

DS668204

STATISTICAL MODEL BASED DIGITAL IMAGE RETRIEVAL AND FRAGILE WATERMARKING

HUA YUAN

B.Sc

Shanghai, P.R.China, 1997

A thesis

Submitted in partial fulfillment
of the requirements for the degree of
Master of Applied Science
in the program of
Electrical and Computer Engineering

Department of Electrical and Computer Engineering
Ryerson University
Toronto, Ontario, Canada

©HUA YUAN 2004

PROPERTY OF
RYERSON UNIVERSITY LIBRARY

UMI Number: EC52989

All rights reserved

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[®]

UMI Microform EC52989

Copyright 2008 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

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]

Statistical Model Based Digital Image Retrieval And Fragile Watermarking

Master of Applied Science 2004

HUA YUAN

Electrical and Computer Engineering

Ryerson University

Abstract

The objective of this thesis is to acquire abstract image features through statistical modelling in the wavelet domain and then based on the extracted image features, develop an effective content-based image retrieval (CBIR) system and a fragile watermarking scheme.

In this thesis, we first present a statistical modelling of images in the wavelet domain through a Gaussian mixture model (GMM) and a generalized Gaussian mixture model (GGMM). An Expectation Maximization (EM) algorithm is developed to help estimate the model parameters. A novel similarity measure based on the Kullback-Leibler divergence is also developed to calculate the distance of two distinct model distributions. We then apply the statistical modelling to two application areas: image retrieval and fragile watermarking. In image retrieval, the model parameters are employed as image features to compose the indexing feature space, while the feature distance of two compared images is computed using the novel similarity measure. The new image retrieval method has a better retrieval performance than most conventional methods. In fragile watermarking, the model parameters are utilized for the watermark embedding. The new watermarking scheme achieves a virtually imperceptible embedding of watermarks because it modifies only a few image data and embeds watermarks at image texture edges. A multiscale embedding of fragile watermarks is given to enhance the embeddability rate and on the other hand, to constitute a semi-fragile approach.

Acknowledgment

I would like to acknowledge the support of everyone in the CASPAL Lab who have made my two years of graduate study at Ryerson University more pleasant and easier. The great support and the invaluable intellectual inspiration provided by Prof. Xiao-Ping Zhang, my supervisor, has helped in every aspect of my graduate studies and I have no doubt that the knowledge and the research skills that he shared with me will benefit my future studies or work.

I would like to thank Prof. Ling Guan, Songnian Li and Xavier Fernando for being the member of my thesis defense committee. I appreciate very much their time on reviewing my thesis draft and the valuable comments and suggestions that they provided on the thesis draft.

I would like to take this opportunity to thank Prof. Ling Guan for his interesting course taught in my first year of study. Discussions with him broadened my understanding in content-based image retrieval and stimulated my research in this area, which has become an important part in this thesis.

I am very grateful to the department of Electrical and Computer Engineering for the funding in my graduate studies. I am also grateful to the Ryerson University for providing me a small, but very pleasant and cozy campus environment.

I express the warmest appreciation to my parents and my wife for their caring and support all the time. They are and will always be the driving force that helps me accomplish the graduate study and all future objectives.

Contents

1	Introduction	1
1.1	Motivation and Objectives	1
1.2	Background	2
1.2.1	Image Retrieval	3
1.2.2	Image Watermarking	6
1.3	Contribution	9
1.4	Structure of Thesis	11
2	Preliminaries	12
2.1	The EM Algorithm for Mixture Models	12
2.1.1	Maximum Likelihood	12
2.1.2	The General EM Algorithm	13
2.1.3	The Estimation of Gaussian Mixture Density Parameters Using EM .	15
2.2	The Kullback-Leibler Divergence	19
3	Statistical Modelling in the Wavelet Domain	21
3.1	New Statistical Models and Related EM Algorithms	21
3.1.1	A Gaussian Mixture Model in the Wavelet Domain	21
3.1.2	An EM Algorithm for the Gaussian Mixture Model	22
3.1.3	A Generalized Gaussian Mixture Model	24
3.2	A Kullback-Leibler Divergence Based Similarity Measure for the Statistical Models	27

3.2.1	General Minkowski Distances	27
3.2.2	Similarity Measure Based on the Kullback-Leibler Divergence	28
3.2.3	Analysis of the New Similarity Measure	29
4	Statistical Model Based Image Retrieval	33
4.1	Overview of the Proposed CBIR System	33
4.2	The Indexing Feature Space for Image Retrieval	34
4.3	Kullback Divergence Based Similarity Measure	35
4.4	Simulation Results	36
4.5	Summary	46
5	Statistical Model Based Fragile Watermarking	47
5.1	Embedding Information into the Statistical Model	47
5.2	Multiscale Embedding of Authentication Messages	52
5.3	Simulation Results	54
5.4	Summary	60
6	Conclusions and Future Work	61
A	The Newton-Raphson Method	65
Vita		71

List of Figures

1.1	General Architecture of A Fragile Watermarking System.	8
4.1	The Proposed CBIR System.	34
4.2	Initial Retrieval of A Query from Class D87.	41
4.3	Second Round Retrieval of the Same Query from Class D87.	41
4.4	Initial Retrieval of A Query from Class D98.	42
4.5	Second Round Retrieval of the Same Query from Class D98.	42
4.6	Initial Retrieval of A Query from Class D111.	43
4.7	Second Round Retrieval of the Same Query from Class D111.	43
4.8	Initial Retrieval of A Query from Class D114.	44
4.9	Second Round Retrieval of the Same Query from Class D114.	44
4.10	Initial Retrieval of A Query Image Using Euclidean Distance.	45
4.11	Initial Retrieval of Same Query Image Using Separate Kullback Distance. . .	45
5.1	An Iterative Approach for Coefficient Modification.	51
5.2	Multiscale Embedding of Authentication Messages.	52
5.3	The Original Lena Image (a) and the Watermarked Lena Image (b).	55
5.4	The HL Wavelet Subspace of Original Lena Image (a) and the HL Wavelet Subspace of Watermarked Lena Image (b).	55
5.5	Message Bits Embedded into the Wavelet Blocks.	56
5.6	The Original Peppers Image (a) and the Watermarked Peppers Image (b) with Lab Logo "CASPAL" Embedded.	57

5.7 The DWT of Original Peppers Image (a) and the DWT of Watermarked Peppers Image (b) with Lab Logo “CASPAL” Embedded.	58
---	----

List of Tables

4.1	Initial Retrieval Rate Based on Db1 And Db2 Filters.	37
4.2	Comparison of Different Decomposition Scales.	37
4.3	Retrieval Results of 4 Different Queries.	38
4.4	Initial Retrieval Rate Based on Similarity Measures.	38
4.5	GMM Compared with GGMM.	39
4.6	GMM & Kullback Compared with Other Traditional Methods.	40
5.1	Code Map for Message Bits Embedding.	53
5.2	Average Parameter Difference Caused by Single Pixel Modification.	58
5.3	Parameter Difference Caused by Noise.	59
5.4	Parameter Difference Caused by Compression.	59
5.5	Parameter Difference Caused by Some Malicious Attacks.	60

Chapter 1

Introduction

1.1 Motivation and Objectives

Many image applications require an accurate modelling of images to have a better understanding and utilization of image contents. Image modelling is aimed at the exploitation of statistical characteristics of the images. The success of some traditional approaches that use mean and variance to interpret image data implies a more complicated Gaussian statistical model, which also has a mean and a variance parameter, would be more effective and natural for image modelling. The primary objective of the thesis is to interpret an image through a Gaussian statistical modelling in the wavelet domain, realizing the fact that the wavelet transform has a multiscale image decomposition and the statistical approach provides an accurate yet concise interpretation of the decomposed image data. The image features obtained from the statistical modelling are applied in two application areas: image retrieval and fragile watermarking.

The statistical model based image retrieval has many attractive features and advantages. The image retrieval is based on the comparison of image features, which represent image statistical characteristics and can therefore be better obtained through the statistical modelling. Moreover, a statistical model based feature extraction is efficient in feature representation and computation because it results in a relatively compact indexing feature space. Thus, our interest lies in the derivation of some new feature extraction methods based on the developed statistical modelling to have a compact indexing feature space. It is also desired that the

extracted image features can reflect some characteristics of human perception and have a sound image retrieval performance.

We are also interested in utilizing the statistical modelling to develop some new fragile watermarking methods. Most traditional fragile watermarking methods inevitably need to modify a large amount of image data for watermark embedding in order to protect the whole image area. They are not efficient and in conflict with the watermarking principle that requires an imperceptible embedding of watermarks. Using the statistical approach, we aim at developing a new fragile watermarking method that modifies a few image data for watermark embedding, while keeping sensitive to any image tampering. Besides, the increasing interest nowadays for a semi-fragile watermark is what we wish to address in our new method.

1.2 Background

With the development of the Internet and digital storage techniques, digital image libraries have been widely used for commercial and research purposes. People want to have an efficient way to browse the library. At the same time, they may require that the information they acquire are secure and authorized to use, in other words, they are authenticated.

Image retrieval provides an effective way to search and browse the image libraries based on their indexing features. Depending on the characteristics of queries, the image retrieval progressed from original text based retrieval to current more effective content based retrieval. The content based image retrieval (CBIR) usually uses color, texture, and shape in the work toward identifying suitable image features. The performance of a CBIR system is directly related to the effectiveness and compactness of the indexing feature space, which is studied under a statistical approach in this thesis.

Image watermarking, on the other hand, provides a solid solution for data hiding and image authentication. The digitized images distributed on the Internet are easy to be manipulated and tampered. In applications people want to make sure their retrieved images are true copy of the author's and can be legally used, the authentication techniques

are required. Some watermarks, after being embedded into a host image, can detect any unauthorized modifications on that image. Therefore they can be used for authentication purposes. These watermarks, commonly known as fragile watermarks, is one focus of the research in this thesis. Traditional methods usually modify a lot of image data and seldom consider human perception characteristics while embedding watermarks. Therefore there is a need for an efficient fragile watermarking scheme using a statistical approach.

1.2.1 Image Retrieval

As more and more images are being captured and digitally stored into a database, an effective image retrieval system is required to make use of the information stored in images. In traditional databases, the approach to indexing and retrieval of images is based on simple attributes such as image number and text description. The performance of this text-based indexing and retrieval technique is quite low because the simple text attributes are not able to describe the features of the images completely and accurately. Besides, text-based retrieval systems also cannot accept content-based queries. To overcome these limitations of text-based indexing and retrieval techniques, content-based indexing and retrieval techniques are pursued.

The earliest use of the term CBIR in literature was by Kato [1992], to describe his experiments into an automatic retrieval of images from a database depending on colour and shape features. The term has since been widely used to describe the process of retrieving desired images from a large collection on the basis of features (such as colour, texture and shape) that can be automatically extracted from the images themselves. CBIR differs from classical information retrieval in that image databases are essentially unstructured, since digitized images consist of arrays of pixel intensities that have no inherent meaning. Therefore One important task is to extract useful information from the image raw data for indexing before any kind of retrieval operation on the image library is possible. The indexing and retrieval in CBIR are defined as follows:

- ◆ Indexing - the computer-assisted data reduction of images into mathematical features;

Indexing may be subdivided into the steps of:

- ◇ Segmentation - the determination of the boundaries of the objects of interest;
 - ◇ Feature extraction - the obtaining of the colour, shape and texture properties of the objects of interest;
 - ◇ Feature vector organization - the organization of the feature vectors in the database for the purpose of searching efficiently;
 - ◇ Classification - the labelling of segmented objects into categories of interest.
- ◆ Retrieval - the user interaction to retrieve desired images from the database; this may include:
- ◇ Query formulation - the method used to specify the query; Most common methods are queries by image examples or queries by sketches;
 - ◇ Query feature extraction - the reduction of the query image (or sketch) into a feature vector compatible with those stored in the database;
 - ◇ Similarity measure - the method used to compare the query with each stored image and to measure the similarity between them.

With the development of Internet and mass media storage techniques, CBIR has been very active in assisting people to browse online digital libraries since 1990s. Some well-known CBIR systems include IBM's QBIC [1], Virage [2], VisualSEEk [3] and Photobook [4]. Being compliant with the above mentioned indexing and retrieval characteristics, the core of a typical CBIR system is mainly composed of feature extraction techniques, similarity measure approaches, and relevance feedback mechanisms.

One of our research objectives focuses on image texture features. They, along with color features and shape features, are part of image low level features that are always the focus of research in image feature extraction. A lot of texture feature extraction tools and methods have been developed. Some researchers used the codebook concept to compose

the indexing feature space, in which the images are divided into blocks and each block is coded through feature extraction and be part of the codebook. Other researchers noticed the multiresolution benefits of image transforms in the compressed domain and developed feature extraction methods employing Gabor filters [5], the pyramid-structured wavelet transform (PWT) and the tree-structured wavelet transform (TWT) [6]. In the TWT based method, a wavelet filter and a scaling filter are used to decompose an image into four subbands and these subbands can be further decomposed to form a tree structure. Energy values of the subbands are calculated and used to extract the texture features. In the PWT based method, only the scaling subband is decomposed by the filters at each scale. In the Gabor filter based method, an image is decomposed by the Gabor transform and the mean and the variance of the coefficients are used to represent image texture features.

Based on the extracted image features, similarity measure computes the distance between the query image and each image in the database so that the top matched images to the query can be retrieved. The similarity measure used to rank the images has direct impact on the retrieval performance. One commonly used approach for similarity measure in many retrieval systems is the norm-based distance between two feature vectors, such as the City-block distance and the Euclidean distance. However, it is far from optimal since the extracted image features are different from each other in global variances. Therefore they are not to be compared on a same scale if the simple Euclidean distance is adopted for similarity measure. In [7], a Kullback divergence similarity measure is proposed to compute the distance of features extracted by a generalized Gaussian density (GGD) model in the wavelet domain. The Kullback divergence approach addresses the above problem by comparing all different features within a same framework of the probability density function (PDF). While the Kullback divergence is more accurate and effective than norm-based distances with regard to the statistical model, the GGD is not optimal for the marginal distribution of wavelet coefficients. With more sophisticated and accurate models applied to the wavelet coefficients, their Kullback divergence forms may become very complicated and their computational complexity may arise.

The features we extract from images such as colors, textures or shapes are often low-level features because most of them are extracted directly from digital representations of images in the database and have little or nothing to do with the human perception. On the other hand, the similarity between two images can be very high-level, or semantic. This requires the system to measure the similarity in a way human being would perceive or recognize. Thus, the gap between low-level features and high-level semantic meanings of the images has been the major obstacle to a better retrieval performance. Various approaches have been proposed to improve the accuracy of CBIR systems. Essentially, these approaches fall into two main categories: to improve the features and to improve the learning of CBIR systems, so that more high level semantics of human perception can be reflected. Researchers have tried many features that are believed to be related with human perception, and they are still working on finding more. On the other hand, when the feature set is fixed, many algorithms have been proposed to measure the similarity in a way human beings might take. Besides, some interactive mechanisms [8][9][10] that involve human as part of the retrieval process, which are called relevance feedback, are introduced into CBIR systems.

1.2.2 Image Watermarking

Usage of digital images has experienced a tremendous growth during the last decade because it has some notable benefits in efficient storage, ease of manipulation and transmission. While they are getting into people's life more and more nowadays, digital images are important commercial assets that need to be efficiently managed, transferred and protected.

Unfortunately, the nature of the digital images makes the work of pirates and tampering easier, since the digital contents are very easy to be reproduced or modified. It has become an especially severe issue with the rapid development of the Internet in recent years. Every year there is a huge loss for those intelligent companies who are owners of issued products. Even worse, the fear of piracy has made digital content creators and distributors hesitant in releasing their assets in the public. For example, it is known that sometimes the new movies can be downloaded from the internet even before they are projected in the cinemas.

Therefore it is generally difficult for product owners or rights holders to control and manage the usage of their works and at the same time, keep them from being improperly used by others.

One of the solutions that addresses the annoying problem is the watermarking process, where information can be hidden into the "essence" of the multimedia object for protection, with condition to be either robust or fragile depending on real applications. The embedded watermarks can be either visible or invisible. However, under most circumstances, the invisible watermarks are preferred because they don't change the visual appearance of an image and they are unseen to prevent potential attacks. The existence of such an invisible watermark can be determined only through a watermark extraction or detection algorithm. Different applications of invisible watermarks divide them into two categories: Fragile Watermark and Robust Watermark, with each term defined as follows:

- ◆ **Fragile Watermark:** A watermark, which is destroyed when the image is manipulated digitally. Such a watermark is useful in proving authenticity of an image or verifying integrity of image content, just as related to the prior example of retrieved images. If the watermark is still intact, then the image has not been modified and can be regarded as an authenticated one. If the watermark has been destroyed, then the image has been tampered with certain image operations such as compression, noising or malicious attacks.
- ◆ **Robust Watermark:** A watermark, which is very resistant to destruction under any image manipulation. This is useful in verifying ownership of an image that is suspected of misappropriation. Digital detection of the watermark would indicate the source of the image.

