A ROBUST REGION-ADAPTIVE DIGITAL IMAGE WATERMARKING SYSTEM

by

CHUNLIN SONG

A thesis submitted to Liverpool John Moores University

in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing & Mathematical Sciences

Liverpool John Moores University

May, 2012

The following figures have been omitted on request of the university –

Fig. 2.1 (p.11)

Fig 6.1 (p.139)

Fig 6.2 (p.141)

Fig 6.3 (p.142)

Fig 6.4 (p.143)

Fig 6.5 (p.145)

TO MY PARENTS

老爸：宋连城

老妈：冯　利

Abstract

Digital image watermarking techniques have drawn the attention of researchers and practitioners as a means of protecting copyright in digital images. The technique involves a subset of information-hiding technologies, which work by embedding information into a host image without perceptually altering the appearance of the host image. Despite progress in digital image watermarking technology, the main objectives of the majority of research in this area remain improvements in the imperceptibility and robustness of the watermark to attacks.

Watermark attacks are often deliberately applied to a watermarked image in order to remove or destroy any watermark signals in the host data. The purpose of the attack is aimed at disabling the copyright protection system offered by watermarking technology. Our research in the area of watermark attacks found a number of different types, which can be classified into a number of categories including removal attacks, geometry attacks, cryptographic attacks and protocol attacks. Our research also found that both pixel domain and transform domain watermarking techniques share similar levels of sensitivity to these attacks.

The experiment conducted to analyse the effects of different attacks on watermarked data provided us with the conclusion that each attack affects the high and low frequency part of the watermarked image spectrum differently. Furthermore, the findings also showed that the effects of an attack can be alleviated by using a watermark image with a similar frequency spectrum to that of the host image. The results of this experiment led us to a hypothesis that would be proven by applying a watermark embedding technique which takes into account all of the above phenomena. We call this technique 'region-adaptive watermarking'.

Region-adaptive watermarking is a novel embedding technique where the watermark data is embedded in different regions of the host image. The embedding algorithms use discrete wavelet transforms and a combination of discrete wavelet transforms and singular value decomposition, respectively. This technique is derived from the earlier hypothesis that the robustness of a watermarking process can be improved by using watermark data in the frequency spectrum that are not too dissimilar to that of the host

data. To facilitate this, the technique utilises dual watermarking technologies and embeds parts of the watermark images into selected regions of the host image. Our experiment shows that our technique improves the robustness of the watermark data to image processing and geometric attacks, thus validating the earlier hypothesis.

In addition to improving the robustness of the watermark to attacks, we can also show a novel use for the region-adaptive watermarking technique as a means of detecting whether certain types of attack have occurred. This is a unique feature of our watermarking algorithm, which separates it from other state-of-the-art techniques. The watermark detection process uses coefficients derived from the region-adaptive watermarking algorithm in a linear classifier. The experiment conducted to validate this feature shows that, on average, 94.5% of all watermark attacks can be correctly detected and identified.

Keywords

Digital image watermark, Watermark attack analysis, Region-adaptive watermarking system, Tamper detection

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

Many parties have helped in the completion of this thesis. Firstly, I would like to express my sincere thanks to my supervisors, Dr Sud Sudirman and Professor Madjid Merabti, for their support and guidance during my PhD. They have given me a very rewarding research experience. Their suggestions on and deep insight into academic matters contributed greatly to the completion of this thesis, as well as to my academic achievements.

Current and former staff at the school of computing and mathematics have been very helpful. Professor Abdennour EI Rhalibi, Dr Llewellyn-Jones, Dr Bo Zhou and Mr Stephen Tang have provided me with valuable advice and support.

Many thanks to my friends: Yi Zhang, Ge Zhang, Chang Su, Siyuan Xu, Tianyi Zhao, Xiaoxiang Su, Yaodong Gu, Chen Chen, Qiaoqiao Wang, Jun Qu, Yiyi Zou, Li Yu, Lu Sun, Jianfeng Luan, Xing Li, Qingdan Feng, Haseeb ur Rahman and Ricardo Duarte. In addition, special thanks to Jesus Christ, my church members and all missionaries for sharing their love and gospel with me.

My parents have been my source of strength and faith throughout these years, and I must thank them for their unconditional love, courage, support and understanding.

CHAPTER 1
INTRODUCTION

Images make up a major component of multimedia content. Examples of images are digital arts, illustrative diagrams, cultural heritage paintings in digitized form and digital photographs. Advances in computing hardware, software, and networks have created threats to copyright protection and content integrity. For instance, nowadays images can be copied, modified and distributed with great ease. Digital watermarking has recently become a popular research area due to the proliferation of digital data in this internet age and the need to find a solution to protect the copyright of these materials. As such, the watermarking process embeds a signal into an image without significantly degrading its visual quality, after which it can be made public or sent to the end user. Later, the extracted watermark can be used for the purposes of copyright protection and content authentication.

Robustness is one of the major characteristics that influence the performance and applications of digital image watermarks and in this context it means the ability of a watermark to resist attack. Watermark systems can be categorised into two major groups based on their robustness, namely robust watermarking systems and fragile watermarking systems. Robust watermarks are required to resist any modifications which do not decrease the commercial value of image. On the other hand, fragile

watermarks are designed to fail when the image is modified, so in the other words they are very sensitive and easily destroyed by image modifications.

Copyright protection concerns the positive identification of content ownership in order to protect the rights of the owner. Robust watermarks can be used in copyright protection because they are persistently associated with an image. When extracted, they can be used to identify the copyright holder of the watermarked image. If an attempt is made to remove the watermark, it should result in a severe degradation of the image's visual quality. A fragile watermark, on the other hand, can be easily altered or destroyed when the host image is modified or attacked. The sensitivity of fragile watermarks to modification leads to their being used in image authentication in that it may be of interest for parties to verify that an image has not been edited, damaged or altered since it was marked.

In this thesis, a robust watermarking algorithm for copyright protection will be investigated. Furthermore, different watermark attacks will be analysed and a novel application that uses a robust watermarking system will be introduced.

In this chapter, an overview of digital image watermarking will be given, including the historical background to the watermarking system. The aims, objectives and novel contributions of the research described in this thesis will be outlined in the last part of this chapter, together with the thesis's organisation.

1.1. Historical Background on Digital Watermarking Technology

Watermarking shares common basic theories with steganography and other related information-hiding techniques. Information-hiding is a wide subject area which covers all techniques that insert hidden messages into or onto a medium such as texts or images in such a way that, apart from the message sender and the intended recipient, no one would realise that such message exists. It has been used for many hundreds of years, and is recorded as far back as the fifth century B.C to a Greek tyrant named Histaiaeus, who wanted to revolt against the Persian King's rule. In order to inform his friend Aristagoras of Miletus that it was time to begin a revolt, he shaved the head of one of his trusted slaves and tattooed the secret message on his scalp. He then waited for the hair to grow back before sending the slave to deliver the message. The message reached his correspondents in Persia safely and the revolt succeeded.

Æneas the Tactician is another famous Greek who contributed to communication security. He developed an information-hiding technique whereby holes representing letters of the Greek alphabet were bored into a wooden disk. Yarn was then threaded through the holes in an order that would spell out the message. The decoder would simply reverse the process, writing the letters down backwards, to reconstruct the message [Ingemar J. Cox, 2008].

Gaspar Schott (1608-1666), a German scientist specialising in the fields of physics, mathematics and natural philosophy, invented a technique that encodes secret information onto musical notes. The technique works by matching letters to specific musical notes (Figure 1.1), although this music would never sound pleasing [Kipper, 2004] .

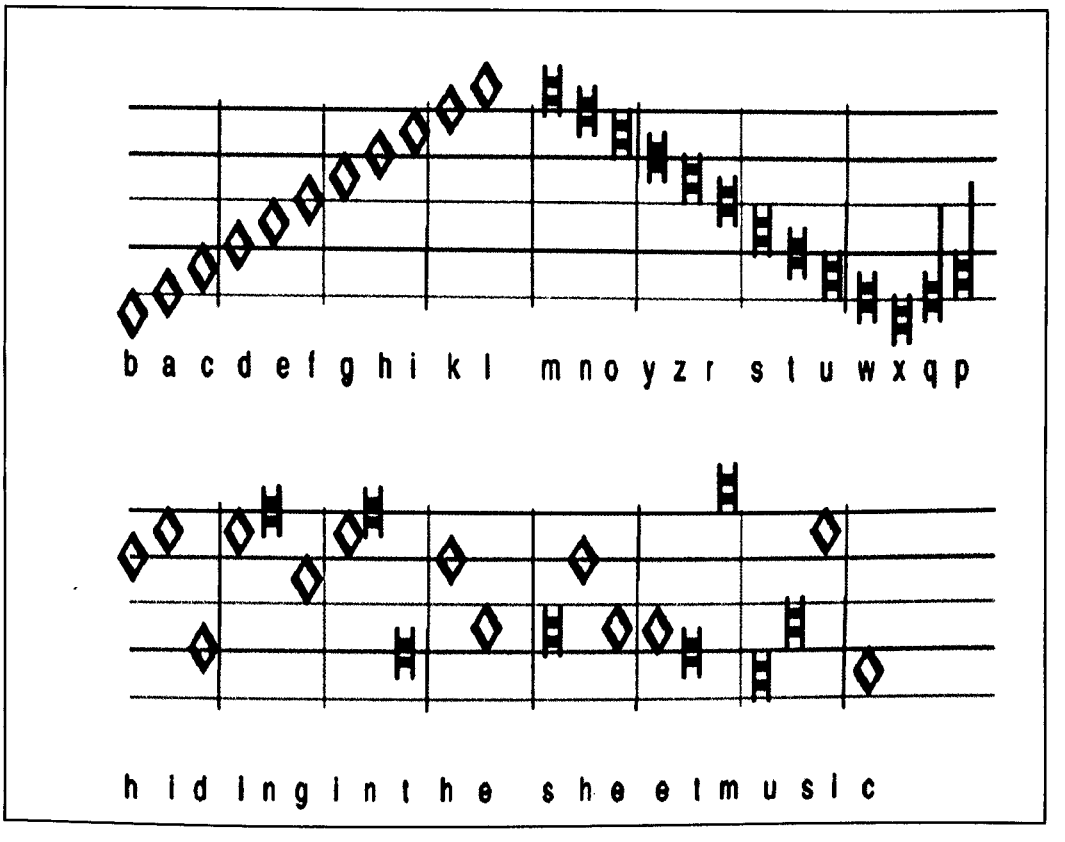


Figure 1.1 Hiding information in music

Another famous name in data-hiding history is Girolamo Cardano, who created the Cardano grille system. The basis of the system is that each recipient has a piece of paper with several holes cut in it. When the "grille" is placed over an innocuous-looking message, its holes line up with words in the larger message to produce the hidden message. This technique was enhanced in World War I and was re-named the

Turning Grille, which looked like a normal grille in that it was a square sheet of cardboard divided into cells, with some of the cells punched out. To use the Turning Grille, the encoder would write the first sequence of letters, then rotate the grille 90 degrees and write the second sequence of letters, rotating the grille after each sequence.

Although it has historical uses, the science of information-hiding is a modern subject that has been rapidly developed in recent years. Digital watermarking technology is a subject that belongs to this area. The first information-hiding workshop [R.Anderson, 1996] , which included digital watermarking as one of its primary topics, was held in 1996. The SPIE began to devote a conference specifically to this subject, named 'Security and Watermarking of Multimedia Contents' [Wong, 1999, Wong, 2000] and beginning in 1999.

In addition, about this time, several organisations began considering watermarking technology for inclusion in various standards. The Copy Protection Technical Working Group (GPTWG) [E.Bell, 1999] tested a system for the protection of video on DVD disks, while the Secure Digital Music Initiative (SDMI) made it a central component of their system for protecting music. Two projects sponsored by the European Union, VIVA [G.Depovere, 1999] and Talisman [Hartung, 1999], tested watermarking for broadcast monitoring, and the International Organisation for Standardization (ISO) took an interest in the technology in the context of designing advanced MPEG standards.

In the late 1990s, several companies were established to market watermarking products. Technology from the Verance Corporation was adopted into the first phase of SDMI and was used by Internet music distributors such as Liquid Audio. In the area of image watermarking, Digimarc bundled its watermark embedders and detectors with Adobe's Photoshop. Most recently, a number of companies have used watermarking technologies for a variety of applications such as digital rights management.

1.2. Overview of Digital Watermarking

This section provides an overview of digital watermarking. It covers some basic concepts, watermark properties, watermark classifications and its applications. A more detailed discussion of these topics is given in Chapter 2.

The list below contains the definition of standard terms used throughout this thesis:

- Host Image: The original image used in watermarking.
- Watermark Image: The image which is to be embedded into the host image.
- Watermarked Image: The result of embedding the watermark image into the host image. It should look similar to the host image.
- Watermark Embedding: The process of encoding a watermark image into the host image.
- Watermark Extraction: The process of decoding the watermark image out of a watermarked image.
- Watermark Attacks: Changes made to a watermarked image to evaluate its robustness.

To understand watermarking methods and to determine their applications, one needs to know understand their properties. Listed below are some fundamental watermark properties.

- Robustness: This refers to its ability to withstand non-malicious and malicious distortions.
- Data Payload and Capacity: Data payload is the encoded message size of a watermark in an image. Multi-bit watermarks can carry textual information. Capacity is an evaluation of how much information can be hidden within a digital host image. If multiple watermarks are embedded into an image, then the watermarking capacity of the image is the sum of all individual watermarks' data payload.
- Transparency: The visual similarity between the watermarked image and the host image.
- Computational Cost: The measure of computing resources and time required to perform watermark embedding or extraction processes. This can be measured by calculating the processing time for a given computer configuration.

There are several ways of classifying watermarking methods, but one of the most widely adopted classifications is based on robustness. Under this classification, a watermark can be grouped into types, namely robust watermarking systems and fragile watermarking systems. Application-wise, robust watermarks are suitable for copyright protection because they can resist common image processing operations. Conversely, fragile watermarks can be used to detect tampering and authenticate an image because they are sensitive to changes. In this thesis, we concentrate on the robust watermarking system.

Beside robustness, watermarks can also be categorised into visible and invisible types whereby visible watermarks are perceptible to a viewer but invisible watermarks are imperceptible and do not change the visual appearance of the host image. In this thesis, we are interested in invisible watermarks because they have a wider range of applications compared to visible watermarks and they do not affect the aesthetic value of an image.

It should be noted that watermarks can be embedded and extracted in different types of domain. The most direct approach is watermarking in the spatial domain, where pixel values are modified to encode the watermark signal; least significant bits (LSB) is an example in a spatial domain. On the other hand, transform domain techniques such as discrete Fourier transform (DFT), discrete cosine transform (DCT) and discrete wavelet transform (DWT) are widely used in image watermarking.

1.3. Aims & Objectives

The aim of the research in this thesis is to advance the state-of-the-art digital image watermarking technology by means of improving its transparency and robustness to different watermark attacks. In addition, the research also aims at providing a novel application of the solution in watermark attack detection.

A fundamental issue in designing a watermarking system is that it should be able to resist different watermark attacks, i.e. have the ability to withstand any distortion or damage to watermarked images. Therefore, in this thesis, a robust watermarking system will be designed accordingly.

Furthermore, application-wise, robust watermarking systems are one of the most widely used techniques for protecting digital images through, for example,

authentication copyright. However, image tamper detection restricts the robustness of a watermarking system, so in this thesis we introduce a novel application to detecting digital image tampering.

In order to fulfil the aim, a number of research specific objectives were formulated as follows:

(1) To investigate the properties of watermark attacks. The rationale for this objective is based on the fact that watermark attacks aim to destroy or create misleading watermark signals. Therefore, by analysing different watermark attacks we can evaluate and categorise them based on their properties, and then design a suitable solution to overcome them.

(2) To design and develop a new scheme to improve the robustness of watermark systems. Despite progress in digital image watermarking technology, existing watermarking systems are not sufficiently stable or strong against different watermark attacks. This is a major driving factor behind the fact that the main objective of the majority of research in this area remains devising improvements in imperceptibility and robustness to attack.

(3) To design a new watermark algorithm in such a way that it not only improves robustness criteria but also makes it possible to use the technique in a novel application.

1.4. Contributions and Thesis Outline

The major research contributions of this thesis are:

(i) The analysis of different watermark attacks. The success of watermarking technology used in a copyright protection or digital rights management system relies heavily on its strength to withstand attacks. However, watermark attacks are aimed at removing or destroying any watermark signals in the host data. This component not only describes categories of watermark attack, but also analyses removal attack and geometric attack by image histogram and image Fourier spectrum.

(ii) The development of robust watermarking methods. This part proposes a novel region-adaptive watermarking technique that can provide

improvements in both the sturdiness and visual quality of the watermarks when compared to the original, non-region-adaptive, embedding technique. The proposed technique, which is derived from the findings of the first contribution, shows that the relative difference in spectral distributions between the watermark data and the host image plays an important role in improving toughness and transparency.

(iii) A novel application for the watermarking algorithm. We have devised a new way in which the region-adaptive watermarking system can be used to detect attacks. The process uses coefficients derived from the region-adaptive watermarking algorithm in a linear classifier. And the experiment conducted to validate this feature shows that all such strikes can be correctly detected and identified.

The research process starts in Chapter 2 with a thorough literature survey on image watermarking. The results illustrate that digital image watermarking techniques have been drawing the attention of researchers and practitioners as a means of protecting copyright in digital images. The technique is a subset of information-hiding technologies, which work by embedding information in a host image, without perceptually altering the appearance of the host image.

In Chapter 3, we analyse different types of watermark attacks, which are often deliberately aimed at a watermarked image in order to remove or destroy any watermark signals in the host data. The purpose of the attack is to disable the copyright protection system offered by current technology. Our research in this area finds a number of different types of incidents, which can be classified into categories such as removal attacks, geometry attacks, cryptographic attacks and protocol attacks. Our research also finds that both pixel domain and transform domain watermarking techniques share similar levels of sensitivity to these attacks. The experiment which was conducted to analyse the effects of different removal attacks on watermarked data provided us with the conclusion that each removal attack affects high and low frequencies of the watermarked image spectrum differently. Chapters 4 and 5 introduce the region-adaptive watermarking algorithm as our solution to the research problems outlined previously. The method is a novel watermark embedding technique

where the watermark data is embedded in different regions of the host image. Two embedding algorithms are used – one uses a discrete wavelet transform (DWT) approach and the other uses a combination of DWT and the singular value decomposition (SVD) approach. These two procedures are described in Chapters 4 and 5 respectively. This technique is derived from the earlier hypothesis that the robustness of a watermarking process can be improved by using watermark data in the frequency spectrum not too dissimilar to that of the host data. To facilitate this, the system utilises dual watermarking technologies and embeds parts of the watermark images into selected regions in the host image. Our experiment shows that our technique improves the resistance of watermark data to image processing attacks as well as geometric attacks, thus validating the earlier hypothesis.

In addition to this improvement, we also present a novel use for the region-adaptive watermarking technique as a means of detecting whether certain types of attacks have occurred, which is described in Chapter 6. This is a unique feature of our watermarking algorithm, which separates it from other state-of-the-art watermarking techniques, because it uses coefficients derived from the region-adaptive watermarking algorithm in a linear classifier. The experiment conducted to validate this feature shows that on average 94.5% of all watermark attacks can be correctly detected and identified.

Conclusions will be drawn in Chapter 7. They will include our achievements in analysing different watermark attacks and designing and applying a robust watermarking system. An appendix at the end of this thesis illustrates the graphic user interface design of watermark algorithms and a description of Markov random field image segmentation.

CHAPTER 2

LITERATURE REVIEW

Digital watermarking is a widely used technique employed to insert hidden information into digital media such as images, documents, sounds and videos. In this chapter, an overview of the basic principles of digital watermarking, the progress, classification, algorithm, properties and application involved will be described. The organisation of this chapter is as follows.

The first subsection describes the basic principle behind the digital watermarking system. Its process is discussed in section 2.2, its classification is described in section 2.3, while section 2.4 provides current algorithms. The last two sections discuss the properties and applications of watermarking systems.

2.1. Introduction

Watermarking is a sub-discipline of information-hiding, which is the process of inserting hidden messages into a collection of data such as texts or images in such a way that, apart from the message sender and the intended recipient, no one would realise that such a message exists. An original classification structure for information-hiding techniques, as show in Figure 2.1, was first proposed in [Fabien A. P. Petitcolas, 1999]. From this classification, it can be seen that information-hiding

covers a wide range of techniques and applications, each with its own preference and requirement.

Figure 2.1The classification of information-hiding technology[Fabien A. P. Petitcolas, 1999]

An analogy of digital watermark is the paper watermark, which is a recognisable image or pattern on a piece of paper that appears as various shades of lightness/darkness when viewed by transmitted light, caused by thickness or density variations in the paper. The concept could extend to the digital world to provide a solution to longstanding problems faced when copyrighting digital data. Digital watermarking is the process of embedding information data into a digital signal that can be detected or extracted later to make an assertion about the data. In digital watermarking, the signal may be audio, images or video. If the signal is copied, then the information also is carried in the copy. A signal may carry several different watermarks at the same time.

The science of digital watermark technology is a modern subject that has developed rapidly in recent years. This is evident through the exponential growth of academic publications in digital watermarking over the years. In addition, research topics are becoming more specialised, e.g. robust watermarking, fragile watermarking, fingerprinting, benchmarking and so on. Furthermore, watermarking technologies have been commercialised. For example, the *Digimarc* watermark was added into

Adobe Photoshop to enable the embedding and detection of digital image watermarks. Furthermore, *Kodak*, *Sony* and *Canon* have produced cameras with image watermarking capabilities.

The research work described in this thesis focuses on digital image watermarking systems in which embedded information is known as a 'watermark image'. In other words, watermarking in digital images is the process by which a discrete data stream is hidden within an image by imposing imperceptible or perceptible changes on the image. The successive stages of the watermarking process are described in section 2.2.

2.2. Steps in the Digital Image Watermarking System

In a digital image watermarking system, the information to be embedded is called a digital watermark image, and the image where the watermark image is to be embedded is called the host image. The general process involved is illustrated in Figure 2.2.

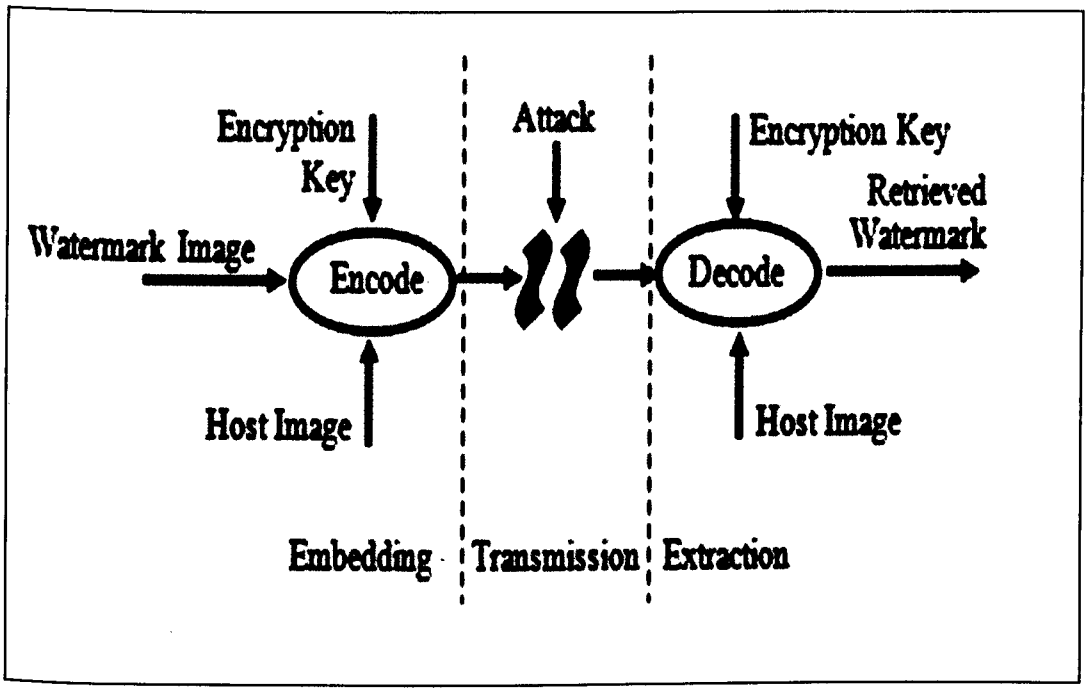


Figure 3.2 The encoding, distribution and authorisation for using watermarked media

In the embedding process, the watermark image may be encoded into the host image using a specific key, which is used to encrypt the watermark image as an additional protection level. An algorithm accepts the host and the data to be embedded, and then produces a watermarked signal.

Then, the watermarked digital signal is transmitted to end user or the public. During this transmission, or just before the decoding stage, the watermarked image may undergo some modification. This process is often referred to or considered as an attack on the watermarked image. While the modification may not be malicious, the term 'attack' arises from the copyright protection application, where software pirates attempt to remove the digital watermark through modification. There are many possible modifications, for example lossy compression of the data, cropping an image or intentionally adding noise.

Extraction is applied to the attacked/original watermarked signal in an attempt to extract the watermark. In strong digital watermarking applications, the extraction algorithm should be able to produce the watermark correctly, even if the modifications are sturdy. In fragile digital watermarking, the extraction algorithm should fail if any change is made to the signal.

2.3. Classification of a Digital Image Watermarking System

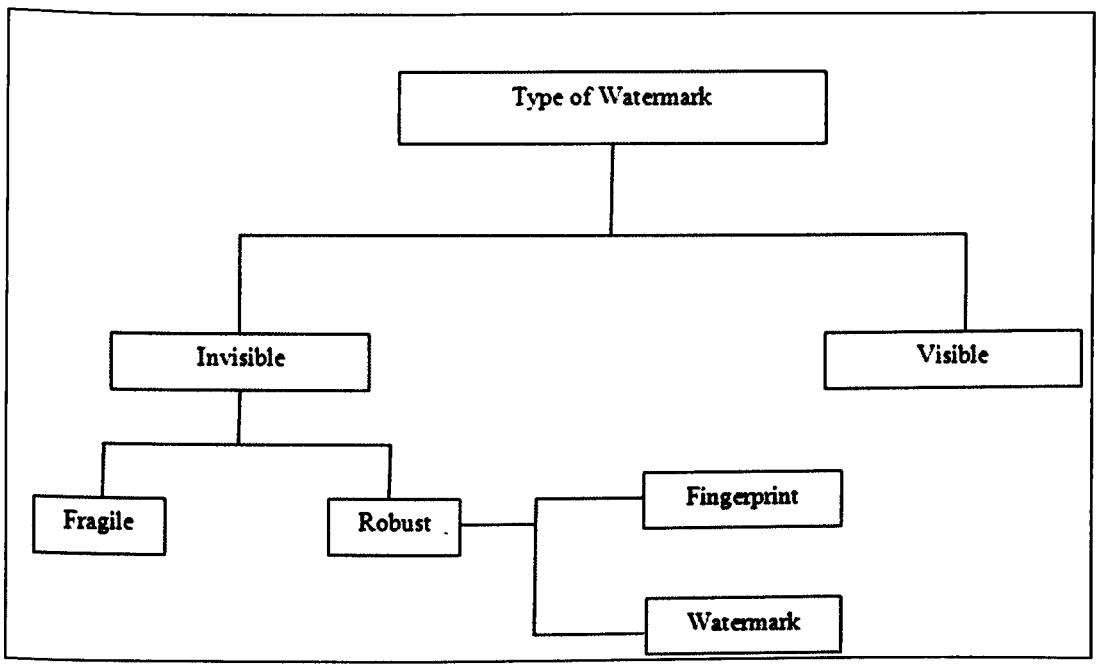


Figure 4.3 A classification of watermarking techniques

Figure 2.3 shows one of the main classification structures of watermarking techniques. In this classification, there are two types of watermarks: visible watermarking, such as different logos either on paper or on a TV screen, and invisible watermarking, which cannot be perceived by the human sensory system. An invisible watermark can be

either robust or fragile. The use of a fragile watermark is important when people want to verify whether the protected media has been tampered with or not. This type of watermark is especially designed to be as delicate as possible, so even the slightest modification to the marked media will destroy it, indicating that someone tampered with the media in question. On the other hand, robust watermarking is designed to withstand such modifications, and its use is mainly to provide proof of ownership of the media in question, even after such media has been subjected to several attempts to remove the watermark.

