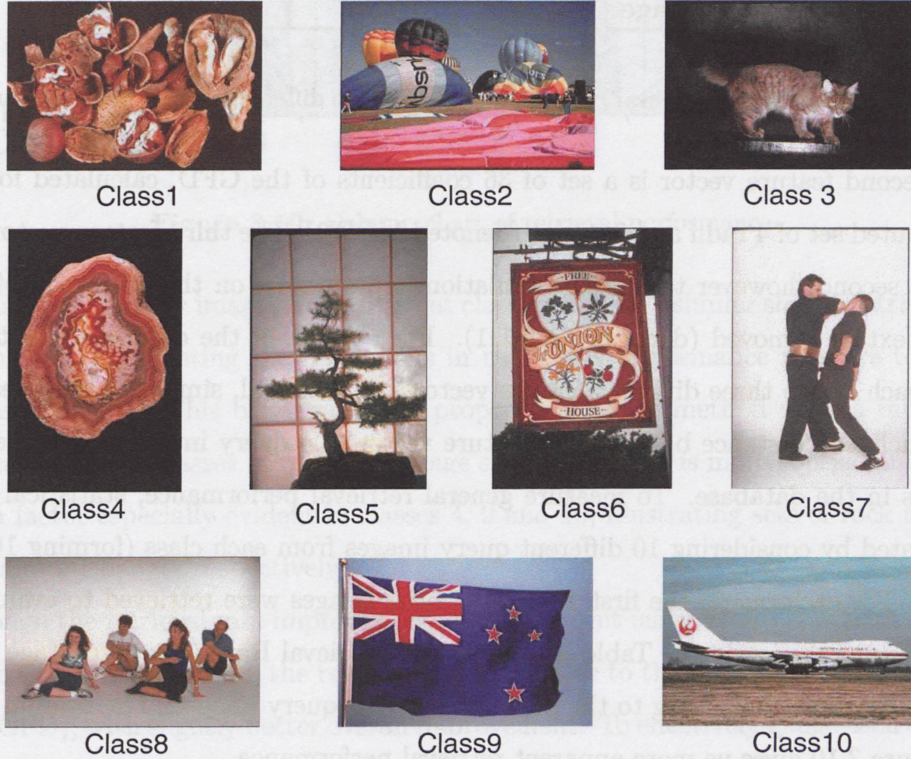


Figure 3.9: One sample image of each class in the database

Simulations were conducted to compare the retrieval effectiveness on this database when indexed with one of three alternative shape descriptors. The first used the lowest 50 coefficients from a standard FD (denoted FD) as a feature vector for each edge mapped image.

Class	FD (%)	MGFD (%)	MGFD1 (%)
1	18.2	30.2	31.0
2	20.7	21.6	23.1
3	21.9	24.1	23.5
4	19.2	52.5	64.2
5	39.4	35.4	33.9
6	24.2	22.1	22.5
7	43.9	47.4	46.5
8	62.6	55.4	55.2
9	37.4	71.8	75.3
10	20.2	72.4	76.8
Average	30.8	43.3	45.2

Table 3.1: Retrieval performance from three different sets of shape features

The second feature vector is a set of 36 coefficients of the GFD, calculated for an equally distributed set of 4 radii and 9 angles (denoted MGFD). The third feature vector is the same as the second, however the GFD calculation is performed on the edge map of each image after texture removed (denoted MGFD1). Each image in the database was then indexed with each of the three different feature vectors. In retrieval, similarity was measured using the Euclidean distance between the feature vector of a query image, and those of all other images in the database. To measure general retrieval performance, statistical results were calculated by considering 10 different query images from each class (forming 100 queries in total). For each query, the first 16 most similar images were retrieved to evaluate the performance of the retrieval. Table 3.1 shows the Retrieval Rate (the percentage of images in the 16 retrieved, belonging to the same class as the query image). The column chart shown in Figure 3.10 gives us more apparent retrieval performance.

Table 3.1 and Figure 3.10 tell us that the retrieval results of most classes using the proposed MGFD and MGFD1 outperform results using FD (Class 1, 2, 3, 4, 7, 9, 10). There were, however, a few classes demonstrating slightly worse performance (Class 5, 6, 8). Such classes exhibit a much higher variation in shape between images considered of the same class

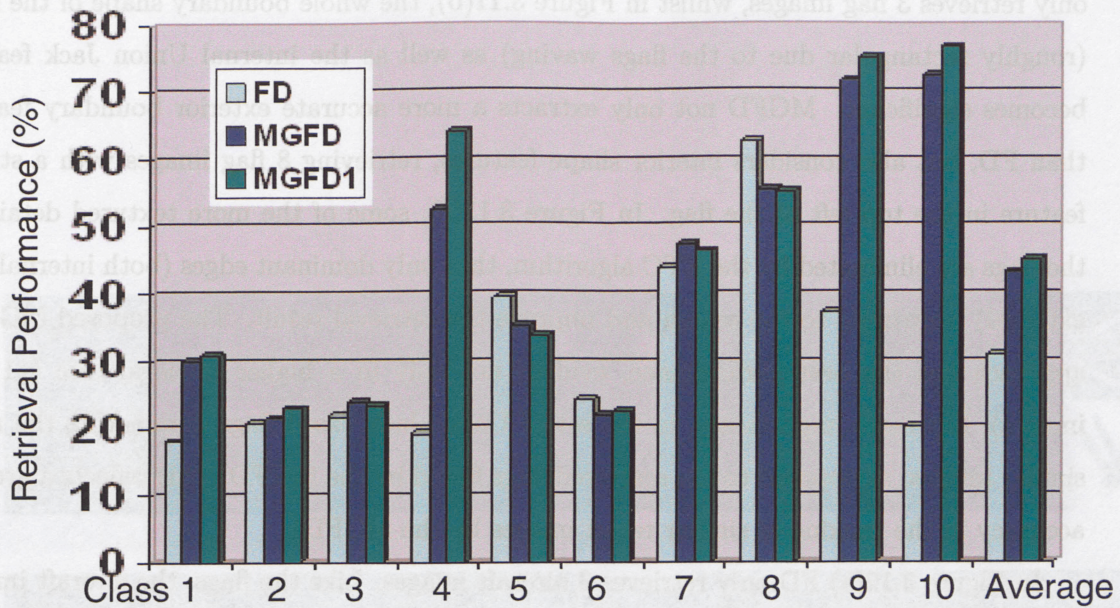


Figure 3.10: Column chart of retrieval performance

conceptually, thus many images from different classes that have similar shape distribution are often confused, highlighting the limitations in using this performance measure to evaluate shape based results. This being said, the proposed MGFD1 method gives a much better representation in the classes in which an image shape structure is more consistent across the images, a factor especially evident in classes 4, 9 and 10, illustrating sets of rock formation, flag and aircraft images respectively.

Although the retrieval rate improves about 14.4 percent using MGFD1 over the standard FD approach, in most classes, the retrieval rate is similar to the edge mapped application of GFD (MGFD), with slightly better overall improvement. To effectively gauge each descriptor more intuitively, we look at some explicit visual results, and offer a more subjective view of their relative success, in terms of the shapes of images retrieved (regardless of class), and their ranking in terms of similarity to the query.

The left top first image in each of the following figures is the query image we selected. The order of similarity ranking is from left to right, top to bottom. In Figure 3.11(a), FD

only retrieves 3 flag images, whilst in Figure 3.11(b), the whole boundary shape of the flags (roughly rectangular due to the flags waving) as well as the internal Union Jack feature becomes significant. MGFD not only extracts a more accurate exterior boundary feature than FD, but also considers interior shape features, retrieving 8 flag images with a strong feature in the top left of the flag. In Figure 3.11(c), some of the more textured details in the flags are eliminated by the ETC algorithm, thus only dominant edges (both internal and external) contribute to the search and improve the retrieval result. The proposed MGFD1 approach has the best performance, evident not only in a higher retrieval rate (11 flag images), but more importantly, in the set of Union Jack based flags dominating the most similar images, as opposed to the scattered flags found by the MGFD. This reflects a greater accuracy in the ranking of similar result images by the MGFD1.

In Figure 3.12(a) FD only retrieves 3 aircraft images. Like the flags, the aircraft images exhibit some regularity in terms of shape, although it isn't captured effectively by FD. In Figure 3.12(b) MGFD retrieves 9 aircraft images, but due to the influence of the internal texture part, the fruit images (the 9th, 10th, 12th and 16th images) are retrieved falsely. Their apparent similarity may be in that the fruit is distributed in an elongated manner, yet internal textures are erratic and confuse the similarity matching. In Figure 3.12(c), MGFD1 removes the texture 'contamination' (the grass and internal part of the plane) and retrieves 14 aircraft images.

In Figure 3.13 the concept of a cat does not necessarily coincide with a consistent silhouette, in fact, in this class different numbers of cats may exist in some images. FD retrieves 3 cat images, MGFD also retrieves 3 cat images, but the 4th result, with 3 kittens is similar to the query. Likewise, the distribution in the tree images (2nd and 3rd), fruit (9th) and rock formations (7th, 15th) are more similar in terms of overall shape than the images with one cat retrieved using the FD method. In the MGFD1 result, not only are 4 cat images retrieved, but the rank of the 3 kittens image is improved from 4th to 2nd, (i.e. it is considered most similar to the query), as opposed to the MGFD result. The experimental results demonstrated that the MGFD1 approach achieves a better result as a shape feature. This



(a) FD



(b) MGFD



(c) MGFD1

Figure 3.11: Flag query - CLASS 9

A grid of 16 small images arranged in two rows of eight. The top row includes: a Concorde airplane on a runway; a small framed picture of a person; a framed picture of a person in a landscape; a framed picture of a dog; a person sitting on a chair; a small potted tree; a large American flag; and a small dog. The bottom row includes: a Concorde airplane on a runway; a Russian flag; a person holding a basketball; a Concorde airplane on a runway; a pear; a small potted tree; a parrot; and a small dog.

A 2x8 grid of 16 images. The top row contains: a Concorde on a runway, a basket of fruit, a jet taking off, a clock tower, a large crowd, a jet in flight, a Concorde on a runway, and a jet in flight. The bottom row contains: a banana and a coconut, a plate of fruit, a jet on a runway, a plate of sliced oranges, a jet on a runway, a large rock, a jet on a runway, and a basket of fruit.

Figure 3.12: Aircraft query - CLASS 10

effect should mean that if MGFD1 is combined with other features (color, etc), we expect that the other features will help capture more images from the same class, allowing MGFD1 to sift out and rank the captured set more accurately.