Our interest of watermark is on the fragile aspect, with an intention to prevent copyright violation and content tampering by designing some authentication techniques. The robust watermark, which features the ability to survive substantial image operations for data hiding or ownership detection purposes, is out of the scope of this thesis and will not be discussed.

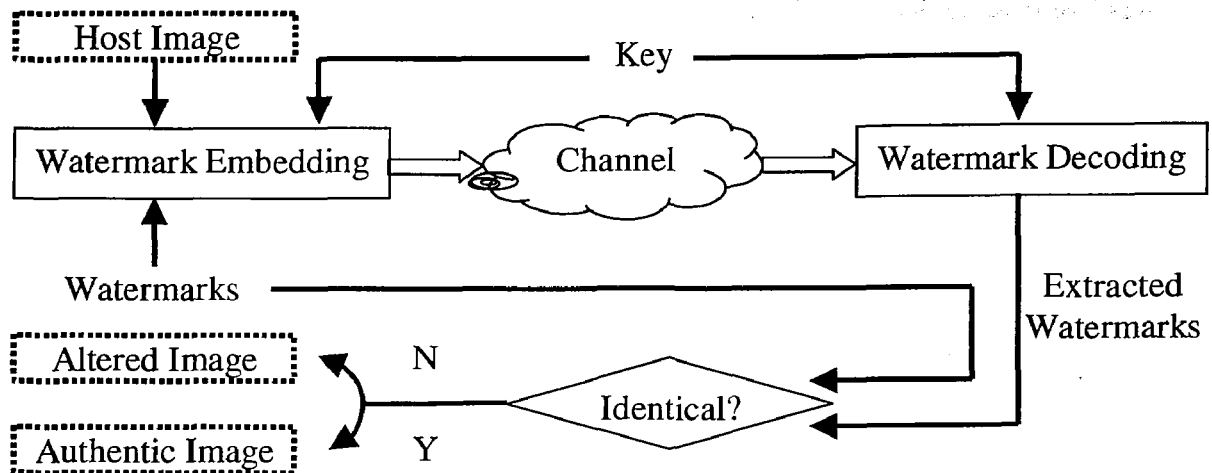


Figure 1.1: General Architecture of A Fragile Watermarking System.

Figure 1.1 displays the general architecture of a fragile watermarking system. There are two major parts in the system: watermark embedding and watermark decoding. In the embedding process, the watermarks are hidden into the host image on a public or private key basis through some watermarking techniques. The embedded watermark can be a number, a text, or even an image and is usually represented by a binary data sequence. Then the watermarked image is transmitted through certain channels to the reception end and experiencing all kinds of distortions or attacks during this phase. At the reception end, the estimation of original watermarks is extracted from the watermarked image and compared with the original ones. If there are any subtle changes induced by compression, noise or attacks during the transmission phase, the extracted watermarks will not be identical with original ones and as a consequence of that, the image obtained will be claimed as not authentic.

There has been a lot of dedications in the search of an efficient fragile watermarking method. The fragile watermarks can be embedded in either the space domain or the compressed domain of an image. With the focus in the space domain, several fragile watermarking methods that utilize the least significant bit (LSB) of the image data were developed. For examples, a technique that inserts a checksum determined by the 7 most significant bits into the LSBs of selected pixels was proposed in [11]. A technique which embeds a digital

signature of the most significant bits of a block of an image into the least significant bits of the same block on a secret user key basis was developed by Wong in [12]. And later he extended his research on a public key scheme [13]. With the focus in the compressed domain, a wavelet-based fragile watermarking method that allows spatial and frequency localization of image tampering is proposed in [14]. Some other researchers noted the constraints of a single fragile watermark and developed a hybrid authentication watermark consisting of a fragile watermark and a robust watermark [15]. The above stated methods are all successful in an imperceptible embedding of watermarks and the detection of potential image content tampering. However, they need to change a large amount of image data to embed watermarks, which is not quite efficient and may reduce the quality of the watermarked image. Their approaches also don't facilitate the work toward a semi-fragile watermark, which is highly favorable in recent applications of fragile watermarks.

1.3 Contribution

In this thesis, we obtain abstract image features by developing a statistical modelling in the wavelet domain. Based on the extracted image features, an image retrieval system and a fragile watermarking method are presented.

The statistical modelling uses mixed Gaussian or generalized Gaussian components to describe the distribution of wavelet coefficients. The modelling has three major contributions as follows:

1. A Gaussian mixture model (GMM) and a generalized Gaussian mixture model (GGMM) are developed for image modelling in the wavelet domain;
2. Two expectation maximization (EM) algorithms are developed to help estimate the model parameters of GMM and GGMM, respectively;
3. A novel similarity measure based on the Kullback-Leibler divergence for GMM and GGMM is developed.

Based on the statistical modelling, a new image retrieval system is presented. It has following characteristics and contributions:

1. GMM or GGMM parameters are used as image feature descriptors and are compliant with certain characteristics of human vision;
2. The indexing feature space is compact and has a multiresolution representation of image texture contents;
3. The novel similarity measure based on the Kullback-Leibler divergence is applied in the distance computation of two feature vectors and proved to be effective in image retrieval.

Based on the statistical modelling, a novel fragile watermarking method that embeds watermarks at multiple wavelet scales is developed. The presented method has four attractive features and contributions:

1. The watermark embedding process modifies only a few image data and conforms to human vision characteristics;
2. The new method constitutes a semi-fragile watermark approach, in which unauthorized changes made by some normal image operations such as compression can be distinguished from those caused by malicious attacks;
3. The new method can embed personal authentication information such as signatures or logos into the host image;
4. The new method is able to not only detect but also localize any slight image tampering.

1.4 Structure of Thesis

In Chapter 2, we introduce some background knowledge that we will employ in later chapters. The EM algorithm is used to estimate statistical model parameters depending on the training data provided. The Kullback divergence is used to measure the similarity of two distribution functions.

In Chapter 3, we develop a statistical modelling of images in the wavelet domain through GMM and GGMM. Some EM algorithms are developed to help estimate the statistical model parameters. A novel similarity measure based on the Kullback divergence for the statistical models is presented.

In Chapter 4, we present a CBIR system that incorporates GMM based image feature extraction and Kullback divergence based similarity measure. GMM parameters from all decomposed wavelet subspaces are employed to compose the indexing feature space. The similarity measure of two indexing feature vectors is based on the novel Kullback divergence approach.

In Chapter 5, we present a new fragile watermarking method. Authentication information is embedded into the statistical model through manipulations of GMM parameters. Later on, a multiscale embedding of fragile watermarks is presented.

In Chapter 6, we conclude the thesis and give some prospects of the future work.

Chapter 2

Preliminaries

In this chapter, we introduce some background knowledge that will be employed later in this thesis. The EM algorithm for mixture models and the Kullback-Leibler divergence are briefly summarized here.

2.1 The EM Algorithm for Mixture Models

We describe the maximum-likelihood parameter estimation problem and how the expectation-maximization (EM) algorithm can be used for its solution [16]. We first present an abstract form of the EM algorithm as it is often given in the literature. We then develop an EM parameter estimation procedure to find the parameters of a mixture of Gaussian densities.

2.1.1 Maximum Likelihood

Recall the definition of the maximum-likelihood estimation problem. We have a density function $p(x|\Theta)$ that is governed by the set of parameters Θ (e.g., p might be a set of Gaussians and Θ could be the means and covariances). We also have a data set of size N , supposedly drawn from this distribution, i.e., $\mathbf{X} = \{x_1, \dots, x_N\}$. We assume that these data vectors are independent and identically distributed with distribution p . Therefore, the resulting density for the samples is:

$$p(\mathbf{X}|\Theta) = \prod_{i=1}^N p(x_i|\Theta) = L(\Theta|\mathbf{X}). \quad (2.1)$$

This function $L(\Theta|\mathbf{X})$ is called the likelihood of the parameters given the data, or just the likelihood function. The likelihood is thought of as a function of the parameters Θ where the data \mathbf{X} is fixed. In the maximum likelihood problem, our goal is to find the Θ that maximizes L . That is, we wish to find Θ^* where

$$\Theta^* = \arg \max_{\Theta} L(\Theta|\mathbf{X}). \quad (2.2)$$

Often we maximize $\log(L(\Theta|\mathbf{X}))$ instead because it is analytically easier.

Depending on the form of $p(x|\Theta)$ this problem can be easy or hard. For example, if $p(x|\Theta)$ is simply a single Gaussian distribution where $\Theta = (\mu, \sigma^2)$, then we can set the derivative of $\log(L(\Theta|\mathbf{X}))$ to zero, and solve directly for μ and σ^2 (this, in fact, results in the standard formulas for the mean and variance of a data set). For many problems, however, it is not possible to find such analytical expressions, and we must resort to more elaborate techniques.

2.1.2 The General EM Algorithm

The EM algorithm is one such elaborate technique. The EM algorithm [17] [18] [19] is a general method of finding the maximum likelihood estimate of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values. The EM algorithm is applied in applications where optimizing the likelihood function is analytically intractable but the likelihood function can be simplified by assuming the existence of (and hence values for) some additional missing (or hidden) parameters.

As before, we assume that data \mathbf{X} is observed and is generated by some distribution. We call \mathbf{X} *the incomplete data*. We assume that a complete data set exists $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ and specify a joint density function:

$$p(\mathbf{z}|\Theta) = p(x, y|\Theta) = p(y|x, \Theta)p(x|\Theta). \quad (2.3)$$

Where does this joint density come from? Often it “arises” from the marginal density function $p(x|\Theta)$ and the assumption of hidden variables and parameter value guesses (just

like in the case of Gaussian mixture). In other cases (e.g., missing data values in samples of a distribution), we also must assume a joint relationship between the missing and observed values.

With this new density function, we can define a new likelihood function, $L(\Theta|\mathbf{Z}) = L(\Theta|\mathbf{X}, \mathbf{Y}) = p(\mathbf{X}, \mathbf{Y}|\Theta)$, called the complete-data likelihood. Note that this function is in fact a random variable since the missing information \mathbf{Y} is unknown, random, and presumably governed by an underlying distribution. That is, we can think of $L(\Theta|\mathbf{X}, \mathbf{Y}) = h_{\mathbf{X}, \Theta}(\mathbf{Y})$ for some function $h_{\mathbf{X}, \Theta}(\cdot)$, where \mathbf{X} and Θ are constant and \mathbf{Y} is a random variable. The original likelihood $L(\Theta|\mathbf{X})$ is referred to as the incomplete-data likelihood function.

The EM algorithm first finds the expected value of the complete-data log-likelihood $\log p(\mathbf{X}, \mathbf{Y}|\Theta)$ with respect to the unknown data \mathbf{Y} given the observed data \mathbf{X} and the current parameter estimates. That is, we define:

$$Q(\Theta, \Theta^{(i-1)}) = E[\log p(\mathbf{X}, \mathbf{Y}|\Theta) | \mathbf{X}, \Theta^{(i-1)}], \quad (2.4)$$

where $\Theta^{(i-1)}$ are the current parameter estimates that we use to evaluate the expectation and Θ are the new parameters that we need to optimize to increase the value of Q .

The key thing to understand (2.4) is that \mathbf{X} and $\Theta^{(i-1)}$ are constants, while Θ is a normal variable we wish to adjust and \mathbf{Y} is a random variable governed by the distribution $f(y|\mathbf{X}, \Theta^{(i-1)})$. Therefore, the right side of (2.4) can be rewritten as:

$$E[\log p(\mathbf{X}, \mathbf{Y}|\Theta) | \mathbf{X}, \Theta^{(i-1)}] = \int_{y \in \Upsilon} \log p(\mathbf{X}, y|\Theta) f(y|\mathbf{X}, \Theta^{(i-1)}) dy. \quad (2.5)$$

Note that $f(y|\mathbf{X}, \Theta^{(i-1)})$ is the marginal distribution of the unobserved data and is dependent on both the observed data \mathbf{X} and on the current parameters $\Theta^{(i-1)}$, and Υ is the space y belongs to. In the best of cases, this marginal distribution is a simple analytical expression of the assumed parameters $\Theta^{(i-1)}$ and perhaps the data. In the worst of cases, this density might be very hard to obtain.

The evaluation of this expectation is called the E-step of the EM algorithm. Notice the meaning of the two arguments in the function $Q(\Theta, \Theta')$. The first argument Θ corresponds

to the parameters that ultimately will be optimized in an attempt to maximize the likelihood. The second argument Θ' corresponds to the parameters that we currently use to evaluate the expectation.

The second step (the M-step) of the EM algorithm is to maximize the expectation we computed in the first step. That is, we find:

$$\Theta^{(i)} = \arg \max_{\Theta} Q(\Theta, \Theta^{(i-1)}). \quad (2.6)$$

These two steps are repeated as necessary. Each iteration is guaranteed to increase the log-likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function.

As presented above, there is not a fixed form to code the algorithm. This is the way, however, that the algorithm is presented in its most general form. The details of the steps required to compute the given quantities are very dependent on the particular application, so they are not discussed here when the algorithm is presented in this general form.

2.1.3 The Estimation of Gaussian Mixture Density Parameters Using EM

The mixture density parameter estimation problem is probably one of the most widely used applications of the EM algorithm in the computational pattern recognition community. In this case, we assume the following probabilistic model:

$$p(x|\Theta) = \sum_{i=1}^M \alpha_i p_i(x|\theta_i), \quad (2.7)$$

where the parameters are $\Theta = (\alpha_1, \dots, \alpha_M, \theta_1, \dots, \theta_M)$ such that $\sum_{i=1}^M \alpha_i = 1$ and each p_i is a density function parameterized by θ_i . In other words, we assume we have M component densities mixed together with M mixing coefficients α_i .

The incomplete-data log-likelihood expression for this density from data \mathbf{X} is given by:

$$\log(L(\Theta|\mathbf{X})) = \log \prod_{i=1}^N p(x_i|\Theta) = \sum_{i=1}^N \log \left(\sum_{j=1}^M \alpha_j p_j(x_i|\theta_j) \right), \quad (2.8)$$

which is difficult to optimize because it contains the log of the sum. However, if we consider \mathbf{X} as incomplete and assume the existence of unobserved data items $\mathbf{Y} = \{y_i\}_{i=1}^N$ whose values inform us which component density “generated” each data item, the likelihood expression can be significantly simplified. That is, we assume that $y_i \in 1, \dots, M$ for each i , and $y_i = k$ if the i^{th} sample was generated by the k^{th} mixture component. If we know the values of \mathbf{Y} , the likelihood becomes:

$$\log(L(\Theta|\mathbf{X}, \mathbf{Y})) = \log(p(\mathbf{X}, \mathbf{Y}|\Theta)) = \sum_{i=1}^N \log(p(x_i|y_i)P(y_i)) = \sum_{i=1}^N \log(\alpha_{y_i} p_{y_i}(x_i|\theta_{y_i})), \quad (2.9)$$

which, given a particular form of the component densities, can be optimized using a variety of techniques.

The problem, of course, is that we do not know the values of \mathbf{Y} . If we assume \mathbf{Y} is a random vector, however, we can proceed.

We must first derive an expression for the distribution of the unobserved data. Let's first guess at parameters for the mixture density, i.e., we guess that $\Theta^g = (\alpha_1^g, \dots, \alpha_M^g, \theta_1^g, \dots, \theta_M^g)$ are the appropriate parameters for the likelihood $L(\Theta^g|\mathbf{X}, \mathbf{Y})$. Given Θ^g , we can easily compute $p_j(x_i|\theta_j^g)$ for each i and j . In addition, the mixing parameters α_j can be thought of as prior probabilities of each mixture component. Therefore, using Bayes rule, we can compute:

$$p(y_i|x_i, \Theta^g) = \frac{\alpha_{y_i}^g p_{y_i}(x_i|\theta_{y_i}^g)}{p(x_i|\Theta^g)} = \frac{\alpha_{y_i}^g p_{y_i}(x_i|\theta_{y_i}^g)}{\sum_{k=1}^M \alpha_k^g p_k(x_i|\theta_k^g)} \quad (2.10)$$

and

$$p(\mathbf{y}|\mathbf{X}, \Theta^g) = \prod_{i=1}^N p(y_i|x_i, \Theta^g), \quad (2.11)$$

where $\mathbf{y} = (y_1, \dots, y_N)$ is an instance of the unobserved data independently drawn. When we now look at (2.5), we see that in this case we have obtained the desired marginal density by

assuming the existence of the hidden variables and making a guess at the initial parameters of their distribution.