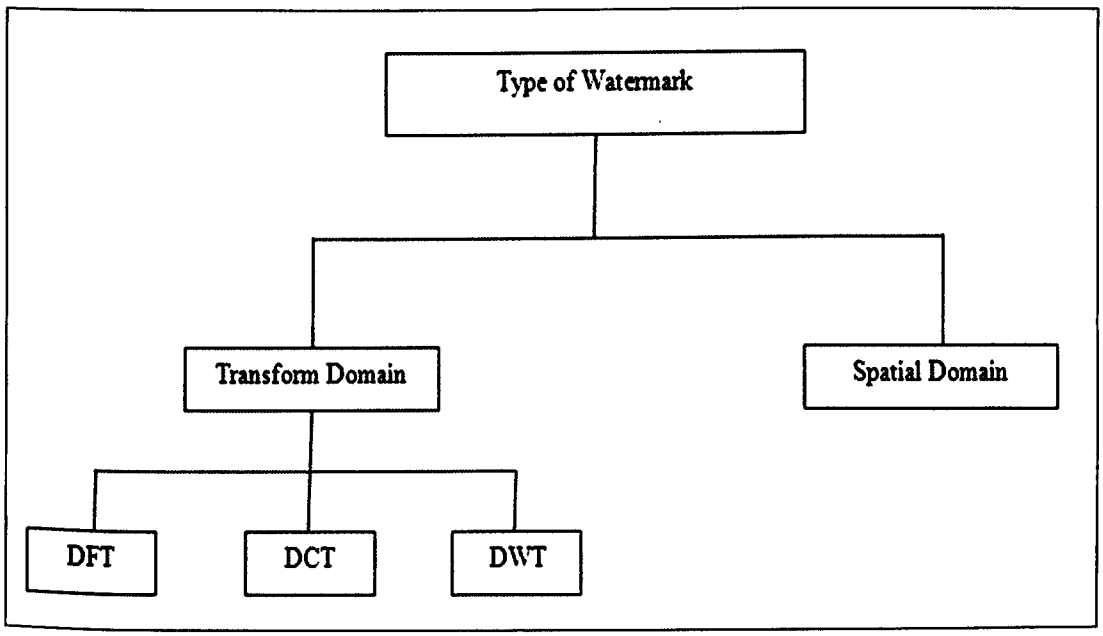


Figure 5.4 A classification of watermarking techniques based on domain type

Digital image watermarking can be also classified into a number of categories based on different sets of criteria [CL.Song, 2009]. Figure 2.4 shows such a classification based on the type of domain in which the data embedding takes place. There are two major domain types, spatial and transform domains. The most famous algorithm in the spatial domain is least significant bits (LSB) [Dey, 2007], and the most popular transform domains are discrete Fourier transform (DFT) [Ren, 2009], discrete cosine transform (DCT) [Hubballi, 2009] and discrete wavelet transform (DWT) [Sun, 2009]. Watermarking in the spatial domain is represented by modifying spatial characteristics directly, such as pixel values. On the other hand, watermarking in the transform domain converts images to the frequency signal first and then modifies the coefficient of certain frequencies.

The classification of the watermarking technique can also be based on the requirement of certain inputs, such as the encryption key, the watermark data or the host image, during the extraction process. It is divided into three types, namely blind, semi-blind and non-blind. Non-blind watermarking requires the host image, the watermark data and the encryption key (if it is used in the embedding process) for the recovery of hidden information [Zaboli, 2007]. The semi-blind system requires only the watermark data and the encryption key, whereas the blind watermarking system requires neither the host image nor the watermark data but only the secret key during the detection process [Eugene, 2007].

2.4. Advances in Digital Image Watermarking Technology

First of all, it should be noted that the reason why digital watermarking is possible, due to the limitations of the human vision system (HVS). Digital watermarks utilise this constraint to make themselves invisible, thus avoiding visual degradation of the original digital products, as well making it difficult to be identified. There have been a number of proposed novel techniques to hide watermarks in digital images. Currently these can be divided into two categories and by embedding method – spatial domain and transform domain technologies. Spatial domain techniques are developed earlier and are easier to implement, but they are more limited in terms of robustness and capacity. Conversely, transform domain techniques, which embed watermark signals in the host's transform domain, are more sophisticated compared to their spatial domain counterparts.

2.4.1 Spatial Domain

An analogue image can be described as a continuous function over a two-dimensional surface. The value of this function at a specific coordinate on the lattice specifies the luminance or brightness of the image at that location. A digital image version of this analogue image contains sampled values of the function at discrete locations or pixels. These values are said to be the representation of the image in the spatial domain or are also often referred to as the pixel domain.

Spatial domain watermarking techniques insert a watermark image directly into image pixels. The oldest and the most commonly used method in this category is the insertion of the watermark into the least significant bits (LSB) [Fabien A. P.

Petitcolas, 1999, Ingemar J. Cox, 2008, Kipper, 2004] of pixel data. The embedding process involved in the LSB technique can be described as follows.

Consider that the system is required to hide watermark number 178 in a 2×2 greyscale (8-bit) image. Let's assume that the image pixels are 234, 222, 190 and 34. In an 8-bit binary format the number 178 is represented as 10110010. Since there are four pixels that can be used to store this data, we can easily decide to embed pairs of bits of the watermark in the last two insignificant bits of the pixels. The process therefore modifies the original bits from 11101010, 11011110, 10111110 and 00100010 to 11101010, 11011111, 10111100 and 00100010, respectively.

One of the major limitations in the spatial domain is the capacity of an image to hold the watermark. In the case of the LSB technique, this capacity can be increased by using more bits for watermark embedding, but this is at the cost of higher detection rates. Alternatively, capacity can be improved by means of lossy embedding the watermark. In the latter approach, the watermark is quantised before the embedding process. Overcoming this limitation seems to be one of the major driving factors in spatial domain research.

LSB is not the only spatial domain watermarking technique. Another spatial domain technique -- colour digital watermarking scheme, described in [Qian-Chuan, 2008], utilises an algorithm based on the Lorenz map and the Arnold cat map. According to the characteristics of the human vision system, the colour space of the colour host image is first converted from RGB to YCbCr, which is then converted to greyscale. Subsequently, the embedding positions of the watermark signals are encrypted by means of discrete Arnold transform. In the final step, the watermark signals are embedded into the host image by using the average grey value and mean square error (MSE) around the embedding point of the host image. The results show that the scheme is secure and robust enough for commonly used image processing functions such as JPEG compression, noise, cropping, rotation, etc.

In [Nasir, 2007], a watermarking algorithm based on colour image in the spatial domain is proposed. This paper presents a new robust watermarking scheme for colour images based on a block probability in the spatial domain. A binary watermark image is permutated using sequence numbers generated by a secret key and grey code, and then embedded four times in different positions by a secret key in order to protect

against cropping attack from the bottom, the top, the left or the right sides of the watermarked image. Each bit of the binary-encoded watermark is embedded by modifying the intensities of a non-overlapping block of 8×8 of the blue component of the host image. The extraction of the watermark is done by comparing the intensities of an 8×8 block of the image and the original image, and then by calculating the probability of detecting 0 or 1. The experiment applied Stirmark 4.0 to the test results, and it indicated that this algorithm is tough against several image processing operations such as filtering, cropping, scaling, compression, rotation and the random removal of some row and column lines.

2.4.2 Transform Domain

Transform domain watermarking techniques embed a watermark image by modifying the transform coefficients of the host image as opposed to its pixel values. Ideally, a transform domain has the effect in the spatial domain of apportioning hidden information through different order bits in a manner that is robust. There are a number of transforms that can be applied to digital images, but there are notably three most commonly used in image watermarking. These are the discrete Fourier transform (DFT), discrete cosine transform (DCT) and discrete wavelet transform (DWT).

A. Discrete Fourier Transform (DFT)

A Fourier transform (FT) is an operation that transforms a continuous function into its frequency components. The equivalent transform for a discrete-valued function requires the discrete Fourier transform (DFT). In digital image processing, even functions that are not periodic can be expressed as the integral of the sine and/or cosine multiplied by a weighing function. This weighing function makes up the coefficients of the Fourier transform of the signal. Fourier transform allows the analysis and processing of the signal in its frequency domain by means of scrutinising and then modifying these coefficients.

In their paper, M.Eskicioglu proposed a watermark algorithm based on DFT [M.Eskicioglu, 2004]. This paper describes a new circular watermark scheme that embeds one watermark in lower frequencies and another in higher frequency components of the cover image. The circularly symmetric watermark is embedded in the DFT domain by considering the magnitude of the DFT coefficient of the host image, the scaling factor and the circular watermark. M.Eskicioglu's paper presented

extensive experimental results to show the performance of the proposed technique given a number of attacks, and it shows that embedding the watermark in both frequency groups can increase the robustness of the watermarking system. In the experiment, 200 watermarked images were attacked using MATLAB with JPEG compression, Gaussian noise, blurring, resizing, histogram equalisation, contrast adjustment, gamma correction, scaling, rotation and cropping. The results show that if the circular watermark is embedded in higher frequencies, the percentage of false negatives (not detecting the watermark in a marked image) is higher for one group of attacks and lower for another group of attacks. Similarly, if the circular watermark is embedded in lower frequencies, the percentage of false negatives is lower for one group of attacks and higher for another group of attacks.

Pereira et al. proposed a method for copyright protection by embedding a digital watermark in the DFT domain [Pereira, 2000]. The properties of this technique are based on polar maps for the accurate and efficient recovery of the template in an image which has undergone a general affine transform. In this technique, the watermark is composed of two parts: one is a template which contains no information in itself but can detect any transformations undergone by the image, and the other one is a spread spectrum message that contains hidden information. The length of the hidden information is supposed to be short, and it is subjected to a pre-processing algorithm to produce the new message length. Prior to embedding the hidden message, the luminance component of the cover image is extracted and is used to calculate the DFT coefficients. The hidden data and the template are then embedded in these coefficients. The template is embedded along two lines in the cover image which go through the original, and its purpose is to detect any attacks (transformation) the image has undergone. The authors applied Stirmark 3.0 to evaluate the algorithm. The result indicated that the algorithm stood firm against most attacks and the major improvement lay in the fact that the algorithm recovered the general affine transform.

Ridzon and Levicky proposed an algorithm in [Ridzon, 2008]. This algorithm is based on the combination of DFT and Log-Polar mapping (LPM) features. The watermark is in the form of matrix which is created depending on the sequence of alphanumeric values. It is then inserted into the original image in the DFT domain. During the embedding process the hash function is used for algorithm security improvement. The

experiment results given in this paper showed that the proposed watermarking method was robust against different watermark attacks, in particular against rotation attack.

In [Hu, 2011], a proposed watermark algorithm is based on a discrete Fourier transform and a discrete wavelet transform. This algorithm is based on two methods in order to solve the synchronisation of the original watermarking images and tested images in wavelet domain watermarking. One method estimates rotation and scaling based on embedding a template in a circle of middle frequency in DFT, while the other method estimates translation based on extracting an invariant centroid from a restricted area inside the image, and in the embedding process a watermark is embedded adaptively in a low band of the DWT domain, according to the conceal quality of the human visual system. The experiment results show robustness against common signal processing attacks, several geometric distortions and some combinations.

In DFT-related watermarking algorithms, there is no clear optimal way of determining how DFT coefficients should be modified. However, no matter how the coefficients are modified to encode a watermark, one must ensure that coefficients of discrete Fourier transform have real components, because images are real signals [Woo, 2007].

B. Discrete Cosine Transform (DCT)

A discrete cosine transform is a widely used technique in image processing and watermarking, and for several reasons. One of these reasons is that all the major compression techniques were developed in the DCT domain, such as JPEG, MJPEG, MPEG1, MPEG2, etc. A great deal of research was carried out in developing various perceptual models for the DCT domain, and these models could be easily applied to watermarking. Furthermore, DCT is related to DFT in a sense that it transforms a time domain signal into its frequency components. The DCT, however, only uses the real parts of the DFT coefficients. A number of DCT watermarking algorithms concentrate information energy in those bands with a low frequency, and therefore are popular in data compression techniques such as JPEG and MPEG, as well as in digital watermarking.

In [Sverdlov, 2004] , Sverdlov et al. created a new, robust, hybrid non-blind watermarking scheme based on discrete cosine transform (DCT) and singular value

decomposition (SVD). In this method, after applying the DCT in the host image, the DCT coefficients are mapped in zigzag order into four quadrants, which represent frequency bands from the lowest to the highest. SVD is then applied to each quadrant. The same process is also applied to the watermark. The technique then modifies the singular values in each quadrant to obtain a set of modified DCT coefficients. The decoding process involves mapping the modified DCT coefficients back to their original positions and applying the inverse equation to produce an original watermarked image. The experiment's results show that watermarking embedding in the lowest frequency is resistant to attacks including Gaussian smoothing, Gaussian noise, pixilation, JPEG compression, JPEG2000 compression and rescaling. Watermark embedding in the highest frequency is resistant to attacks including sharpening, cropping, contrast adjustment, histogram equalisation and gamma correction.

In [Xiaochuan, 2006], a watermarking algorithm based on low luminance smooth blocks in the compressed DCT domain is proposed. The watermark is embedded by setting the sign of a subset of low-frequency DCT coefficients in these smooth blocks. In this algorithm, DCT is applied to a set of 8×8 pixel blocks. The DCT takes such a signal as its input and decomposes it into 64 orthogonal basis signals. The quantised DC and AC coefficients denote the average luminance and the different frequency bank of a block that reflect its texture respectively. Firstly, appropriate low luminance smooth blocks should be selected based on the DC and AC coefficients. Then, the coefficients are quantised in zigzag order. These DCT coefficients are then modified according to a number of robustness measures and watermarking information. The process is repeated until a desired number of smooth blocks are embedded in the watermarks. Experimental results show that the watermarked image looks visually identical to the original, and the watermarks also successfully survive after image processing operations such as image cropping, scaling, filtering and JPEG compression.

Another novel blind watermarking algorithm based on DCT is proposed in [Zhao, 2008]. This paper uses an average value of the DCT coefficient as a threshold to realise watermark embedding. In the embedding process, the watermark image is processed by Arnold transformation (a process involving clipping and splicing that realigns the pixel matrix of the digital image), and the original image is divided into

blocks which are transformed into coefficients in the DCT domain. The average value of a coefficient selected from the same location in DCT blocks is calculated and the selected coefficient in each DCT block is modified into a value not less than the negative threshold or more than the positive threshold according to the Arnold transformed watermark value. The experiment shows that the algorithm has good imperceptibility and is holds firm against common image processing techniques such as JPEG compression, filtering, noise, cropping and scaling.

C. Discrete Wavelet Transform (DWT)

Wavelets have recently become important tools in image processing and watermarking due to the good energy compaction properties they possess, as well as the existence of efficient algorithms for computing the wavelet transform. They also form the basis of the new compression standard JPEG2000 [Skodras, 2006]. These transforms are based on small waves, called wavelets, of varying frequency and limited duration. A wavelet series is a representation of a square-integrable function by a certain orthonormal series generated by a wavelet. Furthermore, the properties of a wavelet can decompose original signals into wavelet transform coefficients which contain position information. The original signal can be completely reconstructed by performing inverse wavelet transformation on these coefficients.

The DWT separates an image into four parts, namely a lower resolution approximation component (LL) as well as horizontal (HL), vertical (LH) and diagonal (HH) detail components. The LL sub-band is the result of low-pass filtering rows and columns, and it contains a rough description of the image. The HH sub-band is high-pass filtered in both directions and contains high-frequency components along the diagonals. HL and LH images are the results of low-pass filtering in one direction and high-pass filtering in another direction.

After the image has been processed by the wavelet transform, some of the information contained in the original image is concentrated into the LL subband. LH contains mostly vertical detail information, which then corresponds to horizontal edges. HL represents horizontal detail information from the vertical edges. The process can be repeated to compute multiple 'scale' wavelet decompositions, as shown in Figures 2.5 and 2.6

Wavelets play an increasingly important role in the contemporary image processing field, as it has lots of special advantages that compare with conventional transforms such as DFT and DCT, which are full frame transform and hence any change in the transform coefficient affects the entire image, except if DCT is implemented using a block-based approach [Cheddad, 2008]. Hence, wavelet-based watermarking techniques are becoming more popular.

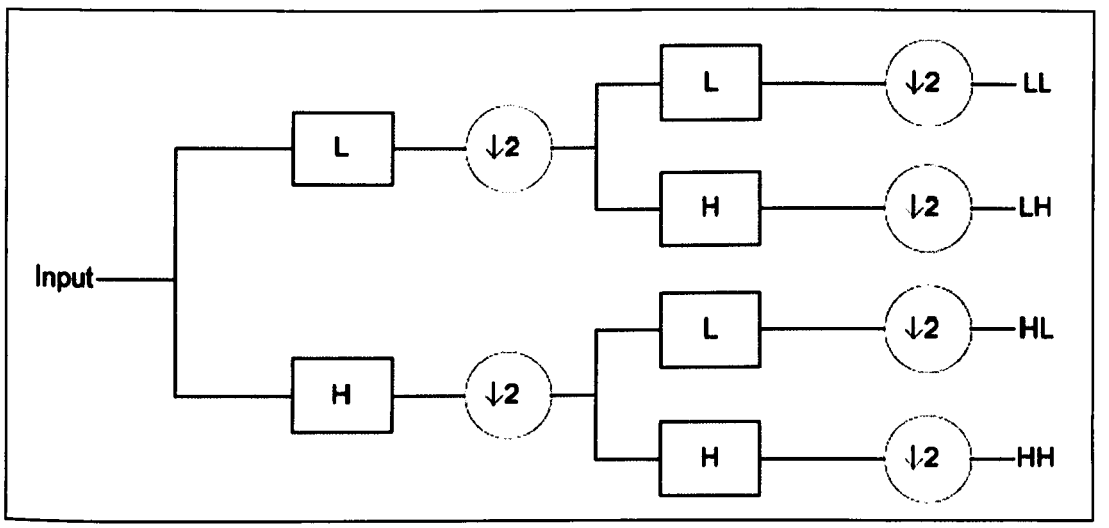


Figure 6.5 The decomposition step of an input image into four sub-bands

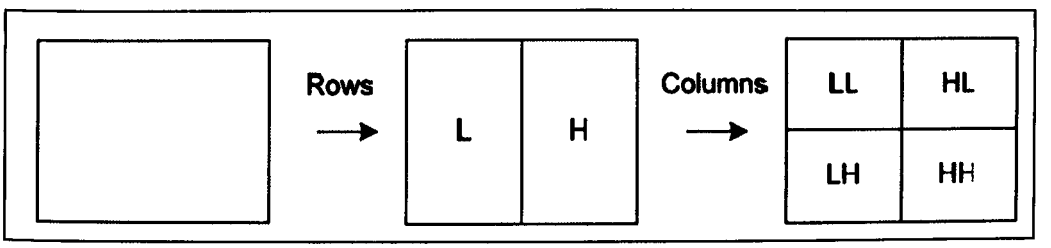


Figure 7.6 One decomposition step

The paper in [CL.Song, 2009] is another argument used to support the above opinion. It surveys recent advances in watermarking techniques in digital images. In addition, it classifies different watermarking techniques into several categories depending upon the domain in which the hidden data is inserted, and it includes LSB, DCT and DWT. An experiment is conducted to further test the robustness of some of these techniques. The results indicate that LSB contains the worst result; however, both DCT and DWT have their strong points.

In [Hwei-Jen, 2010], Hwei proposed a DWT-based digital image watermarking system. This algorithm performs the original image with the DWT and embeds the watermark into the HL and LH sub-bands. The experimental results show that the

proposed approach indeed produces better results than the compared method in terms of the quality of the watermarked image, the extracted watermark with or without attack and time efficiency. The experiment results reported in the paper also suggest that this method is able to achieve ownership protection.

Another novel technique was invented for robust wavelet-based watermarking [Ellinas, 2007]. The technique embeds the signature data in the selected group of wavelet transform coefficients, varying watermark strength according to the sub-band level and the group where the corresponding coefficients reside. Initially, the input image decomposes into four levels by DWT, so we get approximation sub-band with low frequency components and 12 detailed sub-bands with high frequency components. Next, the technique detects the presence of edges in each sub-band by using a Sobel edge detector, thus forming two groups of coefficients. The watermarking embedding process is carried out over the sub-band coefficients that lie on edges, where distortions are less noticeable, with a sub-band level-dependent strength. In addition, the watermark is embedded into selected coefficients around the edges, using a different scale factor for watermark strength, which are captured by a morphological dilation operation. The experiment's evaluation of the proposed method shows very good results in terms of robustness and transparency to various attacks such as median filtering, Gaussian noise, JPEG compression and geometrical transformations.

Another DWT-based watermark scheme is a hybrid scheme based on DWT and singular value decomposition (SVD) [Emir, 2004]. After decomposing the host image into four sub-bands, the algorithm applies SVD to each sub-band and then embeds the same watermark data by modifying the singular values. The experiment indicates that watermark embedding in the LL sub-band is resistant to attacks including Gaussian blur, Gaussian noise, pixilation, JPEG compression, JPEG 2000 compression and rescaling. Watermark embedding in the HH sub-band is resistant to attacks including sharpening, cropping, contrast adjustment, histogram equalisation and gamma correction, while embedding in the LH sub-band is resistant to rotation attack.

2.4.3 Remarks on current existing digital image watermarking techniques

Several watermarking algorithms were described in the previous section. It should be noted that they can be embedded and extracted in different domains. The most direct

approach is watermarking in the spatial domain, where pixel values are modified to encode the signal. Furthermore, frequency domains such as DFT, DCT and DWT are used widely in image watermarking.

In a robust digital watermarking system, algorithms should resist different attacks, including removal attack and geometric attack. So how to develop an algorithm that is impenetrable to different watermarking attacks, especially geometric attack is one of the major objectives in watermarking research. Recently, although there exists a number of watermarking algorithms, only a few of them can resist both removal and geometric attacks.

To design a robust watermarking algorithm, it is necessary to analyse different watermark attacks because we need to:

- Better understand the properties of different watermark attacks including removal attack and geometric attack.
- Hypothesize a watermark algorithm due to the properties of watermark attacks.

2.5. Properties of Digital Image Watermarking Systems

A few important properties associated with watermarking systems concerning digital images are discussed in the following subsections.

2.5.1 Robustness

The robustness of a watermark is its ability to resist non-malicious distortions, which usually include common image processing, geometrical transforms. For example, a watermark is said to be robust against JPEG compression if it can be extracted successfully after image compression. Common image processing operations include noise insertion, contrast adjustment, smoothing, sharpen, and image compression and so on. Geometrical transforms include rotation, scaling, translation, etc. It is desirable to have watermarks that are strong against all possible distortions.

Among the many types of distortions mentioned previously, geometrical distortions remain a major challenge in robust watermarking, as they can be carried easily using image processing software and defeat the purpose of watermarks by making them undetectable. They can also cause serious damage to watermark information through de-synchronisation effects. Most geometrical distortion can be modelled as

combinations of three basic transforms: rotation, scaling and translation (RST). Therefore, a great deal of research into robust watermarking has focused on geometrical robustness, particularly RST robustness.

2.5.2 Capacity and data payload

The number of watermark bits encoded in a message is the data payload [Ingemar J. Cox, 2008], and the maximum repetition of data payload within an image is watermark capacity. The simplest form of watermarks is the one-bit watermark. A watermark may have high capacity but a low data payload. For example, we can have a one-bit watermark embedded many times across the image.

Dual watermarking technology is an example algorithm used to improve capacity and data payload. It embeds into the host image two watermark images, the first of which is called 'primary watermark W1' and the second 'secondary watermark W2'. The dual watermarking process starts by inserting secondary watermark W2 into first watermark W1 to get W1*, followed by inserting W1* into the host image [Congxu, 2006, Sharkas, 2005, Zhiguo, 2007].

2.5.3 Transparency

A watermarking system is of no use if it distorts the host image to the point of being useless, or even highly distracting. Ideally, the watermarked image should look indistinguishable from the host image, even using the highest quality equipment; otherwise distortions in the watermarked image caused by watermark embedding will degrade its aesthetic value. Furthermore, they may cause suspicion and put in danger watermark security. This property is named the 'transparency' of a watermark, and it is sometimes called 'imperceptibility' or 'invisibility'. HVS models can be applied during watermark embedding to enhance watermark transparency and robustness. The model specifies that the visual system of human eyes has certain characteristics in that the eyes are less sensitive to changes made in highly textured regions compared to flat regions.

To evaluate transparency among watermarking methods, a peak-signal-to-noise ratio (PSNR) is generally deployed for comparing transparency performance. The more similar the watermarked image and its host image, the higher the PSNR value between the images[Ingemar J. Cox, 2008]. Assume an image with 8-bit greyscale, the PSNR of a watermarked image compared to its host image is

$$PSNR = 20log_{10}(\frac{I_{MAX}}{RMSE}) \quad (2.1)$$

where I_{MAX} is the maximum grey level of the image. In this case, I_{MAX} can have a maximum value of 255. RMSE is the root mean square error given by

$$RMSE = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[\varphi(m,n) - \omega(m,n)]^2} \quad (2.2)$$

where φ(m,n) and ω(m,n) are the pixel values of the watermarked image and the cover image at position (m,n). In addition, M and N is the number of rows and columns in each image.

2.5.4 Computational Cost

Watermarking methods with highly complex algorithms will incur more computational costs compared to those with low complexity. Although the processing speed and memory size of consumer equipment have been upgraded throughout the years, algorithm complexity has made applications more resource-hungry.

Evaluating computational cost can be done by measuring the execution time of watermark embedding and detection steps using a minimally configured system.

2.5.5 Trade-off between performance factors

Many of the watermark properties explained previously have conflicting characteristics. For example, increasing the robustness of a watermark would normally lower its transparency due to the higher watermark energy imposed on the host image. In addition, higher capacity would compromise its imperceptibility because more modifications to the host image are needed to embed the watermark. Therefore, designing a watermarking algorithm method usually requires finding a balance between these conflicting factors.

Figure 2.7 shows trade-offs between robustness, transparency and capacity – a good watermarking system should balance these three variables.

The requirements regarding each of the watermark properties are application-specific. For example, watermarking medical images requires a high level of transparency to avoid misjudgement during diagnosis. On the other hand, watermarking artistic pictures for copyright protection might give more attention to its robustness compared to its invisibility.

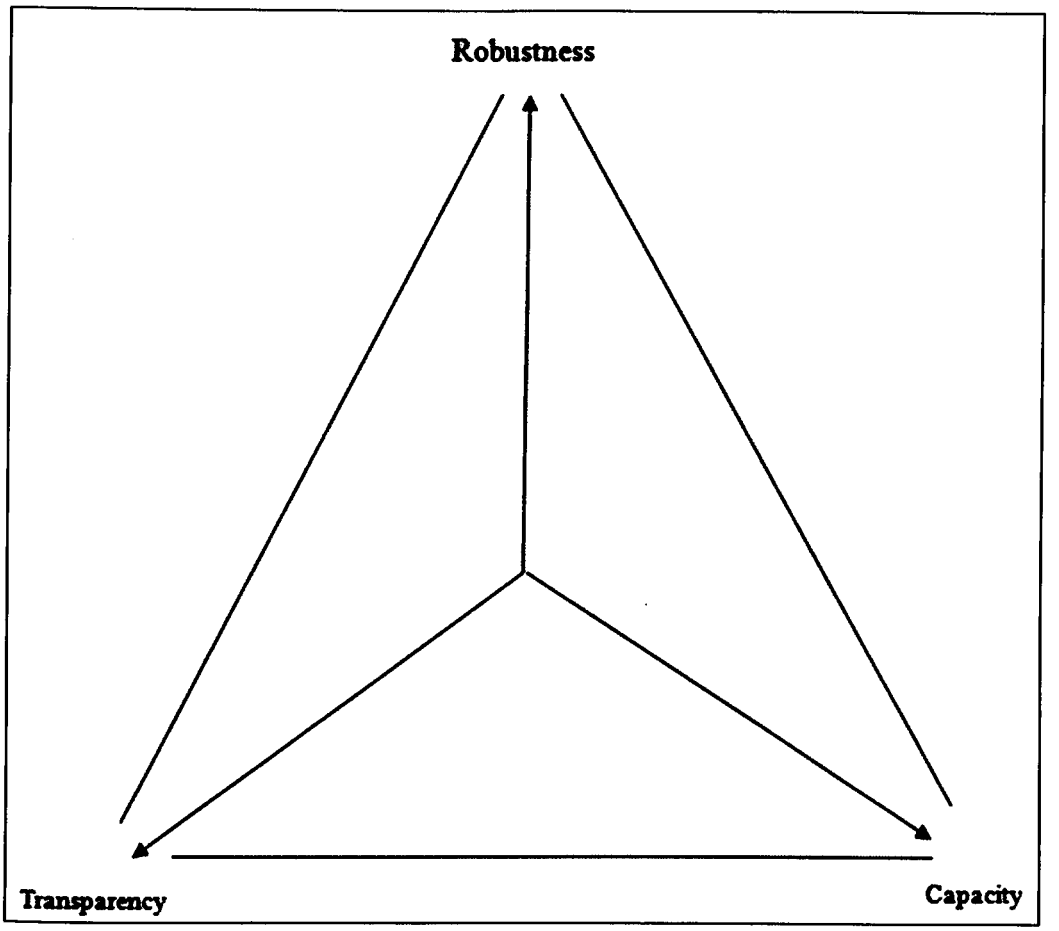


Figure 8.7 Trade-offs between robustness, transparency and capacity.

2.6. Applications of Digital Image Watermarking Systems

Digital watermarking technologies have been proposed for implementation in many applications. This subsection describes some major groups of digital image watermarking applications.

2.6.1 Digital Rights Management

Digital rights management (DRM) can be defined as "the description, identification, trading, protecting, monitoring and tracking of all forms of usages over tangible and intangible assets" [Becker, 2003]. It concerns the management of digital rights and the enforcement of rights digitally.

Many factors have contributed to the rise of DRM because they pose a threat to the protection of digital rights. The list below gives some examples.

- An increased amount of digitized content due to technological advancement, e.g. digital photographs, electronic books, video on demand and downloadable

music.

- Advances in computer networking technologies have created new channels for content distribution in huge quantities and very quickly.
- The sophistication of software functionalities enables end-users to manipulate digital content easily.

DRM is required to support these changes and control the rights to purchase, consume, edit, store and distribute digital content. For example, a DRM system can control access to and the usage and distribution of digital contents.

DRM systems have three major components, namely enabling technologies, the business model and the legislative framework [Becker, 2003]. Concerning technological implementations, DRM systems are normally used to protect the rights of an intellectual property (IP) holder through copyright protection measures. This protection is necessary to thwart mass reproduction of illegal copies. In this case, watermarking is a tool employed to secure digital content. The embedded watermark remains associated with the content wherever it is distributed and duplicated. Furthermore, watermarks applied in DRM systems also enable copyright protection, copy protection, device control, authentication and tamper detection. The technological implementation involves watermarking software, hardware and protocols.

There remain a number of open problems in DRM, and particularly user acceptance, user privacy and user friendliness issues are yet to be solved. For instance, users do not like to be tracked every time they access digital content. In addition, major content providers and distributors need to adopt a set of standards so that the DRM enabling hardware and software can interoperate. Moreover, watermarking methods need to find a balance between robustness, transparency and capacity to be practical. More research is therefore required to overcome these obstacles.