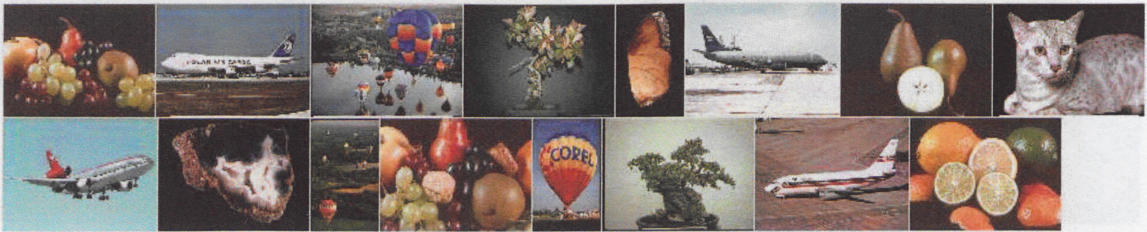
In the same way, the fruit query of Figure 3.14, shows that the clustered balloons are considered to be similar to the clustered fruit objects. The MGFD1 result however, is still



Figure 3.13: Cat query - CLASS 3



(a) FD



(b) MGFD



(c) MGFD1

Figure 3.14: Fruit query - CLASS 1

effect should mean that if MGFD1 is combined with other features (color, etc), we expect that the other features will help capture more images from the same class, allowing MGFD1 to sift out and rank the captured set more accurately.

In the same way, the fruit query of Figure 3.14, shows that the clustered balloons are considered to be similar to the clustered fruit objects. The MGFD1 result however, is still more accurate (same fruit ranked 2th for MGFD1, 12th for MGFD).

3.4 Summary

In this Chapter, we presented our work on extracting effective shape features for CBIR. The Edge-Texture Characterization (ETC) algorithm, morphological operations and GFD are discussed in details. By using ETC method and morphological operations, the texture regions contaminating the edge information are found and removed from the edge map of the image, so that we can get more accurate edge shape information, *i.e.*, shape information related more to boundaries in the scene. The GFD algorithm is also explained in depth. We proposed a modified generic Fourier descriptor (MGFD1) for image retrieval. Compared with the GFD, where the authors extracted the shape feature from the whole shape image in the MPEG-7 region shape database, application to natural images instead of binary trademark images makes the computation very expensive because of high resolution and complicated shape information inherent in natural images. GFD in its original form is not suitable for natural image retrieval. Our proposed MGFD1 overcomes this drawback.

The PFT was applied to the Canny edge maps of images, thereby decreasing the computational complexity. The MGFD1 shape feature improved average image retrieval rate (except where shapes varied dramatically). In such cases however, the similarity rankings were more intuitive. It was argued that in the cases where retrieval rate was lower than that of MGFD, the ultimate performance of the system still showed improvement as images from the correct class, yet with different shape to the query, were rejected or ranked lower in terms of similarity to the query - reflecting an order that more closely fits our perceived

notion of shape.

In Chapter 4, we will discuss CBIR that combines color, texture and shape features for natural image databases, with focus on dynamic weight selection based on different queries for CBIR.

Chapter 4

Dynamic Weight Selection for CBIR using Relevance Feedback

In Chapter 3, we discussed CBIR based on shape features for natural images in detail. We have proposed a modified generic Fourier descriptor (MGFD) to represent shape features, achieving a better result in image retrieval. However, based on very large stored natural image database, the retrieval rate using shape features alone is very low. Furthermore, using only shape to index images is bound to retrieve images from completely different conceptual classes which just happen to have a similar shape boundaries as the query image. An example of this is seen in Figure 3.13(b) and (c). Many classes in an image database may be not conceptually defined by shape only, thus in practical image retrieval, CBIR is often based on a set of low level features representing color, texture and shape. Color histograms and color moments are the popular ways for color representations [70, 71], Gabor wavelet filter provides a good tool for the extraction of texture feature [72].

In image retrieval algorithms, similarity measures between query image and the images in the database are linear combinations of their corresponding lower level similarities [66]. So normalization and weighting of features in a combined low level feature database become two key components in any retrieval algorithm. Relevance feedback builds the bridge between the high level concepts and the low level features, achieving dynamic weight updating and improved image retrieval rate. Figure 4.1 shows a CBIR system for weight selection using relevance feedback. The yellow blocks will be discussed in this Chapter. This Chapter is

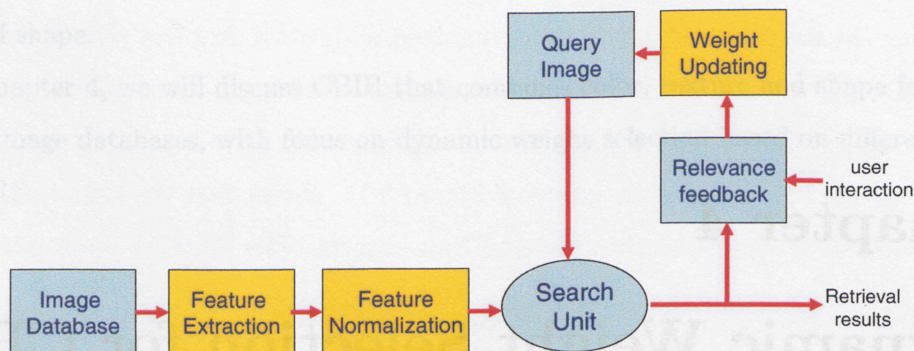


Figure 4.1: The frame for weight selection using relevance feedback

organized as follows. Section 4.1 introduces the extraction of color and texture features. Section 4.2 discusses the normalization of the low level feature database. The dynamically updating weights and query modification based on the user's feedback are given in Section 4.3 and 4.4 respectively. The experimental results with 10,000 natural images are shown in Section 4.5. Finally, we summarize in Section 4.6.

4.1 Feature Extraction

Feature extraction is the basis of content based image retrieval, that is to say, the success of image retrieval significantly depends on the results of feature extraction. Because of the subjectivity of perception, a best representation for a given feature does not exist. The combination of multiple representations for the given features can better characterize features from different viewpoints. In our retrieval system, we extract three kinds of visual features: color, texture and shape. Shape features have been discussed in Chapter 3 thoroughly, in which we demonstrated that the modified generic Fourier descriptor is an effective shape descriptor for natural images. In this section, feature extraction of color and texture will be discussed briefly.

4.1.1 Color Features

Color is a very important visual feature that is more readily perceived by the viewer. It is one of the most widely applied visual features in image retrieval. Color features are very robust to background complication and are independent of the size and orientation of the image. There are many different techniques for extracting color features proposed in the literature [67, 68, 69, 70]. Here, we present two such color feature representations as applied in our image retrieval system: color histograms and color moments.

Color histograms are the most popular representation for color features. M. J. Swain and D. H. Ballard first proposed the notion of color histograms [67]. Color histograms represent the joint probability of the intensities of the three color channels (such as the RGB color space). An advantage of color histograms is that they are invariant to translation and rotation about the viewing axis, and change only slowly under change of angle of view, change in scale and occlusion. Since histograms change slowly with view, a small number of histograms can represent well, the color content of a three-dimensional object. Swain and Ballard also proposed *Histogram Intersection*, as similarity measure of color histograms. In order to solve the location problem, an algorithm called *Histogram Backprojection* was introduced, which showed efficient performance in crowded scenes.

In order to reduce the number of colors and increase the computational efficiency, quantization of the color space plays an important role. However, quantization operations applied to color histograms may produce problems in the presentation of color content in an image. The number of quantization levels should be considered with such condition that perceptually different colors do not fall into the same bin. Generally, a uniform quantization of the color space may be fit for perceptually uniform color spaces. But in case of perceptually non-uniform color spaces such as RGB, CIE-LAB and HSV, a non-uniform quantization algorithm should be found. Researchers have proposed many different ways to quantize the color space. Smith and Chang [68] proposed to partition the HSV color space into 166 color bins. To represent color features, we select HSV color space, due to its independent performance in three color channels and easy transformation from RGB space. Because the V

coordinate is easily affected by the lighting condition, we only use HS coordinates to form 16 X 3 two dimensional histograms. We place more importance on hue (16 levels) than on the value of saturation (only 3 levels).

Besides color histograms, color moments as a color feature representation are also often applied in image retrieval. To overcome the quantization effects in color histograms, M. Stricker and M. Orengo proposed using the color moments approach [70]. Instead of extracting the complete color distribution, the index contains only dominant features, which can be extracted robustly from the images. From probability theory, a probability distribution is uniquely characterized by its moments. Thus, the color distribution of an image can be rendered as a probability distribution and characterized by its moments. Furthermore, because most of the color information is collected within low-order moments, only the first three moments of each color channel are necessary to be extracted as a significant representation of color. The first moment μ is the average value (mean): capturing the average color of the image. The variance σ and the skewness s of each color channel are the second and third central moments. We store the standard deviation σ and the third root s of the skewness of each color channel in the index. The equations of calculating μ , σ and s are shown as follows.

$$\mu_i = \frac{1}{N} \sum_{j=1}^N p_{ij} \quad (4.1)$$

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^2 \right)^{\frac{1}{2}} \quad (4.2)$$

$$s_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^3 \right)^{\frac{1}{3}} \quad (4.3)$$

where in the equations 4.1-4.3, N is the number of all the pixels of each color channels in an image, p_{ij} is the value of the i th color channel at the j th image pixel. So the color moments $f_{moments}$ representing color features in an image are organized as:

$$f_{moments} = \{\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3, s_1, s_2, s_3\} \quad (4.4)$$

The test results show this approach has better performance within a CBIR context [70]: its index is the smallest, the retrieval process is the fastest and it overcomes the quantization effects in color histograms.

4.1.2 Texture Features

Texture is another important feature of images. Although we removed textured regions contaminating the edge content in the description of edge shape information, it is only based on obtaining more accurate edge contour for the extraction of the shape feature (Chapter 3). Texture is still a very powerful visual characteristic of images and often applied in image retrieval as a feature. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment [73]. Due to the importance and usefulness of texture in pattern recognition, computer vision and image retrieval, more and more research achievements are based on texture representations.

In the 1970s, the co-occurrence matrix representation of texture feature was proposed by Garalick et al. [73]. This approach studied the gray spatial dependence of texture. The statistics originally proposed in [73] were evaluated by Gotlieb and Kreyszig, their experimental results showed that, in general, *contrast*, *inverse deference moment* and *entropy* had the biggest discriminatory power [74].

Since the introduction of the wavelet transform in early 1990s, many researchers have applied the wavelet transform to texture feature extraction. Smith and Chang [75] proposed the statistics (mean and variance) obtained from the wavelet sub-bands as the texture representation. It achieved over 90% accuracy on the 112 Brodatz texture images. Chang and Kuo [76] used the tree structured wavelet transformation to explore the middle-band characteristics and further improve the classification accuracy. Ma and Manjunath [72] evaluated the texture image annotation by various wavelet transformation representations, including orthogonal (OWT) and bi-orthogonal (BWT) wavelet transform, tree-structured decomposition using orthogonal (TOF) and bi-orthogonal (TBF) filter banks, and Gabor wavelet transform (GWT). Their experimental results indicated that the Gabor transform was the

best match to the results from human vision studies. In a more recent paper, Chen et al. evaluated the effects of different Gabor filter parameters on image retrieval. They found the best combination of parameters was scale 6 and orientation 4 with 24 filters [77]. In our image retrieval system, the texture feature is extracted by Gabor filter bank proposed in [78].