In this case, (2.4) takes the form:

$$\begin{aligned}
Q(\Theta, \Theta^g) &= \sum_{\mathbf{y} \in \Upsilon} \log(L(\Theta|\mathbf{X}, \mathbf{y})) p(\mathbf{y}|\mathbf{X}, \Theta^g) \\
&= \sum_{\mathbf{y} \in \Upsilon} \sum_{i=1}^N \log(\alpha_{y_i} p_{y_i}(x_i|\theta_{y_i})) \prod_{j=1}^N p(y_j|x_j, \Theta^g) \\
&= \sum_{l=1}^M \sum_{i=1}^N \log(\alpha_l p_l(x_i|\theta_l)) \sum_{y_1=1}^M \sum_{y_2=1}^M \cdots \sum_{y_N=1}^M \delta_{l, y_i} \prod_{j=1}^N p(y_j|x_j, \Theta^g). \quad (2.12)
\end{aligned}$$

After some manipulations, (2.12) can be evolved into:

$$\begin{aligned}
Q(\Theta, \Theta^g) &= \sum_{l=1}^M \sum_{i=1}^N \log(\alpha_l p_l(x_i|\theta_l)) p(l|x_i, \Theta^g) \\
&= \sum_{l=1}^M \sum_{i=1}^N \log(\alpha_l) p(l|x_i, \Theta^g) + \sum_{l=1}^M \sum_{i=1}^N \log(p_l(x_i|\theta_l)) p(l|x_i, \Theta^g). \quad (2.13)
\end{aligned}$$

To maximize this expression, we can maximize the term containing α_l and the term containing θ_l independently since they are not related.

To find the expression for α_l , we introduce the Lagrange multiplier λ with the constraint that $\sum_l \alpha_l = 1$, and solve the following equation:

$$\frac{\partial}{\partial \alpha_l} \left[\sum_{l=1}^M \sum_{i=1}^N \log(\alpha_l) p(l|x_i, \Theta^g) + \lambda \left(\sum_l \alpha_l - 1 \right) \right] = 0. \quad (2.14)$$

As a result, we obtain:

$$\alpha_l = \frac{1}{N} \sum_{i=1}^N p(l|x_i, \Theta^g). \quad (2.15)$$

To find an analytical expression for θ_l , we must first assume a distribution function for the component density. For example, if we assume a Gaussian distribution with mean μ_l and variance σ_l^2 for each component l :

$$p_l(x|\mu_l, \sigma_l^2) = \frac{1}{\sqrt{2\pi}\sigma_l} e^{-\frac{(x-\mu_l)^2}{2\sigma_l^2}}, \quad (2.16)$$

then the distribution parameters $[\mu_l, \sigma_l^2]$ can be derived through following procedures.

Taking the log of (2.16), ignoring any constant terms (since they disappear after taking derivatives), and substituting into the right part of (3.8), we get:

$$\sum_{l=1}^M \sum_{i=1}^N \log(p_l(x_i|\mu_l, \sigma_l^2)) p(l|x_i, \Theta^g) = \sum_{l=1}^M \sum_{i=1}^N \left(-\log \sigma_l - \frac{(x_i - \mu_l)^2}{2\sigma_l^2} \right) p(l|x_i, \Theta^g). \quad (2.17)$$

Taking the derivative of (3.12) with respect to μ_l and setting it to zero, we get:

$$\sum_{i=1}^N \sigma_l^{-2} (x_i - \mu_l) p(l|x_i, \Theta^g) = 0, \quad (2.18)$$

with which we can easily solve for μ_l :

$$\mu_l = \frac{\sum_{i=1}^N x_i p(l|x_i, \Theta^g)}{\sum_{i=1}^N p(l|x_i, \Theta^g)}. \quad (2.19)$$

To find σ_l^2 , we take the derivative of (3.12) with respect to σ_l and setting it to zero:

$$\sum_{i=1}^N \left(-\frac{1}{\sigma_l} + \frac{(x_i - \mu_l)^2}{\sigma_l^3} \right) p(l|x_i, \Theta^g) = 0, \quad (2.20)$$

with which we can solve for σ_l^2 :

$$\sigma_l^2 = \frac{\sum_{i=1}^N (x_i - \mu_l)^2 p(l|x_i, \Theta^g)}{\sum_{i=1}^N p(l|x_i, \Theta^g)}. \quad (2.21)$$

Summarizing, the estimates of the new parameters Θ in terms of the old parameters Θ^g are as follows:

$$\alpha_l = \frac{1}{N} \sum_{i=1}^N p(l|x_i, \Theta^g), \quad (2.22)$$

$$\mu_l = \frac{\sum_{i=1}^N x_i p(l|x_i, \Theta^g)}{\sum_{i=1}^N p(l|x_i, \Theta^g)}, \quad (2.23)$$

$$\sigma_l^2 = \frac{\sum_{i=1}^N (x_i - \mu_l)^2 p(l|x_i, \Theta^g)}{\sum_{i=1}^N p(l|x_i, \Theta^g)}. \quad (2.24)$$

2.2 The Kullback-Leibler Divergence

First, let us have a definition of the *entropy* of a distribution. Assuming we having a distribution $p = \{p(x) : x \in \mathbf{X}\}$, the *entropy* of p , say $H(p)$, is defined as:

$$H(p) = - \sum_{x \in X} p(x) \log p(x) = \mathbf{E}_p[-\log p(\mathbf{X})]. \quad (2.25)$$

The entropy represents the amount of energy in a system that is described by the distribution function p .

Based on the concept of entropy, Kullback and Leibler (1951) introduced a measure of information associated with two probability distributions of a discrete random variable [20]. Assuming the two probability distributions are represented by p and q respectively, the measure (or the discrimination function) of the two distributions is defined as:

$$D(p, q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} = \mathbf{E}_p \left[\log \frac{p(\mathbf{X})}{q(\mathbf{X})} \right]. \quad (2.26)$$

This measure is called the *Kullback-Leibler Divergence*, also called *relative information*, *directed divergence*, *cross entropy* by different authors. It is effective in evaluating the distance of two discrete distributions from a statistical point of view.

Following two properties make the Kullback-Leibler divergence an appropriate choice for distance evaluation:

Property 1 (Nonnegativity). $D(p, q) \geq 0$, with equality if $p = q$.

Property 2 (Additivity). $D(p_1 * p_2, q_1 * q_2) = D(p_1, q_1) + D(p_2, q_2)$,

where $p_1, q_1 \in \Delta_n, p_2, q_2 \in \Delta_m$.

Although very effective in distance evaluation, (2.26) is not symmetric in p and q . Its symmetric version, which is known as J-divergence, is given by:

$$J(p, q) = D(p, q) + D(q, p) = \sum_{x \in X} (p(x) - q(x)) \log \frac{p(x)}{q(x)}. \quad (2.27)$$

If x is a continuous random variable instead, assuming there are two probability density functions $p(x)$ and $q(x)$ with respect to x , then the Kullback-Leibler divergence between them [21] is defined as:

$$D(p, q) = \int_{\mathbf{R}} p(x) \log \frac{p(x)}{q(x)} dx. \quad (2.28)$$

Chapter 3

Statistical Modelling in the Wavelet Domain

In this chapter, we present a statistical model containing multiple Gaussian components to describe the wavelet coefficients. An EM algorithm is developed to help estimate the statistical model parameters. Based on the statistical modelling, a novel similarity measure of two distinct model distributions is developed using the Kullback-Leibler divergence.

3.1 New Statistical Models and Related EM Algorithms

3.1.1 A Gaussian Mixture Model in the Wavelet Domain

The 2-D wavelet transform is known to decompose an image into many wavelet subspaces at different scales. In each decomposed wavelet subspace, the image information is carried in the wavelet coefficients. Therefore an appropriate description of these coefficients can help us with a better understanding of the image and a better utilization of the coefficients in some applications.

The Gaussian mixture model has an accurate description of the wavelet coefficients. Since the wavelet coefficients have a peaky, heavy-tailed marginal distribution [22] and a near zero mean, their probability density function (PDF) can be well expressed through a multi-state Gaussian mixture:

$$\begin{cases} p(w_i) = \sum_{m=1}^M P_m \cdot g(w_i, 0, \sigma_m^2), \\ g(w_i, 0, \sigma_m^2) = \frac{1}{\sqrt{2\pi}\sigma_m} e^{-\frac{w_i^2}{2\sigma_m^2}}, \\ \sum_{m=1}^M P_m = 1, \end{cases} \quad (3.1)$$

where the states of coefficients are represented by subscript “ m ” and the *a priori* probabilities of the M states are represented by P_m . The zero mean Gaussian component $g(w_i, 0, \sigma_m^2)$ corresponding to the state m has the variance σ_m^2 . Note $w_i, i = 1, \dots, K$, represent the wavelet coefficients in a single wavelet subspace.

It is also observed that in this peaky, heavy-tailed marginal distribution, only a few coefficients have large values at the positions where image edges occur, while most others have very small values. Therefore it is reasonable to simplify the GMM into a two-state representation as shown in (3.2). One state is used to describe large coefficient distribution and the other state is used to describe small coefficient distribution.

$$\begin{cases} p(w_i) = P_s \cdot g(w_i, 0, \sigma_s^2) + P_l \cdot g(w_i, 0, \sigma_l^2), \\ g(w_i, 0, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{w_i^2}{2\sigma^2}}, \\ P_s + P_l = 1. \end{cases} \quad (3.2)$$

The state of small coefficients is represented by subscript “ s ” and the state of large coefficients by subscript “ l ”. The *a priori* probabilities of the two states are represented by P_s and P_l , respectively. The zero mean Gaussian component $g(w_i, 0, \sigma_s^2)$ corresponding to the small state has a relatively small variance σ_s^2 , capturing the peakiness around zero (small coefficients), while the component $g(w_i, 0, \sigma_l^2)$ corresponding to the large state has a relatively large variance σ_l^2 , capturing the heavy tails (large coefficients).

3.1.2 An EM Algorithm for the Gaussian Mixture Model

The GMM parameters $[P_s, P_l, \sigma_s^2, \sigma_l^2]$ must first be obtained before they can be used for any purpose. Taking the coefficients of each image as the training data, its GMM parameters

can be obtained through the EM algorithm introduced in Chapter 2. In each iteration of the EM algorithm, there are two steps. The E step calculates the individual state probabilities for each wavelet coefficient $P_{s,i}, P_{l,i}$ and the M step involves simple closed-form updates for the variances $[\sigma_s^2, \sigma_l^2]$ and the overall state probabilities $[P_s, P_l]$. The two steps interact with each other in an iterative process to help obtain a set of final converged GMM parameters:

EM Algorithm for GMM

Step 1) Initilization:

Select an initial model estimate:

$$\Theta(0) = [P_s(0), P_l(0), \sigma_s^2(0), \sigma_l^2(0)], \quad (3.3)$$

where Θ represents the GMM parameter set $[P_s, P_l, \sigma_s^2, \sigma_l^2]$.

Set iteration counter $n = 0$.

Step 2) **E step**: Calculate the state probabilities for each wavelet coefficient w_i :

$$P_{s,i} = \frac{P_s(n) \cdot g(w_i, 0, \sigma_s^2(n))}{P_s(n) \cdot g(w_i, 0, \sigma_s^2(n)) + P_l(n) \cdot g(w_i, 0, \sigma_l^2(n))}, \quad i = 1, \dots, K, \quad (3.4)$$

$$P_{l,i} = \frac{P_l(n) \cdot g(w_i, 0, \sigma_l^2(n))}{P_s(n) \cdot g(w_i, 0, \sigma_s^2(n)) + P_l(n) \cdot g(w_i, 0, \sigma_l^2(n))}, \quad i = 1, \dots, K, \quad (3.5)$$

where K represents the total number of coefficients in the wavelet subspace.

Step 3) **M step**: Update the model parameters to maximize the overall probabilities:

$$\begin{aligned} & \Theta(n) \rightarrow \Theta(n+1) \\ & \Downarrow \\ & \left\{ \begin{array}{l} P_s(n+1) = \frac{1}{K} \sum_{i=1}^K P_{s,i}, \\ P_l(n+1) = \frac{1}{K} \sum_{i=1}^K P_{l,i}, \\ \sigma_s^2(n+1) = \frac{\sum_{i=1}^K w_i^2 \cdot P_{s,i}}{K \cdot P_s(n+1)}, \\ \sigma_l^2(n+1) = \frac{\sum_{i=1}^K w_i^2 \cdot P_{l,i}}{K \cdot P_l(n+1)}. \end{array} \right. \quad (3.6) \end{aligned}$$

Step 4) Set $n = n + 1$. If convergence condition is satisfied, then stop; Otherwise, return to **E step**.

3.1.3 A Generalized Gaussian Mixture Model

The GMM is a specific case of a generalized Gaussian mixture model (GGMM). A two-state representation of GGMM is given as follows:

$$\begin{cases} p(w_i) = P_s \cdot h_s(w_i, \alpha_s, \beta) + P_l \cdot h_l(w_i, \alpha_l, \beta), \\ h(w_i, \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|w_i|/\alpha)^\beta}, \\ P_s + P_l = 1. \end{cases} \quad (3.7)$$

The state of small coefficients is represented by subscript “s” and the state of large coefficients by subscript “l”. The generalized Gaussian component $h_s(w_i, \alpha_s, \beta)$ corresponding to the small state has a relatively small variance represented by α_s , while the component $h_l(w_i, \alpha_l, \beta)$ corresponding to the large state has a relatively large variance represented by α_l . Here $\Gamma(t) = \int_0^\infty u^{t-1} e^{-u} du$ is the Gamma function and β can be any fixed exponent value. When $\beta = 1$, the GGMM becomes a Laplacian mixture model (LMM) and when $\beta = 2$, the GGMM is exactly the GMM. It is possible that some GGMM cases, with their exponent values other than 2, may have a better description of wavelet coefficients than the GMM. Therefore, we need to compare their performance with that of GMM in certain applications. For this purpose, we specifically develop an EM algorithm to help estimate the GGMM parameters $[P_s, P_l, \alpha_s, \alpha_l]$.

As known from Chapter 2, the EM algorithm employs an iterative approach to estimate model parameters. In each iteration (say g), the objective of the EM algorithm is to obtain an updated parameter set Θ from the current parameter set Θ^g by maximizing the following log-likelihood function Q :

$$\begin{aligned} Q(\Theta, \Theta^g) &= \sum_{m=s,l} \sum_{i=1}^K \log(P_m h_m(w_i | \alpha_m)) p(m | w_i, \Theta^g) \\ &= \sum_{m=s,l} \sum_{i=1}^K \log(P_m) p(m | w_i, \Theta^g) + \sum_{m=s,l} \sum_{i=1}^K \log(h_m(w_i | \alpha_m)) p(m | w_i, \Theta^g), \end{aligned} \quad (3.8)$$

where Θ^g represents the current model parameters $[P_s^g, P_l^g, \alpha_s^g, \alpha_l^g]$ and K represents the total

number of coefficients in the wavelet subspace. The two state probabilities for each wavelet coefficient w_i are represented by $p(m|w_i, \Theta^g)$, which can be easily calculated using Bayes's rule:

$$p(m|w_i, \Theta^g) = \frac{P_m^g h_m(w_i|\alpha_m^g)}{p(w_i|\Theta^g)} = \frac{P_m^g h_m(w_i|\alpha_m^g)}{\sum_{k=s,l} P_k^g h_k(w_i|\alpha_k^g)}, \quad m = s, l \quad (3.9)$$

To maximize the log-likelihood expression in (3.8), we can maximize the term containing P_m and the term containing α_m independently since they are not related.

To find the expression for P_m , we introduce the Lagrange multiplier λ with the constraint that $\sum_m P_m = 1$, and solve the following equation:

$$\frac{\partial}{\partial P_m} \left[\sum_{m=s,l} \sum_{i=1}^K \log(P_m) p(m|w_i, \Theta^g) + \lambda \left(\sum_m P_m - 1 \right) \right] = 0. \quad (3.10)$$

As a result, we obtain:

$$P_m = \frac{1}{K} \sum_{i=1}^K p(m|w_i, \Theta^g). \quad (3.11)$$

To find an analytical expression for α_m , we input the generalized Gaussian function $h(w_i, \alpha, \beta)$ as given in (3.7) into the right part of (3.8) and obtain:

$$\begin{aligned} & \sum_{m=s,l} \sum_{i=1}^K \log(h_m(w_i|\alpha_m)) p(m|w_i, \Theta^g) \\ &= \sum_{m=s,l} \sum_{i=1}^K \left(\log \beta - \log 2\alpha_m - \log \Gamma\left(\frac{1}{\beta}\right) - \left(\frac{|w_i|}{\alpha_m}\right)^\beta \right) p(m|w_i, \Theta^g). \end{aligned} \quad (3.12)$$

Taking the derivative of (3.12) with respect to α_m and setting it to zero, we get:

$$\sum_{i=1}^K \left(-\frac{1}{\alpha_m} + \beta |w_i|^\beta \alpha_m^{-(\beta+1)} \right) p(m|w_i, \Theta^g) = 0, \quad (3.13)$$

with which we can solve α_m :

$$\alpha_m = \left(\frac{\sum_{i=1}^K \beta |w_i|^\beta p(m|w_i, \Theta^g)}{\sum_{i=1}^K p(m|w_i, \Theta^g)} \right)^{\frac{1}{\beta}}. \quad (3.14)$$

To be concluded, the update of model parameters in the EM algorithm is an iterative procedure until a final converged set of parameters are found. In each iteration, the EM algorithm has two steps: the E-step and the M-step. The E- step calculates the individual state probabilities for each wavelet coefficient $p(m|w_i, \Theta^g)$ and the M-step involves the updates for the model parameters $[P_s, P_l, \alpha_s, \alpha_l]$. The complete EM algorithm for GGMM is given as follows:

EM Algorithm for GGMM

Step 1) **Initilization:**

Select an initial model estimate:

$$\Theta(0) = [P_s(0), P_l(0), \alpha_s(0), \alpha_l(0)], \quad (3.15)$$

where Θ represents the GGMM parameter set $[P_s, P_l, \alpha_s, \alpha_l]$.