2.6.2 Copyright protection

Copyright protection is an important application of digital watermarking, as it enables the identification of the copyright holder and thus protects his or her rights in content distribution. A watermark is embedded into a host image to protect the rights of the owner. It should be possible to detect the watermark despite common image

processing, geometrical distortion and many other types of image manipulations. Therefore, the deliberate removal of a robust watermark should result in severe degradation of the image's visual appearance. The successful detection of the watermark can also positively identify the owner.

2.6.3 Authentication

Authentication in digital image watermarking refers to the integrity assurance of the image [Ingemar J. Cox, 2008]. An image is said to be 'authentic' if it has not been modified. An integrity check using watermark is advantageous because, first of all, the embedded watermark stays with the image and cannot be removed easily. Secondly, extra space is not required to store watermark information. The authentication of digital images can be useful in insurance claims by ensuring trustworthy photographs for court evidence. Other reported applications related to image authentication are the validation of cultural heritage paintings, medical records and digital artwork.

To determine whether an image is authentic, either robust watermarks or fragile watermarks can be applied. For instance, information extracted from a robust watermark can be compared with the image features to evaluate integrity, while the absence of a fragile watermark in watermark detection indicates that the image has been changed.

2.6.4 Tamper detection and localisation

Tamper detection is used to disclose alterations made to an image, and it is closely related to authentication. If tampering is detected in an image, then the image is considered unauthentic. Tamper localisation enables further investigation of an act of tampering by identifying the tampered regions within the image. This information can assist in media forensics; for example, the severity of the tampering and the motives behind it can be established. Similar to authentication, tamper detection and localisation can be achieved using robust, fragile or semi-fragile watermarks according to the applications in question.

2.6.5 Annotation and privacy control

Multi-bit watermarking can be used to annotate an image. For example, patient records and image details related to a medical image can be carefully inserted into the image, which will not only reduce storage space, but also provide a tight link between

the image and its details. Patient privacy is simply controlled by not keeping sensitive information as clear text in human readable form, and the watermark can be further secured by encryption. Other usages of annotation watermarking are electronic document indexing and automated information retrieval. In these cases, watermark information serves as indices and keywords. Transparency is very important in these cases because the image carries vital information for medical diagnosis, although robustness may not be relevant here if the watermarking system resides in a secured and closed environment.

2.6.6 Other Applications

There are many other applications where digital watermarking methods have been proposed as a technology-enabling tool. Some of them have proved to be useful, while others have been discarded because they were impractical. Some examples are listed below[Ingemar J. Cox, 2008].

- Broadcast monitoring – watermarks embedded into advertisement sections of broadcasts. This is a cost-effective means of monitoring advertisement airtime on television and radio broadcasts, and it is format-independent and does not consume extra bandwidth. The practicality of this application may be limited by watermark imperceptibility.
- Device control – watermarks embedded into radio and television signals can be used to control some feature of a receiver. This has been proven to be a practical usage of watermarks.
- Communication enhancement – watermarks extracted are used to repair error bits in transmission, hence saving time, costs and bandwidth for re-transmission.

2.7. Chapter Summary

Digital image watermarking involves embedding and extracting hidden information in digital media. The increased consumption of digital content and security circumvention technologies necessitates more research into this area.

The discussions in this chapter begin with a digital watermarking model. Then, the processes involved in a digital image watermarking system are described as a sequence of three steps: embedding step, transmission step and extraction step.

Besides that, the classification of a digital image watermarking system is described, as well as criteria of classification based on the type of visible and invisible, type of domain and type of extraction. In addition, the spatial and frequency domains as the main existing watermark algorithms are described. Then, watermark properties and watermark applications were described. The major properties to be considered when creating a digital watermark are robustness, capacity, transparency, computational cost and other practical issues. Watermarking technologies can be applied in DRM, copyright protection, authentication, tamper detection and localisation, and annotation.

In Chapter 3, watermark attack will be analysed, while Chapters 4 and 5 will describe the region-adaptive watermarking algorithms based on the DWT domain and DWT-SVD. Subsequently, a novel watermark attack detection application will be discussed in Chapter 6.

CHAPTER 3

ANALYSIS OF WATERMARK ATTACKS

Robust digital image watermarks can be used in many applications, such as copyright protection. To serve its purpose, a robust watermark should withstand common image processing and other modifications, which are often referred to as watermark attacks. In this chapter, we will describe the classification of watermark attacks first in section 3.1, and then the methodology used for analysing watermark attack will be defined in section 3.2. Subsequently, in section 3.3, the results of analysis on six image processing attacks and geometric attacks will be analysed. The last section will describe the proposed watermark algorithm that was designed through our findings.

3.1. Classification of a Watermark Attack

Voloshynovskiy et al. [Voloshynovskiy, 2001] classify watermark attacks into four distinct categories, namely removal attacks, geometric attacks, cryptographic attacks and protocol attacks. The classification of watermark attacks can be summarised as shown in Figure 3.1.

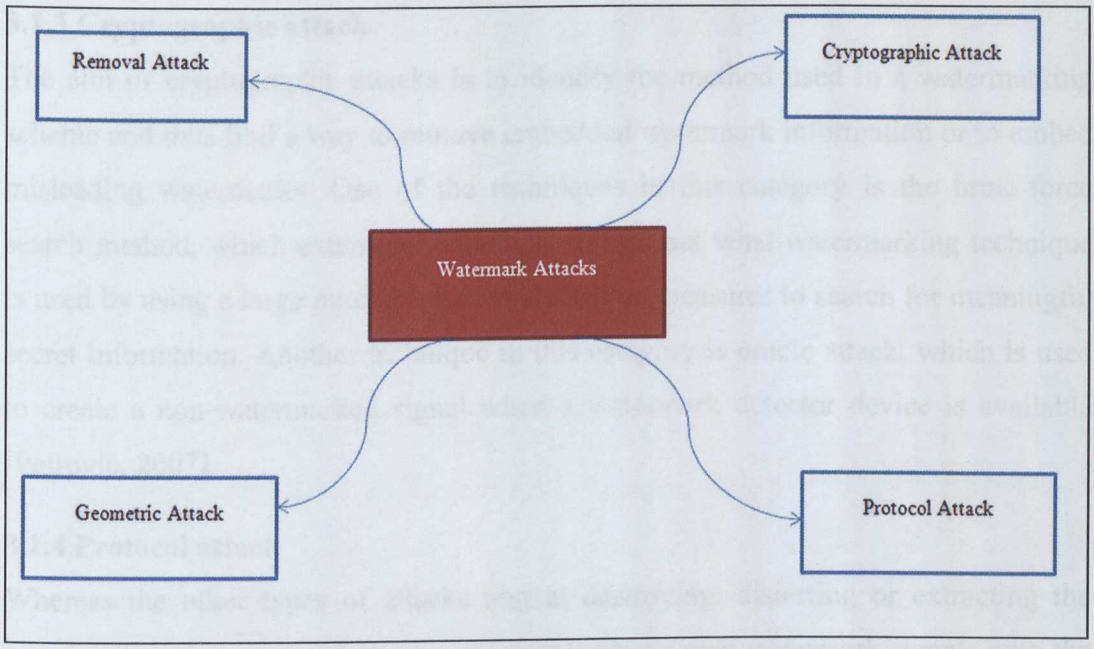


Figure 3.1 The classification of watermark attacks

3.1.1 Removal attack

Removal attacks aim at removing the watermark signal from a watermarked image, without attempting to break the security of the watermarking algorithm. This type of watermark attack does not attempt to find out the encryption techniques used or how the watermark has been embedded. It results in a damaged watermarked image, hence a damaged watermark signal, where no simple post-processing can recover the watermark signal from the attacked data. Included in this category are noising, histogram equalisation, smoothing, sharpen attacks and so on [K.F.Tsang, 2001].

3.1.2 Geometric attack

Geometry attacks are rather different from removal attacks. Instead of aiming to remove or severely damage the watermark signal, this type of attack intends to distort it. Examples of geometric attacks include rotation, scaling, translation, shearing, random bending or change of aspect ratio. It is, however, still theoretically possible for the detector to recover the original watermark if the detail of the geometry attack can be established and a countermeasure applied. The process of correcting this type of attack is often referred to as 'synchronisation'; however, the complexity of the required synchronisation process might be prohibitively expensive and slow [Licks, 2005].

3.1.3 Cryptographic attack

The aim of cryptographic attacks is to identify the method used in a watermarking scheme and thus find a way to remove embedded watermark information or to embed misleading watermarks. One of the techniques in this category is the brute-force search method, which extensively attempts to find out what watermarking technique is used by using a large number of known possible measures to search for meaningful secret information. Another technique in this category is oracle attack, which is used to create a non-watermarked signal when a watermark detector device is available [Petrovic, 2007].

3.1.4 Protocol attack

Whereas the other types of attacks aim at destroying, distorting or extracting the watermark signal, protocol attacks add the attacker's own watermark signals onto the data in question, which results in ambiguities on the true ownership of the data in question. Protocol attacks target the entire concept of using watermarking techniques as a solution to copyright protection. Copy attack also falls into this category, as instead of destroying the watermark, it estimates a watermark from watermarked data and copies it to some other data, called ''target data'. The estimated watermark is adapted to the local features of the target data to satisfy its imperceptibility [Kutter, 2000] .

3.2. Methodology for Analysing a Removal Attack

The objectives of analysing watermark attacks are to show the effects of removal attacks on both the host image and watermark data. By understanding these effects, one can a) identify similarities in effects between different watermark attacks and b) devise a solution to alleviate the effects based on existing techniques for similar phenomena. The analyses are presented using an image histogram [Hadjidemetriou, 2000, Qieshi] and a Fourier spectrum [Kargupta, 2004, Xiaojun, 2008] as tools to better understand the effect of each watermark attack in frequency and spatial domain, respectively.

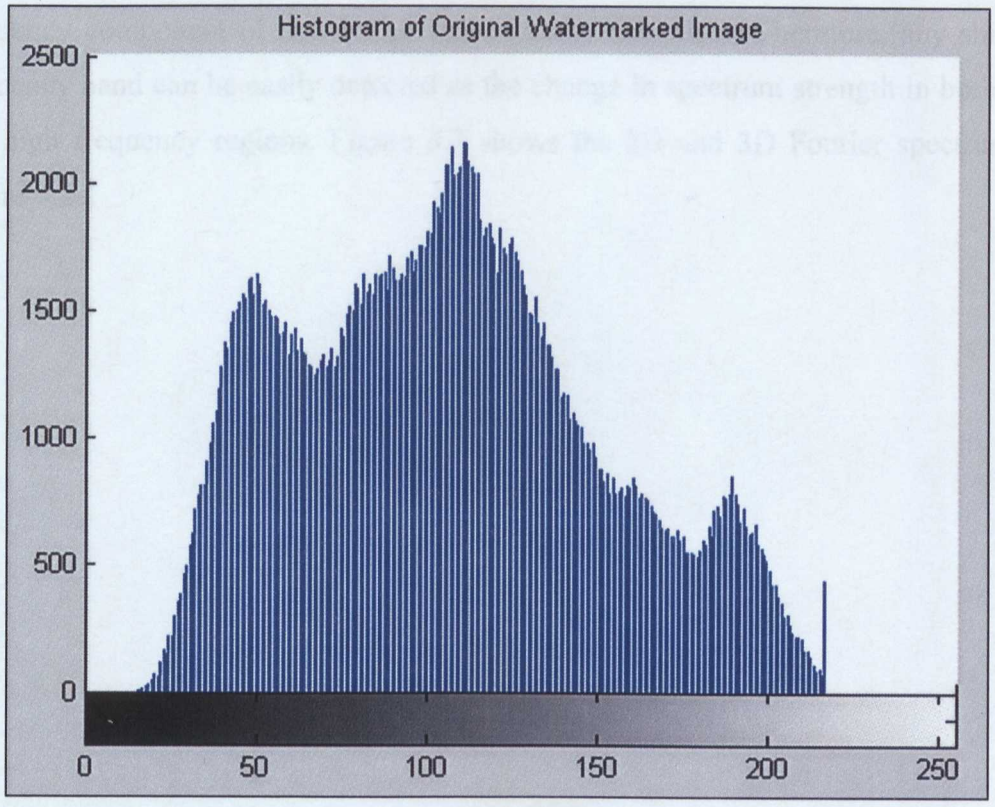
3.2.1 Image Histogram

A histogram is defined as a summary graph showing a count of the data points falling into various ranges. It is an estimate of the probability distribution of a continuous random variable. Histograms are made up of bins, each bin representing a certain

intensity value range. The histogram is computed by examining all pixels in the image and assigning each one to a bin depending on pixel intensity. The final value of a bin is the number of pixels assigned to it. The number of bins in which the whole intensity range is divided is usually in the order of the square root of the number of pixels [Porikli, 2005].

In an image processing context, the histogram of an image refers to a histogram of the pixel intensity values. This histogram is represented as a graph showing the number of pixels in an image at each different intensity value found in that image. For an 8-bit greyscale image there are 256 different possible intensities, and so the histogram will graphically display 256 numbers, thus showing the distribution of pixels amongst those greyscale values. The horizontal axis of the graph represents intensity, while the vertical axis represents the percentage of the number of pixels having that intensity. Figure 3.2 shows a greyscale image and its pixel-based histogram. The horizontal axis of Figure 3.2b represents the variations, while the vertical axis represents the number of pixels.

(a)

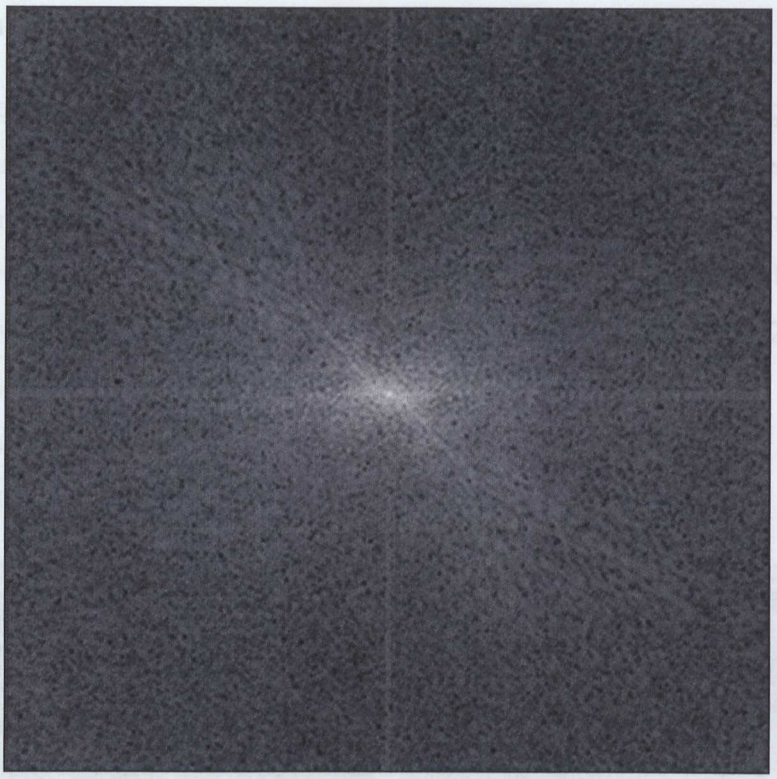


(b)

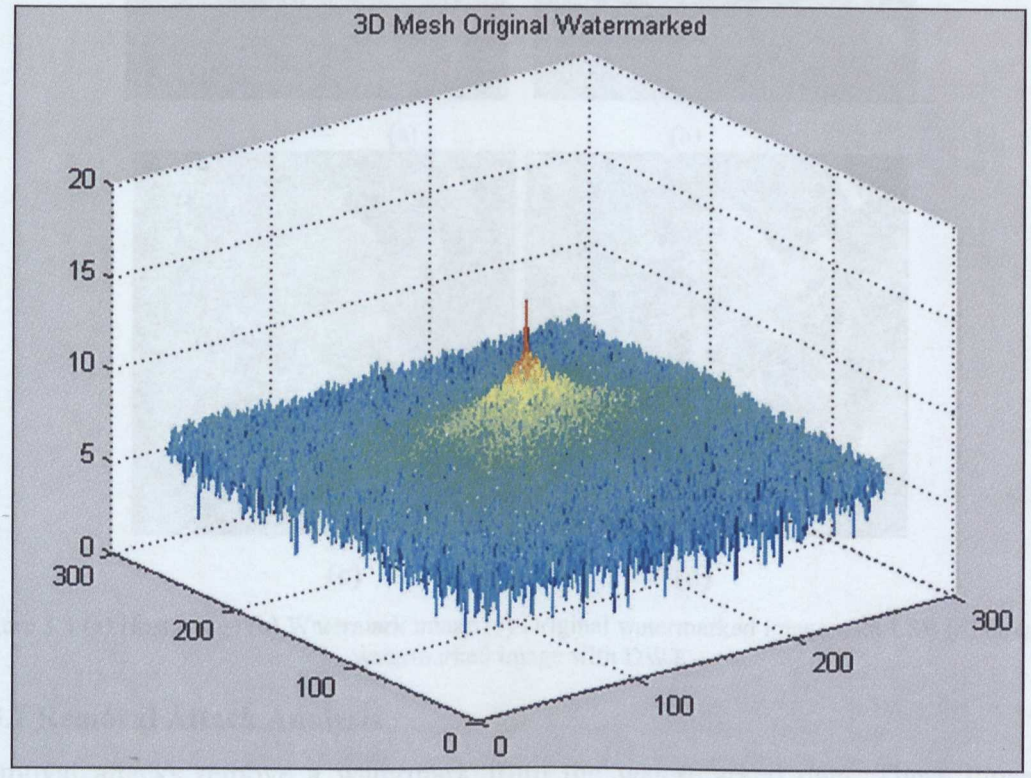
Figure 3.2 (a) Greyscale image (b) Histogram of the greyscale image

3.2.2 Fourier Spectrum

An image transform converts image data into alternative representations that are more amenable for certain types of analysis. The most commonly used image transform is Fourier transform, which takes spatial data and transforms them into their frequency components or spectrum. The Fourier spectrum of an image is a representation of that image in the frequency domain. The Fourier spectrum can be calculated in several ways, but the fast Fourier transform (FFT) method is the most frequently used algorithm. The Fourier spectrum of an image is often visualised as a greyscale image whose intensity corresponds to the strength or magnitude of the spectrum. The coordinates of these pixels correspond to the frequency in x-y directions. The lowest frequency is located at the centre of the spectrum image and the highest frequency is located around the edges, although for clarity in visual inspection the Fourier spectrum is often visualised as a 3D graph as opposed to a 2D greyscale image. The Fourier spectrum has an important characteristic in which its total energy is preserved. This means that both low and high frequency components are complementary. In other words, if the high frequency component of an image increases, the low frequency component of that image subsequently decreases. Therefore, any shift in frequency band can be easily detected as the change in spectrum strength in both low and high frequency regions. Figure 3.3 shows the 2D and 3D Fourier spectrum of Figure 3.2a.



(a)



(b)

Figure 3.3 (a) 2D Fourier spectrum of a greyscale image (b) 3D Fourier spectrum of a greyscale image

3.3. Watermark Attack Analysis

The watermarking process uses in this experiment used two types of images, namely a host image and a watermark image. The watermarking process will attempt to insert the watermark image into the host image to produce a watermarked image using a discrete wavelet transform (DWT) watermarking technique, as described in [Tao, 2004], and the spatial domain of a least significant bit (LSB) [CL.Song, 2009]. Figure 3.4 shows the host image, watermark image and the resulting original watermarked image created by DWT and LSB, respectively. Both images are identical in dimension (512x512 pixels) and depth (256 grey levels).

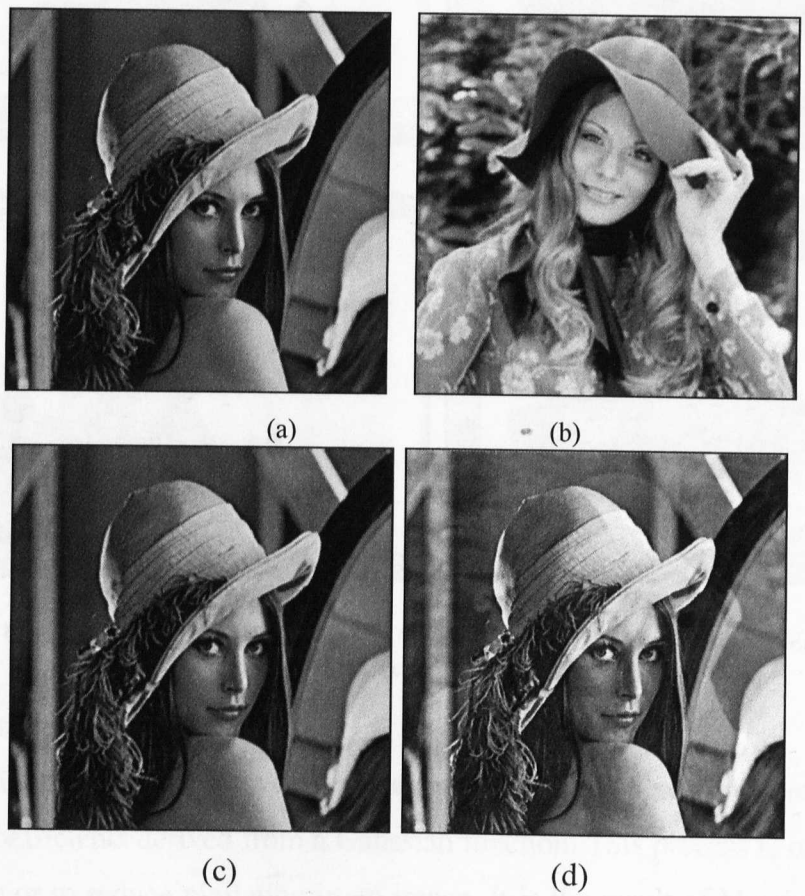


Figure 3.4 (a) Host image (b) Watermark image (c) Original watermarked image with LSB (d) Original watermarked image with DWT

3.3.1 Removal Attack Analysis

Removal attacks remove a watermark from the watermarked data. These types of attacks will be analysed by examining the effect of Gaussian smoothing attack, Gaussian noise attack, salt and pepper noise attack, histogram equalisation attack, sharpen attack and JPEG compression attack. The effects of these attacks on the watermarked image are analysed by comparing the original watermarked image and

the attacked watermarked images in both the spatial and frequency domains by using an image histogram and the Fourier spectrum, respectively. It is understood that the two domains are complementary and provide a complete description of the images or any changes to them. The histogram and the Fourier spectrum of the original watermarked image in the DWT and LSB algorithms are shown in Figure 3.5.

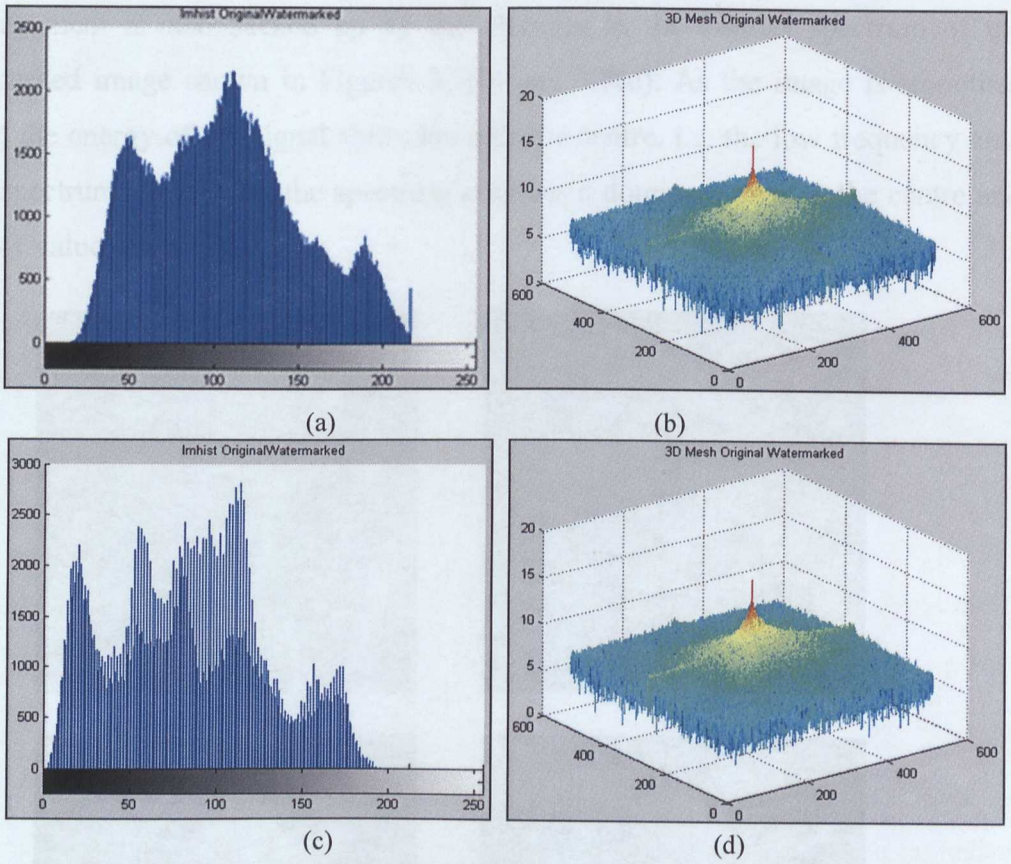


Figure 3.5 (a) & (c) Histogram in DWT and LSB and (b) & (d) Fourier spectrum of the original watermarked image in DWT and LSB.

3.3.1.1 Gaussian Smoothing

Gaussian smoothing is a process that averages the value of pixels over an area using weighting coefficients derived from a Gaussian function. This process is often used to reduce noise or to reduce pixilation in an image. It is the result of blurring an image by a Gaussian function. Visually, the effect of Gaussian smoothing on an image is illustrated in Figure 3.6 in both the spatial and frequency domains. The amount of smoothing can be controlled by adjusting the width of the Gaussian function. In MATLAB, this is implemented by applying a Gaussian filter of width (sigma) to the Fourier coefficients of the image.

The experiment uses a number of different σ values to allow better observation of the effect of Gaussian smoothing on a watermarked image. The σ values used range

between 10 and 200. An observation of the image histogram of the Gaussian smoothed images, shown in Figures 3.7(a) and 3.7(c), shows that as the width of the histogram decreases, the peak of the histogram increases as the level of the smoothing increases. This phenomenon is an indication that smoothing reduces variation in image pixel values.

This argument is also backed up by the changes in the Fourier spectrum of the watermarked image shown in Figures 3.7(b) and 3.7(d). As the image is smoothed further, the energy of the signal shifts towards the centre, i.e. the low frequency area of the spectrum. At $\sigma = 10$, the spectrum contains a dominant peak at the centre and very low value elsewhere.

(a) (b)

(c) (d)

Figure 3.6 Effect of Gaussian smoothing on an image (a) & (c) Original watermarked image in LSB and DWT (b) & (d) Gaussian smoothed image in LSB and DWT.

This phenomenon gives a clear indication that Gaussian smoothing is essentially a low pass filter function [Gonzalez, 2008]. The process reduces the high variation, i.e. high frequency components, of the image and produces low variation pixels or a low frequency signal.

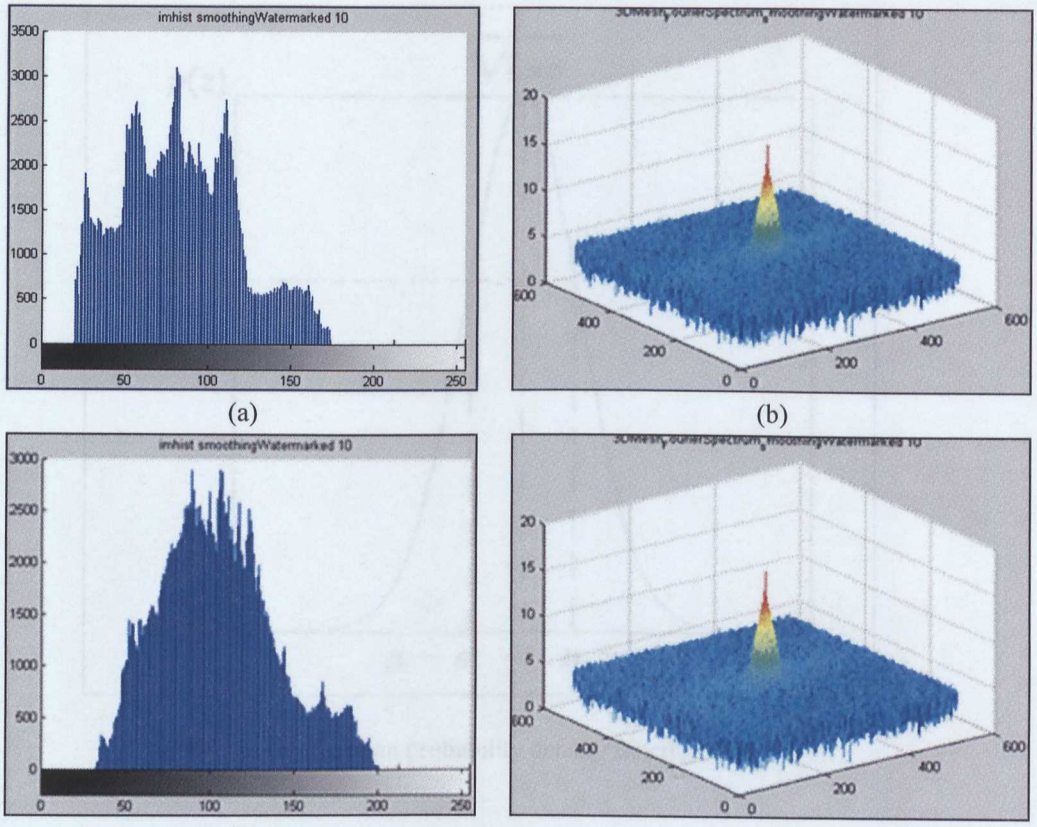


Figure 3.7 (a) & (c) Histogram in LSB and DWT and (b) & (d) Fourier spectrum of the Gaussian smoothed watermarked image in LSB and DWT with $\sigma = 10$.