Gabor Functions

A two dimensional Gabor function $g(x, y)$ and its Fourier transform $G(u, v)$ can be written as follows,

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right] \quad (4.5)$$

$$G(u, v) = \exp\left\{-\frac{1}{2}\left[\frac{(u - W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]\right\} \quad (4.6)$$

where $\sigma_u = 1/2\pi\sigma_x$ and $\sigma_v = 1/2\pi\sigma_y$. W defines the spatial frequency of the sinusoid. Gabor functions form a complete but nonorthogonal basis set. Expanding a signal using this basis provides a localized frequency description. A class of self-similar functions, referred to as *Gabor wavelets*, is now considered. Let $g(x, y)$ be the mother Gabor wavelet, then this self-similar filter dictionary can be obtained by appropriate dilations and rotations of $g(x, y)$ through the generating function defined as:

$$g_{mn}(x, y) = a^{-m}G(x', y'), a > 1, m, n = integer \quad (4.7)$$

$$x' = a^{-m}(x\cos\theta + y\sin\theta) \quad (4.8)$$

$$y' = a^{-m}(-x\sin\theta + y\cos\theta) \quad (4.9)$$

where $\theta = n\pi/K$ and K is the total number of orientations. The scale factor a^{-m} in equation 4.9 is to ensure that the energy is independent of m .

Gabor Filter Dictionary Design

The nonorthogonality of the Gabor wavelet implies that there is redundant information in the filtered images, thus, we have to find a strategy to reduce this redundancy. Let U_l and U_h denote the lower and upper centre frequencies of interest, respectively. Let K denote the number of orientations and S be the number of scales in the multi-resolution decomposition. That is to ensure that the half-peak magnitude support of the filter responses in the frequency spectrum touch each other. The filter parameters σ_u and σ_v are derived from the following formulas:

$$a = \left(\frac{U_h}{U_l}\right)^{-\frac{1}{S-1}} \quad (4.10)$$

$$\sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2\ln 2}} \quad (4.11)$$

$$\sigma_v = \tan\left(\frac{\pi}{2k}\right)[U_h - 2\ln\left(\frac{\sigma_u^2}{U_h}\right)][2\ln 2 - \frac{(2\ln 2)^2 \sigma_u^2}{U_h^2}]^{-\frac{1}{2}} \quad (4.12)$$

where $W = U_h$ and $m = 0, 1, \dots, S-1$. In Figure 4.2 [78], the contours show the half-magnitude of the filter response in the designed Gabor filter dictionary. The filter parameters used are $U_h = 0.4$, $U_l = 0.05$, $K = 6$ and $S = 4$.

Feature Representation

For a given image $I(x, y)$, its Gabor wavelet transform is defined as:

$$W_{mn}(x, y) = \int I(x_1, y_1) g_{mn}^* (x - x_1, y - y_1) dx_1 dy_1 \quad (4.13)$$

where $*$ indicates the complex conjugate. It is assumed that the local texture regions are spatially homogeneous, the mean μ_{mn} and the standard deviation σ_{mn} of the magnitude of the transform coefficients are used to represent the region for classification and retrieval purposes:

$$\mu_{mn} = \int \int |W_{mn}(xy)| dx dy \quad (4.14)$$

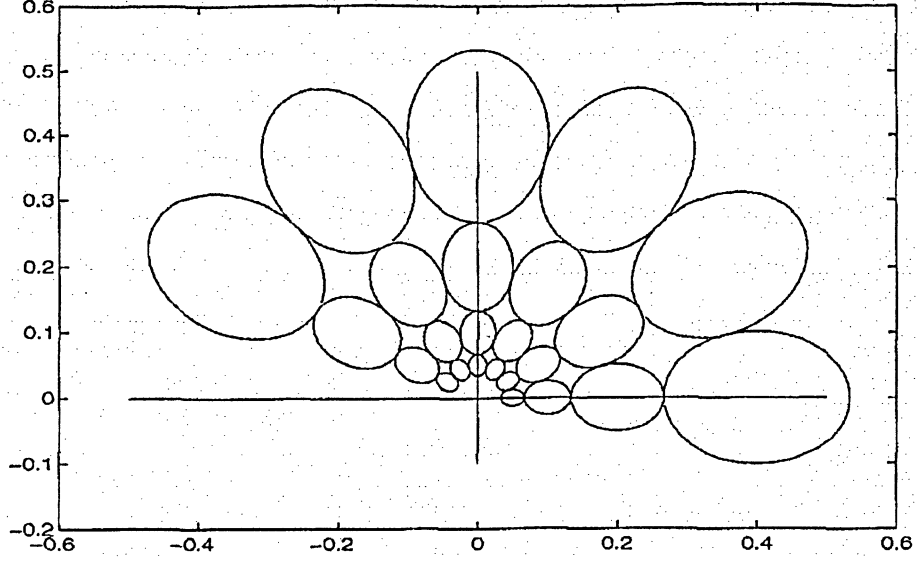


Figure 4.2: Contours of the designed Gabor filter dictionary

$$\sigma_{mn} = \sqrt{\int \int (|W_{mn}(x, y)| - \mu_{mn})^2 dx dy} \quad (4.15)$$

We use u_{mn} and σ_{mn} as feature components to construct a feature vector. Due to the best combination of filtering parameters being scale 6 and orientation 4 [77], for our experiments, the texture feature vector is formed as follows:

$$f_{texture} = \{\mu_{00} \quad \sigma_{00} \quad \mu_{01} \quad \dots \quad \mu_{35} \quad \sigma_{35}\} \quad (4.16)$$

4.2 Normalization

Before performing the similarity measure in image retrieval, the values of the indexed database formed by three kinds of features (color, texture and shape) have to be of the same dynamic range, *i.e.*, from 0 to 1. Otherwise, the linear combination of similarity measure becomes meaningless, because the features owning higher values will overshadow the features owning the lower values. The different components in a vector may be of totally different

physical quantities, thus the feature database should be normalized before applying the similarity calculation. The normalization process emphasizes each component of the feature vector equally [66].

Let V be an $M \times N$ matrix representing the whole feature database of M images, and N elements form the feature vector of an image. $v_{m,n}$ denotes the value of the n th feature of the m th image. Our goal is to normalize the components in each column to the same range so that every individual component gets equal emphasis in similarity measure. One way of normalization to the range $[0, 1]$ is to find the maximum and minimum value for the sequence. The formula can be written as:

$$v_{m,n} = \frac{v_{m,n} - \min_n}{\max_n - \min_n} \quad (4.17)$$

where \min_n and \max_n are the minimum and maximum values of each feature column across the space of all images in the database. Although it is very simple, we can not get desired normalized results. For example, we consider a sequence $[1.0, 1.3, 1.5, 1.6, 200]$, if this sequence is normalized by equation 4.17, most of the $[0, 1]$ range will be taken away by the value 200. The other values $[1.0, 1.3, 1.5, 1.6]$ will be warped into a very narrow range. A better way to normalize the sequence is to use Gaussian normalization. We achieve feature normalization by calculating the mean μ_n and standard deviation σ_n of each column features in the database. The normalized database will be derived by equation 4.18:

$$v_{m,n} = \frac{v_{m,n} - \mu_n}{3\sigma_n} \quad (4.18)$$

The probability of an entry's value within the range $[-1, 1]$ is about 99% (within three standard deviations). The advantage of this normalization process is that a few abnormal values occurring in the sequence (*i.e.* the outlier value 200) will not bias the importance of the other values. Therefore, this normalization approach has been applied our feature database.

4.3 Weight Updating

The image feature database has been normalized according to the normalization approach introduced in the last section, as such, there is equal emphasis on each component within the range $[-1, 1]$ in the normalized database. The initial image retrieval is performed by setting equal weight to every component in the feature database. However, we usually cannot achieve our desired results because of the gap between the high level concepts and the low level features. Relevance feedback based interactive retrieval approaches (introduced in Chapter 2) thus are often used in CBIR. Updating weights based on user's feedback is an efficient approach to improve the retrieval results.

Rui et al. [66] proposed a method which dynamically updated weights based on the query object to model the high level concepts and perception subjectivity. They used inverse standard deviation of feature vectors of relevant images marked by user's feedback as a weight vector of the query image. From the viewpoint of pattern recognition, for a feature to be effective, its variance should be large among all the images in the database and small among relevant images. Aksoy et al. [79] calculated the ratio of standard deviations of feature values between the database and r relevant images selected by user. Their experiments indicated the better performance based on weight updating. The equation for updating weights is written as follows.

$$w_j = \frac{\sigma_j}{\sigma_{r,j}} \quad j = 1, \dots, N \quad (4.19)$$

where $\sigma_j = std(f_j)$, being the standard deviation of values f_j of the j th feature components of the images in the whole feature database F . $\sigma_{r,j} = std(f_{r,j})$, being the standard deviation of values $f_{r,j}$ of the j th feature components of the r relevant images labeled by user's feedback in the feature database F . N is the number of features in a feature vector. In equation 4.19, we also need to normalize w_j in the following formula:

$$w_j = \frac{w_j}{w_{T,j}} \quad (4.20)$$

where $w_{T,j} = \sum_{j=1}^N w_j$. Figure 4.3 shows the graphs of weight vectors based on two different query images, obtained by equation 4.20. In these two graphs, we use different colors corresponding to different sets of feature sequences. *i.e.*, ‘brown’ for color moments, ‘green’ for color histogram, ‘red’ for texture and ‘blue’ for the shape feature set. Their values are completely different, that is to say, the higher values in this weight vector contribute more to the similarity measure than lower values. On the other hand, comparing the weight components in two weight vectors based on different query images, Figure 4.3(a) and (b): the value with the same order is completely different, that means their contributions for the similarity measure are also different, thus dynamically updating weights based on different query objects achieves better performance than fixed weights selection.