Set iteration counter $n = 0$.

Step 2) **E step:** Calculate the state probabilities for each wavelet coefficient $p(m|w_i, \Theta(n))$:

$$p(s|w_i, \Theta(n)) = \frac{P_s(n) \cdot h_s(w_i, \alpha_s(n), \beta)}{P_s(n) \cdot h_s(w_i, \alpha_s(n), \beta) + P_l(n) \cdot h_l(w_i, \alpha_l(n), \beta)}, \quad i = 1, \dots, K, \quad (3.16)$$

$$p(l|w_i, \Theta(n)) = \frac{P_l(n) \cdot h_l(w_i, \alpha_l(n), \beta)}{P_s(n) \cdot h_s(w_i, \alpha_s(n), \beta) + P_l(n) \cdot h_l(w_i, \alpha_l(n), \beta)}, \quad i = 1, \dots, K, \quad (3.17)$$

where K represents the total number of coefficients in the wavelet subspace.

Step 3) **M step:** Update the model parameters to maximize the overall probabilities:

$$\begin{aligned} \Theta(n) &\rightarrow \Theta(n+1) \\ &\Downarrow \\ \left\{ \begin{array}{l} P_s(n+1) = \frac{1}{K} \sum_{i=1}^K p(s|w_i, \Theta(n)), \\ P_l(n+1) = \frac{1}{K} \sum_{i=1}^K p(l|w_i, \Theta(n)), \\ \alpha_s(n+1) = \left(\frac{\beta \sum_{i=1}^K |w_i|^\beta \cdot p(s|w_i, \Theta(n))}{K \cdot P_s(n+1)} \right)^{1/\beta}, \\ \alpha_l(n+1) = \left(\frac{\beta \sum_{i=1}^K |w_i|^\beta \cdot p(l|w_i, \Theta(n))}{K \cdot P_l(n+1)} \right)^{1/\beta}. \end{array} \right. \quad (3.18) \end{aligned}$$

Step 4) Set $n = n + 1$. If convergence condition is satisfied, then stop; Otherwise, return to E step.

3.2 A Kullback-Leibler Divergence Based Similarity Measure for the Statistical Models

3.2.1 General Minkowski Distances

In many applications based on the statistical modelling, there is a necessity to measure the similarity of two distributions that are determined by their model parameters. The most common way to compare two parameter sets is done by the Minkowski distance. Given two N dimensional parameter set X and Y , the general Minkowski distance between X and Y is:

$$d_r(X, Y) = \left(\sum_{i=0}^{N-1} |X_i - Y_i|^r \right)^{\frac{1}{r}}. \quad (3.19)$$

If $r = 1$, the distance is known as the City-block or Manhattan distance. If $r = 2$, the distance is called the Euclidean distance. It can be observed that the general Minkowski distance is not effective because it fails to treat each parameter equally during comparison. Some large value parameters will dominate the distance function $d_r(X, Y)$ regardless of the fact that some small value parameters may have the same importance. For example, the two probability parameters $[P_s, P_l]$ in GMM should be treated equally with the two variance parameters $[\sigma_s^2, \sigma_l^2]$ in distance calculation, however their values are too small and almost neglectable in the Minkowski distance. The normalized Euclidean distance we used in [23] [24] is one way to alleviate the problem such that all parameters have approximately the same influence on the overall distance. However, it is still not optimal because the parameters represent measurements of different image characteristics. It is unlikely to have an objective distance calculation by simply including all parameters with different natures into the Minkowski formula and calculating their overall distance.

3.2.2 Similarity Measure Based on the Kullback-Leibler Divergence

The Kullback-Leibler divergence (or Kullback divergence) provides an effective approach for similarity measure of two distributions. Suppose we have two probability density functions (PDFs) $p_1(x)$ and $p_2(x)$ that are for wavelet coefficients in a wavelet subspace. As known from Chapter 2, their Kullback divergence is calculated as follows:

$$d(p_1(x), p_2(x)) = \int p_1(x) \ln \frac{p_1(x)}{p_2(x)} dx. \quad (3.20)$$

In case we use the GGMM to model the coefficients, the two distributions $p_1(x)$ and $p_2(x)$ take the forms as:

$$\begin{cases} p_1(x) = P_{s_1} h_{s_1}(x, \alpha_{s_1}, \beta) + P_{l_1} h_{l_1}(x, \alpha_{l_1}, \beta), \\ p_2(x) = P_{s_2} h_{s_2}(x, \alpha_{s_2}, \beta) + P_{l_2} h_{l_2}(x, \alpha_{l_2}, \beta). \end{cases} \quad (3.21)$$

By substituting (3.21) into (3.20), we obtain the complete Kullback divergence form:

$$d(p_1(x), p_2(x)) = \int (P_{s_1} h_{s_1}(x, \alpha_{s_1}, \beta) + P_{l_1} h_{l_1}(x, \alpha_{l_1}, \beta)) \ln \frac{(P_{s_1} h_{s_1}(x, \alpha_{s_1}, \beta) + P_{l_1} h_{l_1}(x, \alpha_{l_1}, \beta))}{(P_{s_2} h_{s_2}(x, \alpha_{s_2}, \beta) + P_{l_2} h_{l_2}(x, \alpha_{l_2}, \beta))} dx. \quad (3.22)$$

It can be seen that all GGMM parameters $[P_s, P_l, \alpha_s, \alpha_l]$ are involved in the Kullback divergence computation as given in (3.22) to measure the similarity of two distributions $p_1(x)$ and $p_2(x)$, no matter how different are these parameters in their values and in their meanings of image characteristics.

However, there is no simple closed form for the complete Kullback divergence given in (3.22) so it can only be numerically calculated. The computational complexity is so high that the complete Kullback approach is not practical for a CBIR system. It is observed that by dividing the Gaussian mixture distribution into two separate Gaussian distributions, the simple closed form of the Kullback divergence for each separate Gaussian distribution can be easily computed [25]. The complete Kullback divergence can be approximated by the sum of two separate Kullback divergences. Therefore, we present a new Kullback divergence based similarity measure for GGMM as follows:

$$d_k(p_1(x), p_2(x)) = F_s(x) + F_l(x), \quad (3.23)$$

$$F_s(x) = \int P_{s_1} h_{s_1}(x, \alpha_{s_1}, \beta) \ln \frac{P_{s_1} h_{s_1}(x, \alpha_{s_1}, \beta)}{P_{s_2} h_{s_2}(x, \alpha_{s_2}, \beta)} dx, \quad (3.24)$$

$$F_l(x) = \int P_{l_1} h_{l_1}(x, \alpha_{l_1}, \beta) \ln \frac{P_{l_1} h_{l_1}(x, \alpha_{l_1}, \beta)}{P_{l_2} h_{l_2}(x, \alpha_{l_2}, \beta)} dx. \quad (3.25)$$

The two separate Kullback divergences $F_s(x)$ and $F_l(x)$ have simple closed forms after integral calculation, as given by (3.26) and (3.27):

$$F_s(x) = P_{s_1} \ln \left(\frac{P_{s_1} \alpha_{s_2}}{P_{s_2} \alpha_{s_1}} \right) + \frac{P_{s_1}}{\beta} \left[\left(\frac{\alpha_{s_1}}{\alpha_{s_2}} \right)^\beta - 1 \right], \quad (3.26)$$

$$F_l(x) = P_{l_1} \ln \left(\frac{P_{l_1} \alpha_{l_2}}{P_{l_2} \alpha_{l_1}} \right) + \frac{P_{l_1}}{\beta} \left[\left(\frac{\alpha_{l_1}}{\alpha_{l_2}} \right)^\beta - 1 \right]. \quad (3.27)$$

As for the GMM, the new Kullback divergence based similarity measure take the following form:

$$d_k(p_1(x), p_2(x)) = F_s(x) + F_l(x), \quad (3.28)$$

$$F_s(x) = P_{s_1} \ln \left(\frac{P_{s_1} \sigma_{s_2}}{P_{s_2} \sigma_{s_1}} \right) + \frac{P_{s_1}}{2} \left(\frac{\sigma_{s_1}^2}{\sigma_{s_2}^2} - 1 \right), \quad (3.29)$$

$$F_l(x) = P_{l_1} \ln \left(\frac{P_{l_1} \sigma_{l_2}}{P_{l_2} \sigma_{l_1}} \right) + \frac{P_{l_1}}{2} \left(\frac{\sigma_{l_1}^2}{\sigma_{l_2}^2} - 1 \right). \quad (3.30)$$

The similarity measure based on the proposed separate Kullback divergence approach can be calculated very efficiently using GMM or GGMM parameters. In fact, its computational complexity is retained at the same level as other conventional similarity measures using Minkowski distances. On the other hand, this separate Kullback divergence approach (3.23)-(3.25) has nearly the same efficiency as the complete Kullback computation (3.22), which will be shown later in the simulation results of image retrieval. Both Kullback approaches outperform traditional similarity measures by having a more accurate distance computation and a higher image retrieval rate.

3.2.3 Analysis of the New Similarity Measure

Now it is proposed that the separate Kullback divergence can approximate the complete Kullback divergence in distance computation. The closeness of the two divergences can be justified through a mathematical analysis.

Assume we have two Gaussian mixture distributions:

$$\begin{cases} p_1(x) = P_{s_1}g_{s_1}(x) + P_{l_1}g_{l_1}(x) = A + B, \\ p_2(x) = P_{s_2}g_{s_2}(x) + P_{l_2}g_{l_2}(x) = C + D. \end{cases} \quad (3.31)$$

For simplicity, we use A, B, C, D to denote their four Gaussian components. Their complete Kullback divergence (3.22) can be rewritten as:

$$d(p_1(x), p_2(x)) = \int (A + B) \ln \frac{A + B}{C + D} dx = \int A \ln \frac{A + B}{C + D} dx + \int B \ln \frac{A + B}{C + D} dx, \quad (3.32)$$

while their separate Kullback divergence becomes:

$$d_k(p_1(x), p_2(x)) = \int A \ln \frac{A}{C} dx + \int B \ln \frac{B}{D} dx. \quad (3.33)$$

If we suppose the two Gaussian mixture distributions are similar to each other, which means:

$$\frac{A}{C} \approx \frac{B}{D}, \quad (3.34)$$

we can further derive:

$$\frac{A + B}{C + D} \approx \frac{A}{C}, \quad (3.35)$$

$$\frac{A + B}{C + D} \approx \frac{B}{D}. \quad (3.36)$$

If we substitute (3.35), (3.36) into (3.32) and then compare it with (3.33), we will get the conclusion:

$$d_k(p_1(x), p_2(x)) \approx d(p_1(x), p_2(x)). \quad (3.37)$$

That means the separate Kullback divergence is able to approximate the complete Kullback divergence when the two Gaussian mixture distributions are similar to each other. As far as the application of image retrieval is concerned, since any two images from the same class will have relatively similar Gaussian mixture distributions, their distance can be computed accurately using the separate Kullback divergence instead of the complete Kullback divergence. It is also known that an accurate similarity measure of images within the same class plays a key factor in overall image ranking. From this point of view, the separate Kullback divergence is successful in substituting the complete Kullback divergence for image retrieval.

In case the two Gaussian mixture distributions are different from each other, it is worth of study that how the separate Kullback divergence and the complete Kullback divergence will respond to this difference respectively. To simplify the problem, we assume that they have same Gaussian components but slightly different state probabilities:

$$\begin{cases} p_1(x) = P_s g_s(x) + P_l g_l(x), \\ p_2(x) = (P_s + \delta) g_s(x) + (P_l - \delta) g_l(x), \end{cases} \quad (3.38)$$

where δ is a very small quantum compared with P_s and P_l . The difference between their complete Kullback divergence and separate Kullback divergence is given by:

$$\begin{aligned} T &= d(p_1(x), p_2(x)) - d_k(p_1(x), p_2(x)) \\ &= \int P_s g_s(x) \left[\ln \frac{P_s g_s(x) + P_l g_l(x)}{(P_s + \delta) g_s(x) + (P_l - \delta) g_l(x)} - \ln \frac{P_s g_s(x)}{(P_s + \delta) g_s(x)} \right] dx \\ &\quad + \int P_l g_l(x) \left[\ln \frac{P_s g_s(x) + P_l g_l(x)}{(P_s + \delta) g_s(x) + (P_l - \delta) g_l(x)} - \ln \frac{P_l g_l(x)}{(P_l - \delta) g_l(x)} \right] dx. \end{aligned} \quad (3.39)$$

Applying Taylor's series, since

$$\begin{aligned} &\ln \frac{P_s g_s(x) + P_l g_l(x)}{(P_s + \delta) g_s(x) + (P_l - \delta) g_l(x)} - \ln \frac{P_s g_s(x)}{(P_s + \delta) g_s(x)} \\ &= \ln \left(1 + \frac{\delta}{P_s} \right) - \ln \left(1 + \frac{\delta(g_s(x) - g_l(x))}{P_s g_s(x) + P_l g_l(x)} \right) \\ &= \frac{\delta}{P_s} - \frac{\delta(g_s(x) - g_l(x))}{P_s g_s(x) + P_l g_l(x)} + o(\delta) \\ &= \frac{\delta(P_s + P_l)g_l(x)}{P_s(P_s g_s(x) + P_l g_l(x))} \end{aligned} \quad (3.40)$$

and

$$\begin{aligned} &\ln \frac{P_s g_s(x) + P_l g_l(x)}{(P_s + \delta) g_s(x) + (P_l - \delta) g_l(x)} - \ln \frac{P_l g_l(x)}{(P_l - \delta) g_l(x)} \\ &= \ln \left(1 - \frac{\delta}{P_l} \right) - \ln \left(1 + \frac{\delta(g_s(x) - g_l(x))}{P_s g_s(x) + P_l g_l(x)} \right) \\ &= -\frac{\delta}{P_l} - \frac{\delta(g_s(x) - g_l(x))}{P_s g_s(x) + P_l g_l(x)} + o(\delta) \\ &= \frac{-\delta(P_s + P_l)g_s(x)}{P_l(P_s g_s(x) + P_l g_l(x))}, \end{aligned} \quad (3.41)$$

we substitute (3.40),(3.41) into (3.39) and obtain:

$$\begin{aligned}
 T &= \int P_s g_s(x) \frac{\delta(P_s + P_l) g_l(x)}{P_s(P_s g_s(x) + P_l g_l(x))} dx + \int P_l g_l(x) \frac{-\delta(P_s + P_l) g_s(x)}{P_l(P_s g_s(x) + P_l g_l(x))} dx \\
 &= \int \frac{\delta(P_s + P_l) g_l(x) g_s(x)}{P_s g_s(x) + P_l g_l(x)} dx - \int \frac{\delta(P_s + P_l) g_s(x) g_l(x)}{P_s g_s(x) + P_l g_l(x)} dx \\
 &= 0.
 \end{aligned} \tag{3.42}$$

The result reveals that the separate Kullback divergence is exactly the same as the complete Kullback divergence if the two Gaussian mixture distributions under comparison are slightly different in parameters. Therefore, the separate Kullback divergence can be regarded as a very good approximation to its complete counterpart.

Chapter 4

Statistical Model Based Image Retrieval

The statistical modelling of GMM and GGMM provides a new way to interpret image texture contents through the model parameters, which can be applied very naturally in the application of image retrieval. In this chapter, we present a new CBIR system that incorporates statistical model based image features and a Kullback divergence based similarity measure. The images are modelled by GMM or GGMM in the wavelet domain and the statistical model parameters are employed to construct the indexing feature space for the CBIR system. The similarity measure of two indexing feature vectors is performed based on the novel Kullback divergence approach presented in the previous chapter. Simulations are conducted to demonstrate the effectiveness of the new CBIR system.

4.1 Overview of the Proposed CBIR System

Figure 4.1 shows the architecture of the proposed CBIR system. For each image in the image database, its low-level features (textures, shapes and colors) will be extracted by the GMM and other feature processing techniques and the obtained indexing feature vector will be stored in the feature database. When a query image comes in, its feature vector will be compared with those in the feature database one by one based on the kullback divergence similarity measure. The top M images in the database with smallest feature distances to the query will be retrieved.