3.3.1.2 Gaussian Noise

Additive Gaussian noising is a process that adds a noise signal to an image in order to deliberately corrupt the image and hence reduce its visual quality. The statistical property of this noise follows a Gaussian probability density function (PDF), as shown in Figure 3.8. A Gaussian PDF, often called a *normal* PDF, is a function that describes the relative likelihood for this random variable to occur at a given point. The probability for the random variable to fall within a particular region is given by the integral of this variable's density over the region. The probability density function is non-negative everywhere, and its integral over the entire space is equal to one. In addition, the central limit theorem states that the distribution of the sample means of a sufficiently large number of samples will always approximate to the Gaussian distribution. The Gaussian PDF p(z) of a random variable z is defined as:

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{z^2}{2\sigma^2}} \qquad (3.1)$$

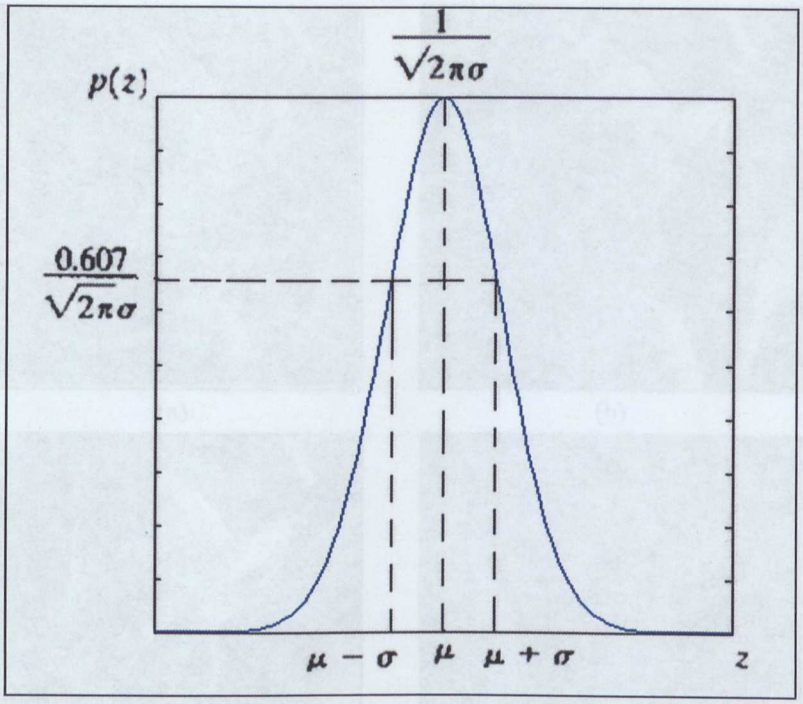


Figure 3.8 Gaussian probability density function of width σ

In MATLAB, controlling the distribution width is achieved by adjusting the fourth input parameter (denoted here as σ) of the *imnoise* function, which takes four parameters, namely *I,'gaussian', m* and *σ*. This function adds Gaussian noise of mean *m* and variance *σ* to the image *I*. In this experiment, zero mean Gaussian noise is used by setting m to 0. Visually, the effect of adding noise into an image can be seen in Figure 3.9.

The experiment uses a number of different σ values to allow better observation of the effect of adding Gaussian noise to the watermarked image. The σ values ranged between 0.005 and 0.1.

(a) (b)

(c) (d)

Figure 3.9 Effect of Gaussian noise on an image (a) & (c) Original watermarked image in LSB and DWT (b) & (d) Gaussian noised image in LSB and DWT.

An observation of the image histogram, as shown in Figure 3.10(a) & (c), shows that as moderate strength noise is added to the image, the width of the histogram increases and the peak of the histogram decreases. At some point when the width reaches the physical limit of the pixel values (i.e., 0 and 255) more pixels with these values are accumulated. Therefore, as the strength of the noise is increased further, there are a large number of black and white pixels in the image, while the rest of the histogram seems to look more uniformly distributed.

Observation of the Fourier spectrum, shown in Figure 3.10(b) & (d), provides another insight into this phenomenon. As stronger noise is used to attack the watermarked image, the Fourier spectrum shows an increase in the strength of the high frequency components of the spectrum and, conversely, a decrease in the strength of the low frequency components. A direct comparison between the original watermarked image and the most severely attacked watermarked image show a significant change in spectrum shape, especially in the high frequency region.

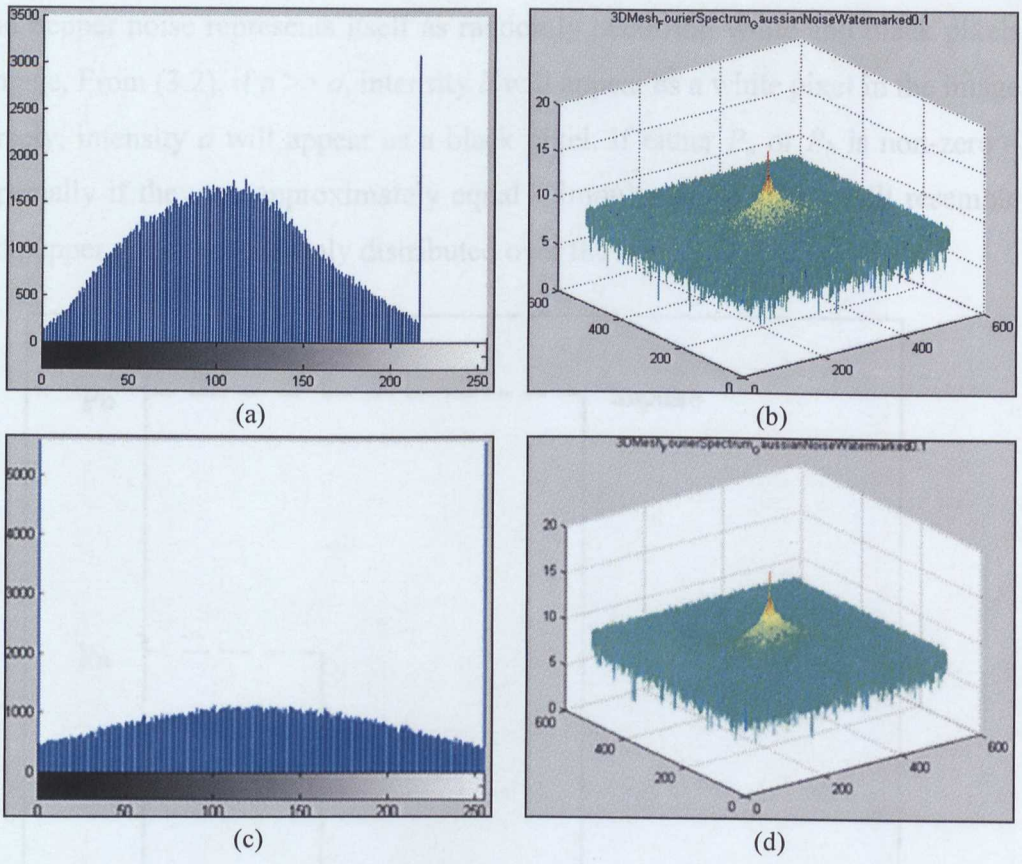


Figure 3.10 (a) & (c) Histogram of the Gaussian noised watermarked image in LSB and DWT and (b) & (d) Fourier spectrum of the Gaussian noised watermarked image in LSB and DWT with σ = 0.1.

This phenomenon gives a clear indication that a Gaussian noise attack is similar to a high pass filter, as the process increases the variation in pixel values to the extent that local edges start appearing in the image. This type of attack has the consequence of removing low frequency watermarks and obscuring high frequency watermarks that may be present in an image.

3.3.1.3 Salt and Pepper Noise

Salt and pepper noise has a very different probability distribution function to Gaussian noise. Its PDF takes the form of two impulse functions at two discrete locations, as illustrated in Figure 3.11. The salt and pepper PDF p(z) can be calculated using the two impulse functions given in Equation (3.2).

$$p(z) = \begin{cases} P_a & z = a \\ P_b & z = b \\ 0 & otherwise \end{cases} \qquad (3.2)$$

Salt and pepper noise represents itself as randomly occurring white and black pixels in an image. From (3.2), if $b >> a$, intensity b will appear as a white pixel in the image. Conversely, intensity a will appear as a black pixel. If either P_a or P_b is non-zero – and especially if they are approximately equal – impulse noise values will resemble salt and pepper granules randomly distributed over the image.

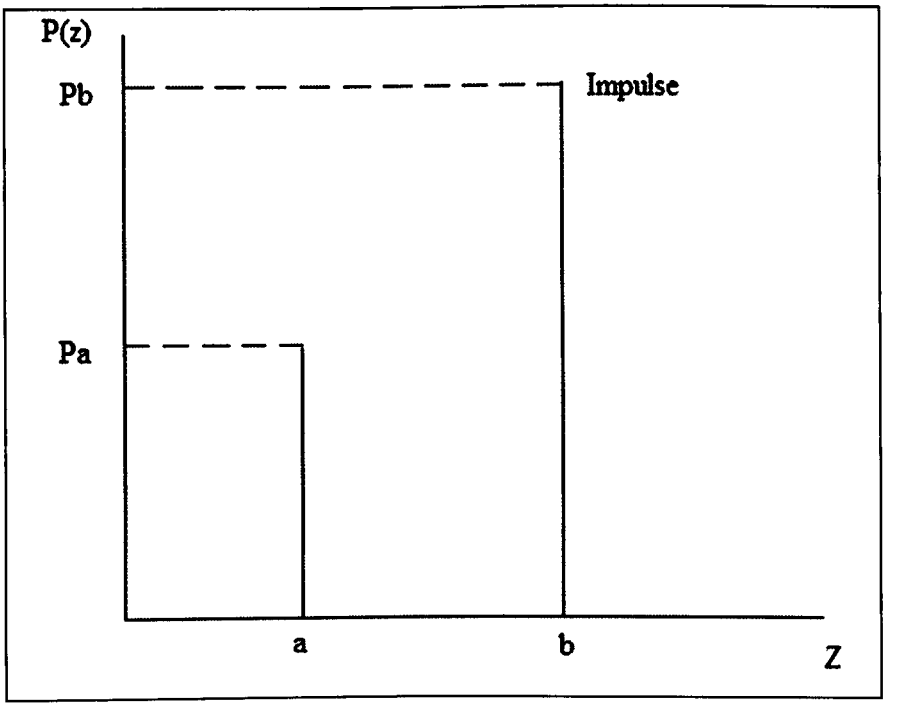


Figure 3.11Impulse probability density function

In MATLAB, the impulse functions are controlled by the third input parameter (denoted here as σ) of the *imnoise* function. To implement salt and pepper noise in MATLAB, imnoise could be represented as *imnoise(I,'salt & pepper',σ)* adding salt and pepper noise to image I, where σ is the noise density. The visual effect of adding salt and pepper noise to the watermarked image can be seen in Figure 3.12.

The experiment uses a number of different σ values to allow better observation of the effect of salt and pepper noise on a watermarked image. The σ values ranged between 0.005 and 0.1

(a) (b)

(c) (d)

Figure 3.12 Effect of adding salt & pepper noise to an image (a) & (c) Original watermarked image in LSB and DWT (b) & (d) Salt & pepper noised image in LSB and DWT.

An observation of the image histogram and Fourier spectrum, as shown in Figure 3.13, yields a similar conclusion to that of Gaussian noise. In some cases, the image histogram and Fourier spectrum show the same indication of an increase in pixel value variation and a high frequency component of the image as stronger noise is added. This gives a clear indication that salt and pepper noise attack is also essentially a high pass filter function.

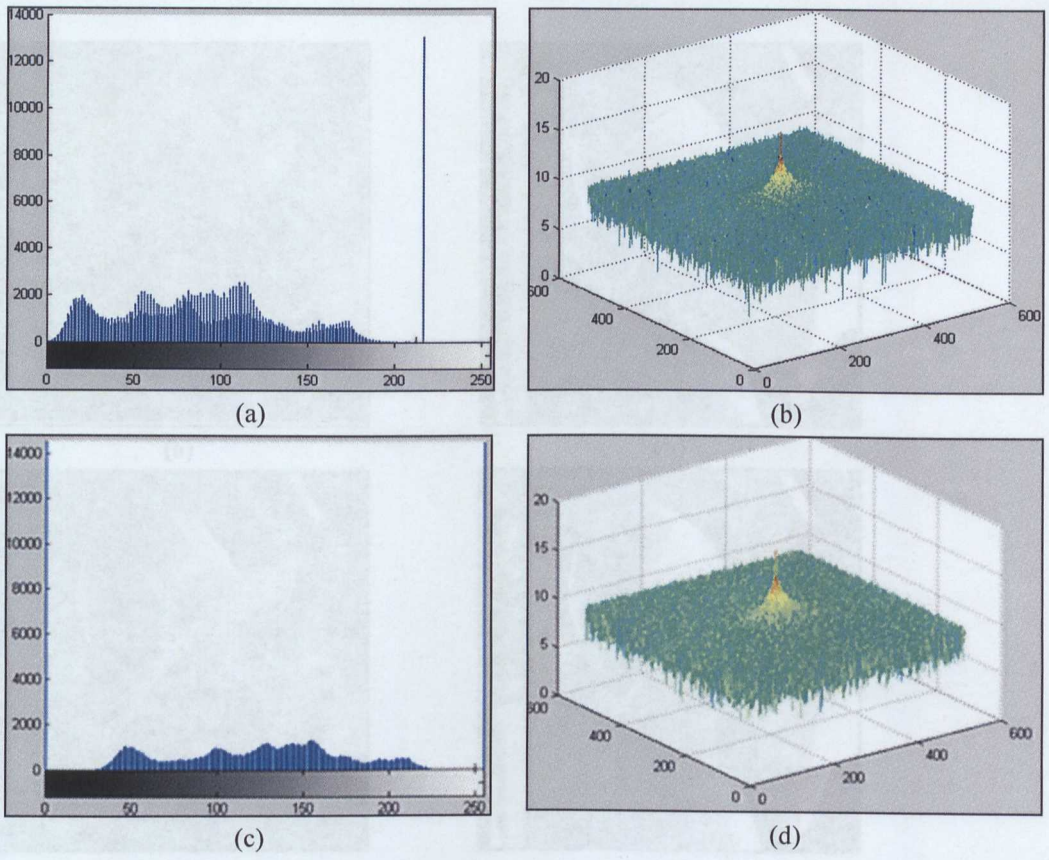


Figure 3.13 (a) & (c) Histogram of the salt & pepper noised watermarked image in LSB and DWT with $\sigma = 0.1$ (b) & (d) Fourier spectrum of the salt & pepper noised watermarked image in LSB and DWT with $\sigma = 0.1$.

3.3.1.4 Histogram equalisation

Histogram equalisation is a method used in image processing for contrast adjustment using the image's histogram, especially when the usable data of the image is represented by close contrast values. This method works by reducing the number of unique grey values in an image and reshaping the histogram to approximate a uniform distribution. In effect, histogram equalisation is controlled by adjusting the desired number of unique grey values. In MATLAB this is the equivalent of adjusting the second input parameter (denoted here as η) of the *histeq* function, which takes η to determine and represents a unique number of colours in an output image. The effect of histogram equalisation is shown as Figure 3.14.

(a) (b)

(c) (d)

Figure 3.14 Effect of histogram equalization on an image (a) & (c) Original watermarked image in LSB and DWT (b) & (d) Histogram equalized image in LSB and DWT with $\eta = 10$.

The experiment uses a number of different η values to allow better observation of the effect of histogram equalisation on a watermarked image. The η values used ranged between 10 and 200.

Since histogram equalisation is a histogram reshaping process, its effect can be seen clearly by comparing the histograms of the original and the attacked watermarked image. Figures 3.15(a) and (c) show that the numbers of grey values have decreased dramatically and the shape of the histogram has also changed drastically.

The changes are not as straightforward in the frequency spectrum as in the spatial domain, which is illustrated in Figures 3.15(b) and (d). However, our observation shows a modest shift in energy from the high frequency region to the lower frequency region as the severity of attack is increased. Nonetheless, it is not sufficient to say that there are any significant changes in this domain to warrant any claim.

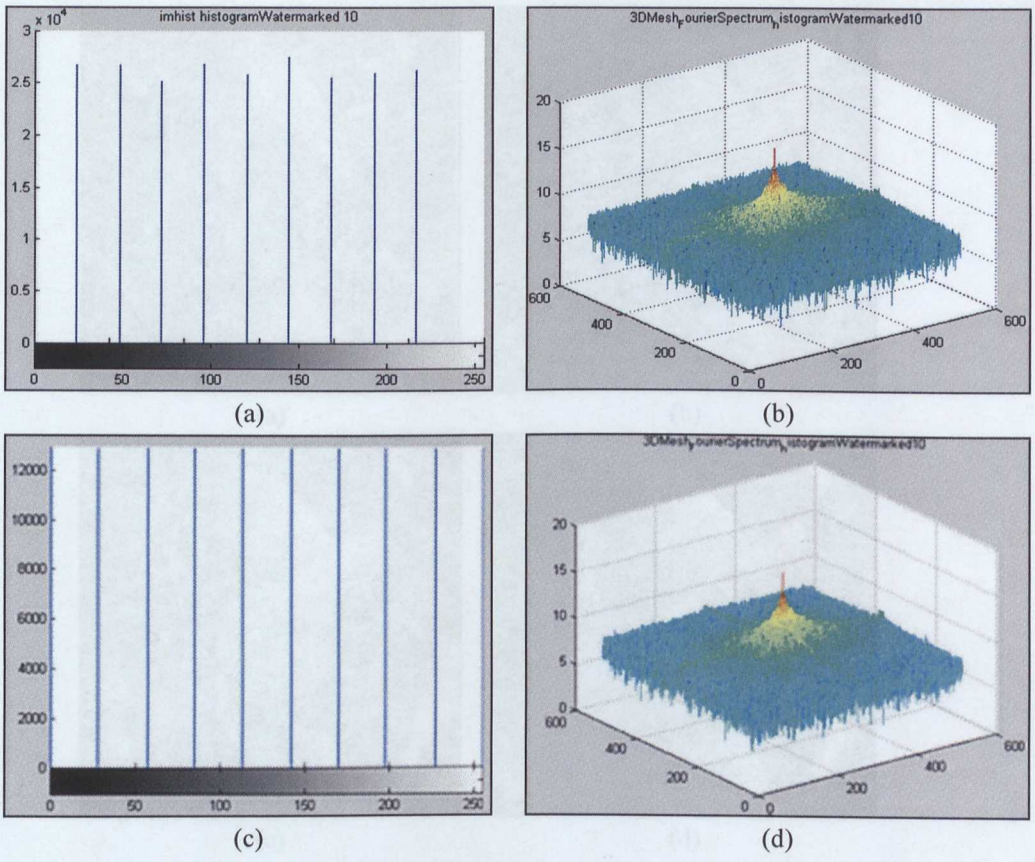


Figure 3.15 (a) & (c) Histogram in LSB and DWT with η = 10and (b) & (d) Fourier spectrum of the histogram equalized watermarked image in LSB and DWT with η = 10.

3.3.1.5 Sharpen

Image sharpening can be seen as the opposite of image smoothing whereby the former amplifies the presence of edges in an image. The process is achieved in the frequency domain by using a high pass filter, which intensifies high frequency components in the Fourier spectrum. A high pass filter $H(\omega)$ is obtained from its low pass filter $L(\omega)$ counterpart and calculated using the $H(\omega) = 1 - L(\omega)$ formula in the frequency domain. The parameter that controls this filter is similar to that of the Gaussian smoothing function. In MATLAB this is the equivalent of adjusting the second input parameter (denoted here as σ) of the *fspecial* function. The visual effect of sharpening can be seen in Figure 3.16.

The experiment uses a number of different σ values to allow better observation of the effect of sharpening on a watermarked image. The σ values ranged between 0.05 and 1.0.

(a) (b)

(c) (d)

Figure 3.16 Effect of sharpen on an image (a) & (c) Original watermarked image in LSB and DWT (b) & (d) sharpened image in LSB and DWT.

The histogram of the sharpened watermark image in Figures 3.17(a) and (c) shows that this attack shifts the histogram a few grey values to the left, which means that the overall intensity of the image is reduced. On the other hand, the shape of the histogram is more or less maintained, with the exception of some peaks which have been eliminated after the attacks. In the frequency domain, the effect of image sharpening is a lot more obvious, as shown in Figures 3.17(b) and (d). As a stronger sharpening coefficient is used, there is a shift in spectrum power from the lower frequency region to the higher frequency ones. This phenomenon in fact proves that sharpening is indeed the inverse of smoothing, i.e. the inverse of a low pass filter.

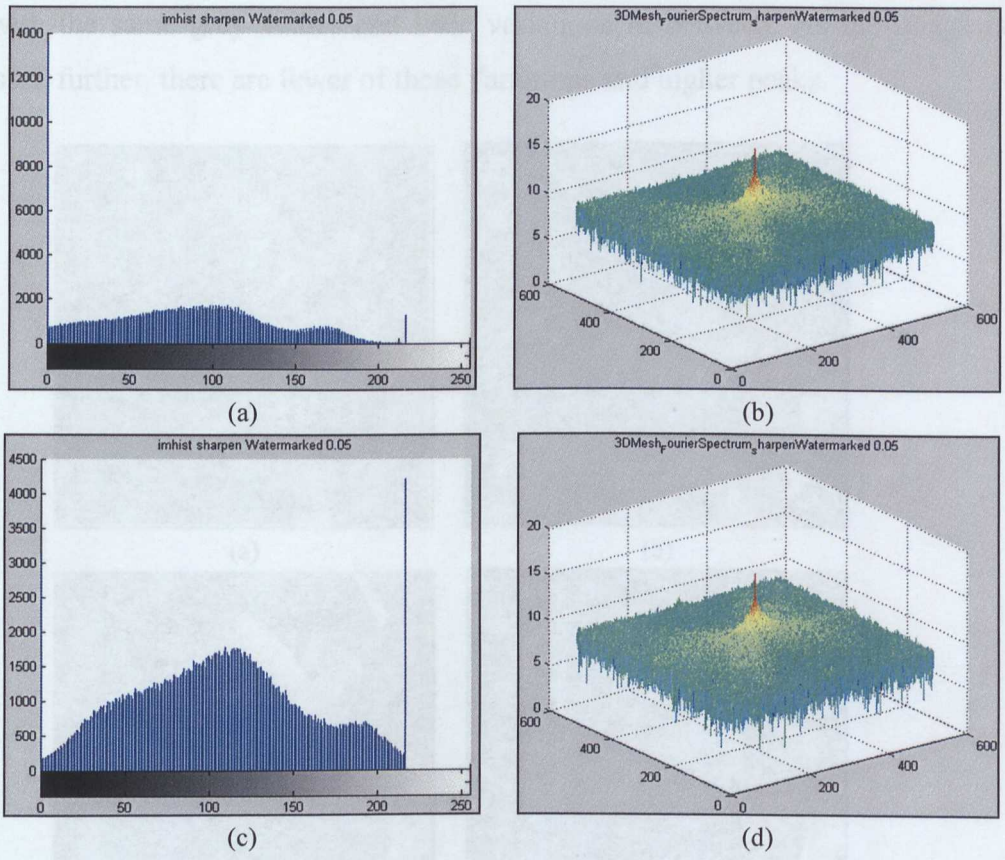


Figure 3.17 (a) Histogram in LSB and DWT with σ = 0.05and (b) Fourier spectrum of the sharpened watermarked image in LSB and DWT with σ = 0.05.

3.3.1.7 JPEG compression

JPEG is the most popular image compression technology due to its versatility and good compression ratio. It can perform both lossless and lossy compression, although it is JPEG lossy compression technology which has made it so versatile and popular. As with other lossy compression technologies, JPEG compression produces artefacts on the compressed image. These artefacts allow JPEG to be used to attack watermarks in an image. The visual effect of JPEG compression is shown in Figure 3.18.

The experiment uses a number of different compression ratios to allow better observation of the effect of JPEG compression on a watermarked image. The compression ratios used ranged between 10 and 90.

Lossy image compression often results in blocky patches, which are the manifestation of regions in the image with similar values. This can be seen from the histogram of the JPEG compressed image in Figures 3.19(a) and (c), which shows a number of

pixels with the same grey values and little variations in between. As the image is compressed further, there are fewer of these variations and higher peaks.

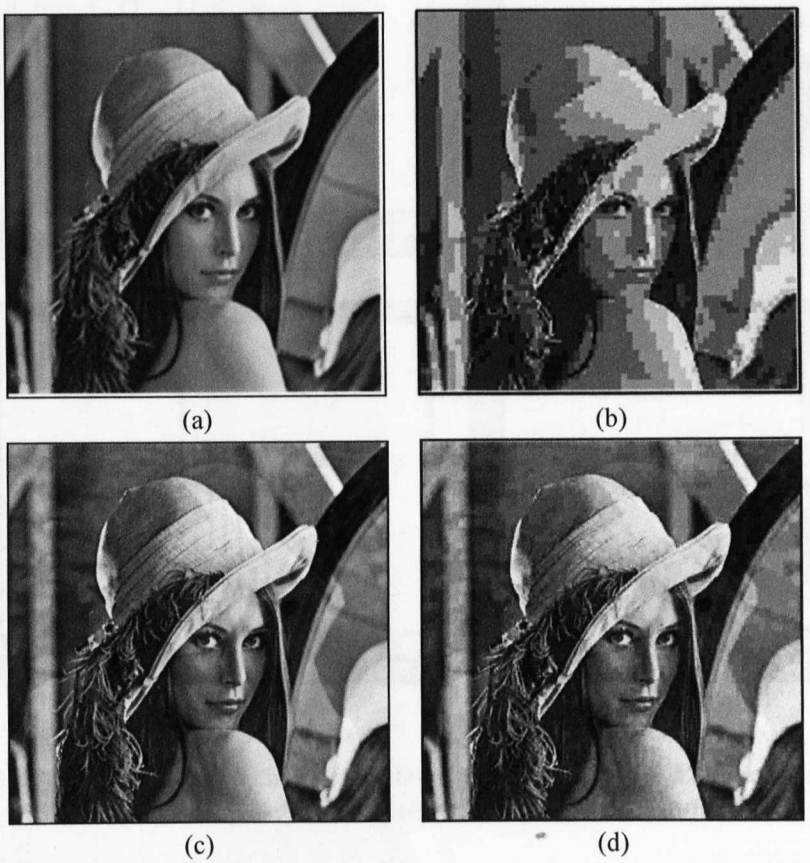


Figure 3.18 Effect of JPEG compression on an image (a) & (c) Original watermarked image (b) & (d) compressed image in LSB and DWT.

Although observation of the spatial domain suggests that the effect of JPEG compression is more akin to a low pass filter, in which image pixel variation is reduced, an observation of the Fourier spectrum, as shown in Figures 3.19(b) & (d), does not give such a clear-cut conclusion. In fact, we noted an increase in strength in the high frequency components of the spectrum and a decrease in strength in the low frequency components [Xu J, 2004].

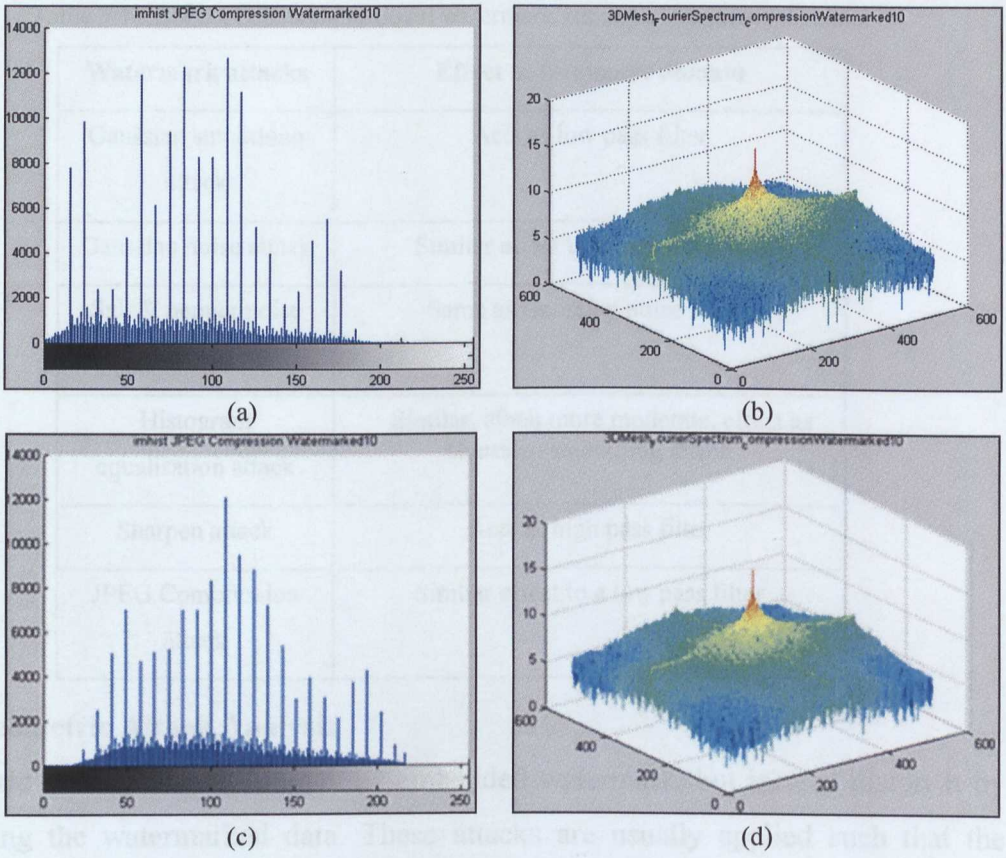


Figure 3.19 (a) & (c) Histogram in LSB and DWT with σ = 10and (b) & (d) Fourier spectrum of the JPEG compressed watermarked in LSB and DWT with σ = 10.