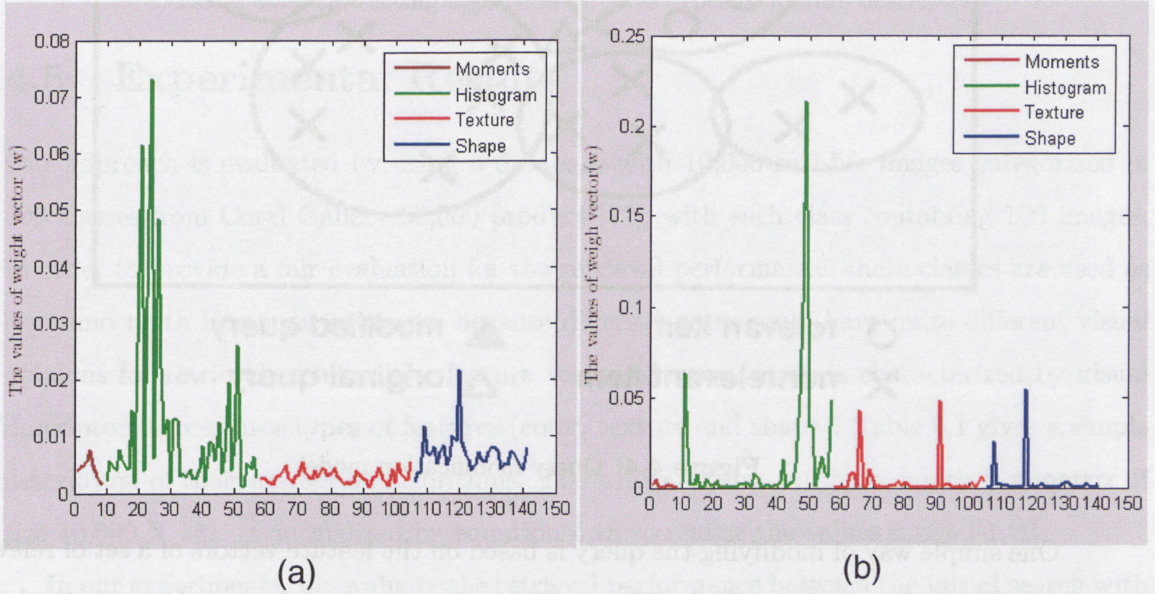


Figure 4.3: The graph of weight vectors based on two different query images

4.4 Query Modification

In the Query by Example (QBE) search paradigm, the user selects a sample image as an initial query to the retrieval system. Let us consider this scenario where the initial query

can not represent the relevant class of images, that is to say, it may be located at the edge of the group of relevant images. Under this scenario, the system is not able to perform significantly. The query modification approach can be used to modify the query based on the user feedback. The modified query is then applied to the next retrieval iteration. The query modification idea was first proposed by Salton for text retrieval [80].

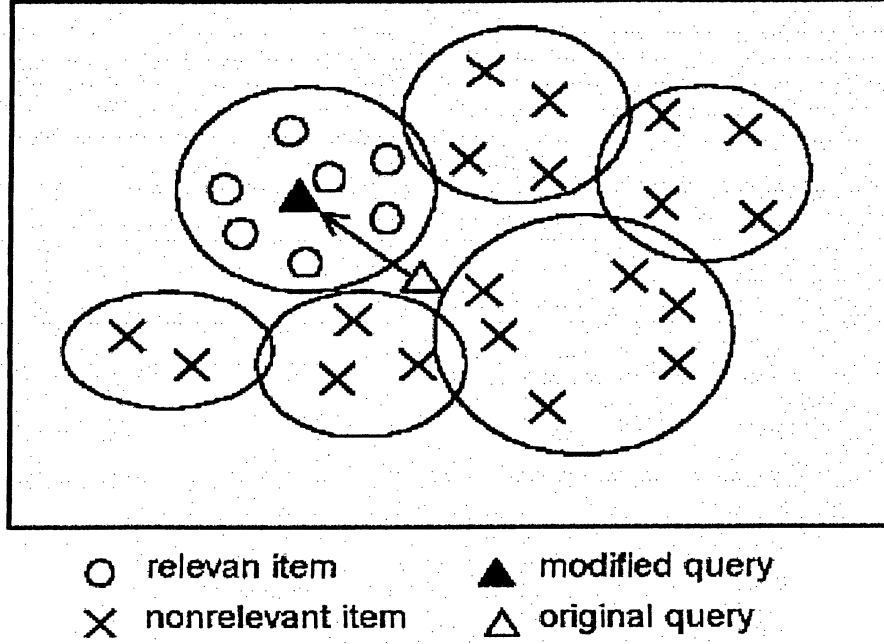


Figure 4.4: Query modification model

One simple way of modifying the query is based on the feature vectors of a set of relevant images labeled by user's feedback. The modified query can be taken as the central value of the relevant sequence of feature vectors [81].

$$F_{mq} = \frac{1}{M} \sum_{m=1}^M X_m \quad (4.21)$$

where F_{mq} is the feature vector of the modified query and X_m is the feature vector of the m th image in the relevant set. This query modification method is illustrated in Figure 4.4.

Color Moments (9)	Extracting mean μ , standard deviation σ and skew s from three color channels (RGB) forms a color feature vector with 9 components.
Color Histograms (48)	48-bin color histograms in HSV color space, H and S are uniformly separated into 16 and 3 scales respectively.
GW Transform (48)	Gabor wavelet (GW) filters with 4 scales and 6 orientations are applied to the gray level image resized into 128 x 128 pixels in size. The mean and standard deviation of the GW coefficients form 48 features.
Modified GFD (36)	36 coefficients of the generic Fourier descriptor, for calculating an equally distributed set of 4 radius and 9 angles. (introduced in Chapter 3)

Table 4.1: Extracting algorithm of three types of feature descriptors

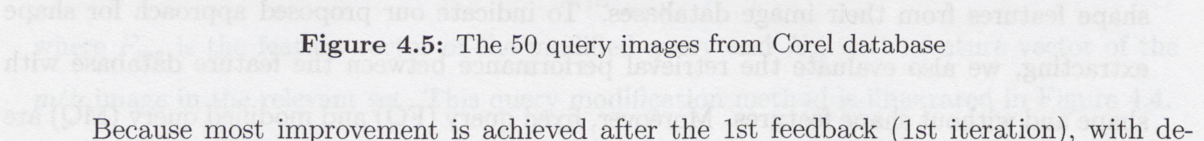
4.5 Experimental Results

Our approach is evaluated by using a database with 10,000 real-life images categorized in 100 classes from Corel Gallery 65,000 product [82], with each class containing 100 images. In order to provide a fair evaluation for the retrieval performance, these classes are used as a ground truth in our experiments, because different users could have quite different visual opinions for retrieval results. The feature vector of every image is characterized by visual descriptors using three types of features (color, texture and shape). Table 4.1 gives a simple description of their extraction algorithms. The whole feature database, which is a matrix of size 10,000 X 141, is normalized by equation 4.18 to realize the values range [-1, 1].

In our experiments, we evaluate the retrieval performance between the initial search with equal weights and iterative search after user's feedback and weight updating. Because of the heterogeneity of the Corel test set, the extraction of some features is difficult, such as shape [66]. In their retrieval experiments [66] and [79], the authors did not extract shape features from their image databases. To indicate our proposed approach for shape extracting, we also evaluate the retrieval performance between the feature database with shape and without shape features. Moreover, fixed query (FQ) and modified query (MQ) are

$$Rate = \frac{\sum_{i=1}^{300} (n_i/N_i)}{300} \quad (4.22)$$

where N_i is the number of the top 16 similar images based on the i th query, n_i is the number of images as same class as the i th query in N_i .



Because most improvement is achieved after the 1st feedback (1st iteration), with de-

Options	0 Iteration	1 Iteration	2 Iteration	3 Iteration
FQ-NS	39.5	49.8	51.1	52.8
FQ-S	42.3	52.4	54.9	57.5
MQ-NS	39.5	56.7	58.4	61.3
MQ-S	42.3	61.8	63.7	65.1

Table 4.2: Average retrieval rate (in percentage) for 300 queries in 10,000 image database

creased rate of improvement after the 2nd and 3rd iteration, we only show the results of first 3 iterations in Table 4.2. We have summarized the results obtained by using the feature weighting scheme in the feedback loop. These results also show the comparison of the performance based on 4 different scenarios. In the initial search, it is observed that using the feature database with shape performs better than without shape and improves 2.8 per cent. The retrieval results after user's feedback are improved significantly compared with the initial search (almost 23%). Query modification approach gives us a better performance in image retrieval compared with fixed query, as MQ-S gives a 7.6 per cent higher retrieval rate than FQ-S. Furthermore, this is especially apparent with the shape feature extracted by our proposed method (MGFD) as MQ-S performs 3.8 per cent better than MQ-NS after 3 iterations. Figure 4.6 depicts the average retrieval performance of the system for the different scenarios.

Figures 4.7-4.12 show the retrieval performance of the system for a few example queries which are modified in the feedback loop. In every example, we use a graphical user interface (GUI) to display the top 16 similar images (ranking from left to right and from top to bottom), the left-top single image is the query. Figure 4.7 and 4.8 show the performance based on 'Rock' query including and excluding the shape features. As the main area appears to be dominated by shape in 'Rock' image, Figure 4.7(a) shows 12 relevant images retrieved in an initial search, Figure 4.7(b) illustrates that the 16 'Rock' images are retrieved after user's feedback. Especially for cat images (the 5th, 9th and 14th), the shape between 'cat' and 'rock' is different, due to improved contribution of shape after weight updating, thus

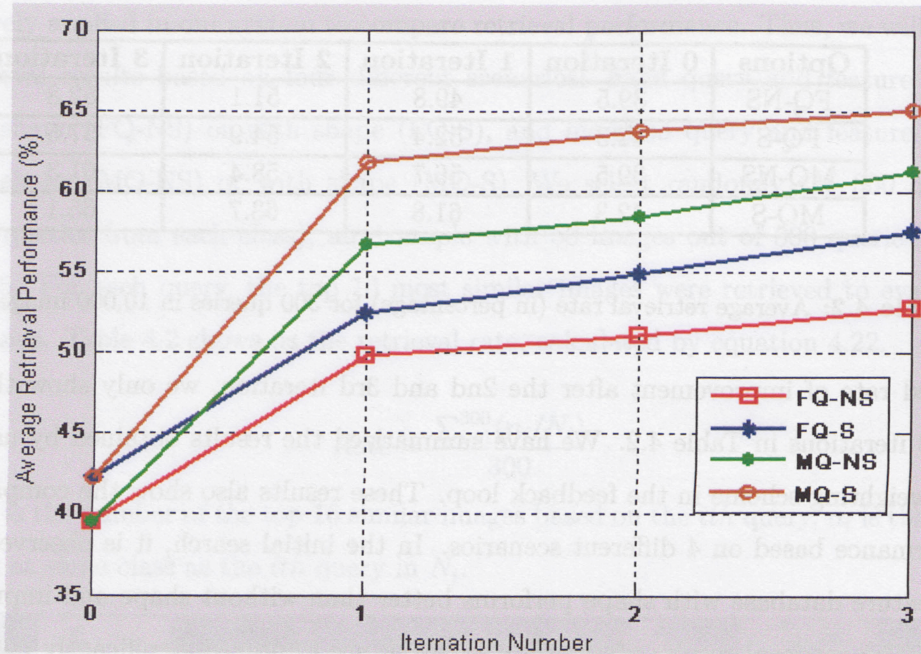


Figure 4.6: Retrieval performance

the cat images are rejected. In Figure 4.8(a), due to loss of the shape information, only 8 rock images are retrieved, furthermore, the cat images (3rd and 4th) having similar color background and texture foreground are also retrieved. Also one more relevant image is retrieved by user's first feedback.