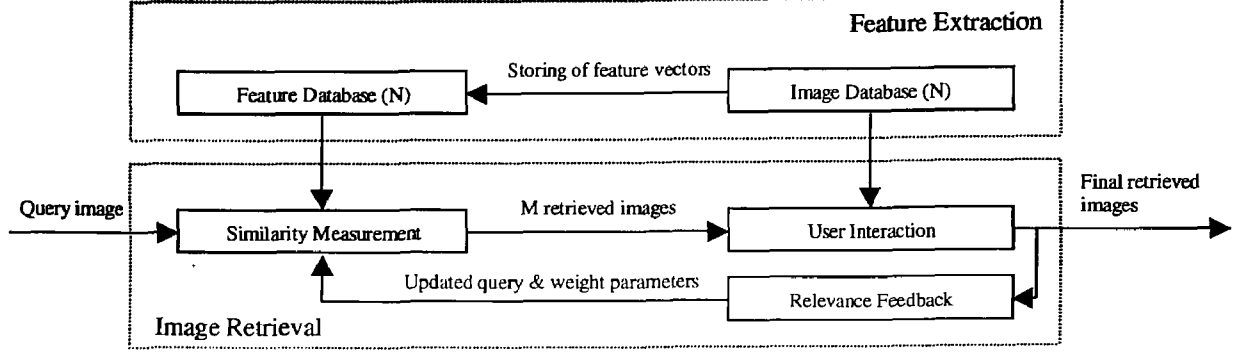


Figure 4.1: The Proposed CBIR System.

Since there is always a gap between low-level image features and high-level semantics of human perception, the M retrieved images may not be the optimal results. Therefore the relevance feedback mechanism, which gets human involved in the retrieval process, is introduced in the CBIR system. The user will compare the M retrieved images with the query and determine which one is relevant and which one is irrelevant based on his visual discrimination. Based on the discrimination results, the query and the feature weight parameters will be updated and applied in the next round of similarity measure. Through this iterative process, the gap between low-level image features and human perception can be compensated and accordingly the retrieval performance will be improved. The iterative process will stop if there is no further improvement compared with the previous round of retrieval or the optimal retrieval result is achieved.

4.2 The Indexing Feature Space for Image Retrieval

While constructing the indexing feature space, we consider the GMM parameters $[P_s, P_l, \sigma_s^2, \sigma_l^2]$ as legitimate image texture features. From a human perception point of view, since large coefficients indicate singularity such as image edges or visible texture patterns, the two *a priori* probabilities P_s and P_l are able to represent the denseness of such singularity within an image while the two Gaussian variances σ_s^2 and σ_l^2 are able to represent the depth of such singularity. Therefore, they all have significant meaning in image texture content

interpretation and are very good candidates of image texture features.

The indexing feature space consists of GMM parameters obtained from all decomposed wavelet subspaces. As known, the 2-D wavelet transform decomposes an image into three wavelet subspaces (horizontal, vertical and diagonal) at each scale. The decomposed wavelet subspaces contain different image texture information, therefore their respective GMM parameters should be all incorporated into the indexing feature space, which has a following form of representation:

$$F = [W_{1H}, W_{1V}, W_{1D}, \dots, W_{RH}, W_{RV}, W_{RD}], \quad (4.1)$$

where W represents the GMM parameter set $[P_s, P_l, \sigma_s^2, \sigma_l^2]$ of a single wavelet subspace. Subscripts H , V , and D represent the three different directions (Horizontal, Vertical and Diagonal) of the wavelet transform at each scale and subscript R represents the number of scales the image is decomposed. As for the GGMM, its indexing feature space has a same form of representation given by (4.1), with W representing the GGMM parameter set $[P_s, P_l, \alpha_s, \alpha_l]$ instead.

4.3 Kullback Divergence Based Similarity Measure

In order to measure the similarity of an image and the query, we need to compute their overall feature distance. Using the separate Kullback divergence approach given in Chapter 3, we compute the distance of the pair of Gaussian mixture distributions in each decomposed wavelet subspace and then sum them up to get an overall distance, which can be shown as follows:

$$D(U, V) = \sum_{i=1}^R \sum_{j=H,V,D} d_k(U_{ij}(x), V_{ij}(x)), \quad (4.2)$$

where U and V represent the feature vectors of the image and the query, respectively. Their gaussian mixture distributions in each decomposed wavelet subspace are represented by $U_{ij}(x)$ and $V_{ij}(x)$. The separate Kullback divergence $d_k(U_{ij}(x), V_{ij}(x))$ can be calculated using (3.23)-(3.30).

4.4 Simulation Results

The Brodatz image database is used to demonstrate the effectiveness of the proposed feature extraction method. The database consists of 1,856 images in 116 different classes, with each class containing 16 similar images. Given a query from any class, the ideal condition is that all 16 images in the same class as the query are retrieved. The retrieval performance of the presented new method is evaluated by the overall retrieval rate which is defined as the average percentage of images belonging to the same class as the query in the top 16 matched [5]. Besides, since the first round of retrieval (or initial retrieval) reflects objectively the effectiveness of the applied feature extraction technique and similarity measure, we use the initial retrieval rate to benchmark the retrieval performance.

We consider the following two factors will affect the retrieval performance: 1) How many wavelet scales each image is decomposed, 2) which wavelet and scaling filters are used in the wavelet decomposition. With more levels of wavelet decomposition, the retrieval performance will get better because more image texture information will be represented by the indexing feature space. However at the same time, the feature number will increase accordingly, which will lower the efficiency of the whole feature space. Besides the decomposition levels, different wavelet and scaling filters have different characteristics in decomposition and will also affect the retrieval performance.

In order to test the filters' impact on the retrieval performance, the db1 (Haar) filters and the db2 (Daubechies 2) filters are used in the wavelet decomposition, respectively. Images are decomposed into two wavelet scales in this test. A comparison of their retrieval performance is given in TABLE 4.1. We can see that db2 filters slightly outperform the db1 filters in our experiment. Although the db2 filters are more complex and ought to have a better decomposition of image texture contents, its retrieval performance is not significantly better than that of the db1 filters.

We further compare the retrieval performance of our method by decomposing the images into two wavelet scales and three wavelet scales, respectively. In the two-scale decomposition, we get a total of 6 wavelet subspaces and 1 scaling subspace. Since each wavelet subspace

	Db1 Filters	Db2 Filters
Initial Retrieval Rate	68.96%	69.12%
Wavelet Scales	2	2

Table 4.1: Initial Retrieval Rate Based on Db1 And Db2 Filters.

has 4 features, the total number of features in the indexing feature space is $6 \times 4 = 24$. It is also obvious that the feature number in the three-scale decomposition is $9 \times 4 = 36$. TABLE 4.2 lists the comparison result. We can see that with three scales of decomposition, the retrieval rate is enhanced by 3 percent. But at the same time, the number of features also increases by nearly 50%.

Scales of Decomposition	Initial Retrieval Rate	Number of Features
2	69.12%	24
3	73.72%	36

Table 4.2: Comparison of Different Decomposition Scales.

To illustrate the performance of the new method, we select four different query images from the Brodatz database and perform the retrieval. Figure 4.2 - 4.9 display the retrieval results. All the four query images are full of edge and texture information therefore are ideal to evaluate the retrieval performance. TABLE 4.3 lists the retrieval results of the four queries. As can be seen, the initial retrieval result is very satisfactory and the performance is further improved in the second round of retrieval by involving human interaction. This result indicates that the features we extracted are appropriate and effective. Besides, the image retrieval rate can be significantly enhanced if we introduce the relevance feedback mechanism and perform more iterations of retrieval.

We continue the experiment to compare the Kullback divergence based similarity measure with other similarity measure schemes. In this experiment, two levels and three levels of wavelet decomposition are performed respectively. According to (4.1), the two-level decom-

Query Class	Correctly Retrieved Images Among Top 16 Matched	
	Initial Retrieval	Second Round Retrieval
87	14	16
98	14	15
111	13	15
114	12	16

Table 4.3: Retrieval Results of 4 Different Queries.

position generates six wavelet subspaces and has twenty-four features (four features for each subspace) in the feature vector, while the three-level decomposition generates nine wavelet subspaces and has thirty-six features in the feature vector. Then three different kinds of similarity measures (normalized Euclidean distance, complete Kullback divergence and separate Kullback divergence) are applied on the texture features to compare their retrieval performances, as shown in TABLE 4.4.

Scales of Decomposition	Type of Similarity Measure		
	Normalized Euclidean	Complete Kullback	Separate Kullback
2 scales	69.12%	71.63%	71.87%
3 scales	73.72%	75.68%	75.50%

Table 4.4: Initial Retrieval Rate Based on Similarity Measures.

When calculating the complete Kullback divergence, we use a discrete integral approach with integral range $[-1500, 1500]$ and integral step 1 to obtain the result. As can be seen from TABLE 4.4, the similarity measure based on the Kullback divergence achieves a better retrieval rate than the normalized Euclidean distance approach, because the GMM is a statistical model that uses PDF to describe image texture features. The closeness of two images, represented by the similarity of their PDFs, can be more accurately measured using Kullback divergence than other approaches. Figure 4.10 - 4.11 show the top 16 matched images of the query "d12.1" based on the Euclidean distance and the Kullback divergence,

respectively. The correctly retrieved images are marked with check boxes. It can be seen in this example that the similarity measure based on Kullback divergence has a better performance than that based on Euclidean distance. It can also be seen from TABLE 4.4 that the presented new similarity measure based on separate Kullback calculation has the same effectiveness as that based on the complete Kullback computation. Meanwhile, it has much lower computational complexity and is practical in a CBIR system.

In order to observe the impact of different exponent values in GGMM on the retrieval performance, we compare different GGMM cases with the exponent value β ranging from 1.0 to 3.5. The model becomes LMM if $\beta = 1$ and GMM if $\beta = 2$. Some researchers studied wavelet coefficients using a uni component GGD model and suggested the appropriate exponent is around 0.5 for low or middle frequency subspaces [26]. In fact, the appropriate exponent value selected for retrieval is dependent on the model applied, the subspaces decomposed and the image database that is tested. In our case, as far as GGMM is adopted and some high frequency subspaces are used for feature extraction, the appropriate exponent value for retrieval is around 2.0. In our experiment, image features from three wavelet scales are extracted and the similarity measure is based on the separate kullback divergence. It can be concluded from Table 4.5 that the retrieval rate achieves the highest when β takes a value between 2.0 and 2.5, which means, the GMM is appropriate and near optimal among all GGMM cases when the experiment is conducted on the brodatz image database.

	GMM	GGMM				
	$\beta = 2.0$	$\beta = 1.0$	$\beta = 1.5$	$\beta = 2.5$	$\beta = 3.0$	$\beta = 3.5$
Retrieval Rate	75.50%	69.15%	73.75%	75.80%	73.63%	68.44%

Table 4.5: GMM Compared with GGMM.

We then compare the GMM and separate Kullback based approach with other traditional methods, such as the PWT (Pyramid Wavelet Transform) based and the Gabor filter based methods. Like GMM, these methods also extract image texture features from the compressed domain. We also compare our method with the GGD model based Kullback divergence

approach presented in [7]. The GGD is another statistical model in the wavelet domain, but it's a uni model that contains only one generalized Gaussian component. In our experiment, we apply the whole Brodatz image database (1,856 images in 116 classes) instead of using the portion of it (640 images in 40 classes) as done in [7]. TABLE 4.6 lists the comparison result. The new method achieves a higher retrieval rate than the PWT and Gabor methods, with equal or fewer features in the feature vector. When compared with the GGD model based Kullback approach, the new method has more features but also a much higher retrieval rate. It proves that GMM is more accurate for the marginal distribution of wavelet coefficients than GGD.

	GMM & Kullback		GGD & Kullback		PWT	Gabor
	2 scales	3 scales	2 scales	3 scales		
Retrieval Rate	71.87%	75.50%	49.82%	56.26%	68.70%	74.37%
Feature Number	24	36	12	18	24	48

Table 4.6: GMM & Kullback Compared with Other Traditional Methods.

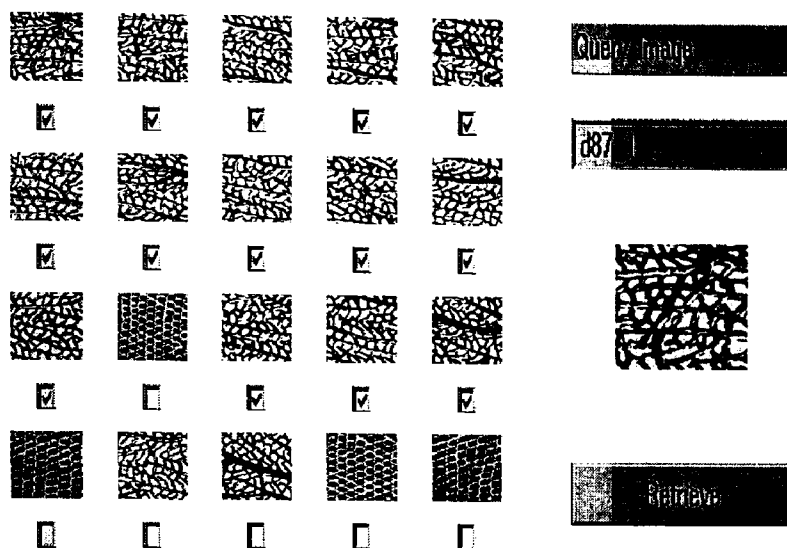


Figure 4.2: Initial Retrieval of A Query from Class D87.

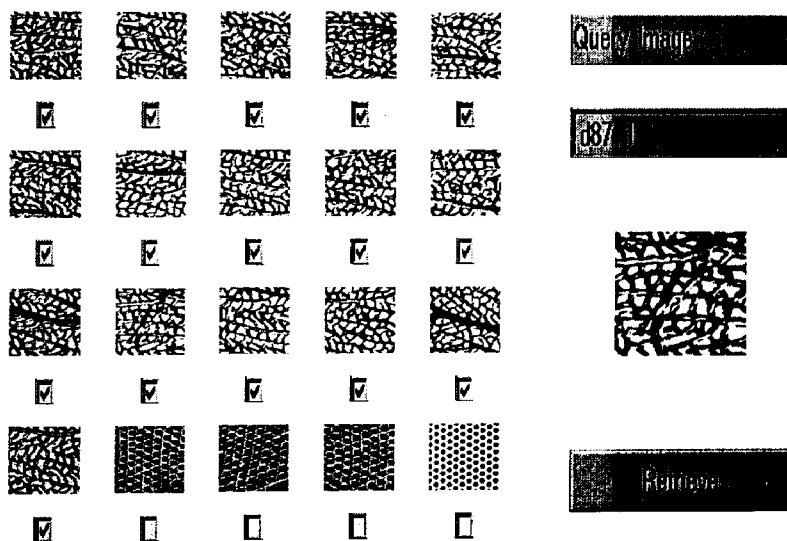


Figure 4.3: Second Round Retrieval of the Same Query from Class D87.

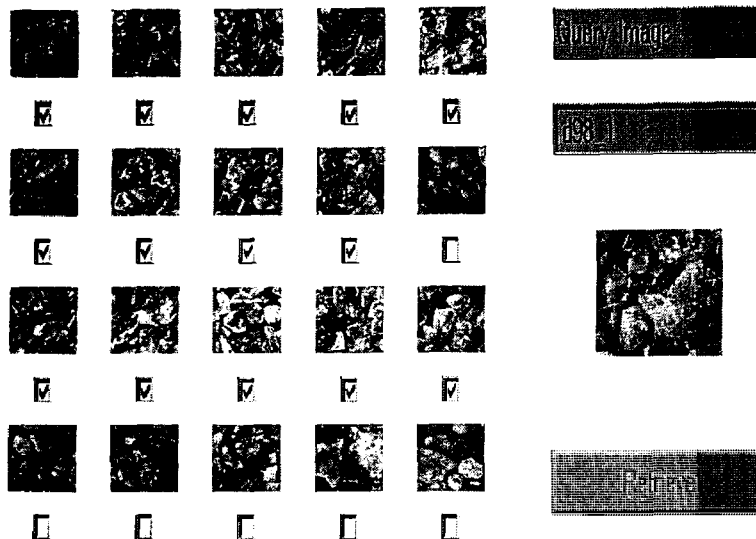


Figure 4.4: Initial Retrieval of A Query from Class D98.