3.3.1.8 Summary

The analysis of removal attacks on watermarks is summarised in Table 3.1. Our analysis indicates that watermark attacks can be divided into two general categories: high frequency (HF) and low frequency (LF). The former category attacks the high frequency part of the signal while keeping most of the low frequency component relatively unchanged, and the latter category attacks vice versa. Gaussian noise attack, salt and pepper noise attack, sharpen attack and JPEG compression attack belong to HF watermark attacks, whereas Gaussian smoothing attack and histogram equalization attack are more akin to the LF watermark attack.

Table 3.1 Effects of different removal watermark attacks in frequency domains

Watermark attacks	Effect in frequency domain
Gaussian smoothing attack	Acts as low pass filter
Gaussian noise attack	Similar effect to a high pass filter
Salt & pepper noise attack	Same as Gaussian noise attack
Histogram equalisation attack	Similar, albeit more moderate, effect as Gaussian smoothing attack
Sharpen attack	Acts as high pass filter
JPEG Compression attack	Similar effect to a low pass filter

3.3.2 Geometric Attack Analysis

Geometric attacks do not remove the embedded watermark, but instead distort it by alternating the watermarked data. These attacks are usually applied such that the watermark detector loses synchronisation with the embedded information. Three geometric attacks will be analysed in this section, namely translation, rotation and scaling. The effects of attacks are analysed by comparing the original watermarked image and the attacked watermarked images.

3.3.2.1 Translation

The position of each image pixel in an original image is mapped to a new position in an output image and is named a 'translation', which performs a geometric transformation. Under translation, an image pixel located at (x_1, y_1) in the original is shifted to a new position (x_2, y_2) in the corresponding output image by displacing it through a user-specified translation ratio $(ß_x, ß_y)$. The performance of the translation operation is listed below:

$$x_2 = x_1 + \beta_x \tag{3.3}$$

$$y_2 = y_1 + \beta_y \tag{3.4}$$

In MATLAB this is the equivalent of adjusting the parameter of the *imtransform* function. The visual effect of a translation attack is shown in Figure 3.20. The translation ratios used ranged between 10 and 100.

(a) (b)

(c) (d)

Figure 3.20 Effect of Translation on an image (a) & (c) Original watermarked image (b) & (d) translated watermarked image in LSB and DWT with translation ratio = 20.

3.3.2.2 Rotation

Image rotation is performed by computing the inverse transformation for every destination pixel. It maps the position (x_1, y_1) of a pixel in an original image onto a position (x_2, y_2) in an output image by rotating it through a user-specified angle θ. In some implementations, output locations (x_2, y_2) which are outside the boundary of the image are ignored. The performance of the rotation operation is listed below:

$$x_2 = x_1 cos\theta - y_1 sin\theta \quad (3.5)$$

$$y_2 = x_1 sin\theta + y_1 cos\theta \quad (3.6)$$

In MATLAB this is the equivalent of adjusting the second input parameter of the *imrotate* function (denoted here as θ). It represents as *imrotate(A, θ)*, which rotates image A by angle degrees in a counter-clockwise direction. The visual effect of the rotation attack is shown in Figure 3.21. The angle θ ranged between 30 and 90.

(a) (b)

(c) (d)

Figure 3.21 Effect of Rotation on an image (a) & (c) Original watermarked image (b) & (c) rotated watermarked image in LSB and DWT with rotate angle = 30.

3.3.2.3 Scaling

In computer graphics, image scaling is the process of resizing a digital image, and it can be used to shrink or enlarge the size of an image. Image reduction is performed by replacement or by interpolating between pixel values in a local neighbourhood. Image zooming is achieved by pixel replication or by interpolation.

In MATLAB this is the equivalent of adjusting the parameter of the *imresize* function. It represents as *imresize(A,m,method)* and returns an image that is m times the size of A using the interpolation method. The visual effect of a scaling attack is shown in Figure 3.22. The scaled watermarked image is twice the size of the original watermarked image.

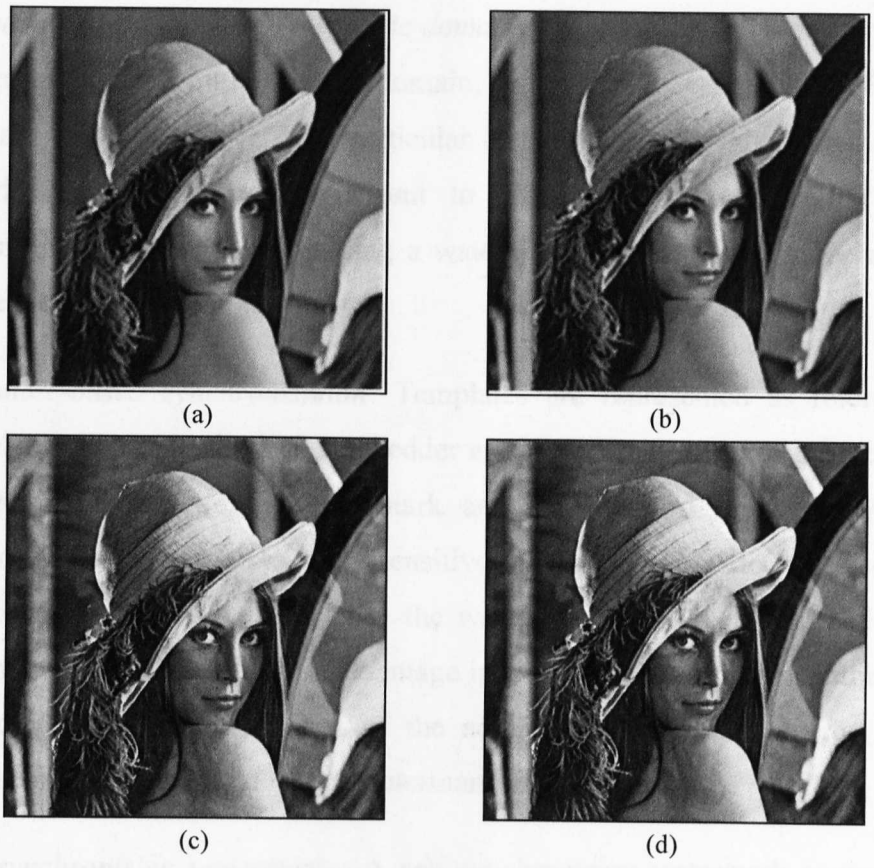


Figure 3.22 Effect of scaling on an image (a) & (c) Original watermarked image (b) & (d) Scaled watermarked image in LSB and DWT.

In conclusion, every geometric attack is defined by a set of parameters that determines the operation performed over the target image. Consider a typical case, in which the original image x is watermarked, yielding y. Let y suffers a geometric attack conditioned to the vector of attack parameters $\xi = (\xi_1 \ldots \xi_L)$, that is $z = T(y, \xi)$, where z represents the watermarked image attacked by a geometric transformation T. At the receiver side, the decoder only has access to z such that a conservative decoding strategy assumes that the image has been tampered with, and it therefore performs a countermeasure in an attempt to invert the attack. The decoder estimates the attack parameters based on the observed sample z. We can formulate the attack inversion procedure, or synchronization, as $y' = T^{-1}(z, \xi')$.

Licks [Licks, 2005] surveyed a number of image watermarking techniques that are designed to be robust against geometric attack. Successful methods are generally: (1) embedding information in a specific domain, (2) using templates, (3) using a self-synchronising watermark and (4) using features points to achieve synchronisations.

Embedding information in a specific domain: There are several algorithms that embed watermark images into a specific domain, and one such example is the Fourier-Mellin domain [Joseph, 1998] . This particular technique exploits the fact that an image's Fourier-Mellin domain is resistant to global rotation, translation and scaling transformations, hence embedding a watermark in this domain is be robust against these attacks as well.

Template-based synchronisation: Templates are represented as reference patterns which are known by both the embedder and the detector. Templates are added to the image in addition to the watermark and are intended only for synchronisation purposes that do not contain any sensitive information. The decoder uses knowledge about the template to synchronise the watermark with its reference. However, this technique has a significant disadvantage in that it might compromise the watermark's capacity, because it uses part of the admissible embedding energy to guarantee synchronisation and not to carry watermark information.

Self-synchronising watermarks: A self-synchronising watermark approach does not use templates but instead relies on the watermark's autocorrelation properties to achieve synchronisation. Generally, a self-synchronising watermark algorithm is an autocorrelation approach which contains several peaks. On the extraction stage, the decoder correlates the received watermarked image with itself and uses knowledge about the autocorrelation function's periodic nature to synchronise the watermark.

Synchronisation based on image features: Identifying certain feature points in an image before and after an attack is a characteristic of synchronisation based on image features. It is possible to invert the attack that distorted the watermarked image if enough of these points establish correspondence. This method does not rely on the presence of a template, which an attacker can erase to confuse the watermark decoder, but rather on intrinsic image features.

3.4. Chapter Summary

Watermark attacks are often deliberately applied to a watermarked image in order to remove or destroy any watermark signals in the host data. The purpose of the attack is aimed at disabling the copyright protection system offered by the watermarking technology. Our research in this area found a number of different types of watermark

attacks, which can be classified into a number of categories including removal attacks, geometry attacks, cryptographic attacks and protocol attacks. After that, we presented a thorough analysis of a number of attack types on image watermarking. The analysis was carried out using two image analysis tools, namely an image histogram and the Fourier spectrum, for spatial and frequency domain analysis, respectively. The results also uncovered a number of common similarities between different types of watermark attack. This is a property which could be exploited when designing a new solution for a more robust digital image watermarking technique

In the remainder of this thesis, Chapters 4 and 5 will describe region-adaptive watermarking algorithms based on the DWT domain and DWT-SVD. Furthermore, watermark attack detection as a novel watermark application will be discussed in Chapter 6.

CHAPTER 4

DESIGN OF A ROBUST REGION-ADAPTIVE WATERMARKING SCHEME

The previous chapter analysed different watermark attacks including removal and geometric attacks. Image histograms and the Fourier spectrum were used to carry out the analysis. The experiment and results indicate that removal watermark attack can be divided into two categories, namely a high frequency part watermark attack and a low frequency part watermark attack. In this chapter, a watermarking algorithm named the 'region-adaptive watermarking system' will be described. In section 4.3, we will present image segmentation, and then quad-tree image partition will be defined in section 4.4. After that, the region-adaptive watermark algorithm will be introduced followed by an experiment. The last section will describe the weaknesses of the region-adaptive watermarking system.

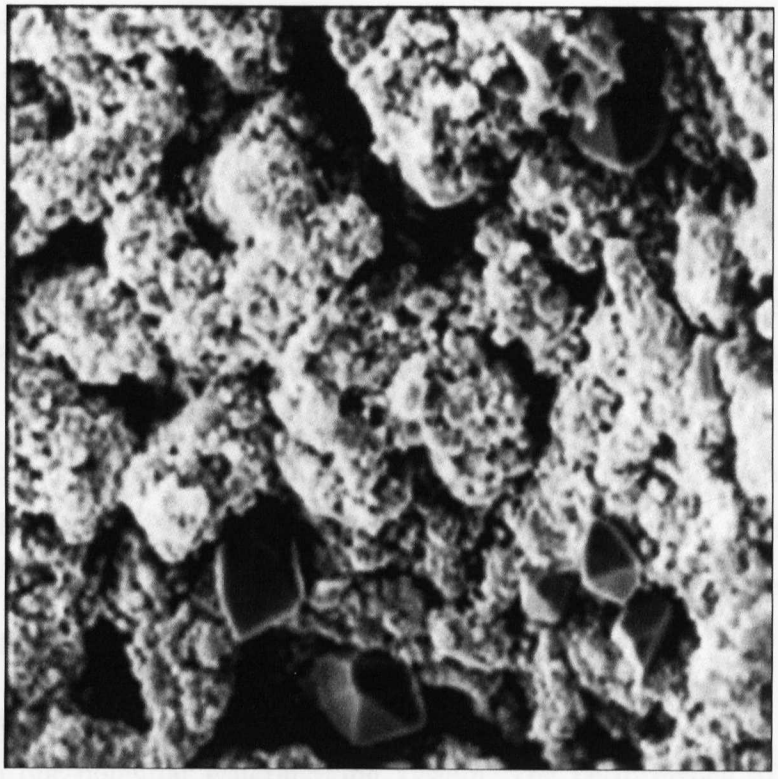
4.1. Rationale of the Region-Adaptive Watermarking Algorithm

Our previous work dealt with determining the effect of different watermark attacks on watermark signals. The experiment conducted involved subjecting watermarked images to a number of removal attacks, and then observing the results in both spatial and frequency domains.

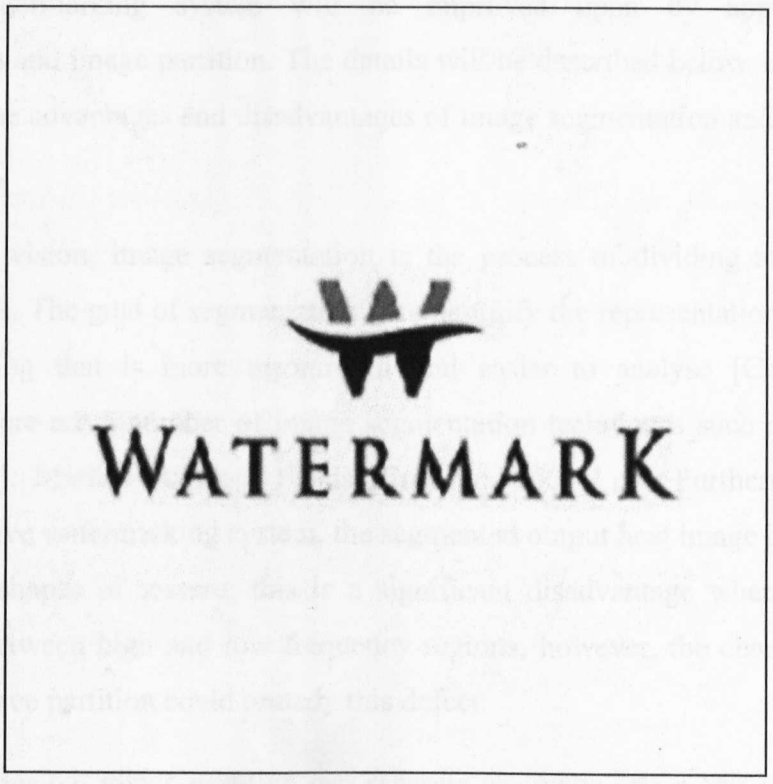
In this work, we analysed the effect of six different watermark attacks, namely Gaussian noise attack, salt and pepper noise attack, Gaussian smoothing attack, sharpen attack, histogram equalisation attack and JPEG compression attack.

The analysis was carried out using two image analysis tools, namely an image histogram and Fourier transforms in both spatial and frequency domain respectively. The results in Table 3.1 identified some common similarities between different types of watermark attack, which could be divided into two general categories: high frequency (HF) and low frequency (LF). Gaussian noise attack, salt and pepper noise attack, sharpen attack and JPEG compression attack belong to the HF watermark attack category, whereas the Gaussian smoothing attack, and histogram equalisation attack are more akin to the LF watermark attack.

To test this hypothesis we developed our proposed region-adaptive watermark technique, which employs two unique solutions, namely a) it uses two watermark images, each with strong high frequency or low frequency components, and b) it embeds different parts of the watermark images into the host image based on the difference between their spectral distributions. Examples of high frequency (HF) and low frequency (LF) watermark images are shown in Figure 4.1. The details of our region-adaptive watermarking technique will be described in the next few sections.



(a)



(b)

Figure 4.1 (a) the HF watermark image and (b) the LF watermark image

4.2. Analysis and Design of the Solution

To test and subsequently prove our initial hypothesis, we need to design a novel image watermarking algorithm that addresses a number of problems.

Firstly, our hypothesis states that in order to improve robustness there is a need to match the frequency spectrum of the host image with that of the watermark. This necessitates the identification of regions in the host and watermark images which can be classified as high and low frequency areas. For the watermark image, this requirement can be satisfied by having dual watermarks using two watermark images with distinct high and low frequency spectrums. For the host image, there are several approaches that can be used to identify different frequency regions, such as image partition, image segmentation, etc.

Image segmentation and image partition are of equal importance and complement each other in our potential watermarking system. The performance of the region-adaptive watermarking system will be improved upon by applying image segmentation and image partition. The details will be described below, and Table 4.1 summaries the advantages and disadvantages of image segmentation and image quad-tree partition.

In computer vision, image segmentation is the process of dividing an image into multiple parts. The goal of segmentation is to simplify the representation of an image into something that is more meaningful and easier to analyse [Carson, 2002]. Currently, there are a number of image segmentation techniques such as Blobworld [Chad, 1999], Markov Random Fields [Erdogan, 2001], etc. Furthermore, in our region-adaptive watermarking system, the segmented output host image is represented as arbitrary shapes of texture; this is a significant disadvantage when making the distinction between high and low frequency regions; however, the characteristics of image quad-tree partition could remedy this defect.

In computer vision, image partition can separate an image into a tree structure. The whole square-sized image is represented by the root of the tree; the image is then slit into several quadrants that correspond to children of the root node. The splitting process is recursively iterated on each leaf in the tree, resulting in many small regions corresponding to the condition that is being approximated. This indicates that the

techniques can potentially be used in our application to separate high and low frequency regions in the segmented host image. Therefore, this approach will be used as a means of identifying different frequency regions.

Table 4.1 Advantages and disadvantages of image segmentation and quad-tree partition

	Image Segmentation	Image Partition
Advantage	• Simplify host image • Separate texture regions	• Separate high & low frequency regions • Square-sized region obtained
Disadvantage	Arbitrary shape of texture region	The result is not accurate if applying quad-tree partition to host image directly

After that, an embedding algorithm needs to be considered. As we recall in section 2.4, the transform domain is more popular than the spatial domain, and a discrete wavelet transform in the transform domain has many special advantages that compare with conventional transforms such as DFT and DCT. One of the most significant advantages is that DWT has spatial frequency locality. So in this watermarking algorithm, the transform domain DWT will be applied.

In conclusion, the structure and process of the proposed region-adaptive watermarking algorithm are as follows:

1. Dual watermark images will be used to provide high frequency and low frequency watermark components
2. High and low frequency regions in the host image will be identified by
    a. Segmenting the host image
    b. Quad-tree partitioning the segmented host image
3. The embedding process to be done in the transform domain.

The resulting technique can be classified as a region-adaptive watermarking algorithm because it inserts watermark data into selected regions of the host image instead of the entire image. In most conventional watermarking techniques, the watermark is embedded into an entire image for full coverage/protection [Hubballi, 2009, Ren, 2009, Sudirman, 2009], but it imposes some limitations on applications that are involved with multi-stage information processing, such as target identification. Selective watermarking is more effective than full-image watermarking in protecting the authenticity of images [Di, 2006]. Most image analysis applications work on specific parts of an image rather than the entire image regions. These specific parts

are referred to as region of interests. Most region-adaptive watermarking techniques start by identifying a number of features in the host image hence locating the regions of interest. This then followed by watermark embedding onto these regions [k.Zebbiche, 2008, Parameswaran, 2008, Phadikar, 2009, Verma, 2009].

4.3. Image Segmentation

Image segmentation refers to the process of partitioning a digital image into multiple segments. There are a number of image segmentation algorithms that can be used such as Blobworld [Chad, 1999] , Gabor filters [Jain, 1990], k-means clusters [Ng, 2006], image histogram thresholding [Tobias, 2002] and Markov random fields [Erdogan, 2001], to name a few. Different image segmentation algorithms have different characteristics.

To select the most suitable image segmentation technique for our region-adaptive watermarking algorithm, we experiment with two image segmentation techniques for comparison purposes, namely histogram thresholding and Markov random fields.

4.3.1 Image histogram thresholding segmentation

Suppose that the image histogram in Figure 4.2 corresponds with an image f(x, y), composed of light objects on a dark background, in such a way that object and background pixels have intensity values grouped into two dominant modes. One obvious way to extract the objects from the background is to select a threshold, T, which separates these modes. The segmented image g(x, y) is given by:

$$g(x,y) = \begin{cases} 1 & if\ f\ (x,y) > T \\ 0 & if\ f(x,y) \leq T \end{cases} \qquad (4.1)$$

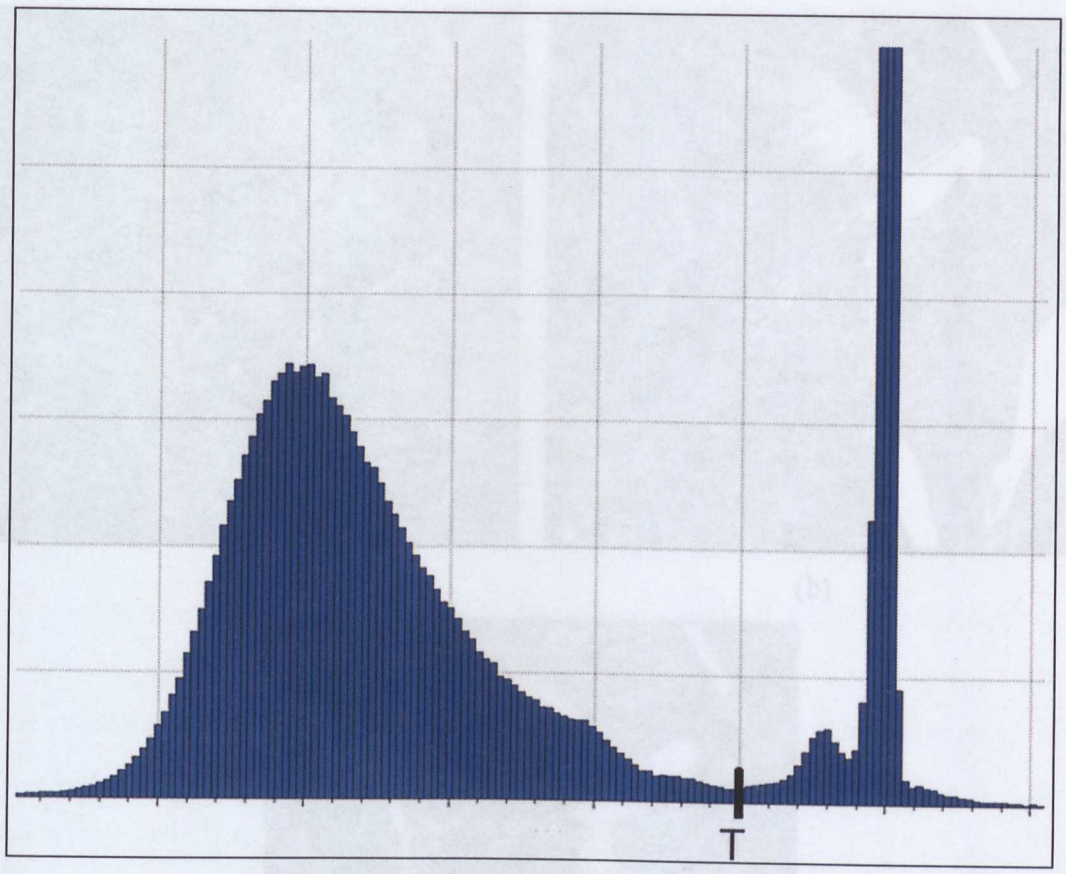


Figure 4.2 Image histogram that can be partitioned by a single threshold

4.3.2 Markov Random Field Image Segmentation

Image segmentation can be optimally posed as Bayesian labelling, in which the solution to a problem is defined as the maximum a posteriori (MAP) probability estimate of the true labelling. The posterior probability is usually derived from a prior model and a likelihood model. The latter relates to how data is observed and is problem domain-dependent. The former depends on how various prior constraints are expressed. The Markov random field (MRF) models theory is a tool used to encode contextual constraints into the prior probability [Erdogan, 2001].

4.3.3 Experiment and Selection

To select the most suitable image segmentation algorithm, we compare an image histogram thresholding segmentation algorithm and an MRF image segmentation algorithm through experimentation, which is shown in Figure 4.2. Figure 4.2a is an original image, Figure 4.2b and 4.2c represent an output image produced by the histogram thresholding segmentation algorithm and the MRF image segmentation algorithm, respectively.

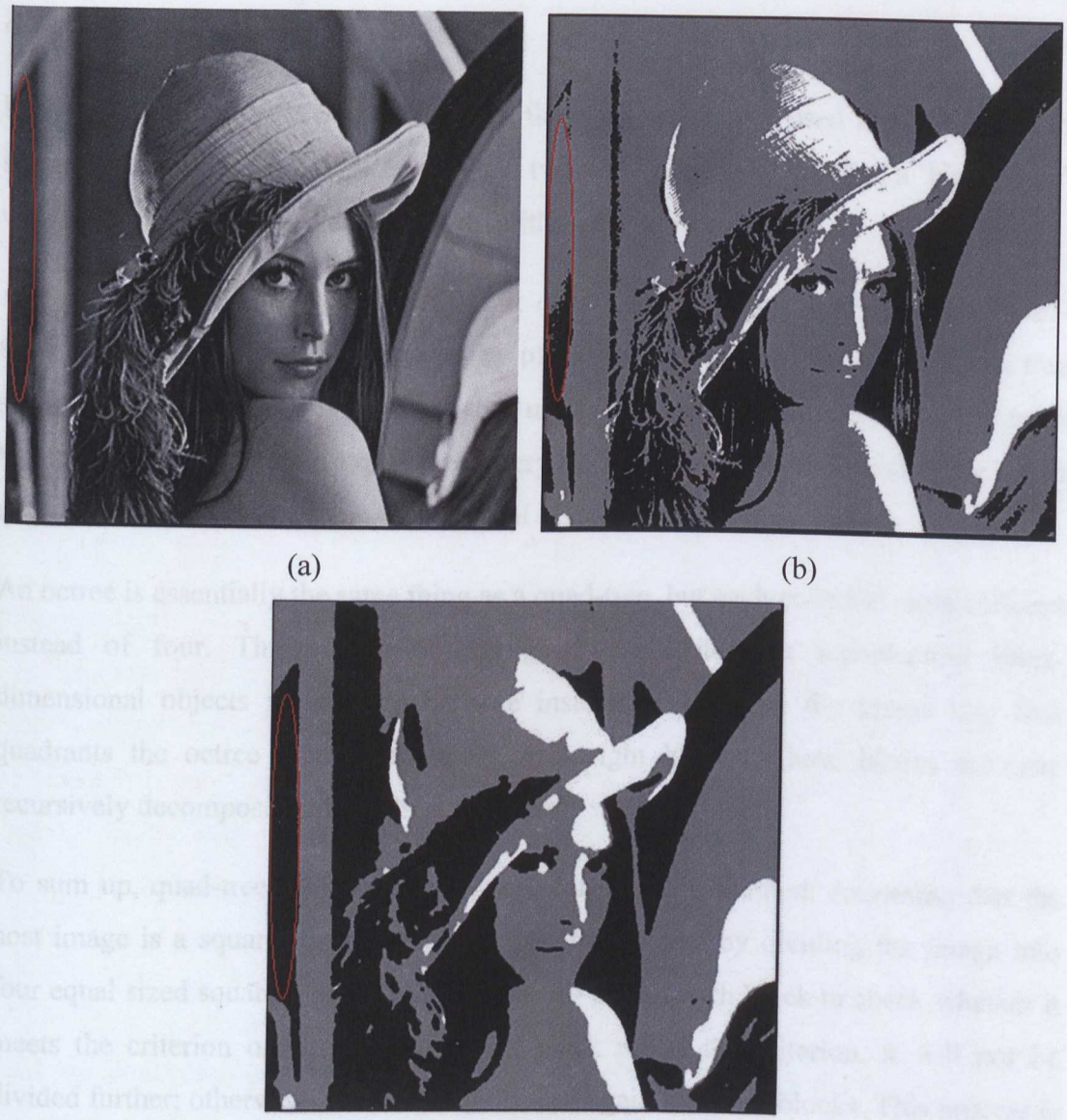


Figure 4.3 (a). Original image (b). Image histogram thresholding segmentation (c). Markov random field image segmentation

As can be seen from Figure 4.3, we found similar results in Figures 4.3b and 4.3c, whereby the input image is segmented into three different kinds of tints, although Figure 4.3c looks more accurate than Figure 4.3b. For example, in the red circle of the original image, histogram thresholding image segmentation divides them into two regions – grey and black – however, in the MRF image segmentation, only black pixels are clustered. So, Markov random field image segmentation is selected for our region-adaptive watermarking system.

4.4. Quad-Tree Image Partition

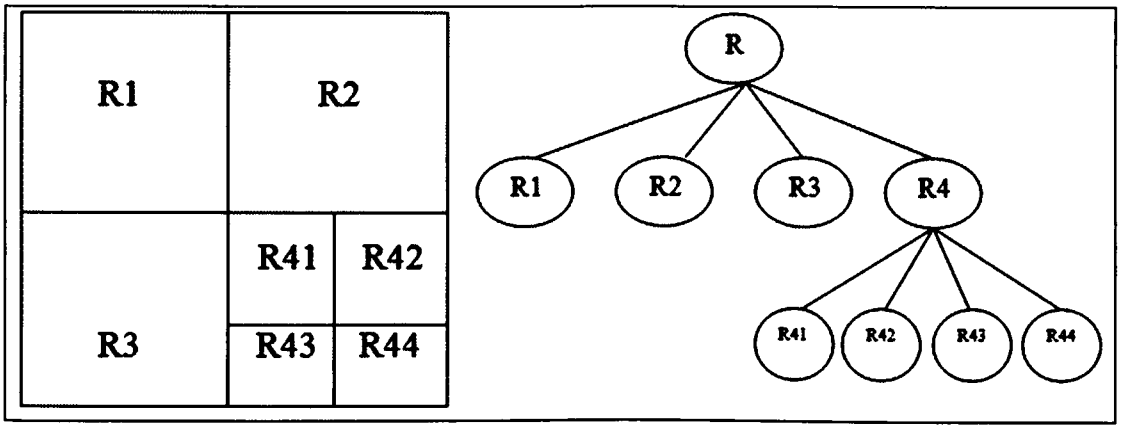
Binary trees, quad-trees and octrees are the most frequently used tree structures in image partitioning. Each node has at most two child nodes and is called a 'binary tree', which can be applied to data searching, arithmetic expression, etc.