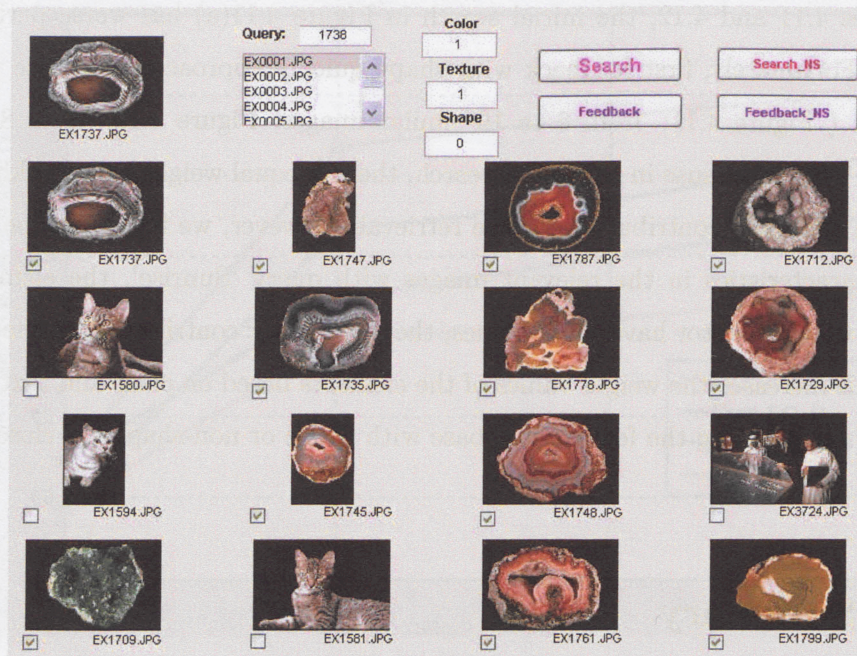
The retrieved results for a query of a plane flying in the cloudy sky are shown in Figures 4.9 and 4.10. An initial search finds 6 plane images, because one plane image (11th) depicting "taking-off" doesn't belong to the same class with query. According to ground truth, only five similar images are considered to be retrieved. Due to the contribution of dynamically updated weights through relevance feedback, Figure 4.9(b) retrieves 10 plane images, and ranking is improved (the first 6 images are now planes). When ignoring shape information of the plane, the cloud becomes the dominant component in color and texture, therefore, only 3 planes are retrieved, whilst 6 cloud images belonging to other classes are found to be similar images. Figure 4.10 (b) shows only mild improvement in the ranking of plane images.

Similarly, as the dominant features of the query 'Sunrise' appear to be color and texture

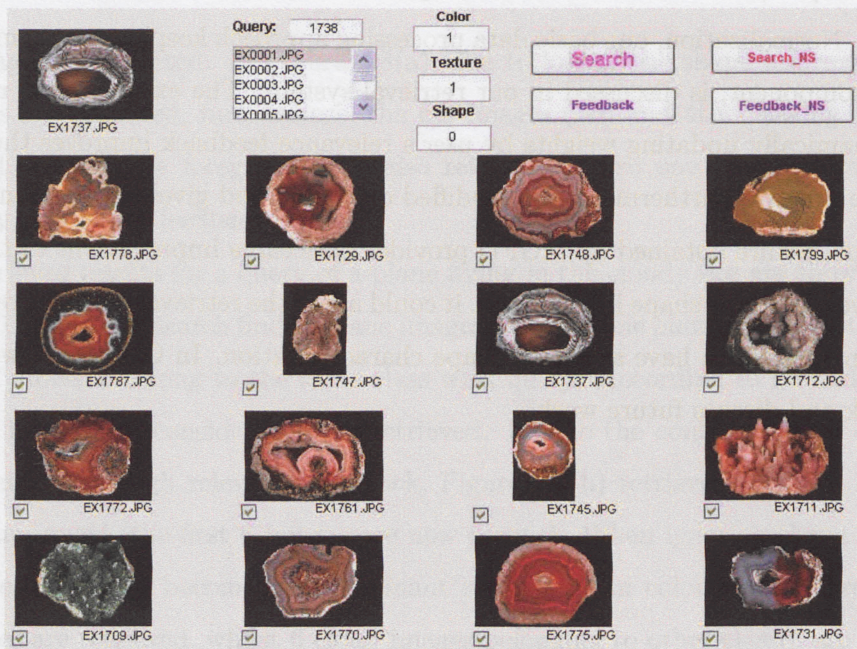
in Figures 4.11 and 4.12, the initial search in Figure 4.11(a) has worse performance than 4.12(a). Fortunately, first feedback with shape quickly approaches the case without shape features. (Figure 4.11: from 6 to 10 similar images; Figure 4.12: from 8 to 11 similar images). This is because in the initial search, there is equal weight value: 3 kinds of features are given the same contribution for the retrieval. However, we find that due to having less shape characteristics in the relevant images with query 'Sunrise', the elements based on shape in weights vector have small values, they give lower contribution in the first feedback, whilst this increases the weight values of the elements based on color and texture. Thus, the retrieval results using the feature database with shape or non-shape are close after the first feedback.

4.6 Summary

In this Chapter, we first introduced the representations of two primary features (color and texture). Normalization, one basic data processing approach keeping equal emphasis on each feature component, is discussed in our retrieval system. The experimental results indicate that dynamically updating weights by user's relevance feedback improves the retrieval performance greatly. Furthermore, the modified query method gives an apparent contribution. The shape feature obtained by MGFD provides noticeable improvement to the retrieval result. If we ignore the shape information, it could affect the retrieval results directly, especially for the images which have apparent shape characterization. In Chapter 5, we will conclude this work and discuss future work.

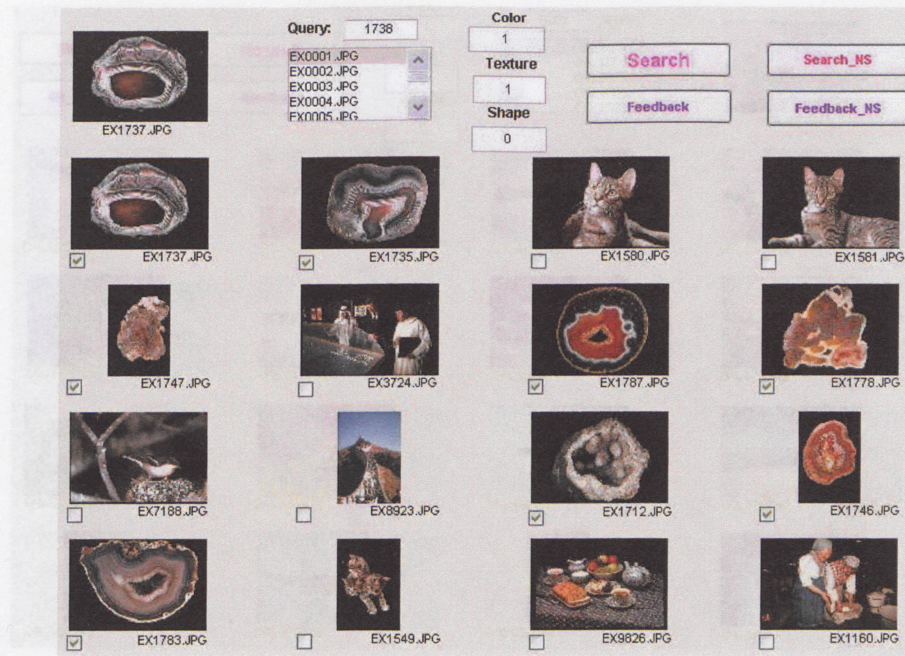


(a) Initial search

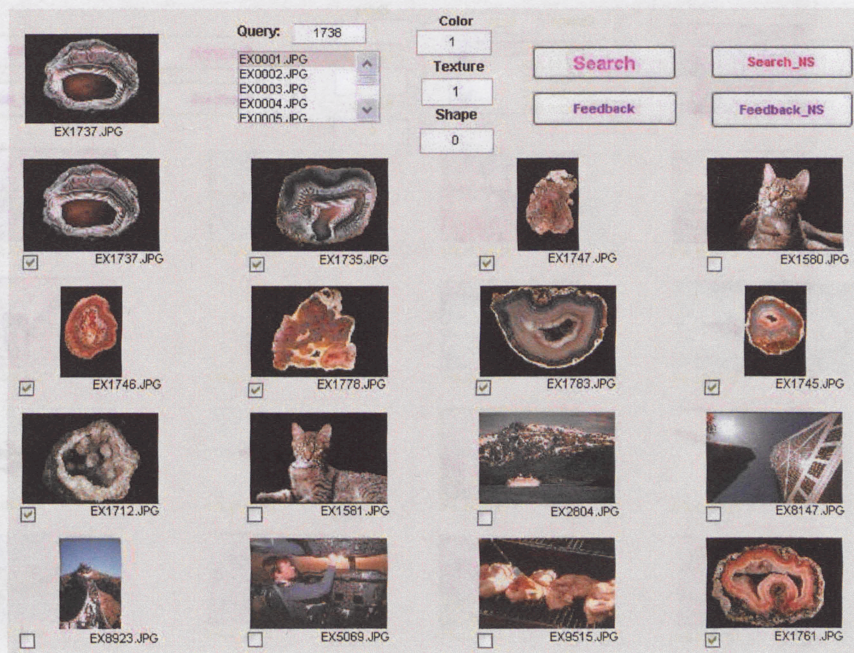


(b) First feedback

Figure 4.7: The retrieval performance based on 'Rock' query (including shape features)