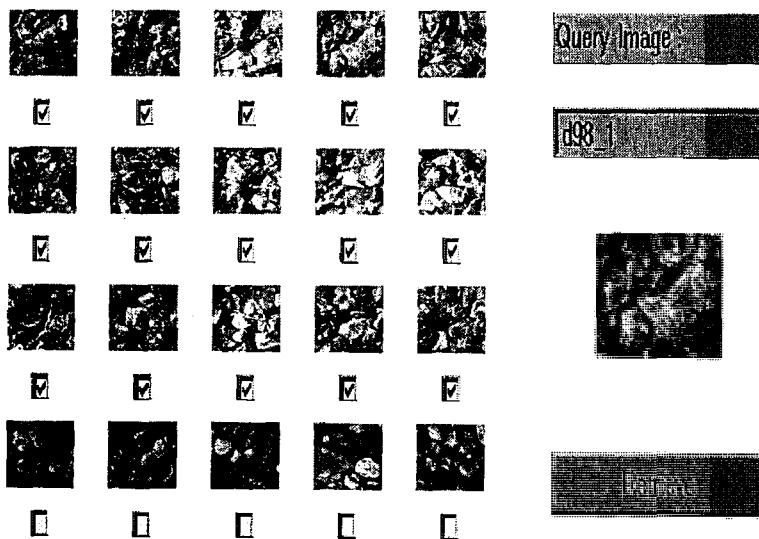


Figure 4.5: Second Round Retrieval of the Same Query from Class D98.

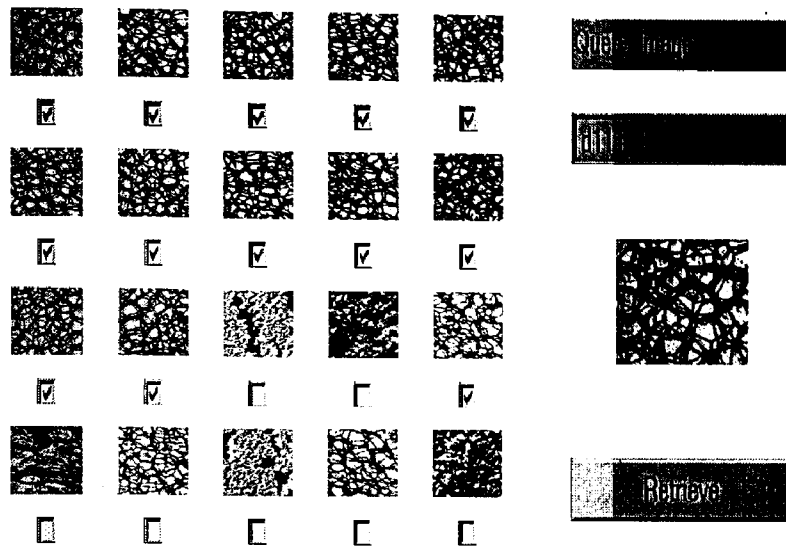


Figure 4.6: Initial Retrieval of A Query from Class D111.

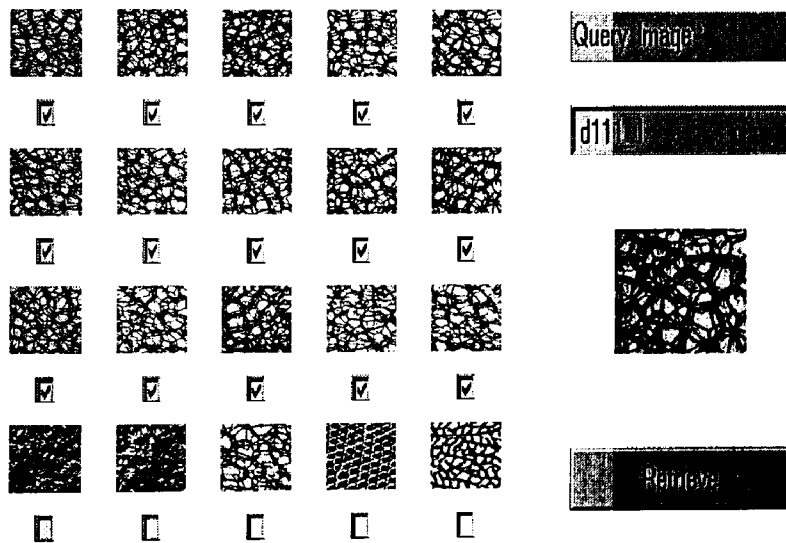


Figure 4.7: Second Round Retrieval of the Same Query from Class D111.

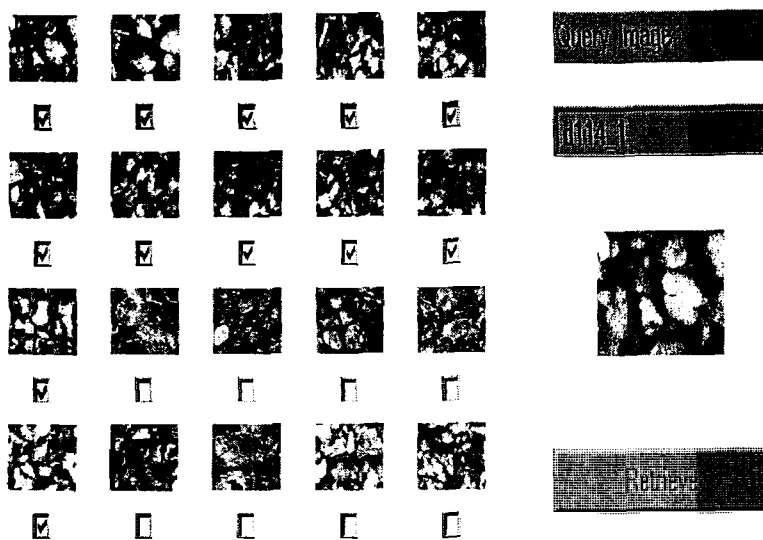


Figure 4.8: Initial Retrieval of A Query from Class D114.

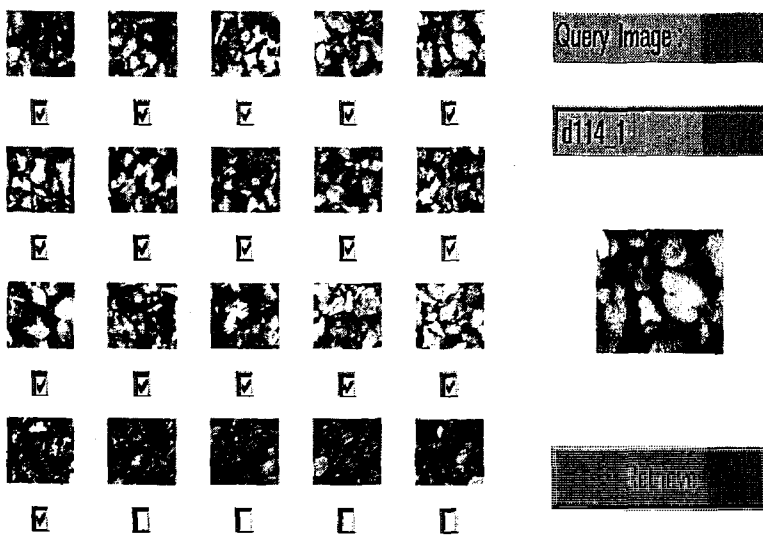


Figure 4.9: Second Round Retrieval of the Same Query from Class D114.

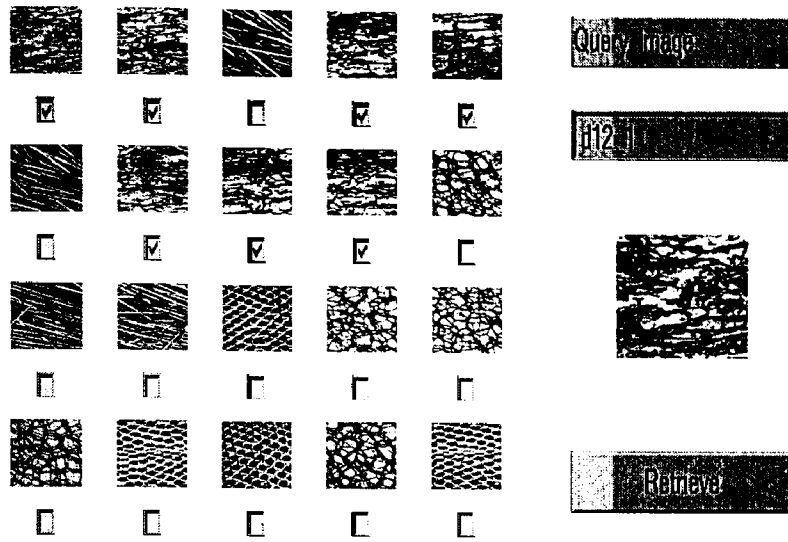


Figure 4.10: Initial Retrieval of A Query Image Using Euclidean Distance.

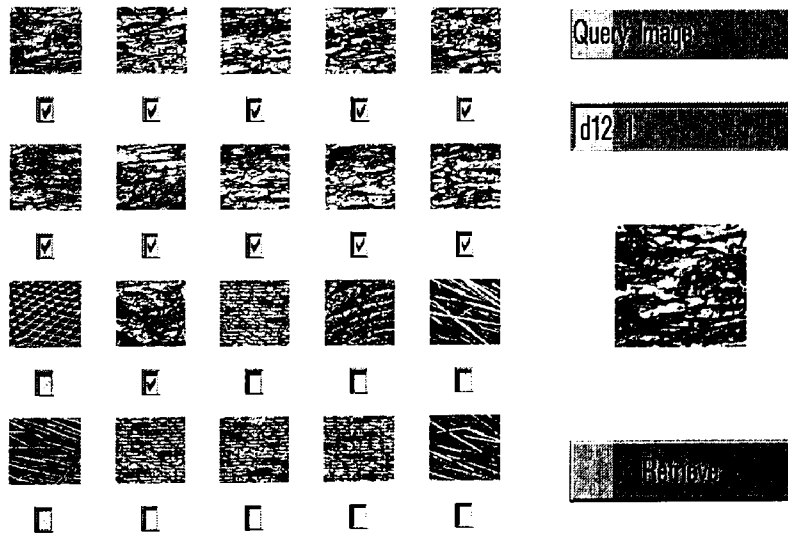


Figure 4.11: Initial Retrieval of Same Query Image Using Separate Kullback Distance.

4.5 Summary

In this chapter, a new CBIR system is presented for image retrieval based on GMM or GGMM extracted image features and a Kullback divergence similarity measure. The obtained image features are effective in capturing image characteristics. Besides, the Kullback divergence is an appropriate and efficient way to measure the similarity of features extracted by statistical models such as the GMM and GGMM. Simulation results indicate that the Kullback divergence based similarity measure achieves a higher image retrieval rate, while keeping the same level of computational complexity as those Minkowski distance based similarity measures. It is shown that the new CBIR system with the combination of GMM or GGMM extracted image features and the Kullback divergence similarity measure outperforms many other image retrieval methods.

Chapter 5

Statistical Model Based Fragile Watermarking

The statistical modelling of GMM and GGMM describes image statistical characteristics through the model parameters, which can be utilized for image watermarking purposes. In this chapter, a novel fragile watermarking method based on the statistical modelling is developed. First, a preliminary watermarking scheme is presented to embed authentication information into the statistical model. The embedding is analyzed to achieve the least distortion on the host image. Then a multiscale fragile watermarking scheme is presented to embed authentication messages such as personal signatures or logos into the host image. Simulation results are given to demonstrate the effectiveness and advantages of the new method.

5.1 Embedding Information into the Statistical Model

As known, the 2-D wavelet transform decomposes an image into three wavelet subspaces (horizontal, vertical and diagonal) at each scale. If the Gaussian mixture model and the EM algorithm are applied to these three subspaces, three different sets of model parameters will be obtained. We can modify the large coefficients of a single wavelet subspace so that its large variance parameter σ_l^2 will have the same value as that of another wavelet subspace $\sigma_l'^2$ [27]. This specially formed relationship serves as the basis of the proposed fragile watermarking method. Any image operations or malicious attacks will inevitably change the wavelet

coefficients, therefore they will break this relationship and be detected. The modification of only large coefficients brings two advantages. First, The large coefficients usually represent image edges in the space domain which, when modified, are generally difficult to be detected by human vision. Second, large coefficients are not so many in a wavelet subspace therefore the changes made on them will not introduce much image distortion.

To make the large variance parameter σ_l^2 the same value as $\sigma_l'^2$, each large coefficient w_i will be modified by a certain amount Δw_i . The modification is guided by the following principle:

$$\sum_{i=1}^P [(w_i + \Delta w_i)^2 - w_i^2] = K (\sigma_l'^2 - \sigma_l^2), \quad (5.1)$$

where P is the number of modified coefficients and K is the total number of coefficients in the wavelet subspace. Theory behind the principle is that the variance difference contributed by the coefficients modification should be equal to the overall parameter difference.

Since the modification of large coefficients are independent from one another, there seems to be numerous solutions satisfying (5.1). Therefore, it is worth of analysis that among all possible solutions, which one can lead to a minimum image distortion, which is a basic requirement of all watermarking systems. We are able to analyze this problem and get an optimal solution by *Theorem 1*.

Theorem 1 : Assume σ_l^2 and $\sigma_l'^2$ are the large variance parameters of two wavelet subspaces; Assume $w_i, i = 1, \dots, P$ represent the P coefficients with the largest absolute values in the wavelet subspace of σ_l^2 , and the total number of coefficients in that wavelet subspace is K ; If each large coefficient w_i is to be modified by a respective amount Δw_i in order to make σ_l^2 and $\sigma_l'^2$ equal, i.e., (5.1) is satisfied, then the optimal way of modification with least image distortion is to minimize the following mean square error (MSE):

$$J(\Delta w_i | i = 1, \dots, P) = \sum_{i=1}^P \Delta w_i^2, \quad (5.2)$$

which leads to the conclusion that each coefficient w_i must be modified with a constant pro-

portional rate α , that is:

$$\Delta w_i = \alpha w_i. \quad (5.3)$$

Proof : Let the image distortion defined as the MSE after modification:

$$J(\Delta w_i | i = 1, \dots, P) = \sum_{i=1}^P \Delta w_i^2; \quad (5.4)$$

Let the modification principle (5.1) be a constraint condition:

$$D(\Delta w_i | i = 1, \dots, P) = \sum_{i=1}^P [(w_i + \Delta w_i)^2 - w_i^2] - K(\sigma_l'^2 - \sigma_l^2) = 0, \quad (5.5)$$

under which we seek a way to minimize the image distortion. Using the Lagrangian approach, we construct a Lagrangian relaxation function as follows:

$$\begin{aligned} & J'(\Delta w_i | i = 1, \dots, P) \\ &= J(\Delta w_i | i = 1, \dots, P) + \lambda D(\Delta w_i | i = 1, \dots, P) \\ &= \sum_{i=1}^P [(1 + \lambda) \Delta w_i^2 + 2\lambda w_i \Delta w_i] - \lambda K(\sigma_l'^2 - \sigma_l^2). \end{aligned} \quad (5.6)$$

A minimization of the Lagrangian relaxation:

$$\begin{aligned} & \operatorname{argmin} \left\{ J'(\Delta w_i | i = 1, \dots, P) \right\} \\ & \Rightarrow \frac{\partial J'(\Delta w_i | i = 1, \dots, P)}{\partial (\Delta w_i)} = 0 \\ & \Rightarrow \Delta w_i = \frac{-\lambda w_i}{1 + \lambda} = \alpha w_i \end{aligned} \quad (5.7)$$

tells us: when all large coefficients w_i are modified with a constant proportional rate α , it achieves the least image distortion in terms of MSE.

Substituting the conclusion of (5.7) into (5.1), we obtain an updated modification principle:

$$\sum_{i=1}^P \{ [w_i(1 + \alpha)]^2 - w_i^2 \} = K(\sigma_l'^2 - \sigma_l^2), \quad (5.8)$$

which can be further evolved into a quadratic equation regarding the proportional rate α :

$$\left(\sum_{i=1}^P w_i^2\right) \alpha^2 + \left(\sum_{i=1}^P 2w_i^2\right) \alpha + K \left(\sigma_l^2 - \sigma_l'^2\right) = 0. \quad (5.9)$$

Therefore, α value can be easily calculated.

It is noticed that the two large variance parameters σ_l^2 and $\sigma_l'^2$ are obtained through the EM algorithm, while in the modification principle, the gap between them are compensated using a simple statistical approach other than the EM algorithm. In fact, after the modification guided by (5.8), there is still a small discrepancy between the updated parameter of σ_l^2 and the target parameter $\sigma_l'^2$. Therefore an iterative approach involving the modification and the EM algorithm in each single step is required to finally adjust the large variance parameter σ_l^2 to the target value $\sigma_l'^2$, as demonstrated in Figure 5.1.