A quad-tree is a hierarchical data structure often used for image representation. Quad-trees encode two-dimensional images by placing subsections of the image into a tree structure – much like a binary tree – but unlike a binary tree where every tree node has up to two children, in a quad-tree every node can have up to four children. Each node stores a particular piece of the overall image [Mark de Berg, 2000].

An octree is essentially the same thing as a quad-tree, but each node has eight children instead of four. This makes octrees good candidates for representing three-dimensional objects in memory because instead of breaking the image into four quadrants the octree breaks the image into eight blocks. These blocks are then recursively decomposed into eight sub-blocks.

To sum up, quad-tree partition is the most considerable method. Assuming that the host image is a square image, quad-tree partition works by dividing the image into four equal sized square blocks. A test is carried out on each block to check whether it meets the criterion of homogeneity. If a block meets the criterion, it will not be divided further; otherwise, it will be subdivided again into four blocks. This process is repeated until every block meets the criterion. The result may be blocks of several different sizes. Figure 4.4 illustrates the quad-tree partition of an image up to two levels. The parent node in the tree represents the entire image region, and its four descendants' nodes represent the large disjointed sub-regions.

The criterion used in deciding the partitioning of a region takes into account the pixel characteristics of that region. In general, the test checks whether the characteristics of the pixels in that region are too heterogeneous, in which case the test will yield negative and vice versa. To achieve this, an image segmentation algorithm is applied to the host image prior to the partitioning process. Each segment of the host image should represent areas in which pixel characteristics are homogenous. By performing this process, we also avoid the computationally expensive task of recalculating pixel characteristics in each region during the partitioning process.



(a) (b)

Figure 4.4 Quad-tree partition (a) Resulting blocks, (b) The quad-tree structure

4.5. Region-adaptive Watermarking Algorithm

Region-adaptive watermarking is a novel embedding technique where the watermark data is embedded on different regions of the host image using discrete wavelet transform methods. This technique is derived from the hypothesis that the robustness of a watermarking process can be improved by matching the watermark and host image regions with similar spectral distributions. To facilitate this, the technique utilises dual watermarking technologies and applies image segmentation and partitioning to create a selected region, and then it embeds parts of the watermark images into selected regions in the host image.

4.5.1 Recent Advances in the Region-Adaptive Watermarking Algorithm

Zebbiche proposed a region-adaptive watermarking algorithm in 2008. The watermark should insert into the most relevant parts of the image, which is known as the region of interest, to overcome attacks such as cropping [k.Zebbiche, 2008]. Figure 4.5 shows the personalised watermarking system. It shows how an original host image is divided into two separate images. The first image is background image and second image is ROI image, the watermark image will embed into ROI to form a watermarked ROI image. And then the background image and watermarked ROI would merge together to become whole watermarked image. The experiment indicated that the algorithm would remain robust against most attacks.

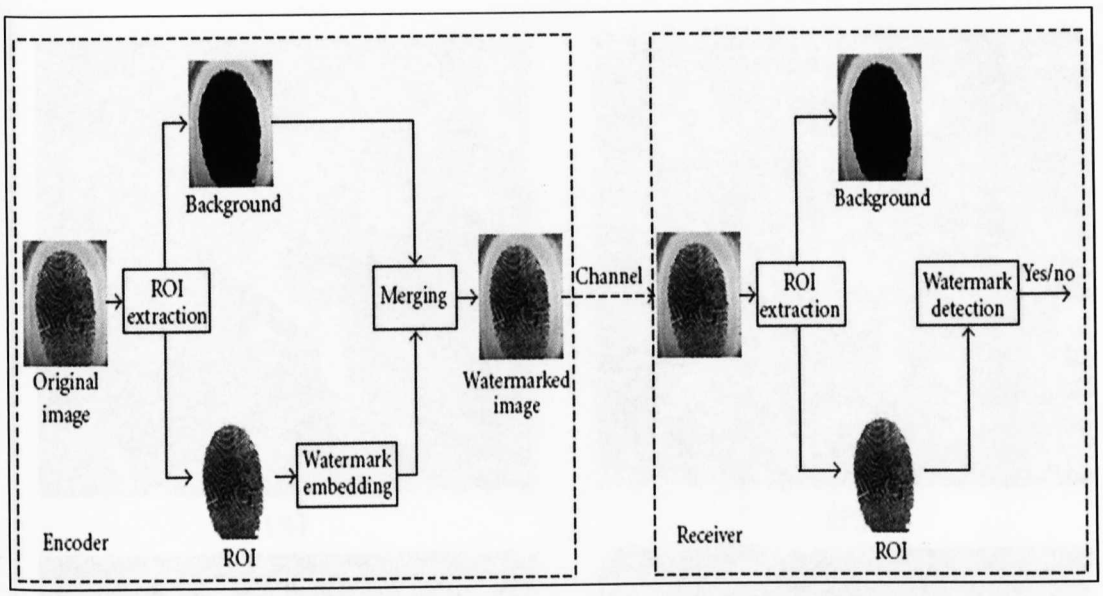


Figure 4.5 Personalized watermarking system applied to a fingerprint image

Figure 4.6 indicates another region-adaptive watermarking approach, which is described in [Nikolaidis, 2001]. The first stage is to find a segmentation or clustering technique that will provide us with a robust region representation under image processing, in the sense that it will not be seriously affected by usual geometric image manipulations. After the segmentation process, the resulting regions are ordered according to their size, excluding those that are on the image boundaries, in order to avoid problems arising from image cropping along boarders. The largest regions are preferred for watermarking, in order to preserve as much of the watermark's power as possible. The experiment tests the robustness of the approach by applying several processing attacks to numerous images. The results denote the fact that although the geometric handling of the prototype watermark ensured the robustness of the method to certain geometric manipulations, notably rotation, scaling, cropping and translation, the watermark proved to be robust to other attacks like compression and filtering.

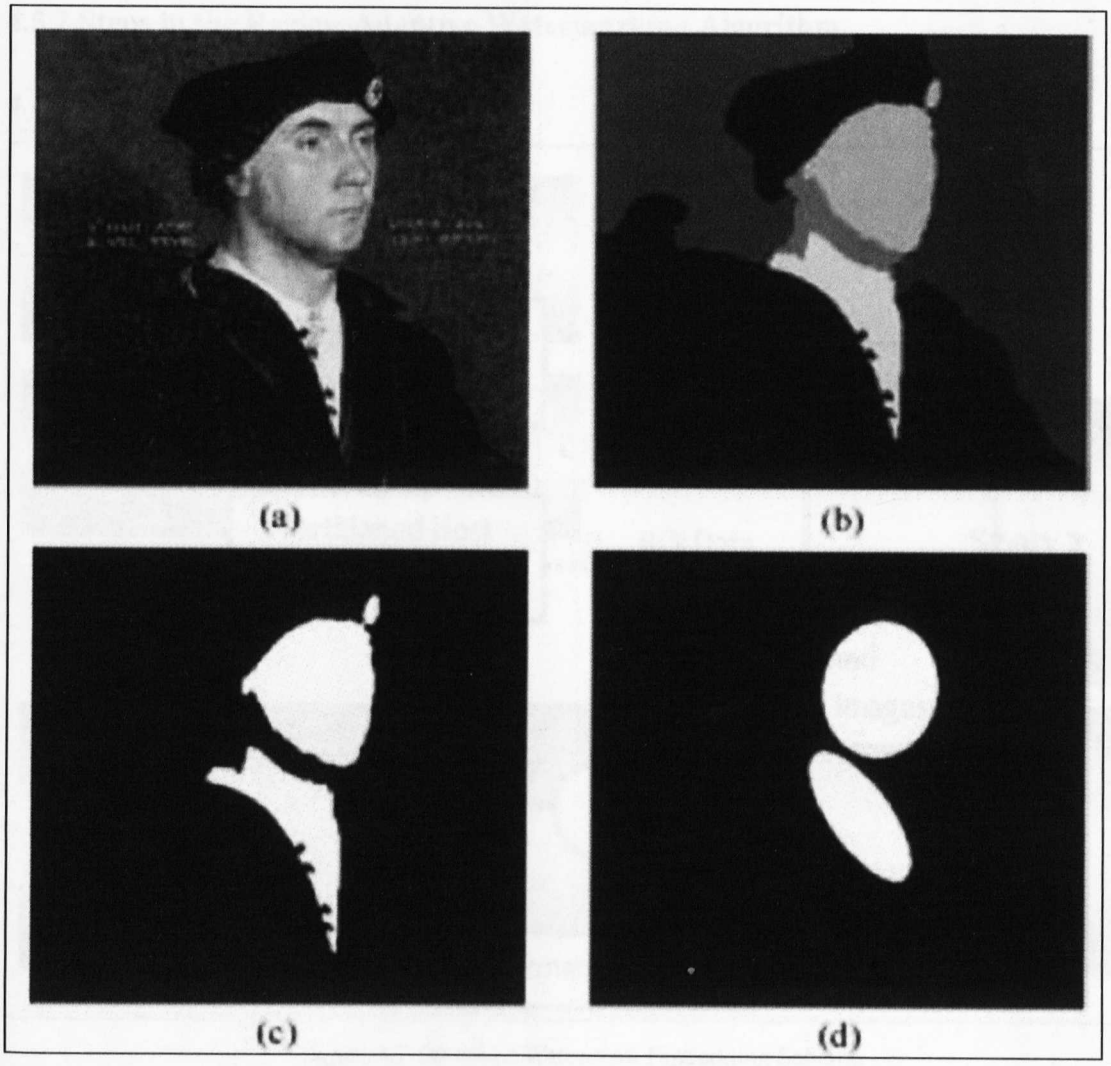


Figure 4.6 Segmentation and region approximation by ellipsoid

The above region-adaptive watermarking algorithms have some disadvantages. The most significant disadvantage is that the existing region-based watermarking algorithm could not be applied to all watermark images. However, the region-adaptive watermarking technique we propose in this chapter works by embedding parts of the watermark image into selected regions in the host image. The selection process works by matching the watermark and host image regions with similar spectral distributions. To improve the watermark embedding and extraction speed, we decided not to use regions with arbitrary shape and orientation. Instead, we will use non-overlapping squares of varying sizes as our regions.

4.5.2 Steps in the Region-Adaptive Watermarking Algorithm

4.4.2.1 Watermark Insertion

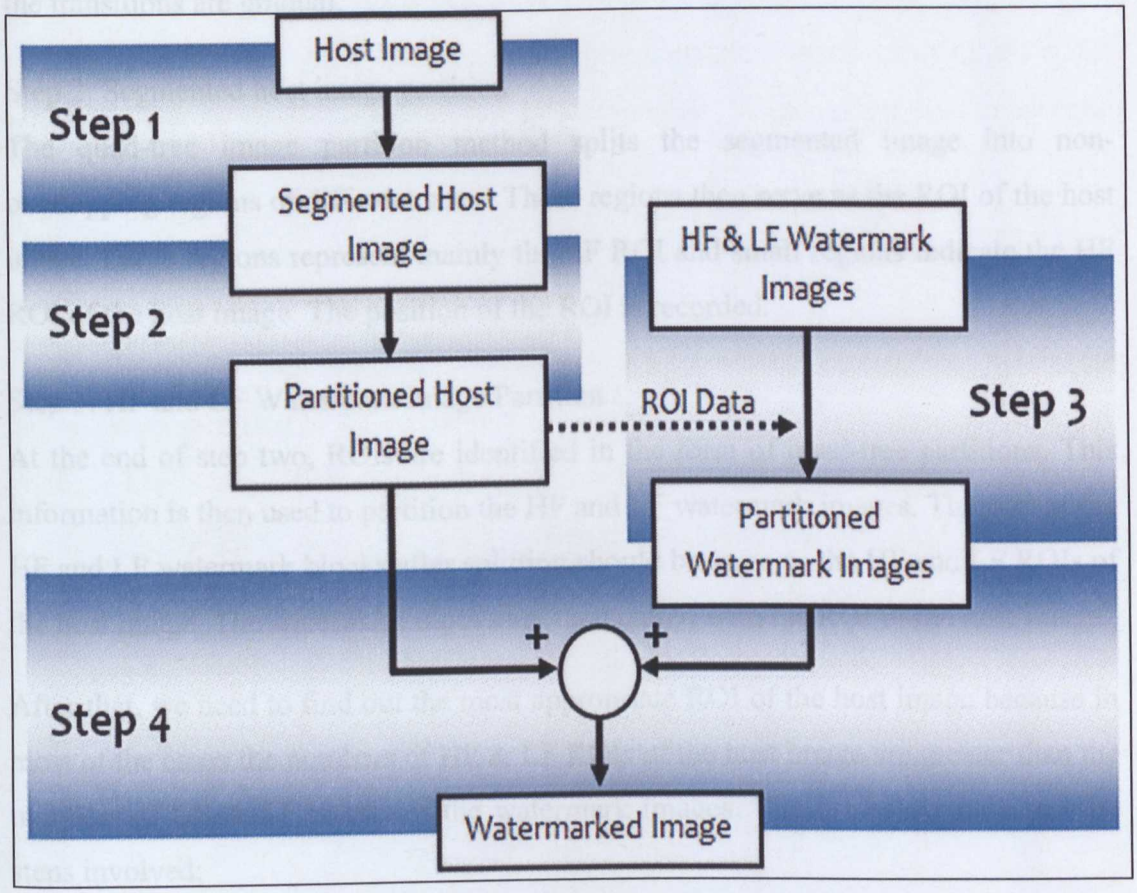


Figure 4.7 Proposed Watermark Embedding Scheme

Figure 4.7 describes the region-adaptive watermark embedding scheme, which can be divided into four main steps. The first step is image segmentation by MRF and is followed by the second step, which is the application of a quad-tree image partition to divide the host image into a block-based square. The third step splits the HF and LF watermark images into non-overlapping squares, and the final step applies DWT to the host and watermark images' blocks to insert watermark data.

Step 1: Host Image Segmentation

In this step the Markov random field image segmentation technique is used due to its computation speed and statistical categorisation of image textures. The goal of image segmentation in this step is to simplify the representation of the host image into a segmented image and to make it easier for the subsequent process to identify the HF and LF regions of interest (ROI) in the host image. HF regions can be seen as having

a great deal of abrupt tonal transitions in a small space and, conversely, LF regions are those where the tone remains relatively constant throughout these small areas and the transitions are gradual.

Step 2: Segmented host image partition

The quad-tree image partition method splits the segmented image into non-overlapping regions of different sizes. These regions then serve as the ROI of the host image. Large regions represent mainly the LF ROI and small regions indicate the HF ROI of the host image. The position of the ROI is recorded.

Step 3: HF and LF Watermark Image Partition

At the end of step two, ROIs are identified in the form of quad-tree partitions. This information is then used to partition the HF and LF watermark images. The size of the HF and LF watermark blocks after splitting should be same as the HF and LF ROIs of the host image. The watermark blocks are then paired with the ROI of the host image.

After that, we need to find out the most appropriate ROI of the host image because in most of the cases the numbers of HF & LF ROIs of the host image are greater than the numbers of HF & LF blocks of the watermark images. The list below indicates the steps involved:

1. Decompose ROI of host image by DWT algorithm to get ROI_ha, ROI_hh, ROI_hv, ROI_hd .
2. Decompose HF and LF watermark blocks to get W_a, W_h, W_v, W_d.
3. Loop 1: x = number of ROI of host image
   Loop 2 : y = number of watermark blocks

```
comparion_value _matrix (x,y) = sum(abs(ROI_ha - W_a)
                                + abs(ROI_hh - W_h)
                                + abs(ROI_hv - W_v)
                                + abs(ROI_hd - W_d);
```

   end

   end

4. Find out the min values in each row to get a new matrix min_value.
5. Sort out these values by ascend order in matrix min_value.
6. Find out the best regions

Step 4: Region-Adaptive Watermark insertion Algorithm

An additive-embedding strategy is applied in the region-adaptive watermarking system because it is simple and fast. It takes the form $I' = I + \alpha m$ where I' is the watermarked image, I the host image, α the embedding strength and m the watermark

signal [Sabbag, 2006]. At the end of step 3, we have regions in the host image matched with regions in HF and LF watermark image. The watermark data contained in each region of the watermark images is then inserted into the corresponding region in the host image using the additive-embedding approach, which is facilitated in the DWT domain by first applying DWT to each region before adding the watermark. The sub-steps are listed below:

1. Using the two-dimension separable dyadic DWT, obtain the first level decomposition of the host image region I.
2. Using the two-dimension separable dyadic DWT, obtain the first level decomposition of the watermark image W.
3. Modify the DWT coeffiecients V_{ij} in the LL, HL, LH and HH sub-bands: $V_{w,ij}^k = V_{ij}^k + \alpha W_{ij}$ where $V_{w,ij}^k$ represents as watermarked image's coefficient, V_{ij}^k is host image's coefficient, α is embedding strength, W_{ij} represents as watermark image's coefficient.
4. Apply inverse DWT to obtain the watermarked image.

The HF and LF ROI blocks are then combined to create the final watermarked image.

4.4.2.2 Watermark Extraction

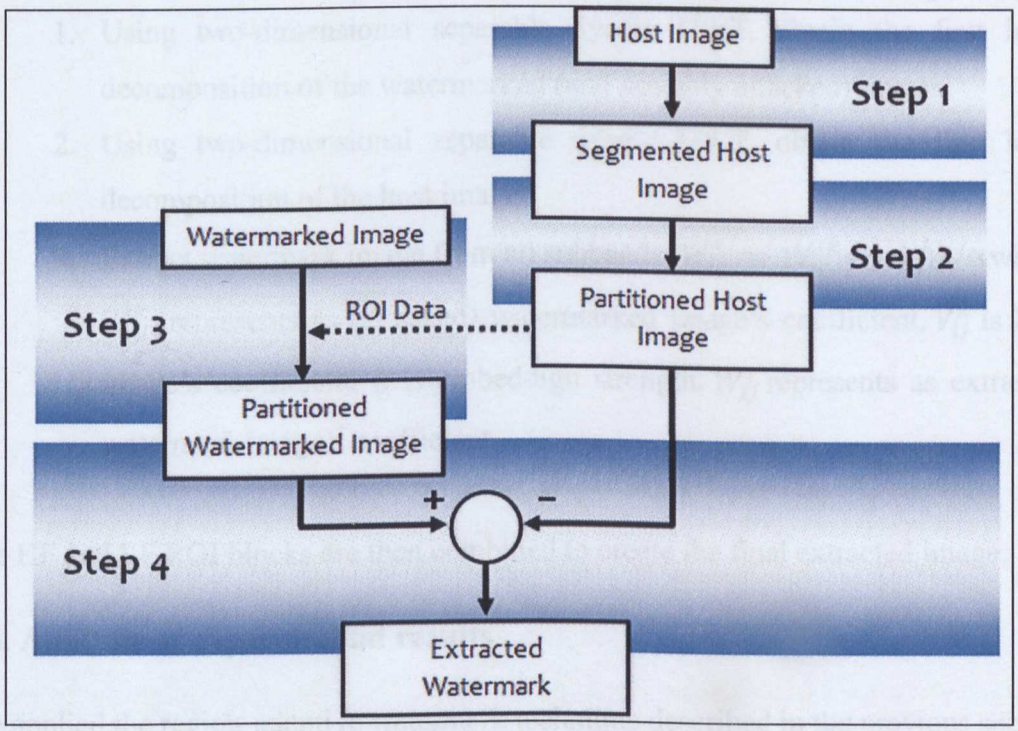


Figure 4.8 Proposed watermark extraction scheme

Figure 4.8 describes the watermark extraction scheme. There are four main steps in the extraction process and they are described below:

Steps 1 & 2: Host image segmentation and partition

Watermark extraction requires the original un-watermarked host image to calculate the ROI of the watermarked image. The process begins with the segmentation and partitioning of the un-watermarked host image, which is similar to the first two steps of the watermark insertion process described previously. The end result of these steps will be the sizes and locations of the ROI.

Step 3: Watermarked image partition

The sizes and locations of the ROI calculated from the previous step are used to partition the watermarked image. At the end of this step we will have the matching ROI of both the watermarked original host images.

Step 4: Region-Adaptive Watermark Extraction Algorithm

The partitioned watermarked image is first decomposed through DWT in four levels, and then we apply DWT to the all sub-band image. Next, HF and LF watermark image blocks are extracted out of the HF and LF ROI of the watermarked image by applying an extraction algorithm directly. The sub-steps are listed below:

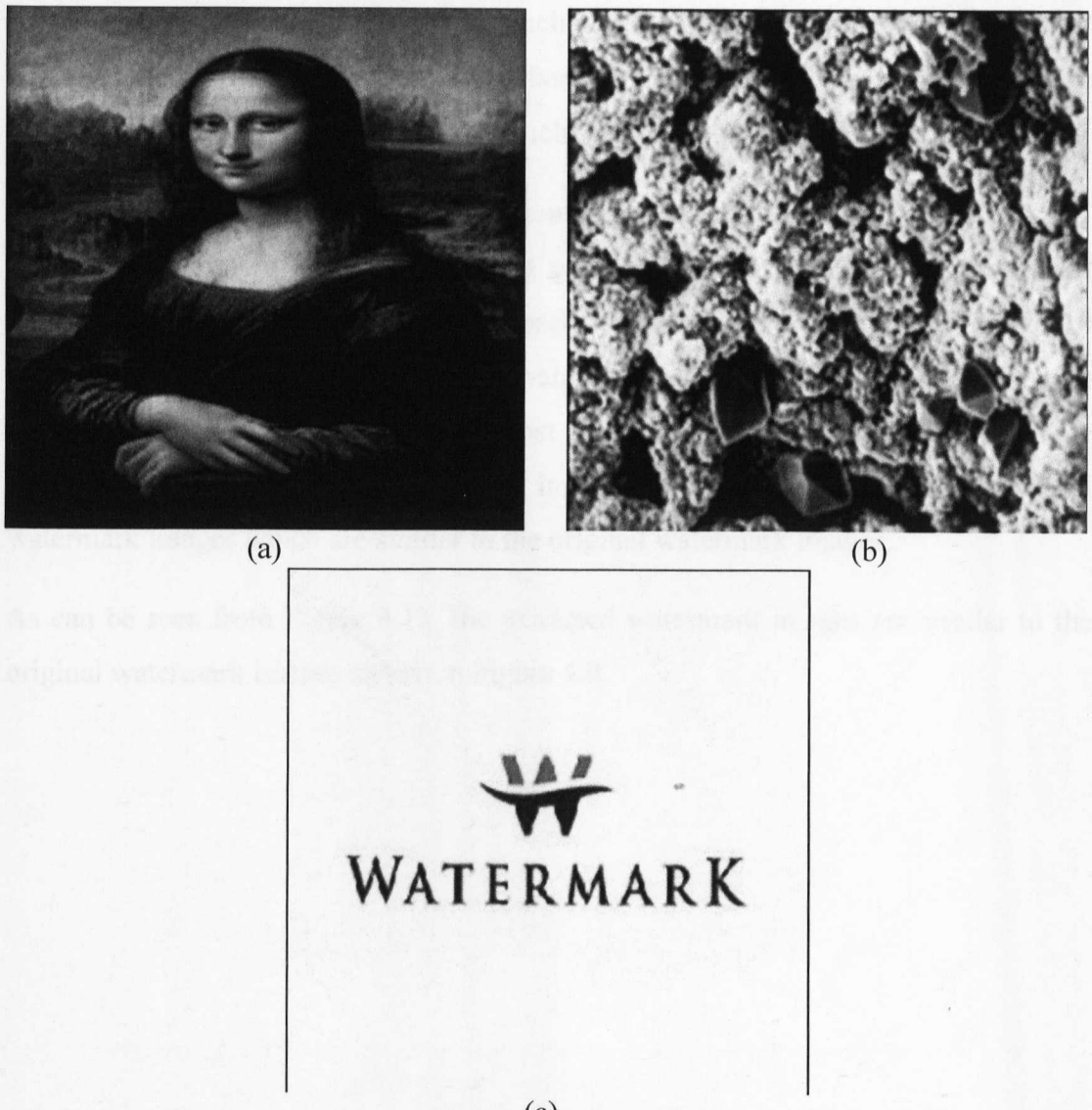
1. Using two-dimensional separable dyadic DWT, obtain the first level decomposition of the watermarked (and possibly attacked) image.
2. Using two-dimensional separable dyadic DWT, obtain the first level decomposition of the host image.
3. Extract watermark image from all subbands: $W_{ij}^{*} = (V_{w,ij}^{*k} - V_{ij}^{k})/\propto$where $V_{w,ij}^{*k}$ represents as (attacked) watermarked image's coefficient, V_{ij}^{k} is host image's coefficient, α is embeddign strength, W_{ij}^{*} represents as extracted watermark image's coefficient.

The HF and LF ROI blocks are then combined to create the final extracted image.

4.6. Analysis of experimental results

We applied the region-adaptive watermark technique described in the previous section to the images shown in Figure 4.9. The host image has dimensions of 512×512 pixels

and both of HF and LF watermark images have dimensions of 256×256 pixels. The HF watermark image is divided into 64 blocks of 32×32 pixels and the LF watermark image is divided into 16 blocks of 64×64 pixels.



(a) (b)

(c)

Figure 4.9 (a) Host image, (b) HF watermark image and (c) LF watermark image

As a quantitative measure of the degradation effect caused by the attacks we use the peak-signal-to-noise ratio (PSNR). The formula used to calculate the PSNR between the original and attacked watermarked signals can be found in Chapter 2. High PSNR values indicate lower degradation, hence signifying that the watermarking technique is more robust to that type of attack.

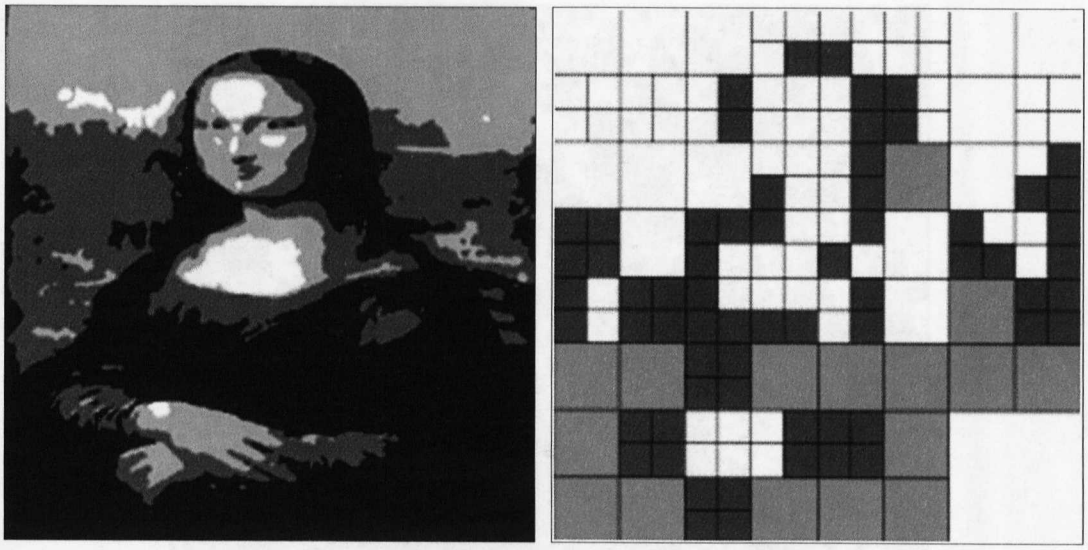
Two experiments were conducted to test the algorithm. The first experiment was aimed at verifying that inserted watermark images can be extracted with minimal distortion. The second experiment measured the robustness of the proposed region-

adaptive technique and compared it to the original DWT algorithm [Tao, 2004]. This algorithm embeds a watermark in all four bands at decomposition levels. The finding of the paper suggests that a watermark inserted in the lowest frequencies is most robust against one group of attacks such as JPEG compression, smoothing and Gaussian noise, while a watermark embedded in the highest frequencies is more robust against another groups of attacks including histogram equalisation etc.

4.6.1 Watermark insertion and extraction verification

The first experiment tried to prove that watermark images could be inserted and extracted completely by using the proposed region-adaptive technique. Figure 4.10 shows some intermediate results of the watermark insertion process, highlighting the segmented host image, the parts of the host image where the different watermarks are inserted and the resulting watermarked image. Figure 4.11 shows those extracted watermark images which are similar to the original watermark images.

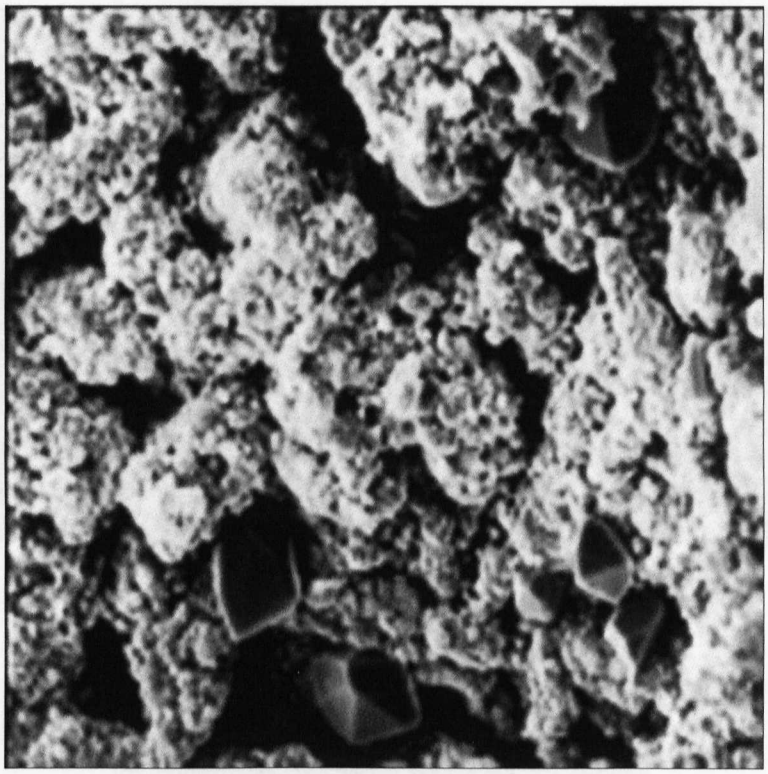
As can be seen from Figure 4.11, the extracted watermark images are similar to the original watermark images shown in Figure 4.9.