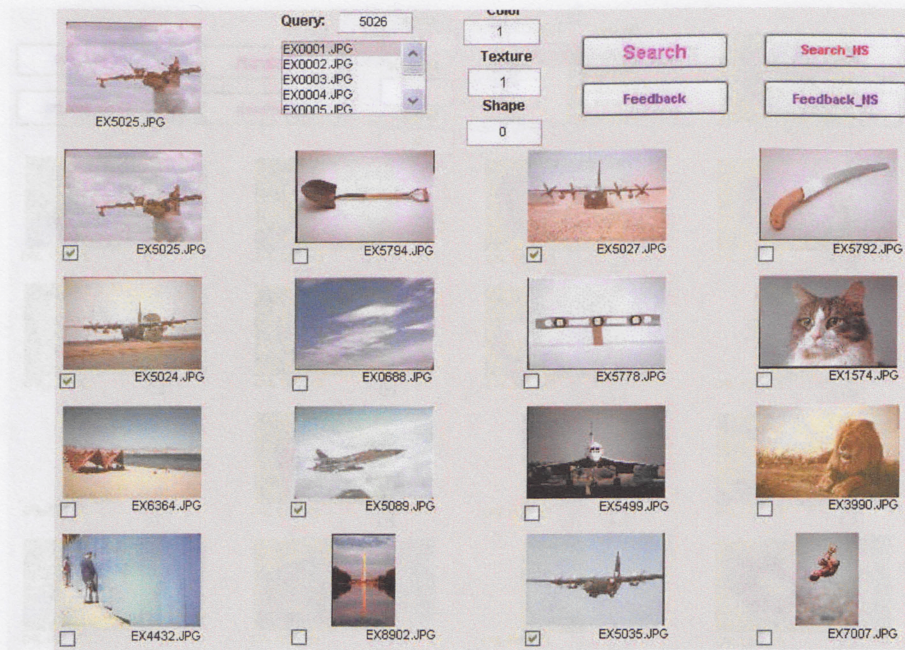


(a) Initial search

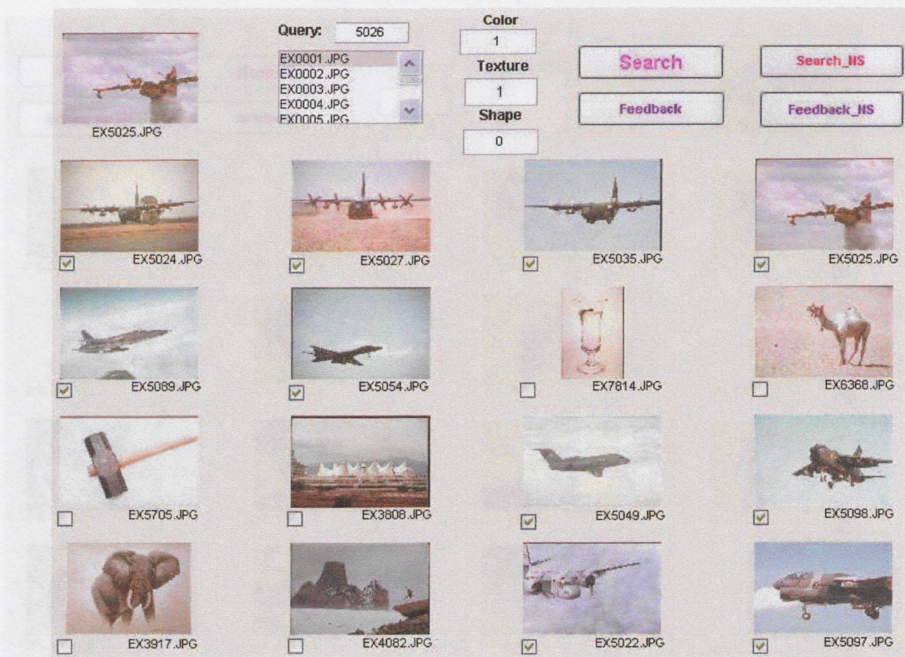


(b) First feedback

Figure 4.8: The retrieval performance based on 'Rock' query (excluding shape features)

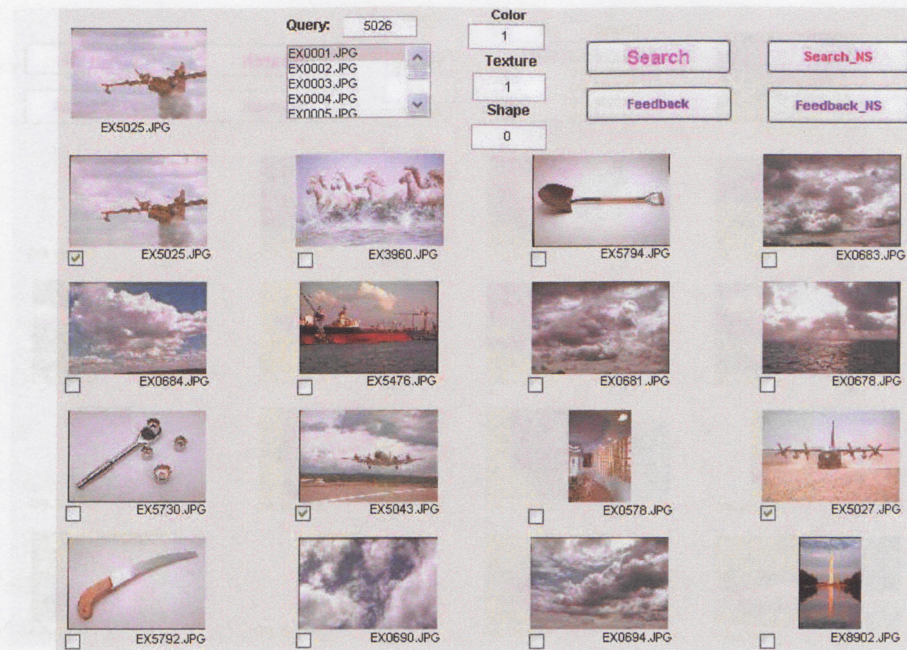


(a) Initial search

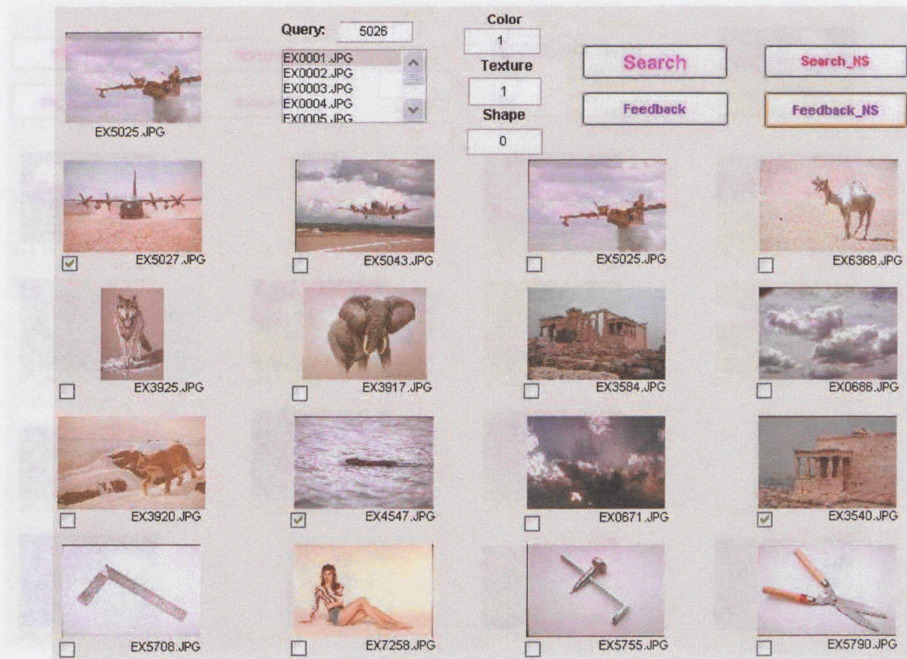


(b) First feedback

Figure 4.9: The retrieval performance based on 'Plane' query (including shape features)



(a) Initial search

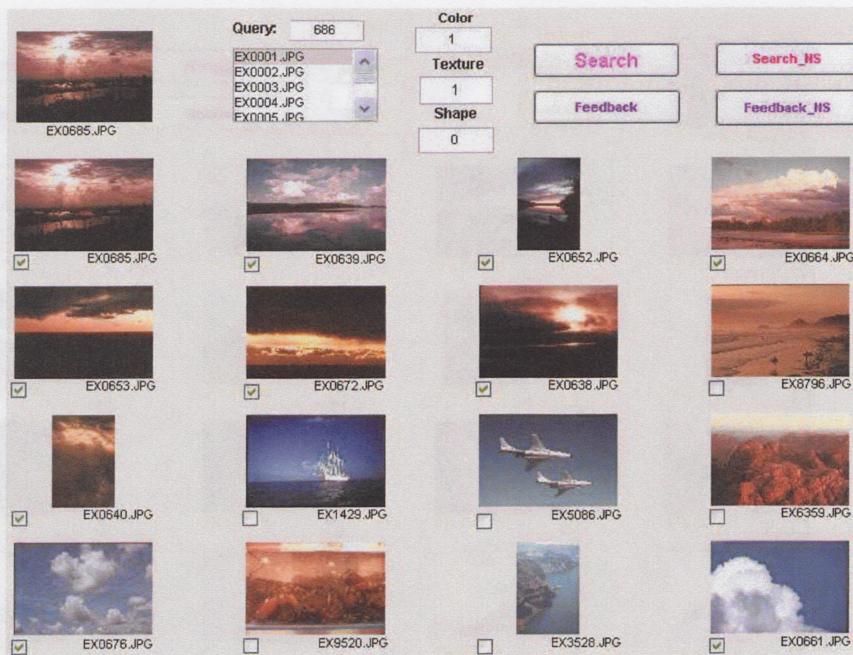


(b) First feedback

Figure 4.10: The retrieval performance based on 'Plane' query (excluding shape features)

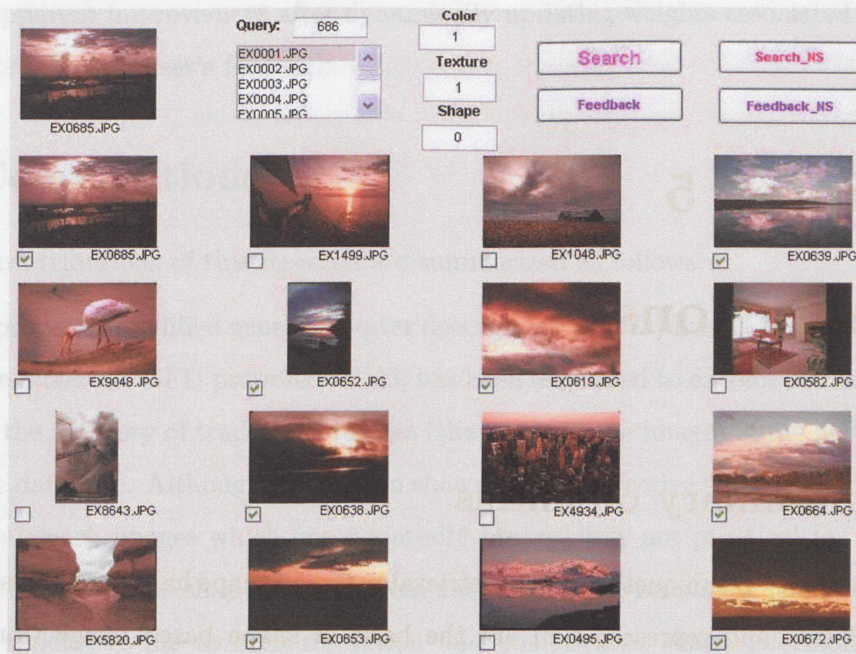


(a) Initial search

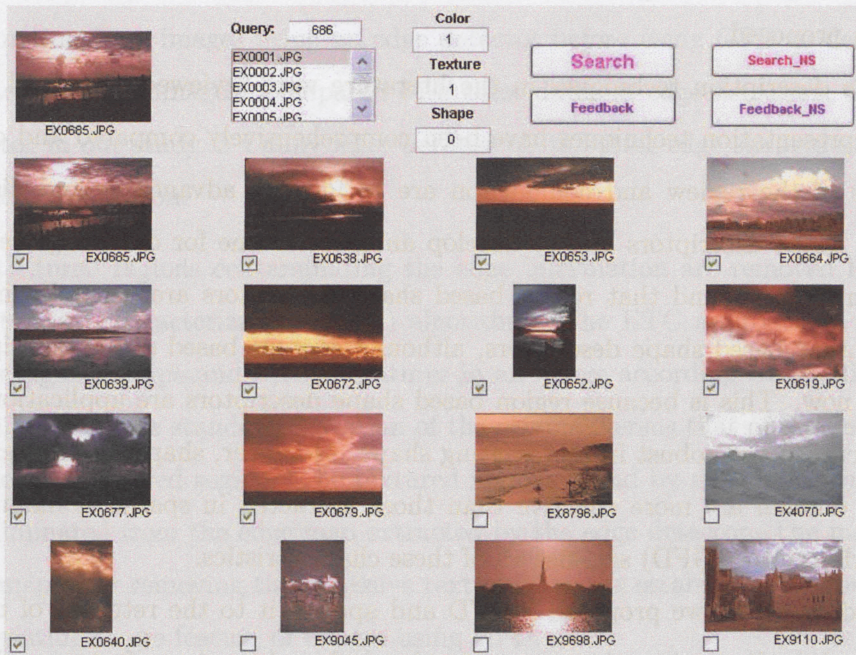


(b) First feedback

Figure 4.11: The retrieval performance based on 'Sunrise' query (including shape features)



(a) Initial search



(b) First feedback

Figure 4.12: The retrieval performance based on 'Sunrise' query (excluding shape features)

Chapter 5

Conclusions

5.1 Summary of Thesis

IN this thesis, techniques for image retrieval based on shape have been investigated. Shape description and representation are the basis of shape based image retrieval. In this research, an effective shape descriptor called the modified generic Fourier descriptor (MGFD) has been proposed.