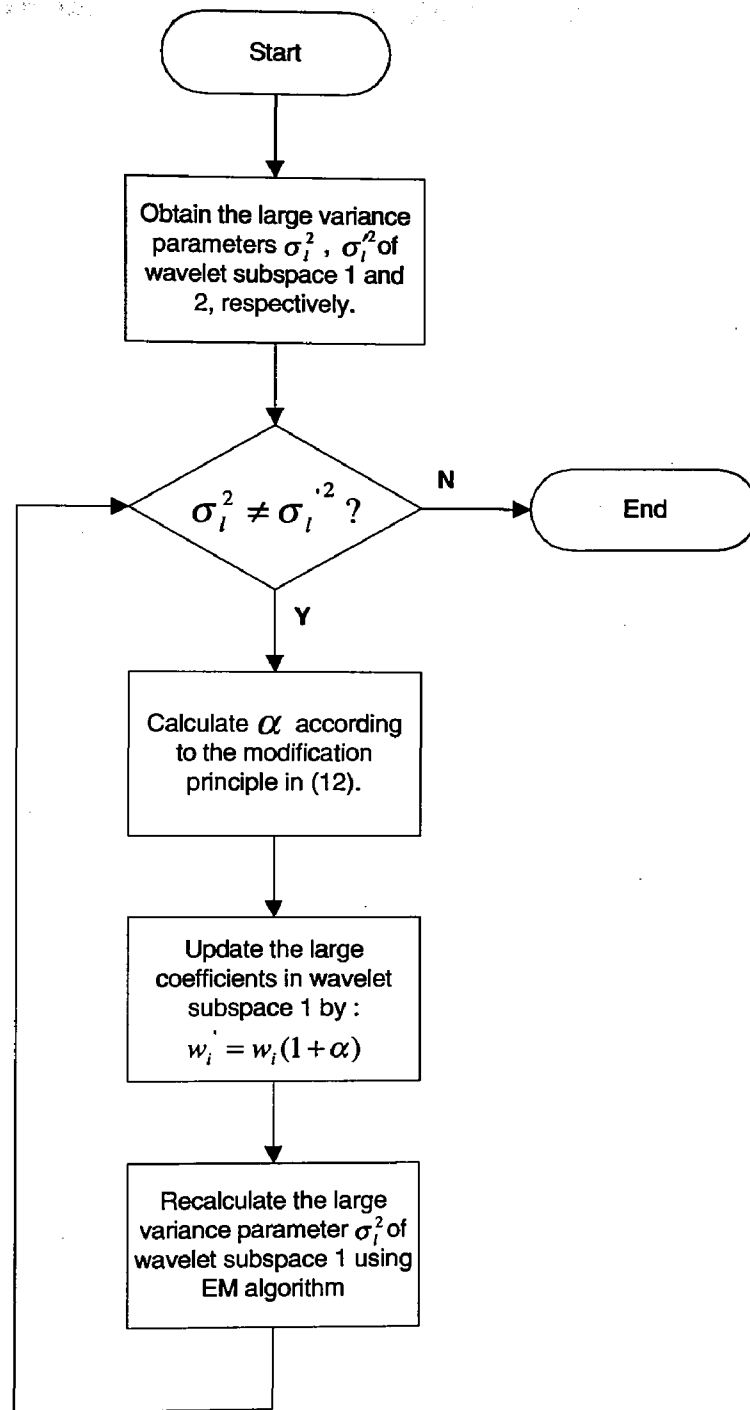


Figure 5.1: An Iterative Approach for Coefficient Modification.

5.2 Multiscale Embedding of Authentication Messages

We consider the following three aspects very attractive for a fragile watermarking system: 1) Can embed some authentication messages into the host image. 2) Can localize the image tampering if there is any. 3) Can distinguish some normal image operations from malicious attacks. Integrated with some coding techniques and implemented at multiple wavelet scales, our proposed method is able to achieve the above objectives.

Figure 5.2 has an overview of the watermark embedding process. Authentication message are initially translated into some binary bit streams. Then the wavelet subspaces at multiple scales are divided into a number of wavelet blocks depending on how many message bits being embedded and how many wavelet scales these bits will spread into. The binary bit streams are finally embedded into the wavelet blocks on a private key basis by forming some special relationships specified by the code map.

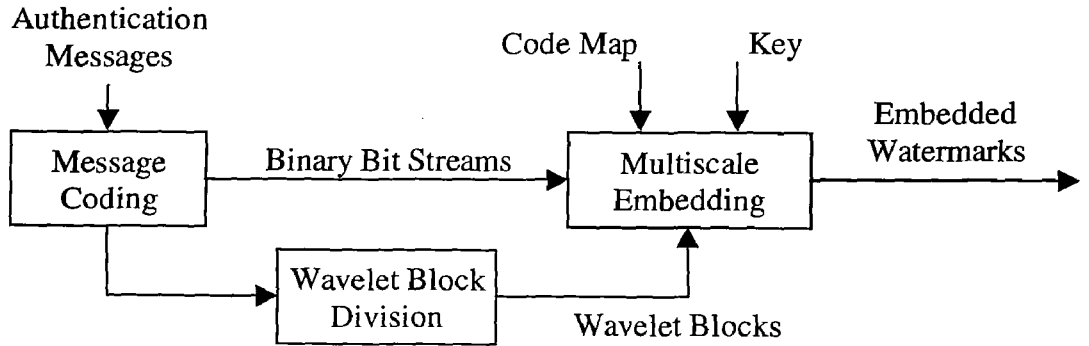


Figure 5.2: Multiscale Embedding of Authentication Messages.

To embed a binary bit stream into the wavelet subspace instead of a single bit, we must not use the entire wavelet subspace but divide it into many wavelet blocks [28]. Every three wavelet blocks obtained at the same position from the wavelet subspaces (Horizontal, Vertical, Diagonal) can form a special relationship to encode two message bits. An example is shown in TABLE 5.1.

The parameters $\sigma_{i,H}^2, \sigma_{i,V}^2, \sigma_{i,D}^2$ represent the variances of the large coefficients of the three

Formed Relationship	Coded Bits
$\sigma_{l,V}^2 = \sigma_{l,H}^2$	00
$\sigma_{l,V}^2 = \sigma_{l,D}^2$	01
$\sigma_{l,H}^2 = \sigma_{l,D}^2$	10
$\sigma_{l,V}^2 = \sigma_{l,H}^2 = \sigma_{l,D}^2$	11

Table 5.1: Code Map for Message Bits Embedding.

wavelet blocks obtained from horizontal subspace, vertical subspace and diagonal subspace respectively. Various parameter equity relationships among $\sigma_{l,H}^2$, $\sigma_{l,V}^2$ and $\sigma_{l,D}^2$ can be formed in the way shown in TABLE 5.1 to encode different two bits into these three wavelet blocks. Since there are N^2 groups of such wavelet blocks, at most $2N^2$ bits at a single wavelet scale can be embedded. Any unauthorized changes made in a specific area of the watermarked image will destroy the corresponding relationship and message bits, therefore the tampering can be detected and localized.

The new method can be used to embed message bits into multiple wavelet scales so that the watermark embeddability can be further enhanced. Furthermore, it can help us to distinguish some normal image operations such as image compression from malicious attacks. Therefore we may determine the source of tampering. As will be shown in the simulation results, the compression has a gradually decreased impact on wavelet coefficients and fragile watermarks as the wavelet scale increases. Other malicious attacks do not have this characteristic.

Using the proposed method, the introduced image distortion is imperceptible because of the statistical approach to embed watermarks. Only a few image data in the wavelet domain is modified, no matter the watermarks are embedded at high frequency or low frequency scales.

5.3 Simulation Results

Two 512×512 images are used to demonstrate the effectiveness of the presented fragile watermarking method. One is the Lena image and the other is the peppers image. The Lena image is used to present a preliminary scheme of the fragile watermarking method, while the peppers image is applied in a more complicated scheme to embed watermarks into multiple wavelet scales.

In the experiment of the Lena image, we select the vertical wavelet subspace (or the HL wavelet subspace) at the first scale to embed the authentication information. The wavelet subspace is divided into 4×4 equal-size blocks. We name the first block at the left upper corner as the watermark block where authentication information is embedded and the part consisting of the remaining fifteen blocks as the reference block. According to (5.1), we modify the large coefficients in the watermark block so that its large variance parameter has the same value as that of the reference block after modification. This specially formed relationship is used by us to authenticate an image or to detect image tampering.

Figure 5.3 shows the original Lena image and the watermarked image for quality comparison in the spatial domain. The left upper corner enclosed by two marking lines is where the authentication information is embedded. As can be seen, the modifications are hardly detected by human vision. The watermarked image looks just like a perfect copy of the original image.

Figure 5.4 displays the HL wavelet subspace before and after watermarking. Before watermarking, the large variance parameter is 41.32 for the watermark block and 246.70 for the reference block. After watermarking, the parameter of the watermark block is adjusted to 246.70, the same value as that of the reference block. It can be observed from Figure 5.4 that only a few coefficients in the watermark block are modified. Actually, the total number of modified coefficients is 44, which ensures an imperceptible alteration of original image contents after watermarking.

In the experiment of the peppers image, our lab logo "CASPAL" is embedded as the fragile watermark. The logo is first translated into a binary bit stream. Since at least five



Figure 5.3: The Original Lena Image (a) and the Watermarked Lena Image (b).

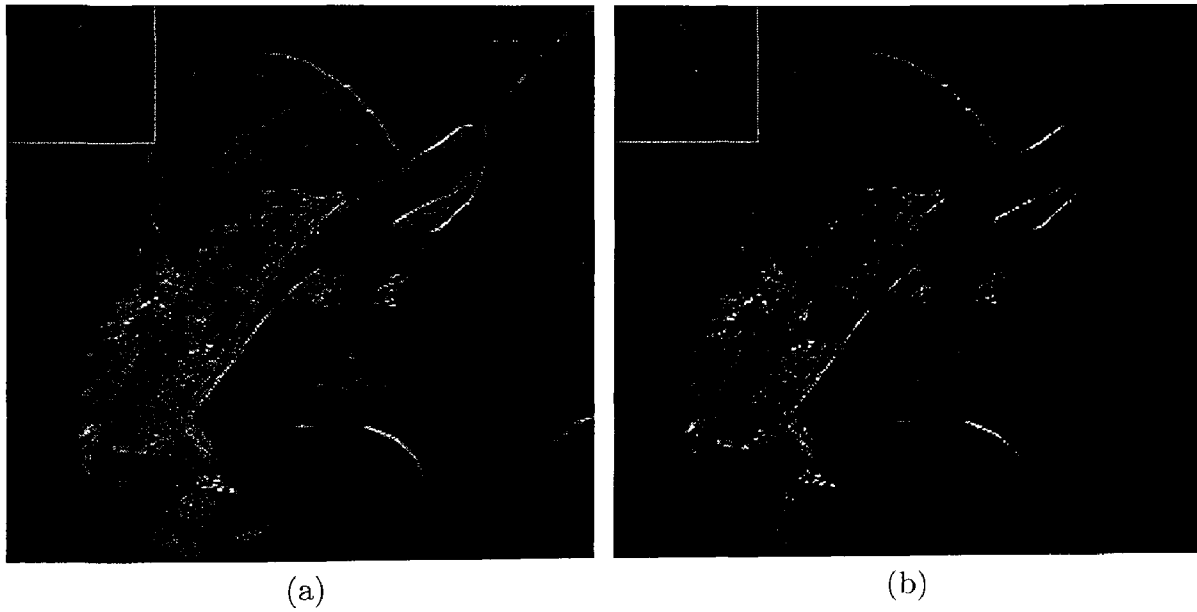


Figure 5.4: The HL Wavelet Subspace of Original Lena Image (a) and the HL Wavelet Subspace of Watermarked Lena Image (b).

bits are required to encode the alphabet (00001 for A, 00010 for B, so on ...), the total number of bits required to represent the logo is $5 \times 6 = 30$. According to TABLE 5.1, at least 15 blocks are needed at each wavelet subspace. To facilitate the operation, each wavelet subspace is divided into 16 blocks so that 32 message bits representing the logo are embedded. Figure 5.5 shows the wavelet subspaces with 32 message bits embedded into 16 divided wavelet blocks. Every three wavelet blocks obtained at the same position from the wavelet subspaces are used to embed two message bits. For example, the three shaded wavelet blocks embed message bits "00", which are the initial bits of the letter "C", using the relationship shown in TABLE 5.1.

				Vertical Subspace			
				00	01	10	00
				01	10	01	11
				00	00	00	00
				10	11	00	00
00	01	10	00	00	01	10	00
01	10	01	11	01	10	01	11
00	00	00	00	00	00	00	00
10	11	00	00	10	11	00	00
Horizontal Subspace				Diagonal Subspace			

Figure 5.5: Message Bits Embedded into the Wavelet Blocks.

Figure 5.6 displays the original peppers image and the watermarked peppers image. As can be seen, the embedding of the lab logo "CASPAL" doesn't cause any perceptible distortion in the watermarked image.

The main advantage of the proposed fragile watermarking method is that it doesn't need to modify much image data and doesn't have much image distortion. Figure 5.7 plots

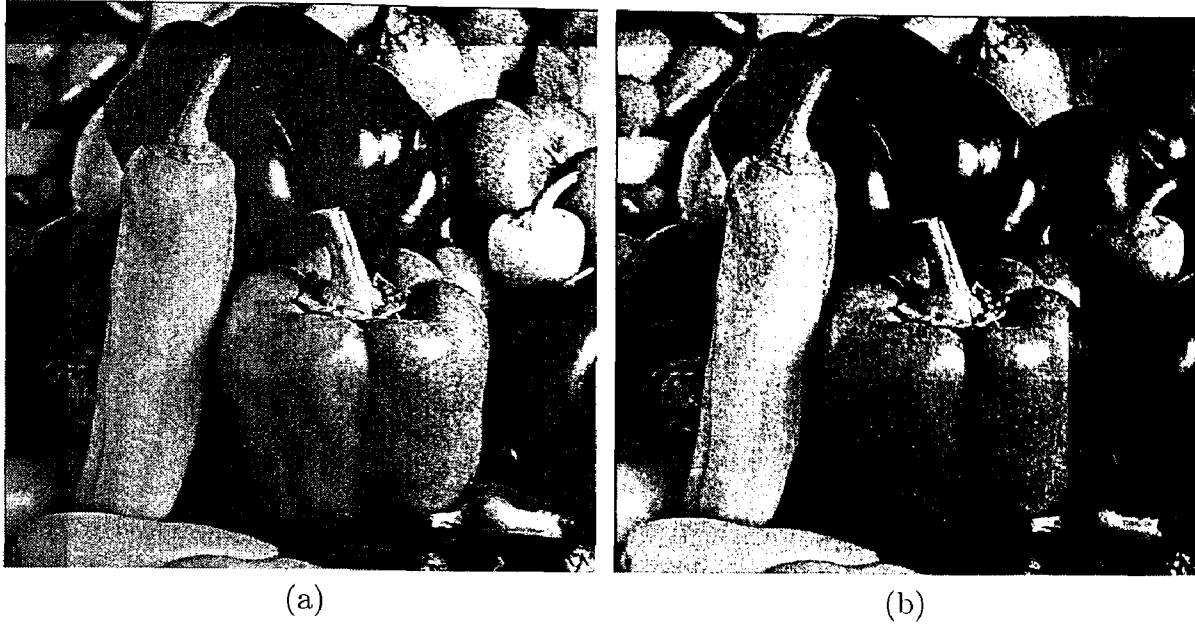


Figure 5.6: The Original Peppers Image (a) and the Watermarked Peppers Image (b) with Lab Logo “CASPAL” Embedded.

the wavelet coefficients of the original peppers image and the watermarked peppers image, respectively. The changes made on the wavelet coefficients can be observed by comparing the two plots. It is calculated that the total number of modified coefficients in the watermarked image is 680 (out of 512×512 image data). Compared with some conventional fragile watermarking methods [11] [13] [14] that modify nearly half of the image pixels, the statistical model based approach modifies much fewer image data. Besides, the changes are only made on large coefficients that represent image edges in the space domain. When watermarks are embedded at image edges, they are generally more imperceptible by human vision, as shown in Figure 5.6. If we think of watermarks as a kind of noise introduced in the host image, the PSNR (peak-signal-to-noise-ratio) is an appropriate indicator of how much image distortion is involved in the watermarking process. In our experiment, the PSNR is 52.12 db, which indicates very few image distortion involved.