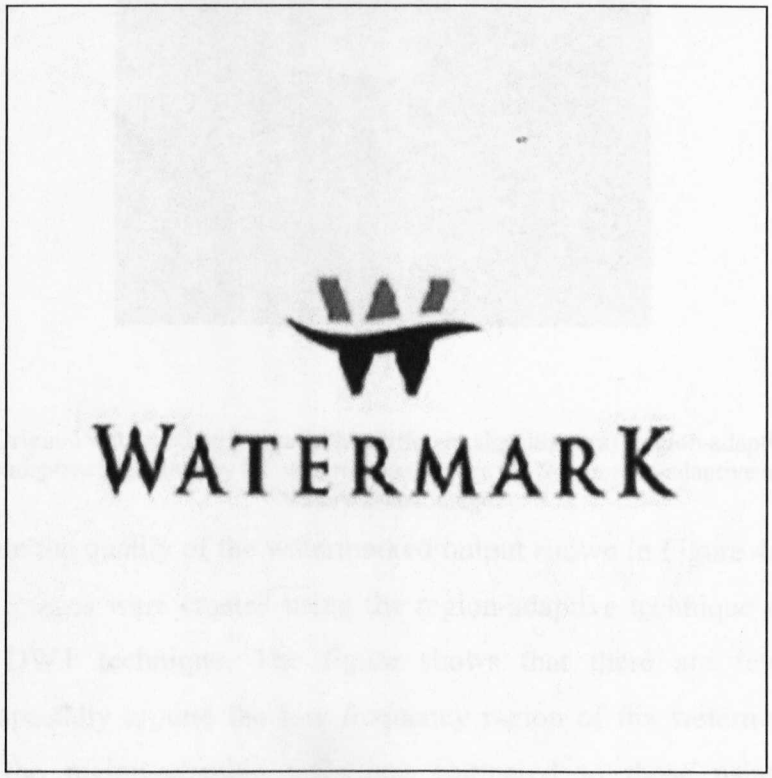
(a) (b)

(c)

Figure 4.10 MRF segementated host image, (b) watermark insertion region (dark grey – HF and light grey – LF) and (c) watermarked image



(a)



(b)

Figure 4.11 Extracted (a) HF watermark and (b) LF watermark images

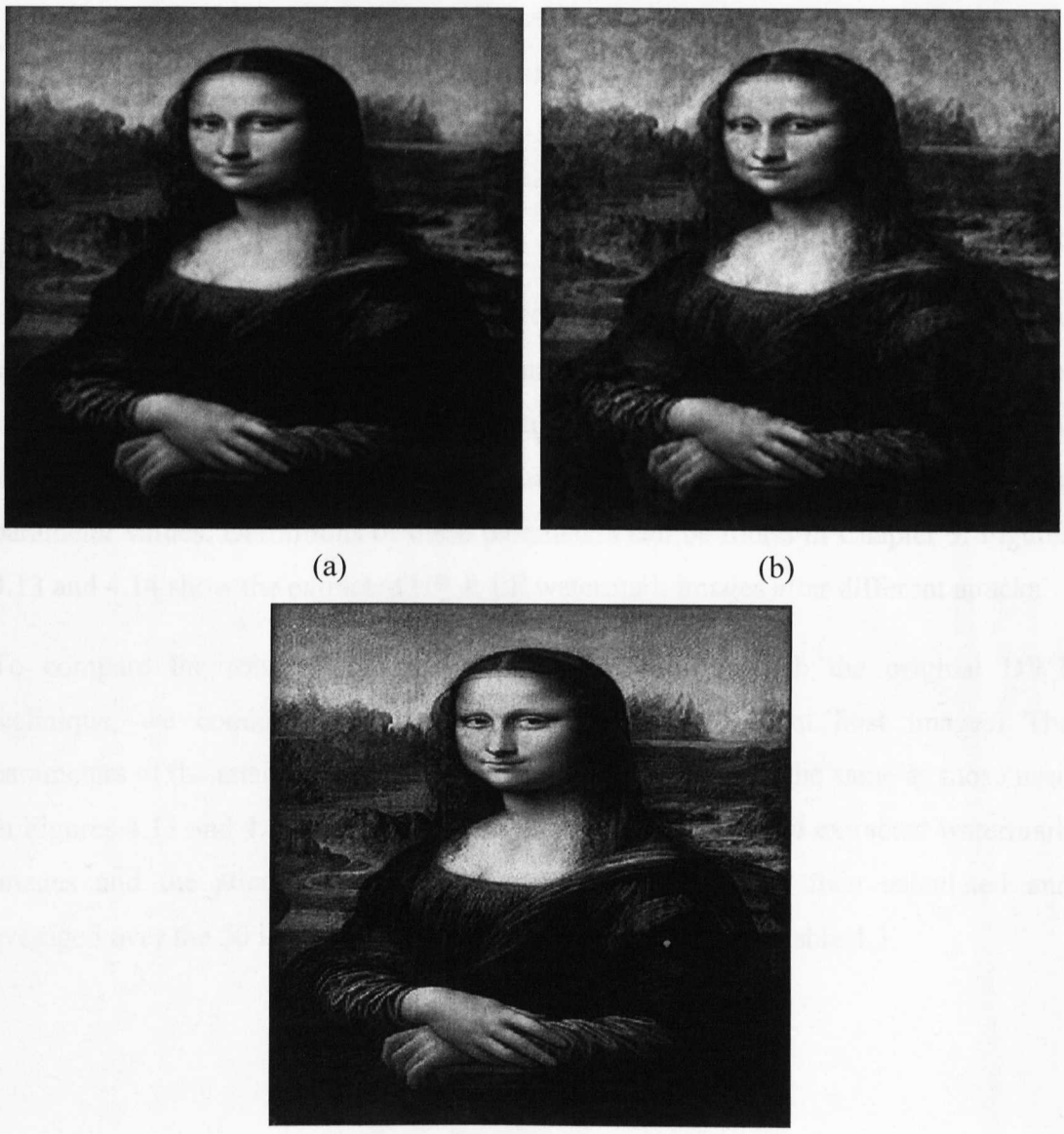


Figure 4.12 Original watermarked image with a different algorithm: (a) Region-adaptive algorithm. (b)Non-region-adaptive algorithm by HF watermarked image (c) Non-region-adaptive algorithm by LF watermarked image

To demonstrate the quality of the watermarked output shown in Figure 4.12, different watermarked images were created using the region-adaptive technique and by using the original DWT technique. The figure shows that there are fewer artefacts introduced, especially around the low frequency region of the watermarked image, when using the region-adaptive technique compared to those using the DWT technique. This result can also be quantitatively measured by the higher PSNR values of the region-adaptive technique, as shown in Table 4.2

Table 4.2 PSNR values of the Unmodified Watermarked Image

Region Adaptive	DWT (HF watermarked)	DWT (LF watermarked)
27.45	24.13	25.7

4.6.2 Robustness Comparison

To test the robustness of the proposed watermarking scheme, six watermark removal attacks were applied to the watermarked image: Gaussian noise, salt and pepper noise, sharpen, smoothing, histogram equalisation and JPEG compression attack. The severity of these attacks could be adjusted by modifying their corresponding parameter values. Definitions of these parameters can be found in Chapter 3. Figures 4.13 and 4.14 show the extracted HF & LF watermark images after different attacks.

To compare the robustness of the proposed technique with the original DWT technique, we conducted the experiment using 50 different host images. The parameters of the attack algorithms used in the experiment are the same as those used in Figures 4.13 and 4.14. The PSNR values between the original extracted watermark images and the attacked extracted watermark images were then calculated and averaged over the 50 images. These results are summarised in Table 4.3.

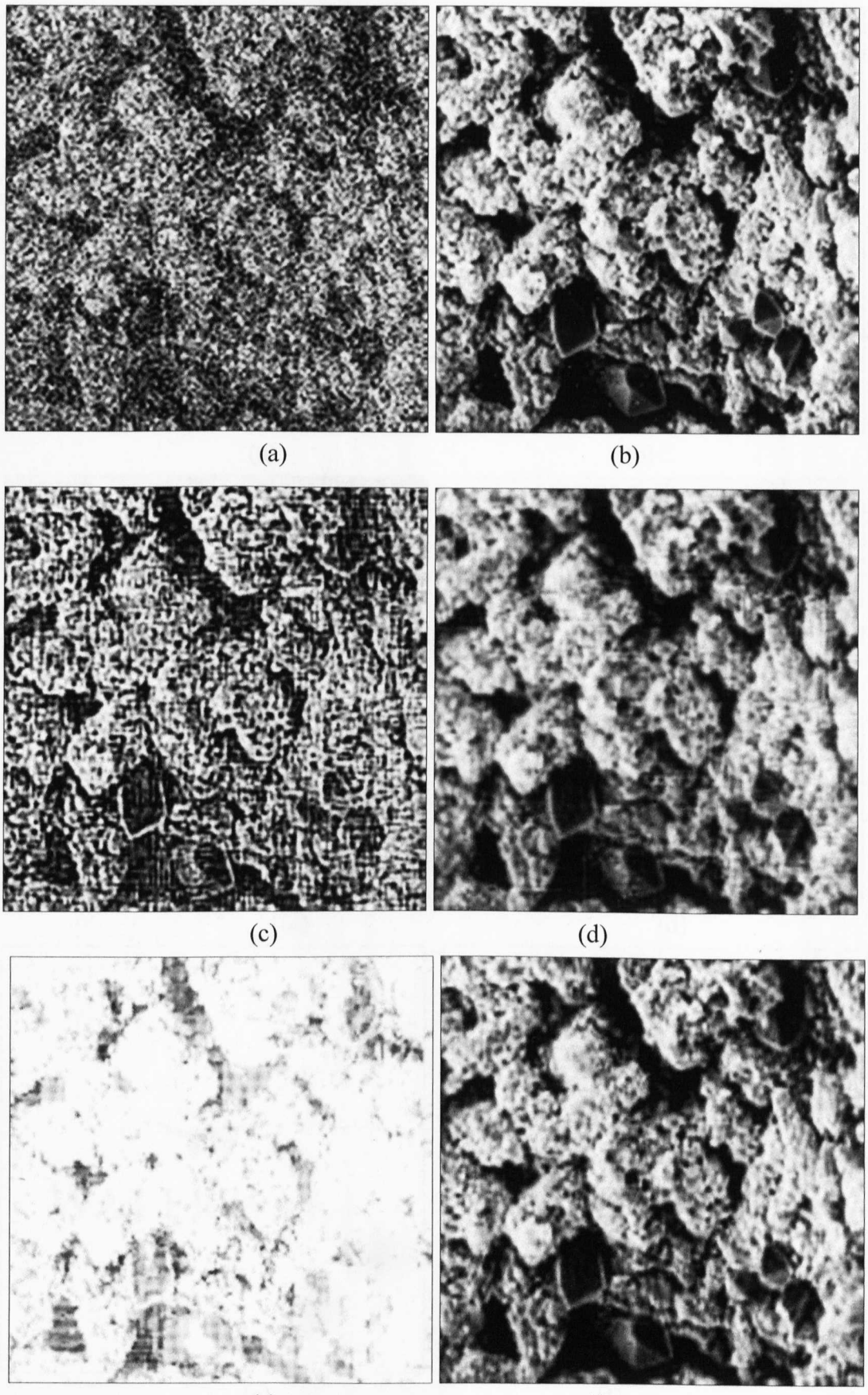


Figure 4.13 Extracted HF watermark image (a) Gaussian noise with $\sigma = 0.005$ (b) Salt and pepper noise with $\sigma = 0.005$ (c) Sharpen with $\sigma = 1.0$ (d) Smoothing with $\sigma = 200$ (e) Histogram equalisation with $\eta = 200$ (f) JPEG compression with $\sigma = 90$

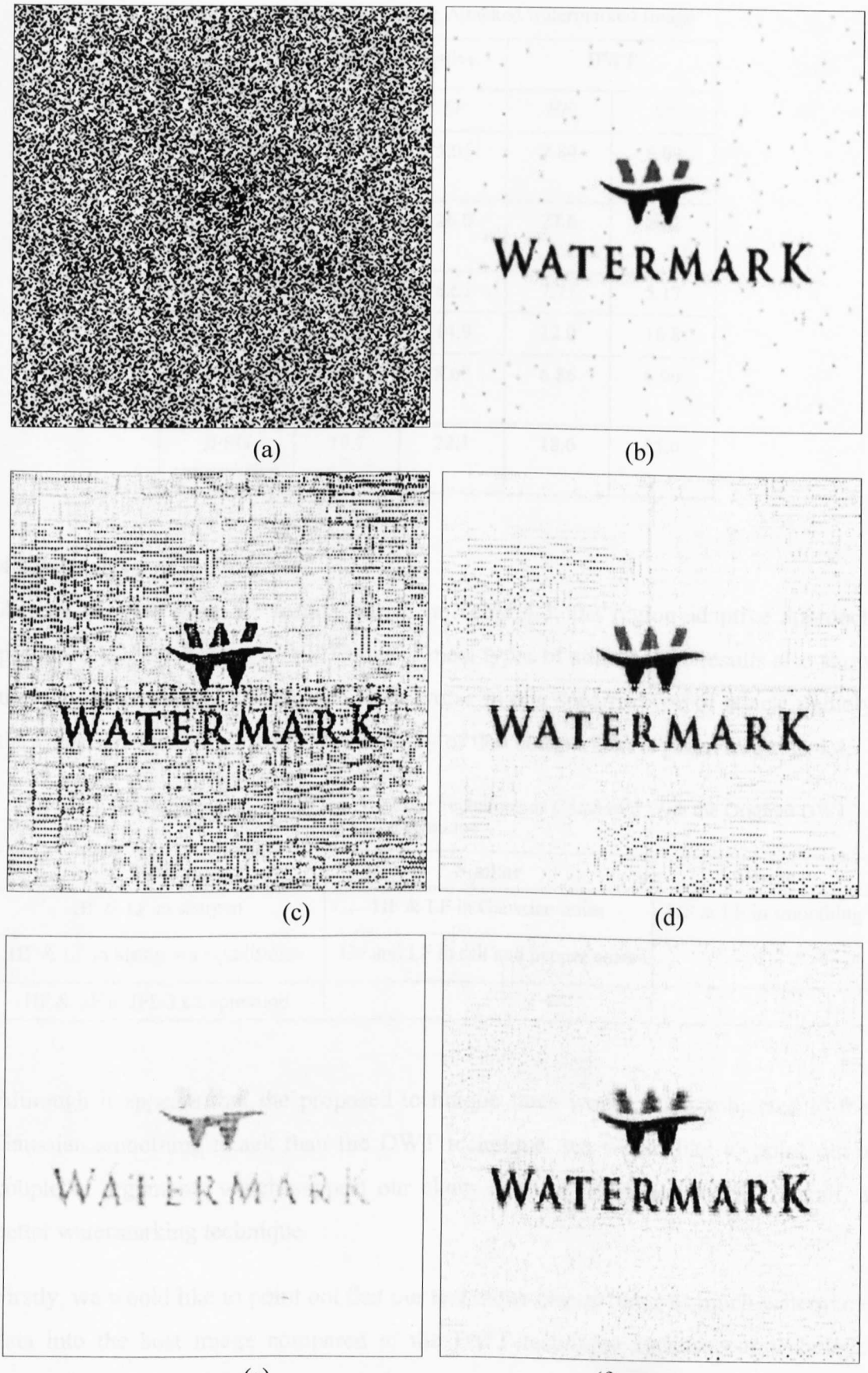


Figure 4.14 Extracted LF watermark image (a) Gaussian noise with $\sigma = 0.005$ (b) Salt and pepper noise with $\sigma = 0.005$ (c) Sharpen with $\sigma = 1.0$ (d) Smoothing with $\sigma = 200$ (e) Histogram equalisation with $\eta = 200$ (f) JPEG compression with $\sigma = 90$

Table 4.3 PSNR Values of the Attacked watermarked Image

Attack Method	Region Adaptive		DWT	
	HF	*LF*	*HF*	*LF*
Gaussian noise	7.89	5.06	7.84	5.09
Salt & pepper noise	27.8	26.0	27.6	26.2
Sharpen	8.27	6.61	7.71	5.17
Smoothing	14.9	14.9	22.0	18.8
Histogram equalisation	6.91	8.68	6.86	6.90
JPEG compression	19.7	22.1	18.6	15.6

4.6.3 Analysis of the Results

As can be seen from the results shown in Table 4.3, the region-adaptive approach produced higher or equal PSNR values in most types of attacks. The results also show that the region-adaptive technique fared worse in one specific type of attack, namely the Gaussian smoothing attack. A summary of this comparison is given in Table 4.4.

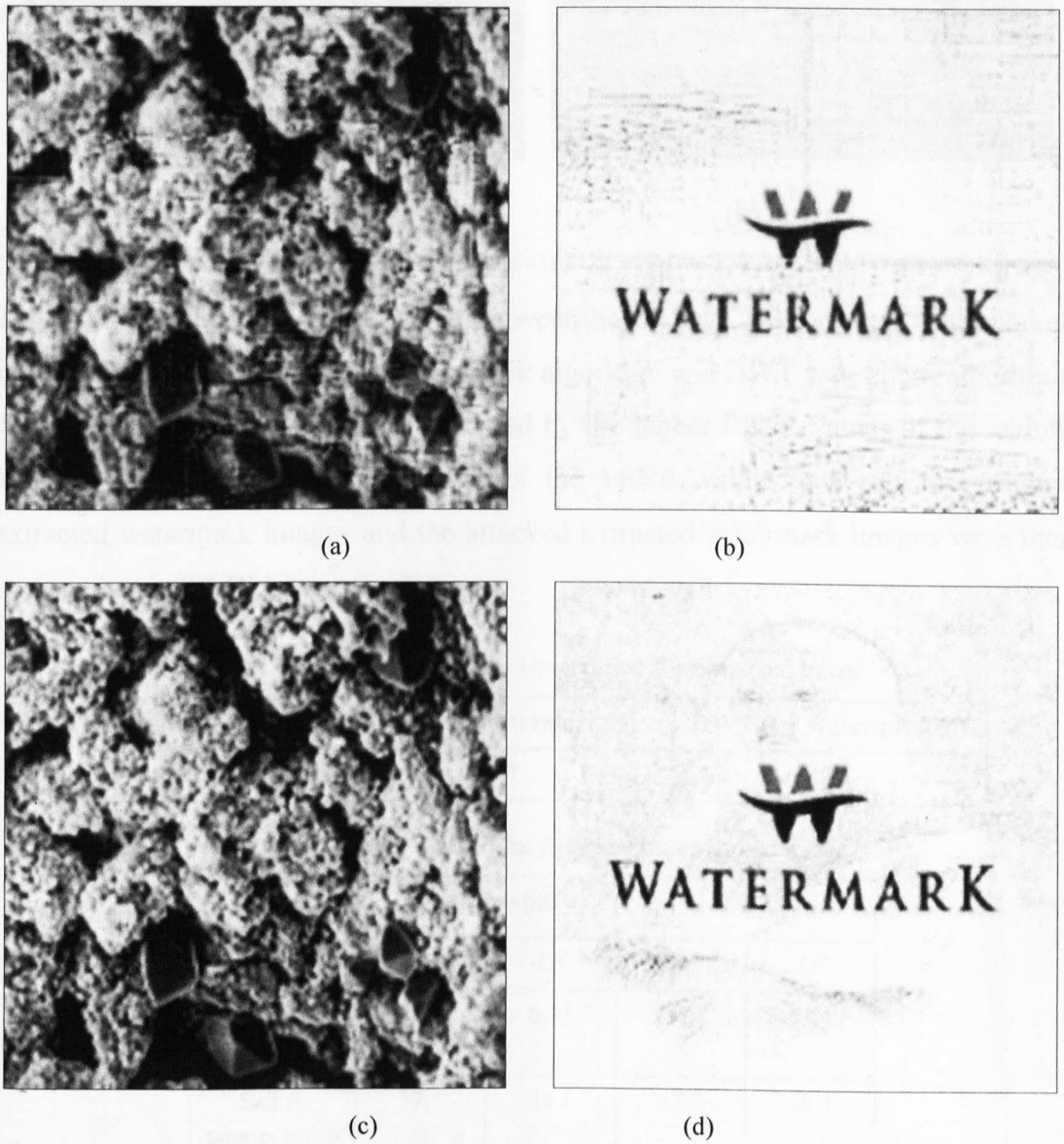
Table 4.4 The Robustness Level of the Proposed Technique as Compared With the Original DWT Technique

Better	Similar	Worse
HF & LF in sharpen	HF & LF in Gaussian noise	HF & LF in smoothing
HF & LF in histogram equalisation	HF and LF in salt and pepper noise	
HF & LF in JPEG compression		

Although it appears that the proposed technique fares worse when subjected to the Gaussian smoothing attack than the DWT technique, we would like to point out a couple of arguments which support our claim that our technique is still, overall, a better watermarking technique.

Firstly, we would like to point out that our technique inserts twice as much watermark data into the host image compared to the DWT technique because our technique inserts both HF and LF watermark images at the same time, whereas the DWT technique stores one image at a time.

Furthermore, we can show that the extracted watermark images stored using our technique was less distorted than those using the DWT technique. An example of our claim can be seen in Figure 4.15, which shows that the smoothing attack imprints some significant artefacts on the extracted watermark image when it is inserted using the DWT technique. This can be seen clearly as the face contour in the extracted LF image.

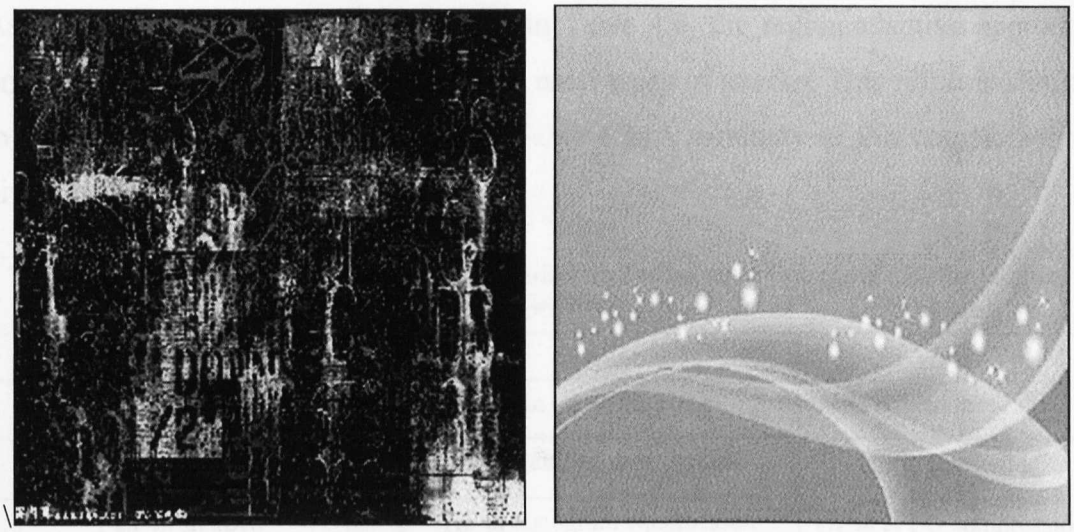


(a) (b)

(c) (d)

Figure 4.15 Extracted HF & LF watermark image after smoothing attack in (a)(b) region-adaptive algorithm and (c)(d) DWT

To further test the hypothesis of region-adaptive watermark algorithm, we apply another pair of watermark images which is shown in Figure 4.16 and embed them into 50 different host images.



(a) (b)

Figure 4.16the alternative of watermark images (a) HF watermark image (b) LF watermark image

Table 4.5 indicates the PSNR value between host image and original watermarked image by using region-adaptive watermark algorithm and DWT watermark algorithm. This result can be quantitatively measured by the higher PSNR values of the region-adaptive technique. Table 4.6 indicates the PSNR values between the original extracted watermark images and the attacked extracted watermark images were then averaged over the 50 images

Table 4.5 PSNR values of the Unmodified Watermarked Image

Region adaptive	**DWT (HF watermarked)**	**DWT (LF watermarked)**
32.4	30.1	28.7

Table 4.6 PSNR Values of the Attacked watermarked Image

Attack Method	**Region adaptive**		**DWT**	
	HF	*LF*	*HF*	*LF*
Gaussian noise	7.84	6.41	7.74	6.41
Salt & pepper noise	27.5	28.1	27.3	28.1
Sharpen	7.96	7.71	8.43	7.37
Smoothing	14.9	14.3	16.8	13.4
Histogram equalisation	5.03	8.14	5.20	7.59
JPEG compression	19.9	20.4	21.5	19.8

As can be seen from the results shown in Table 4.6, the region-adaptive approach produced higher or equal PSNR values in most types of attacks. This result is similar to that from the previous experiment in Table 4.3. A summary of this comparison is given in Table 4.7.

Table 4.7 The Robustness Level of the Region-Adaptive Technique as Compared with the Original DWT Technique

Better	Similar	Worse
HF in Gaussian noise	LF in Gaussian noise	HF in sharpen
HF in salt and pepper noise	LF in salt and pepper noise	HF in smoothing
LF in sharpen	HF in histogram equalization	HF in JPEG compression
LF in smoothing		
LF in histogram equalization		
LF in JPEG compression		

The result from Table 4.5, 4.6 and 4.7 shows that the region-adaptive watermarking algorithm produces better results than original DWT algorithm by applying another pair of watermark images. This argument proof a further point that the region-adaptive watermark algorithm could resist on different watermark attacks. This result is also similar to the conclusion from the previous experiment.

4.7. Discussion

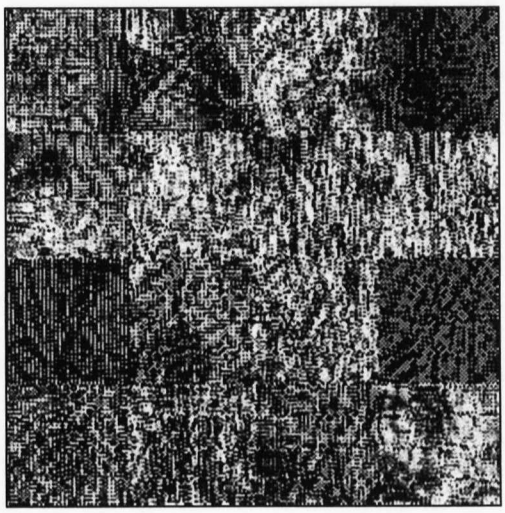
4.7.1 Advantage

As we described in the previous section of this chapter, we found the region-adaptive watermarking system improved the robustness of watermarking system while comparing with the original DWT-based watermarking [Tao, 2004]. First of all, we showed in the first experiment that the technique can extract the watermark image successfully. Using the second experiment we also showed that the region-adaptive watermarking system is better than the original DWT-based algorithm in most of attacks. The two techniques have similar robustness levels against Gaussian noise attack and salt and pepper noise attack. We also showed some evidence that, despite its lower PSNR value when Gaussian smoothing attack is applied, the region-adaptive technique can produce a visually better extracted watermark image than the DWT technique.

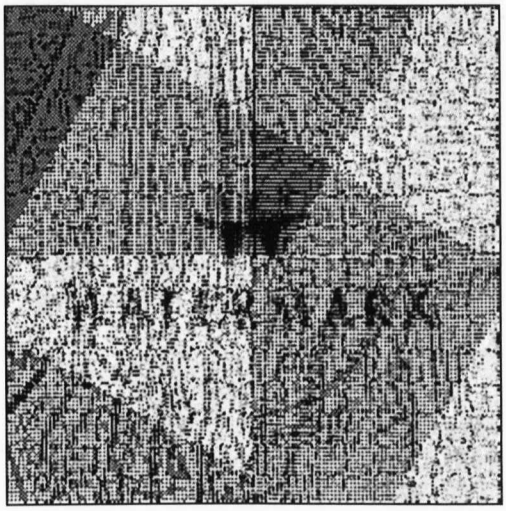
4.7.2 Disadvantages

So far we have not taken into account a geometric attack in our experiment. This is because the initial design of this technique does not yet include ways to counter these types of attacks. We did however experiment by applying these types of attacks to both region-adaptive and DWT techniques. The results of this experiment are shown in Figures 4.17 and 4.18.

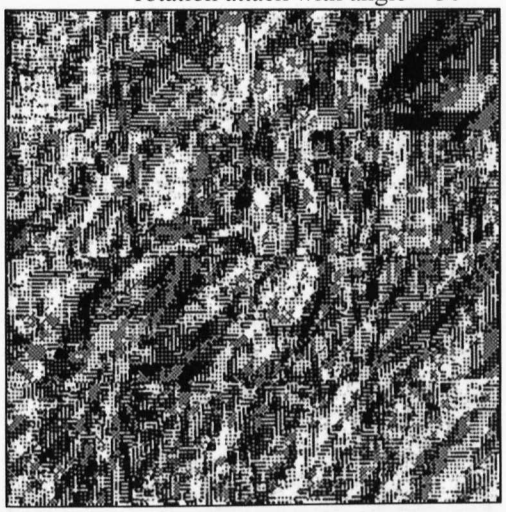
As can be seen from the results, both techniques fare badly against the geometric attacks. Figure 4.17 shows extracted HF & LF watermark images after a geometric attack by the region-adaptive watermarking algorithm, and Figure 4.18 shows extracted HF & LF watermark images after a geometric attack by the original DWT watermarking algorithm.