Shape description techniques in the literature were reviewed. Some of the important shape representation techniques have been comprehensively compared and discussed. The purposes of the review and comparison are to identify advantages and disadvantages of different shape descriptors and to develop an effective one for our image retrieval system. The comparison found that region based shape descriptors are generally more promising than contour based shape descriptors, although contour based shape descriptors are more popular now. This is because region based shape descriptors are application independent, more accurate and robust in representing shape. Moreover, shape features extracted in the spectral domain are more effective than those extracted in spatial domain. The Generic Fourier descriptor (GFD) satisfies all of these characteristics.

Based on GFD, we proposed MGFD and applied it to the retrieval of natural images. Because of the limitation of image retrieval only based on shape feature, color and texture features are combined with shape to form the feature database. The retrieval performance

shows an apparent improvement after dynamically updating weights associated with feature components based on user's feedback.

5.2 Contributions

The main contributions of this research are summarized as follows:

- The proposed modified generic Fourier descriptor (MGFD) has been applied to natural image retrieval. GFD proposed in [43] has been only used to extract the shape features from the category of trademark images (the whole shape image) in the MPEG-7 region shape database. Although it has been shown GFD is effective to extract shape features in trademark images which are essentially binary, it is not practical for the GFD to directly extract the shape features from natural images, because it makes computation very expensive due to high resolution and complicated shape information inherent in natural images. MGFD overcomes this drawback by extracting the edge shape information from images using an edge detector before using GFD to describe shape characteristics of images. Compared with three different edge detectors, the best edge detector in our study is Canny filter because of its better edge shape description and simple computation.
- The textured regions contaminating the edge information are removed by using the edge-texture characterization (ETC) algorithm. The ETC approach is mainly used to distinguish shape and texture features in an image according to the different ratio range between the standard deviation of the original versus that of blurred intensities in a local windowed region. The textured regions found by morphological operations are eliminated from the edge map extracted by the edge detector. The more accurate edge map after removing the excessive textured regions establishes the foundation for representing shape feature of images using GFD.
- Combining color moments, color histograms, Gabor texture and the proposed shape features for image retrieval based on 10,000 images selected from Corel image database

has been investigated. Relevance feedback with the scheme of dynamically updating weights has also been adopted to bridge the gap between high level concepts and low level visual features. The experimental results show that the combination of the shape features extracted by MGFD and the dynamic updating of weights by user's feedback has a significant effect on improving retrieval performance.

5.3 Future Research Extensions

In content based image retrieval system, humans tend to compare the retrieval results using high-level concepts. However, image similarity measures are based on low-level visual features extracted automatically from images. In order to reduce this semantic gap, relevance feedback techniques have been applied in our system. We only use positive examples from user's feedback to decide the weights selection and ignore the negative examples. We should combine both examples to adjust weight vectors. Moreover, in relevant images, the users might consider some of them as more relevant than others. The multi-class input obtained from the users should be extended to relevance feedback.

In our system, the images relevant to the query have been found by the users in every iteration, this increases the users' workload greatly. An algorithm to automatically distinguish positive and negative examples from the retrieved images, instead of human participation, is an attractive approach. The self-organizing tree map (SOTM) is a suitable method to implement automatic machine interaction for CBIR [83]. We will use SOTM approach to automatically determine relevant and irrelevant images, minimizing users' subjectivity.

Our image retrieval system is only based on the characteristics of the whole image. The user may be interested in a particular object (region of interest) in a query image, so that the retrieval system should allow the user to feed-back the object rather than the whole image. The shape information of the object may be very significant, and MGFD might be the best shape descriptor to represent the shape of the object. However, the extraction of the object is a challenging task. Image segmentation algorithm will play an

important role in the extraction of the object. 'GrabCut' [84] and 'GraphCut' [85], recent developments in interactive computer vision, may provide an effective tool to interactively solve the segmentation problem.

Bibliography

- [1] Lycos, <http://www.lycos.com>.
- [2] Google, <http://www.google.com>.
- [3] Altavista, <http://www.Altavista.com>.
- [4] J. Li, J. Z. Wang and G. Wiederhold, "Intergrated region matching for image retrieval", *Proc. International Conference ACM on Multimedia*, pp. 147-156, 2000.
- [5] Sven Loncaric, "A survey of shape analysis techniques", *Pattern Recognition*, vol. 31, no. 8, pp. 983-1001, 1998.
- [6] M. Safar, C. Shahabi and C. H. Tan, "Image retrieval by shape: a comparative study", *Proc. IEEE International Conference on Multimedia and Expo*, vol. 1, pp. 141-144, Jul. 2000.
- [7] T. Pavlidis, "Algorithms for shape analysis of contours and waveforms", *IEEE trans. PAMI*, vol. 2, pp. 301-312, 1980.
- [8] H. Freeman and A. Saghri, "Generalized chain codes for planar curves", *Proc. the 4th International Joint Conference on Pattern Recognition*, pp. 701-703, Kyoto, Japan, Nov. 1978.
- [9] R. Mehrotra and J. E. Gary, "Similar-shape retrieval in shape data management", *IEEE computer*, vol. 28, no. 9, pp. 57-62, Sep. 1995.

- [10] C. L. Huang and D. H. Huang, "A Content-based Image Retrieval System", *Image and Vision Computing*, vol. 16, pp. 149-163, 1998.
- [11] D. M. Squire and T. M. Caelli, "Invariance Signature: Characterizing Contours by Their Departures from Invariance", *Computer Vision and Image Understanding*, vol. 77, no. 3, pp. 284-316, 2000.
- [12] I. Yong, J. Walker and J. Bowie, "An analysis technique for biological shape", *Computer Graphics and Image Processing*, vol. 25, pp. 357-370, 1974.
- [13] M. Peura and J. Iivarinen, "Efficiency of simple shape descriptors", *Proc. the 3rd International Workshop on Visual Form*, pp. 443-451, May 1997.
- [14] H. Asada and M. Brady, "The Curvature Primal Sketch", *IEEE Trans. on PAMI*, vol. 8, pp. 2-14, 1986.
- [15] F. Mokhtarian, S. Abbasi and J. Kittler, "Robust and Efficient Shape Indexing Through Curvature Scale Space", *Proc. British Machine Vision Conference*, pp. 53-62, Edinburgh, UK, 1996.
- [16] M. Daoudi and S. Matusiak, "Visual Image Retrieval by Multiscale Description of User Sketches", *Journal of Visual Language and Computing*, vol. 11, no. 3, pp. 287-301, 2000.
- [17] E. Persoon and K. S. Fu, "Shape Discrimination using Fourier Descriptors", *IEEE Trans. on System, Man and Cybernetics*, vol. 7, no. 3, pp. 170-179, 1977.
- [18] R. Chellappa and R. Bagdazian, "Fourier Coding of Image Boundaries," *IEEE Trans. on PAMI*, vol. 6, no. 1, pp. 102-105, 1984.
- [19] C. C. Lin and R. Chellappa, "Classification of Partial 2D Shapes using Fourier Descriptors", *IEEE Trans. on PAMI*, vol. 9, no. 5, pp. 686-690, 1987.
- [20] C. S. Lin and C. L. Hwang, "New Forms of Shape Invariants from Elliptic Fourier Descriptors", *Pattern recognition*, vol. 20, no. 5, pp. 535-545, 1987.

- [21] A. Krzyzak, S. Y. Leung and C. Y. Suen, "Reconstruction of two Dimensional Patterns from Fourier Descriptors", *Machine Vision and Applications*, vol. 2, no. 2, pp, 123-140, 1989.
- [22] K. Arbter, W. E. Snyder, H. Burkhardt and G. Hirzinger, "Application of Affine-Invariant Fourier Descriptors to Recognition of 3-D Objects", *IEEE Trans. on PAMI*, vol. 12, no. 7, pp. 640-647, 1990.
- [23] H. Kauppinen, T. Seppanen and M. Pietikainen, "A Experimental Comparison of Autoregressive and Fourier-based Descriptors in 2D Shape Classification", *IEEE Trans. on PAMI*, vol. 17, no. 2, pp. 201-207, 1995.
- [24] T. W. Rauber, "Two-Dimensional Shape Description", *Technical Report: GR UNINOVA-RT-10-94*, University Nova de Lisboa, Portugal, 1994.
- [25] B. M. Mehtre, M. S. Kankanhalli and W. F. Lee, "Shape measures for Content Based Image Retrieval: A Comparison", *Information Processing & Management*, vol. 33, pp. 319-337, 1997.
- [26] Y. Rui, A. C. She and T. S. Huang, "Modified Fourier descriptors for shape representation - a practical approach", *Proc. First International Workshop on Image Database and Multimedia Search*, 1996.
- [27] F. J. S. Marine, "Automatic recognition of biological shapes with and without representation of shape", *Artificial Intelligence in Medicine*, vol. 18, pp. 173-186, 2000.
- [28] Q. M. Tieng and W. W. Boles, "Recognition of 2D object contours using the wavelet Transform zero-crossing representation", *IEEE Trans. on PAMI*, vol. 19, no. 8, pp. 910-916, 1997.
- [29] 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.