To test the sensitivity of the watermark detection, first we perform a single pixel tampering experiment by 20 times, in each of which a randomly selected pixel is modified by

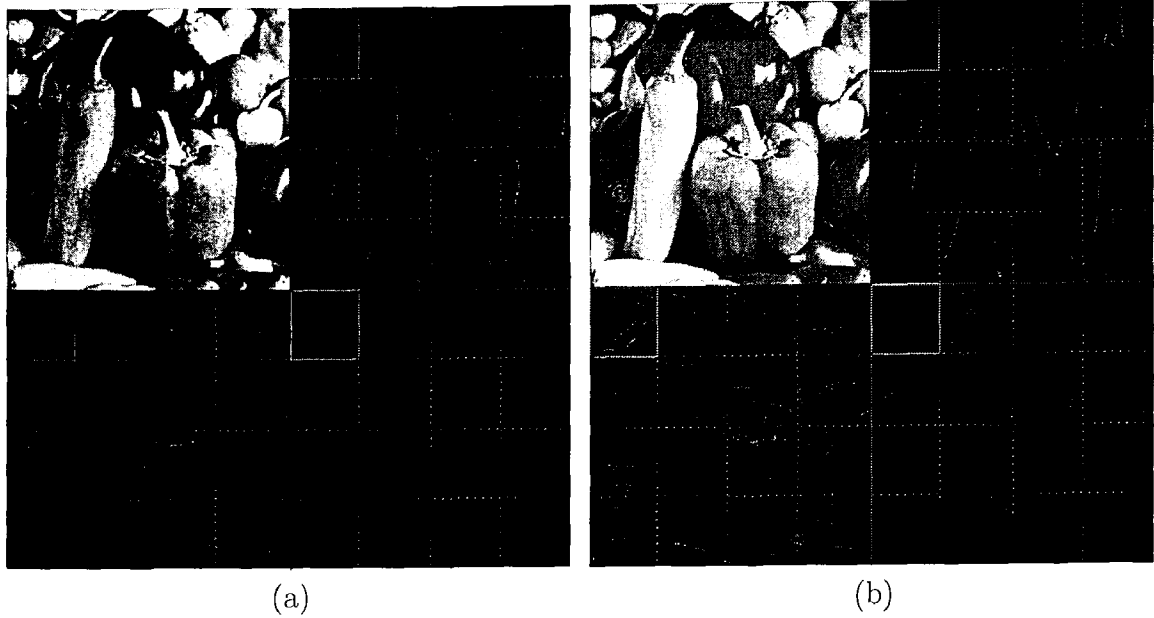


Figure 5.7: The DWT of Original Peppers Image (a) and the DWT of Watermarked Peppers Image (b) with Lab Logo “CASPAL” Embedded.

a small amount and its impact to the embedded watermark is recorded. TABLE 5.2 shows the mean value of relative parameter differences deviated from the constructed parameter equity relationship as in TABLE 5.1, which can represent the sensitivity of the watermark detection.

Number of Trials	20
Average Parameter Difference Off Balance	0.11%
Corresponding Message Bits Destroyed	Yes

Table 5.2: Average Parameter Difference Caused by Single Pixel Modification.

As can be seen, no matter how slight the tampering is, it will be detected because it destroys the formed parameter equity relationship by a noticeable amount. The location of the tampering can be determined since it only destroys the message bits at positions of the tampering.

TABLE 5.3 lists the relative parameter differences caused by the additive noise with different variances using the new watermark embedding method at two scales. As can be seen, no matter how slight the tampering is, the previously formed parameter equity relationship will be broken and the tampering will be detected. Moreover, the parameter differences tend to become larger with the increase of the extent of tampering.

		Noise Variance			
Scale Level		0.0001	0.0002	0.0005	0.0010
	1	1.25%	2.37%	6.06%	19.42%
	2	4.65%	8.59%	10.27%	17.49%

Table 5.3: Parameter Difference Caused by Noise.

By embedding the fragile watermarks at multiple wavelet scales, we can distinguish some normal image operations such as image compression from malicious attacks. TABLE 5.4 shows the impact of the JPEG compression on the watermarked image that has watermarks embedded at scales 1 to 4. The numbers in Table 5.4 represent relative parameter differences. It can be seen that at the same compression level, the relative parameter difference decreases as the wavelet scale increases. On the other hand, some malicious attacks, including additive Gaussian white noise and deliberate slight change of image contents, are simulated. The simulation results are shown in TABLE 5.5. The parameter changes due to these malicious attacks do not have the same characteristic as the image compression. Therefore we may use this feature to distinguish image compression from these malicious attacks.

		Compression Ratio			
Scale Level		60%	38%	25%	15%
	1	1.47%	2.36%	2.86%	4.53%
	2	0.84%	1.27%	1.76%	2.75%
	3	0.21%	0.31%	1.18%	1.99%
	4	0.09%	0.23%	0.54%	0.81%

Table 5.4: Parameter Difference Caused by Compression.

Malicious Attacks			
Scale Level		Gaussian White Noise	Content Change
	1	3.06%	1.65%
	2	7.27%	0.34%
	3	6.13%	0.98%
	4	14.75%	1.13%

Table 5.5: Parameter Difference Caused by Some Malicious Attacks.

5.4 Summary

In this chapter, a new fragile watermarking method is developed. The new method utilizes statistical model parameters to embed watermarks for image authentication. The new method modifies only a very small amount of image data and has a virtually imperceptible alteration of the original image. Integrated with some coding techniques, the new method can easily embed authentication messages such as personal signatures or logos into the host image. Any unauthorized changes that remove the embedded watermarks will be detected and localized. Embedding of watermarks at multiple wavelet scales is able to enhance the robustness of tampering detection. On the other hand, it can help distinguish some normal image operations, such as compression, from those deliberate or malicious attacks.

Chapter 6

Conclusions and Future Work

Many image applications are based on the image characteristics obtained from image modelling. Since the wavelet transform has a multiresolution image decomposition that is quite compliant with human vision characteristics, an efficient image interpretation through image modelling in the wavelet domain is preferred and sought in our research. In this thesis, we notice the wavelet coefficients have a peaky, heavy tailed marginal distribution and hence develop a statistical modelling method based on GMM and GGMM. The statistical modelling is able to depict wavelet coefficients flexibly and accurately through a variety of Gaussian components being employed. As for statistical model parameters, they are usually estimated by using an EM algorithm approach. In this thesis, we specifically develop some new EM algorithms for GMM and GGMM to help estimate their model parameters.

Based on the statistical modelling and the obtained model parameters, a new image retrieval system and a novel fragile watermarking method are developed.

The image retrieval system employs a new feature extraction method based on the statistical modelling and a new similarity measure based on the Kullback divergence. It is observed that the extracted image features are able to represent image texture contents effectively and the new similarity measure outperforms traditional measures using Minkowski distances. We notice that the traditional Minkowski distances are not very effective for the comparison of features extracted from statistical modelling. Therefore we transform the problem of measuring the similarity of two feature vectors to the problem of measuring the

similarity of two distinct model distributions, where the Kullback divergence can be applied. Simulation results indicate that the Kullback divergence based similarity measure achieves a higher image retrieval rate than the Minkowski distance based similarity measures. Besides, since we develop a separate Kullback approach for distance computation, the computational complexity is retained at the same level as that of the Minkowski distances. The new image retrieval system is also compared with some other traditional methods and it can be observed from the experiments that the new system achieves a better retrieval performance.

In the fragile watermarking, the major concern in our research is how to have an efficient embedding of watermarks to authenticate an image. Most conventional methods are able to satisfy the basic authentication requirement of detecting and localizing unauthorized image tampering. However, they seldom address the 'efficient' embedding problem. The efficient embedding can be interpreted as modifying as fewer image data as possible to embed the watermarks imperceptibly into the host image. In our research, we develop a novel fragile watermarking method by utilizing the statistical model parameters. We manage to set up some special relationships among the model parameters to authenticate an image. One attractive advantage of the novel method is that it can embed practical authentication messages such as personal signatures or logos into the host image. At the same time, the embedded messages are able to detect and localize any image tampering. The new method has an efficient embedding of watermarks. Since the watermarking process uses a statistical approach and modifies only large coefficients, the image data modified for watermark embedding is far less than most traditional methods. Besides, the modification made on large coefficients are virtually imperceptible because large coefficients usually represent image texture edges that when modified, are generally unnoticed by human vision. In our research, we also successfully develop a semi-fragile watermarking approach by embedding the watermarks at multiple wavelet scales. The interest for a semi-fragile watermark lies in the application where people wish the watermarks are not so 'fragile' to certain image operations such as compression, while keeping fragile to malicious attacks. The semi-fragile application is quite practical nowadays, considering people frequently wish to distribute and

transmit the watermarked images over the Internet where the images are often stored in an compressed format.

In the future, we are interested in furthering our studies in the following areas:

1. In the statistical modelling of wavelet coefficients using GGMM, the exponent β has a pre-determined or fixed value for all mixed Gaussian components. On the other hand, if we take β as another unknown variable for each Gaussian component, that will lead to a fully flexible GGMM. Therefore, in the two state representation of GGMM, the Gaussian component with a small variance will have a relatively large β value, while the Gaussian component with a large variance will have a relatively small β value. The problem now is how to develop an EM algorithm to obtain the model parameters. The EM algorithm involving the estimation of β is extremely complicated, which remains to be studied in our future research. We hope that the new GGMM with flexible Gaussian exponents has a more accurate description of wavelet coefficients. When applied to image retrieval, it can further improve the retrieval performance.
2. The robust watermarks can survive any malicious attacks but have no way to tell where and how the image is attacked. On the contrary, the fragile watermarks are not resistant to image tampering but can detect and localize it if there's any. There is some applications where the two types of watermarks can be integrated together as a hybrid watermark into a host image. The attacking information provided by the fragile watermarks can help the robust watermarks have a better encoding and decoding strategy. Besides, the two watermarks can share some important information such as a secret key with each other so that the whole watermarking scheme is more secure and reliable;
3. The JPEG 2000 compression standard may become a major image compression tool, especially for those images distributed on the Internet. From an application point of view, there is a need to further develop our fragile watermarking scheme to make it

compliant with JPEG 2000 compression, which means, the embedded watermarks are able to resist the JPEG 2000 compression but keep fragile to other image operations and tampering at the same time. The fact that the JPEG 2000 compression standard and our fragile watermarking scheme both work in the wavelet domain will facilitate our work to achieve this objective based on the same statistical modelling.

Appendix A

The Newton-Raphson Method

The Newton-Raphson method is a widely used root-finding algorithm which uses the first few terms of the Taylor series of a function to numerically search for its solutions. The Taylor series of $f(x)$ about the point $x = x_0 + \epsilon$ is given by:

$$f(x_0 + \epsilon) = f(x_0) + f'(x_0)\epsilon + \frac{1}{2}f''(x_0)\epsilon^2 + \dots \quad (\text{A.1})$$

Keeping terms only to first order,

$$f(x_0 + \epsilon) \approx f(x_0) + f'(x_0)\epsilon. \quad (\text{A.2})$$

This expression can be used to estimate the amount of offset ϵ needed to land closer to the root starting from an initial guess x_0 . Setting $f(x_0 + \epsilon) = 0$ and solving (A.2) for $\epsilon \equiv \epsilon_0$ gives:

$$\epsilon_0 = -\frac{f(x_0)}{f'(x_0)}, \quad (\text{A.3})$$

which is the first-order adjustment to the root's position. By letting $x_1 = x_0 + \epsilon_0$, calculating a new ϵ_1 , and so on, the process can be repeated until it converges to a root using:

$$\epsilon_n = -\frac{f(x_n)}{f'(x_n)}. \quad (\text{A.4})$$

Unfortunately, this procedure can be unstable near a horizontal asymptote or a local extremum. However, with a good initial choice of the root's position, the algorithm can be applied iteratively to obtain:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (\text{A.5})$$

for $n = 1, 2, 3, \dots$. An initial point x_0 that provides safe convergence of Newton's method is called an approximate zero.

Bibliography

- [1] M. Flickner, H. Sawhney, W. Niblack, and *et al.*, "Query by image and video content: the QBIC system," *IEEE Computer*, vol. 3, no. 9, pp. 23–32, 1995.
- [2] J. Bach, C. Fuller, A. Gupta, and *et al.*, "The virage image search engine: An open framework for image management," *Proc. SPIE Storage and Retrieval for Image and Video Databases*, vol. 2670, pp. 76–87, 1996.
- [3] J. R. Smith and S. Chang, "Visualeek: a fully automated content-based image query system," *ACM Multimedia*, pp. 87–98, 1996.
- [4] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: content-based manipulation of image databases," *International Journal of Computer Vision*, vol. 18, no. 3, pp. 233–254, June 1996.
- [5] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. on PAMI*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
- [6] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. on Image Processing*, vol. 2, pp. 429–441, Oct. 1993.
- [7] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. on Image Processing*, vol. 11, no. 2, pp. 146–158, Feb. 2002.
- [8] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: past, present, and future," *Journal of Visual Communication and Image Representation*, vol. 10, pp. 1–23, 1999.

- [9] R. Picard and T. P. Minka, "Vision texture for annotation," *Journal of Multimedia Systems*, vol. 3, no. 1, pp. 3–14, 1995.
- [10] S. Sclaroff, L. Taycher, and M. L. Cascia, "Imagerover: A content-based image browser for the world wide web," *IEEE Workshop on Content-based Access of Image and Video Libraries*, pp. 2–9, June 1997.
- [11] S. Walton, "Information authentication for a slippery new age," *Dr. Dobbs Journal*, vol. 20, no. 4, pp. 18–26, Apr. 1995.
- [12] P. W. Wong, "A watermark for image integrity and ownership verification," *Proc. IS&T PIC Conf.*, pp. 374–379, May 1998.
- [13] P. W. Wong, "A public key watermark for image verification and authentication," *Proc. IEEE Int. Conf. on Image Processing*, vol. 1, pp. 445–449, Oct. 1998.
- [14] A. Paquet and R. Ward, "Wavelet-based digital watermarking for image authentication," *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering*, vol. 2, pp. 879–884, May 2002.
- [15] J. Fridrich, "A hybrid watermark for tamper detection in digital images," *Proc. Int. Symposium on Signal Processing and its Applications*, pp. 301–304, Aug. 1999.
- [16] J. Bilmes, "A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," 1997.
- [17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society Series B*, vol. 39, no. 1, pp. 1–38, Nov. 1977.
- [18] R. A. Rednet and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195–239, Apr. 1984.

- [19] M. Jordan and R. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Neural Computation*, vol. 6, no. 2, pp. 181–214, Mar. 1994.
- [20] S. Kullback and R. A. Leibler, "On information and suiciency," *Annals of Math. Statistics*, vol. 22, pp. 79–86, 1951.
- [21] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications. John Wiley & Sons, 1991.
- [22] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. on Image Processing*, vol. 10, no. 7, pp. 1056–1068, July 2001.
- [23] H. Yuan, X.-P. Zhang, and L. Guan, "A statistical approach for image feature extraction in the wavelet domain," *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering*, vol. 2, pp. 1159–1162, May 2003.
- [24] H. Yuan, X.-P. Zhang, and L. Guan, "Content-based image retrieval using a Gaussian mixture model in the wavelet domain," *Proc. SPIE Visual Communications and Image Processing*, vol. 5150, pp. 422–429, June 2003.
- [25] H. Yuan and X. P. Zhang, "Texture image retrieval based on a Gaussian mixture model and similarity measure using a Kullback divergence," *Proc. IEEE Int. Conf. on Multimedia & Expo*, June 2004.
- [26] Q. Cheng and T. S. Huang, "Robust optimum detection of transform domain multiplicative watermarks," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 906–924, Apr. 2003.
- [27] H. Yuan and X. P. Zhang, "Fragile watermark based on the Gaussian mixture model in the wavelet domain for image authentication," *Proc. IEEE Int. Conf. on Image Processing*, vol. 1, pp. 505–508, Sept. 2003.

- [28] H. Yuan and X. P. Zhang, "A multiscale fragile watermark based on the Gaussian mixture model in the wavelet domain," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May 2004.

VITA

NAME: Hua Yuan

PLACE OF BIRTH: Shanghai, China

YEAR OF BIRTH: 1974

POST-SECONDARY EDUCATION
AND DEGREES: Fudan University
Shanghai, China
1992-1997, B.Sc

HONORS AND AWARDS: The People's Scholarships
1994-1996

RELATED WORK EXPERIENCE: Software Engineer
Alcatel Shanghai Bell Corporation
1997-2001

PUBLICATIONS

1. H. Yuan, X.-P. Zhang and L. Guan, "A statistical approach for image feature extraction in the wavelet domain," *Proc. IEEE Canadian Conf. on Electrical and Computer Engineering (CCECE)*, vol. 2, pp. 1159-1162, May 2003.
2. H. Yuan, X.-P. Zhang and L. Guan, "Content-based image retrieval using a Gaussian mixture model in the wavelet domain," *Proc. SPIE Visual Communications and Image Processing (VCIP)*, vol. 5150, pp. 422-429, June 2003.
3. H. Yuan and X.-P. Zhang, "Fragile watermark based on the Gaussian mixture model in the wavelet domain for image authentication," *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, vol. 1, pp. 505-508, Sep. 2003.
4. H. Yuan and X.-P. Zhang, "A multiscale fragile watermark based on the Gaussian mixture model in the wavelet domain," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 17-21, 2004.
5. H. Yuan and X.-P. Zhang, "Texture image retrieval based on a Gaussian mixture model and similarity measure using a Kullback divergence," *Proc. IEEE Int. Conf. on Multimedia & Expo (ICME)*, Taipei, Taiwan, June 27-30, 2004.

1944

1944

1944

1944

1944

1944

1944

1944

1944

1944

1944

1944

1944

1944

1944

1944

200006-196