- [30] J. R. Ohm, F. B. Bunjamin, W. Liebsch, B. Makai, K. Muller, A. Somlic and D. Zier, "A Set of Visual Feature Descriptors and Their Combination in a Low-level Description Scheme", *Signal Processing: Image Communication*, vol. 16, pp. 157-179, 2000.
- [31] G. Eichmann et al. "Shape representation by Gabor expansion", *SPIE, Hybrid Image and Signal Processing II*, vol. 1297, pp. 86-94, 1990.
- [32] D. S. Zhang and G. Lu, "A Comparison of Shape Retrieval Using Fourier Descriptors and Short-time Fourier Descriptors", *Proc. the Second IEEE Pacific-Rim Conference on Multimedia*, pp. 855-860, Beijing, China, Oct. 2001.
- [33] E. R. Davies, *Machine vision: Theory, Algorithms, practicalities*, Academic Press, 1997.
- [34] M. K. Hu, "Visual pattern recognition by moment invariants", *IRE Trans. on Information Theory*, vol. 8, pp. 179-187, 1962.
- [35] S. O. Belkasim, M. Shridhar and M. Ahmadi, "Pattern Recognition with Moment Invariants: A Comparative Study and New Results", *Pattern Recognition*, vol. 24, no. 12, pp. 1117-1138, 1991.
- [36] B. Jahne, *Digital Image Processing - Concepts, Algorithms and Scientific Applications*, Springer Verlag, Berlin, Heidelberg, 1997.
- [37] M. R. Teague, "Image Analysis via the General Theory of Moments", *Journal of Optical Society of America*, vol. 77, no. 80, pp. 920-930.
- [38] M. E. Celebi and Y. A. Aslandogan, "A Comparative Study of Three Moment-based Shape Descriptors", *Proc. IEEE International Conference on Information Technology: Coding and Computing*, vol. 1, pp. 788-793, Mar. 2005.
- [39] G. J. Lu and A. Sajjanhar, "Region-based Shape Representation and Similarity Measure Suitable for Content-based Image Retrieval", *Multimedia systems*, vol. 7, no. 2, pp. 165-174, 1999.

- [40] K. Chakrabarti et al. "Similar Shape Retrieval in MARS", *Proc. IEEE International Conference on Multimedia and Expo*, pp. 709-712, New York, USA, 2000.
- [41] R. Mukundan, S. H. Ong and P. A. Lee, "Image Analysis by Tchebichef Moments", *IEEE Trans. on Image Processing*, vol. 10, no. 9, pp. 1357-1364, 2001.
- [42] D. S. Zhang and G. Lu, "Enhanced Generic Fourier Descriptors for Object-based Image Retrieval," *Proc. IEEE International conference on Acoustics, Speech and Signal Processing*, vol. 4, pp: 3668-3671, Orland, USA, May 2002.
- [43] D. S. Zhang and G. Lu, "Generic Fourier Descriptors for Shape-based Image Retrieval", *Proc. IEEE International Conference on Multimedia and Expo*, vol. 1, pp. 425-428, 2002.
- [44] S. K. Setarehdan and J. J. Soraghan, "Fuzzy multiscale edge detection (FMED) applied to automatic left ventricle boundary extraction", *Proc. the 7th International Conference on Image Processing and Applications*, vol. 2, pp. 552-556, Jul. 1999.
- [45] Y. Uchiyama, M. Haseyama and H. Kitajima, "Hopfield Neural Network for Edge Detection", *Proc. IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 608-611, May 2001.
- [46] P. Bao and L. Zhang, "Scale Correlation-based Edge Detection," *Proc. IEEE International Symposium on Video/Image Processing and Multimedia Communications*, pp. 345-350, Jun. 2002.
- [47] J. F. Canny, "A Computational Approach to Edge Detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 112-131, 1986.
- [48] P. Kovesi, "Edge Are not Just Steps", *Proc. the 5th Asian Conference on Computer Vision*, pp. 822-827, Melbourne, Australia.
- [49] T. Kato, "Database Architecture for Content-based Image Retrieval", *Proc. SPIE in Image Storage and Retrieval systems*, pp. 112-123, 1992.

- [50] M. Flickner et al. "Query by Image and Video Content: the QBIC System", *IEEE computer Magazine*, vol. 28, pp. 23-32, Sep. 1995.
- [51] A. Pentland, R. W. Picard and S. Sclaroff, "Photobook: Content-based manipulation of image database", *International Journal of Computer Vision*, vol. 18, pp. 233-254, 1996.
- [52] J. R. Smith and S. F. Chang, "Visualseek: A Fully Automated Content-based Image Query System", *Proc. ACM Multimedia Conference*, pp. 87-98, 1996.
- [53] J. R. Smith and S. F. Chang, "Visually searching the web for content", *IEEE Multimedia Magazine*, vol. 4, pp. 12-20, 1997.
- [54] T. S. Huang, S. Mehrotra and K. Ramchandran, "Multimedia analysis and retrieval system (MARS) project", *Proc. 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval*, 1996.
- [55] Y. Rui, T. S. Huang and S. Mehrotra, "Content-based Image Retrieval with Relevance Feedback in MARS", *Proc. IEEE International Conference on Image Processing*, vol. 2, pp. 815-818, 1997.
- [56] Q. Tian, P. Hong and T. S. Huang, "Update Relevant Image Weights for Content-based Image Retrieval using Support Vector Machines", *Proc. IEEE International Conference on Multimedia and Expo*, vol. 2, pp. 1199-1202, 2000.
- [57] Z. Su, H. Zhang, S. Li and S. Ma, "Relevance Feedback in Content-based Image Retrieval: Bayesian Framework, Feature Subspaces and Progressive Learning," *IEEE Trans. on Image Processing*, vol. 12, pp. 924-937, 2003.
- [58] Y. Zhuang, J. Yang, Q. Li and Y. Pan, "A Graphic-Theoretic Model for Incremental Relevance Feedback in Image Retrieval", *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 413-416, 2002.

- [59] K. Chung and C. Fung, "Multiple Layer Kernel-based Approach in Relevance Feedback Content-based Image Retrieval System", *Proc. IEEE International Conference on Machine Learning and Cybernetics*, vol. 1, pp. 405-409, 2005.
- [60] P. Muneesawang and L. Guan, "Interactive CBIR using RBF-based Relevance Feedback for WT/VQ Coded Images," *Proc. IEEE International Conference on Acoust., Speechh, Signal Processing*, pp. 1641-1644, May 2001.
- [61] K. H. Yap and K. Wu, "Fuzzy Relevance Feedback in Content-based Image Retrieval System using Radial Basis Function Network", *Proc. IEEE the 4th Pacific Rim Conference on Multimedia*, pp. 177-180, 2005.
- [62] H. S. Wang and L. Guan, "A neural learning approach for adaptive image restoration using a fuzzy model-based network architecture", *IEEE Trans. on Neural Networks*, vol. 12, pp. 516-531, May, 2001.
- [63] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing, 2nd Edition*, Prentice Hall, 2002.
- [64] K. Jarrah, P. Muneesawang, I. Lee and L. Guan, "Miniminzing human-machine interactions in automatic image retrieval", *proc. canadian conference on Electrical and Computer engineering*, vol. 3, pp. 1589-1592, Niagra Falls, canada, May 2004.
- [65] D. S. Zhang and G. Lu, "A comparative study of three region shape descriptors", *DICTA 2002: Digital Image Computing techniques and Applications*, Melbourne, Australia, Jan. 2002.
- [66] Y. Rui, T. S. Huang, M. Ortega and S. Mehrotra, "Relevance Feedback: a power tool for interactive content-based image retrieval", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, pp. 644-655, Apr. 1998.
- [67] M. J. Swain and D. H. Ballard, "Indexing Via Color Histograms," *Porc. the 3rd Intena-tional Conference on Computer Vision*, pp. 11-32, Apr. 1991.

- [68] J. R. Smith and D. H. Chang, "Single Color Extraction and Image Query," *Proc. of International conference on Image processing(ICIP-95)*, Washington, USA, Oct. 1995.
- [69] J. R. Smith and D. H. Chang, "Tools and techniques for color image retrieval", in *IS&T/SPIE Proc. Storage & Retrieval for Image and Video databases IV*, vol. 2670, pp. 426-437, 1995.
- [70] M. Stricker and M. Orengo, "Similarity of color images", *Proc. IEEE International Conference on Image Processing*, vol. 2420, pp. 381-392, San Jose, USA, 1995.
- [71] M. J. Swain and D. H. Ballard, "Color indexing", *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.
- [72] W. Y. Ma and B. S. Manjunath, "A comparison of wavelet transform features for texture image annotation", *Proc. IEEE International Conference on Image Processing*, vol. 2, pp. 256-259, Washington, USA, Sep. 1995.
- [73] R. M. Haralick, K. Shaanmugam and I. Dinstein, "Texture features for image classification", *IEEE Trans. on Sys, Man, and Cyb*, vol. 3, no. 6, pp. 610-621, 1973.
- [74] C. C. Gotlieb and H. E. Kreyszig, "Texture descriptors based on co-occurrence matrices", *Proc. Computer Vision, Graphics and Image*, vol. 51, pp. 70-86, 1990.
- [75] J. R. Smith and S. F. Chang, "Automated binary texture feature sets for image retrieval", *Proc. IEEE International Conference on Accoustics,Speech and Signal Processing*, vol. 4, pp. 2239-2242, Atlanta, USA, 1996.
- [76] T. Chang and C. C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform", *IEEE Trans. on Image Processing*, vol. 2, no. 4, pp. 429-441, Oct. 1993.
- [77] L. Chen, G. Lu and D. S. Zhang, "Effects of differnet Gabor filters parameters on image retrieval by texture", *Proc. IEEE the 10th Intenational Multimedia Modeling Conference*, pp. 273-278, Jan. 2004

- [78] B. S. Manjunath and W. Y. Ma, "Texture feature for browsing and retrieval of image data", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 837-842, Aug. 1996.
- [79] S. Aksoy, R. M. haralick, F. A. Cheikh and M. Gabbouj, "A weighted distance approach to relevance feedback", *Proc. IEEE International Conference on Pattern Recognition*, vol. 4, pp: 812-815, Sep. 2000.
- [80] G. Salton and M. J. McGill, *Introction to Modern Information Retrieval*, NY, MaGram-Hill Book Company, 1983.
- [81] P. Muneesawang and L. Guan, "An interactive approach for CBIR using a network of radial basis functions", *IEEE Trans. on Multimedia*, vol. 6, pp. 703-716, 2004.
- [82] Corel Gallery Magic 65000 (1999) [online], Available: <http://www.corel.com/>.
- [83] P. Muneesawang and L. Guan, "Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture", *IEEE Trans. on Neural Network*, vol. 13, pp. 821-834, Jul. 2002.
- [84] C. Rother, V. Kolmogorov and A. Blake, "GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts", *ACM Trans. on Graphics (SIGGRAPH)*, vol. 23, pp. 309-314, Aug. 2004.
- [85] Y. Y. Boykov and M. P. Jolly, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images", *Proc. International Conference on Computer Vision*, vol. I, pp. 105-112, Vancouver, Canada, 2001.

Appendix A

List of Publications

- Yupeng Li and Ling Guan, “An Effective Shape Descriptor for the Retrieval of Natural Image Collections”, submitted to Canadian Conference on Electrical and Computer Engineering, Ottawa, May, 2006.
- Yupeng Li, Matthew J. Kyan and Ling Guan, “Improving Shape-based CBIR for Natural Image Content Using a Modified GFD”, Proc. the 2nd International Conference on Image Analysis and Recognition, pp. 593-600, Toronto, Canada, Sep. 2005.