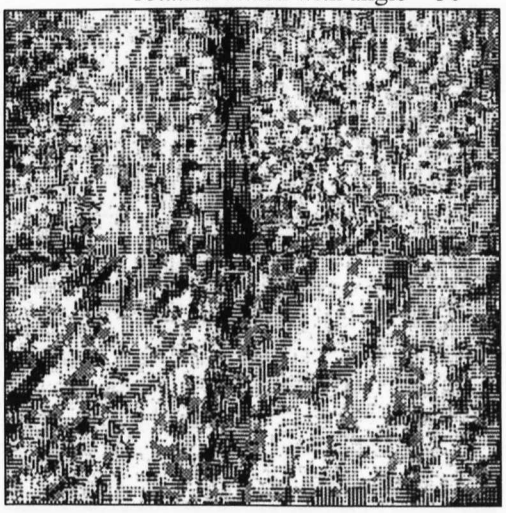
(a) Extracted HF watermark image after a rotation attack with angle = 30



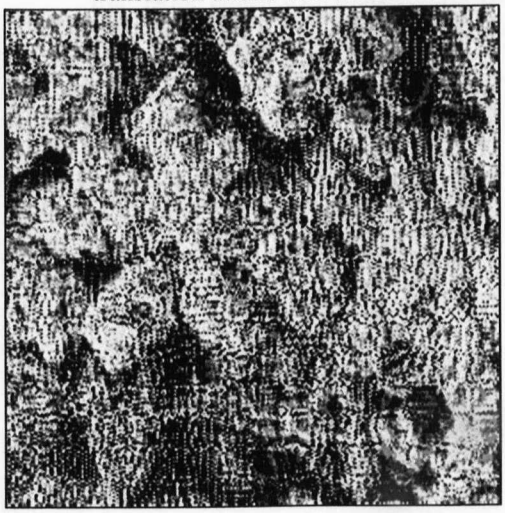
(b) Extracted LF watermark image after a rotation attack with angle = 30



(c) Extracted HF watermark image after a translation attack with translation = 10



(d) Extracted LF watermark image after a translation attack with translation = 10



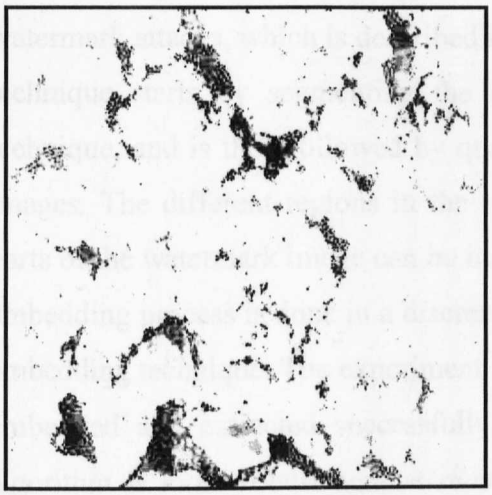
(e) Extracted HF watermark image after a scale attack

(f) Extracted LF watermark image after a scale attack

Figure 4.17 Extracted HF & LF watermark image in the region-adaptive watermarking algorithm after a geometric attack

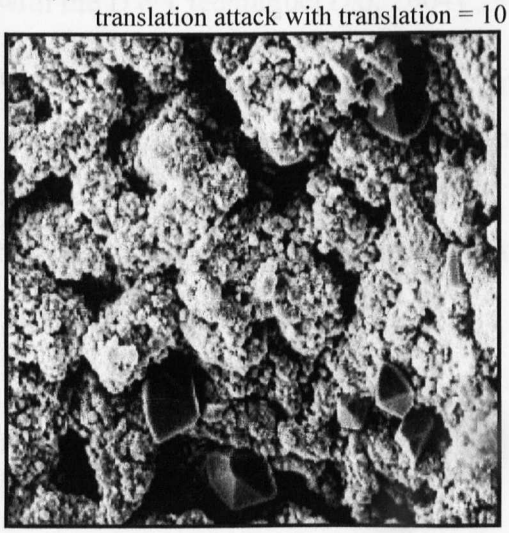
(a) Extracted HF watermark image after a rotation attack with angle = 30

(b) Extracted LF watermark image after a rotation attack with angle = 30

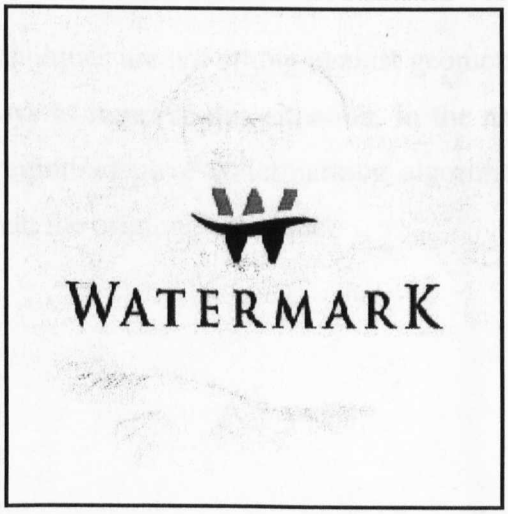


(c) Extracted HF Watermark image after a translation attack with translation = 10

(d) Extracted LF watermark image after a translation attack with translation = 10



(e) Extracted HF watermark image after a scale attack



(f) Extracted LF watermark image after a scale attack

Figure 4.18 Extracted HF & LF watermark images in the region-adaptive watermarking algorithm after a geometric attack

As we can see from Figure 4.17 and 4.18, we found that the proposed region-adaptive watermarking system is not good enough to withstand geometric attack, especially rotation and translation attacks, as both of the extracted HF & LF watermark images are distorted totally. Therefore, in the next chapter we will describe an improvement to the region-adaptive watermarking algorithm using a combination of the discrete wavelet transform and the singular value decomposition (DWT-SVD). It will be applied as a novel embedding process instead of DWT alone [Tao, 2004].

4.8. Summary

In this chapter, we have covered the design process behind the region-adaptive watermarking scheme. The approach is the result of previous work analysing watermark attacks, which is described in Chapter 3. The embedding process using this technique starts by segmenting the host image using the Markov random field technique, and is then followed by quad-tree partitioning of the host and watermark images. The different regions in the two images are then matched to decide which parts of the watermark image can be inserted into a certain part of the host image. The embedding process is done in a discrete wavelet transform domain using the additive-embedding technique. The experiment shows that the HF & LF watermark images are embedded and extracted successfully, and it also indicates that the region-based algorithm is more robust against most types of watermark attacks when compared with the DWT technique[Tao, 2004] .

We have also discussed the fact that both techniques are not strong against geometric attacks, so made suggestions on possible ways to improve this situation. In the next chapter, we will describe a variant of the region-adaptive watermarking algorithms that can withstand geometric attacks better than the original technique.

CHAPTER 5

REGION-ADAPTIVE WATERMARKING SCHEME USING A DISCRETE WAVELET TRANSFORM AND A SINGULAR VALUE DECOMPOSITION

The previous chapter described a region-adaptive watermarking system using a DWT algorithm. Two different watermark images, namely the high frequency watermark image and the low frequency watermark image, are applied to the region-adaptive watermarking scheme, which attempts to embed different watermark images into high low frequency regions. The experiment indicates that the watermark system is robust against most watermark attacks. However, there is one significant disadvantage of the region-adaptive watermarking system, which is that it cannot resist geometric attack, e.g. rotation, translation and scale. Therefore, in this chapter, we propose an improved robust watermarking algorithm that will be designed to withstand both removal attack and geometric attack. Section 5.2 will describe singular value decomposition and section 5.3 will show the experiment's results.

5.1. Introduction

The previous chapter described the region-adaptive DWT watermarking algorithm. This technique employs two watermark images, each with strong high frequency or

low frequency components named HF & LF watermark images, and it embeds different parts of the watermark images into the host image based on the differences between their spectral distributions. Furthermore, the experiment proves that the region-adaptive DWT watermarking algorithm can resist removal attack. However, the experiment and discussion in the previous chapter also argue that it is not impregnable to geometric attack.

Therefore, in this chapter, an improved region-adaptive watermarking system will be introduced. Compared with previous watermarking algorithms, our improved system employs a combination of discrete wavelet transform and singular value decomposition (SVD), which could be robust against both removal attack and geometric attack.

Recently, geometric attacks have been some of the most popular research projects in the field of digital watermarking. Unlike removal attacks, geometric attacks are effective in that they can destroy the synchronisation in a watermarked image. Licks [Licks, 2005] surveyed a number of image watermarking techniques designed to be robust against geometric attack, which are described in Chapter 3. One of these techniques is called singular value decomposition (SVD).

Recently, an SVD-based algorithm was explored for watermarking. The SVD for square matrices was discovered independently by Beltrami in 1873 and Jordan in 1874, and then extended to rectangular matrices by Eckart and Young in the 1930s [Emir, 2004]. It was not used as a computational tool until the 1960s because of the need for sophisticated numerical techniques. In later years, Gene Golub demonstrated its usefulness and feasibility as a tool in a variety of applications [D.Lahaner, 1989]. SVD is one of the most useful tools of linear algebra, with several applications in image compression [H.C.Andrews, 1976, N.Garguir, 1979, S.O.Aase, 1999], watermarking [Ayangar, 2010 , Bhatnagar, 2008] and other signal processing fields [R.Karkarala, 2001].

A discrete wavelet transform is one of the most rapidly developing algorithms in the field of watermarking techniques. One of the characteristics of DWT is that is can split into four sub-bands in the host image where a watermark can be embedded effectively. In general, in a DWT-based watermarking algorithm, most of the image energy is concentrated at the lower frequency sub-bands [Chin-Chen, , Sudirman,

2009, Tao, 2004]; however embedding watermarks in low frequency sub-bands may degrade image quality significantly. On the other hand, embedding in low frequency sub-bands could increase robustness significantly. Furthermore, the high frequency sub-band includes the edges and textures of the image, and the human eye is not generally sensitive to changes in such a sub-band. The compromise adopted by many DWT-based watermarking algorithms is to embed the watermark in the middle frequency bands LH and HL, where acceptable performance of imperceptibility and robustness can be achieved [Hwei-Jen, 2010, Ying-Hua, 2007] .

A recent paper [Ayangar, 2010] on DWT-SVD-based watermarking systems argues that embedding a visual watermarking in both the HL and LH sub-bands results in a strong scheme that can resist different kinds of attacks. Experimental results of the proposed techniques have shown both significant improvements in transparency and robustness under different attacks such as Gaussian noise, JPEG compression, histogram equalisation, etc. Therefore, in this chapter, DWT-SVD as an embedding and extraction algorithm will apply. Furthermore, the two most frequently applied sub-bands are the LL and LH sub-bands.

5.2. Singular Value Decomposition

Prior to the embedding process, the wavelet coefficients of the watermark and host images are further decomposed using the SVD technique in a set of uncorrelated coefficients that better expose the various relationships among the original data. The singular value decomposition of any real matrix A can be explained as follows.

Let us denote A as a product of three matrices, $A = U\acute{O}V^T$, where U and V are orthogonal matrices, $U^TU = I$, $V^TV = I$ and $\acute{O}$ = diag ($\ddot{e}_1$, $\ddot{e}_2$, ...). The diagonal entries of $\acute{O}$ are called the singular values of A, the columns of U are called the left singular vectors of A and the columns of V are called the right singular vectors of A. In addition, according to equations *(5.1) and (5.2)*, we would like to say the columns of U are called the eigenvectors of AA^T, and the columns of V are called the eigenvectors of A^TA.

$$AA^T = U\acute{O}V^TV\acute{O}^TU^T = U\acute{O}\acute{O}^TU^T \qquad (5.1)$$

$$A^TA = V\acute{O}^TU^TU\acute{O}V^T = V\acute{O}\acute{O}^TV^T \qquad (5.2)$$

This decomposition of A can then be further expressed as:

$$A = ë_1 U_1 V_1^T + ë_2 U_2 V_2^T + \ldots + ë_r U_r V_r^T, \qquad (5.3)$$

where r is the rank of matrix A. It is important to note that each singular value specifies the luminance of an image layer, while the corresponding pair of singular vectors specifies the layer's geometry.

In SVD-based watermarking, several approaches are possible. A common approach is to apply the SVD to the entire host image, and then modify all singular values by embedding them with the watermark data. An important property of SVD-based watermarking is that the largest of the modified singular values change very little for most types of attacks. A theoretical analysis of the effects of ordinary geometric distortions on the singular values of an image is provided in [Chih-Chin Lai, 2010, E.Ganic, 2004, Zhou, 2004]. In these papers, five geometric attacks were analysed – transpose attack, flip attack, rotation attack, scaling attack and translation attack. The findings are as follows. Let us denote a real matrix A as the original watermarked image and real matrix A' as the watermarked image after a geometric attack T, such that $A' = T(A)$. It is shown that both A and A' have identical non-zero singular values when T belongs to any one of the above geometric attack types. This finding indicates that SVD-based watermark algorithm is impregnable to these geometric attacks.

Using SVD in a watermarking algorithm to improve its robustness against different geometric attacks has some further advantages. First, SVD transformation can be applied to an image with arbitrary sizes, i.e. not restricted to square images. Secondly, singular values possess intrinsic algebraic image properties, in other words image can be represented as a matric view, which can activate a variety of algebraic transformations or the decomposition of the matrix.

5.3. Region-Adaptive Watermarking with DWT-SVD

Region-adaptive watermarking DWT-SVD in this chapter is an ascendant of region-adaptive DWT in the previous chapter, in order to further improve robustness against different watermark attacks, especially a geometric attack. The main structure is similar to that described in section 4.4, although there are still some significant changes, which we will describe below.

5.3.1 Embedding and Extraction Algorithm

Previously, the region-adaptive technique embedded the watermark in DWT coefficients of the host image. The technique described in this chapter instead embeds the watermark after applying SVD to the DWT coefficients. In addition, in the region-adaptive technique described in the previous chapter, the DWT embedding and extraction processes are applied to the all sub-band. However, in this chapter, we investigate the difference between using LL and LH sub-bands, to compare against their robustness and to find out the most suitable sub-band which could apply to the embedding process.

The steps of the embedding processes are listed below:

1. Using DWT, decompose the HF & LF ROI of the host image into four sub-bands: LL, HL, LH and HH.

2. Apply SVD to LL/LH sub-band of HF&LF ROI of the host image: $I^H = U^H * S^H * V^H$.

3. Apply SVD to HF and LF blocks of the visual watermark image W; $W = U^W * S^W * V^W$.

4. Modify the singular values of the HF&LF of the host image in the LL/LH sub-band with the singular values of the HF&LF of the visual watermark: $S^{WM} = S^H + \alpha S^W$

5. Obtain the modified LL/LH image: $I^{WM} = U^H * S^{WM} * V^{HT}$.

   Apply the inverse DWT using a modified DWT coefficient to produce the watermarked image.

And the following list describes the extraction:

1. Using DWT, decompose the watermarked image I^{WM} into four sub-bands: LL, HL, LH and HH.

2. Apply SVD to the LL/LH sub-band image: $I^{WM} = U^{WM} * S^{WM} * V^{WM}$

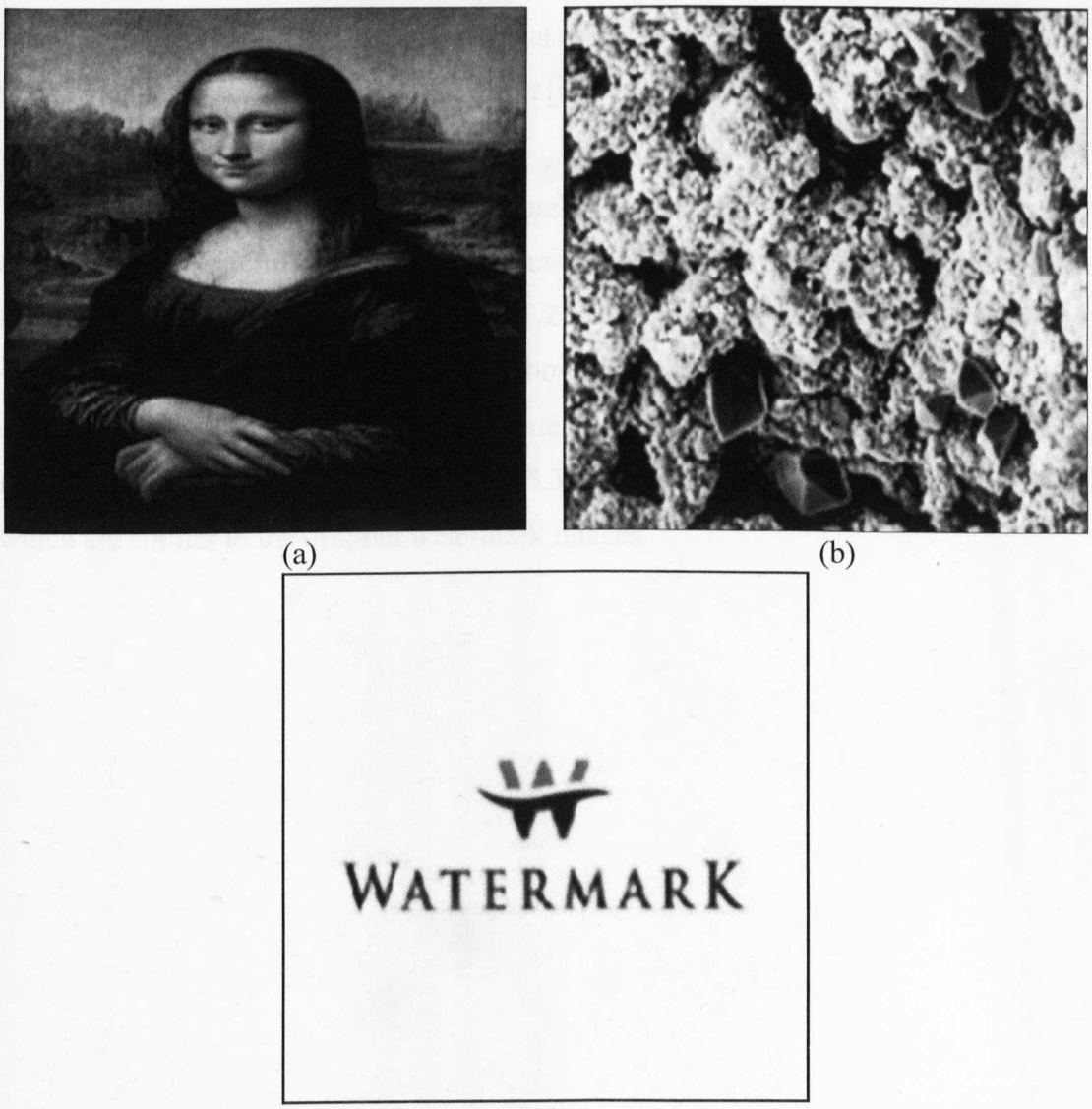
3. Apply SVD to HF & LF ROI of the watermark image W: $W = U^W * S^W * V^W$

4. Extract the singular values of watermarking: $S^{W'} = (S^{WM} - S^{H})/\alpha$ where S^{H} are the singulars of original HF & LF ROI of the host image

5. Recover the watermarking $W' = U^{W} * S^{W'} * V^{W}$.

5.4. Experimental Results

We applied the region-adaptive watermark technique described in the previous section to the images shown in Figure 5.1. The original image has dimensions of 1024x1024 pixels, and both HF and LF watermark images have dimensions of 256x256 pixels. The HF watermark image is divided into 64 blocks of 32x32 pixels and the LF watermark image is divided into 16 blocks of 64x64 pixels.



(a) (b)

(c)

Figure 5.1 (a) The host image, (b) the HF watermark image and (c) the LF watermark image

As a quantitative measure of the degradation effect caused by the attacks we use the peak-signal-to-noise ratio (PSNR). The formulation between the original and the attacked watermarked signals is described in Chapter 2. High PSNR values indicate lower degradation, hence demonstrating that the watermarking technique is more robust to that type of attack.

5.4.1 Comparison between an LL Sub-band and an LH Sub-band

In this experiment, we apply two sub-bands including the low frequency LL sub-band and the mid frequency LH sub-band. Two experiments were conducted to test the algorithm. The first experiment aimed at verifying that inserted watermarks images can be extracted with minimal distortion when there are no attacks applied. The second experiment measured the robustness of the region-adaptive technique in withstanding attacks, including geometrical attacks, and compared it with the original non-region-adaptive DWT-SVD algorithm [Emir, 2004].

A. Watermark insertion and extraction verification

The first experiment shows that watermark images can be inserted and extracted completely by using region-adaptive technique. The embedding strength is 0.013 in both the LL and LH sub-bands. Figure 5.2 shows some intermediate results for the watermark insertion process. The figure shows the segmented host image, the parts of the host image where the different watermarks are inserted and the original resulting LL and LH watermarked image. Figure 5.3 shows the extracted watermark images, which are similar to the original watermark images.

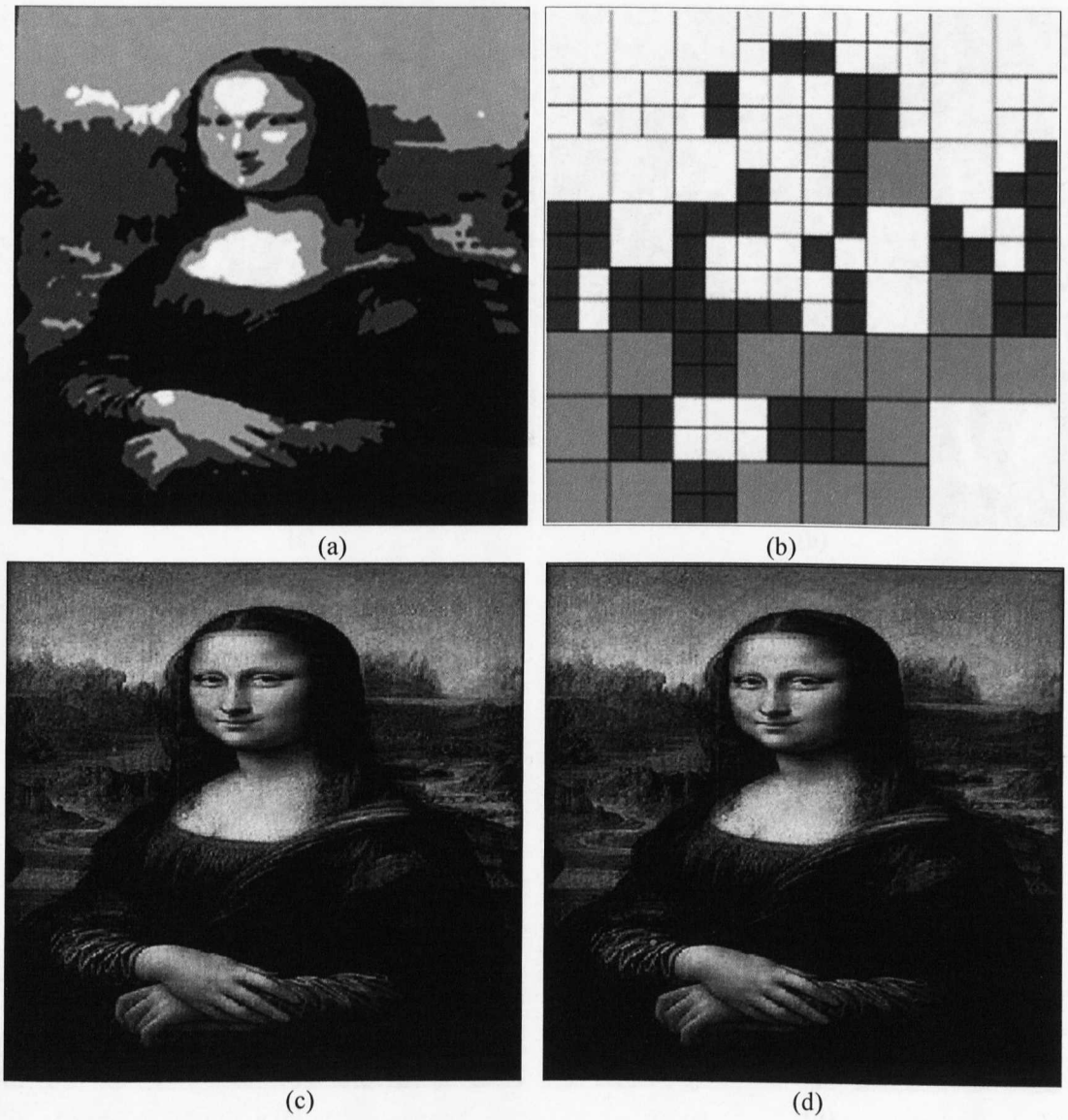
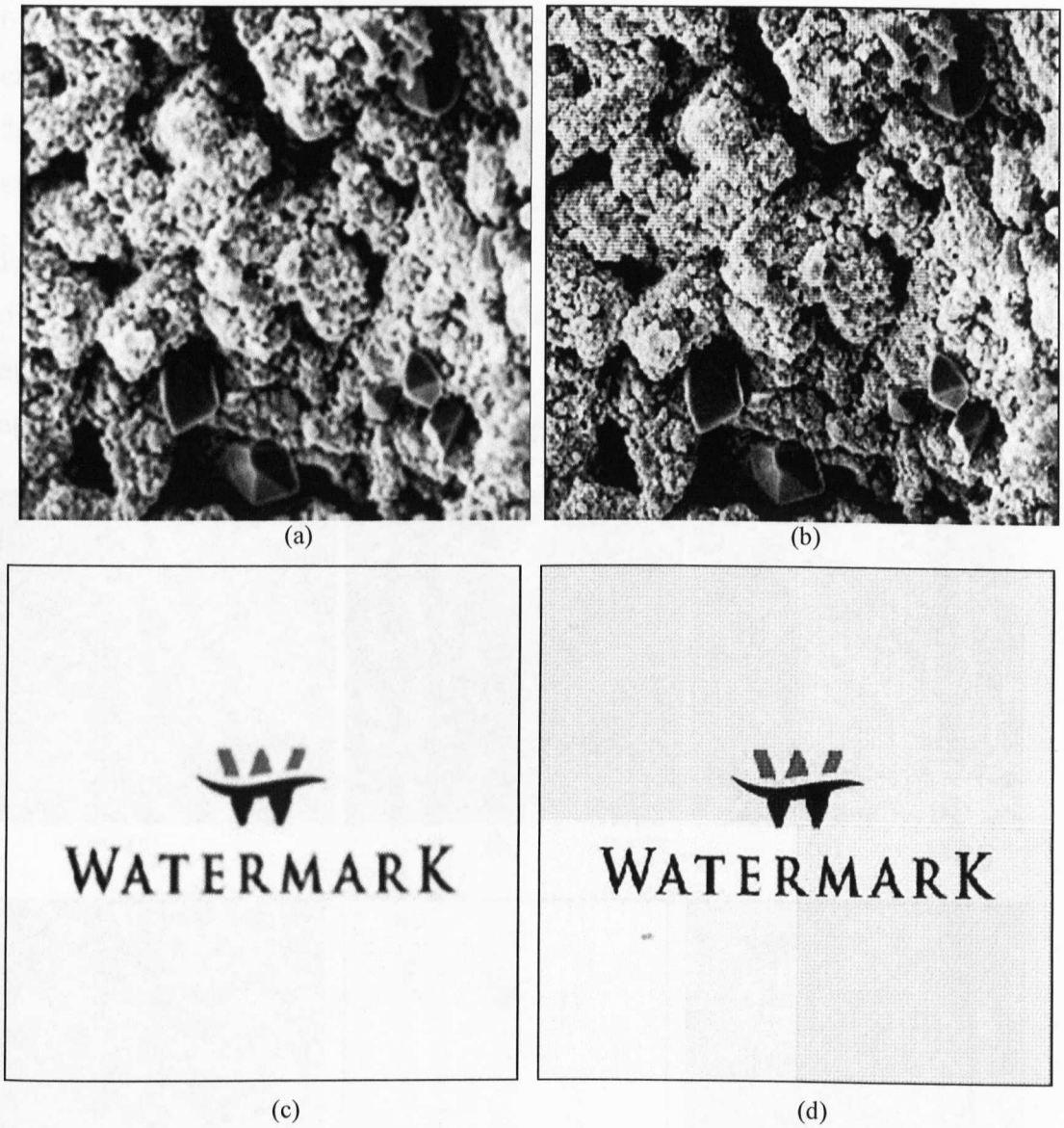


Figure 5.2 MRF segementated host image, (b) watermark insertion region (dark grey – HF and light grey – LF) and (c) original LL watermarked image (d) original LH watermarked image



(a) (b)

(c) (d)

Figure 5.3 Extracted (a) HF watermark of the LL sub-band, (b) HF watermark of the LH sub-band (c) LF watermark of the LL sub-band and (d) LF watermark of the LH sub-band

In conclusion, based on our experiment, we found that two watermark images embedded into the host image successfully, in both the LL and LH sub-band; furthermore, the results also show that the HF & LF watermark images are extracted completely.

B. Robustness Comparison

To test the robustness of the region-adaptive watermarking scheme, nine watermark attacks are applied to the watermarked image. They are Gaussian noise, salt and pepper noise, sharpen, smoothing, histogram equalisation, JPEG compression, rotation, translation and scaling attacks. The severity of these attacks can be adjusted

by modifying their corresponding parameter values. Figures 5.4 and 5.5 shows the extracted watermark images after different attacks in the LL sub-band, while Figures 5.6 and 5.7 show the extracted watermark images after different attacks in the LH sub-band.

From Figure 5.4 to 5.7, it is evident that all watermark images are extracted after different attacks; we can also clearly distinguish both HF and LF watermark images after different attacks, which proves the proposed watermark algorithm is robust against various watermark attacks, including removal and geometric attack.

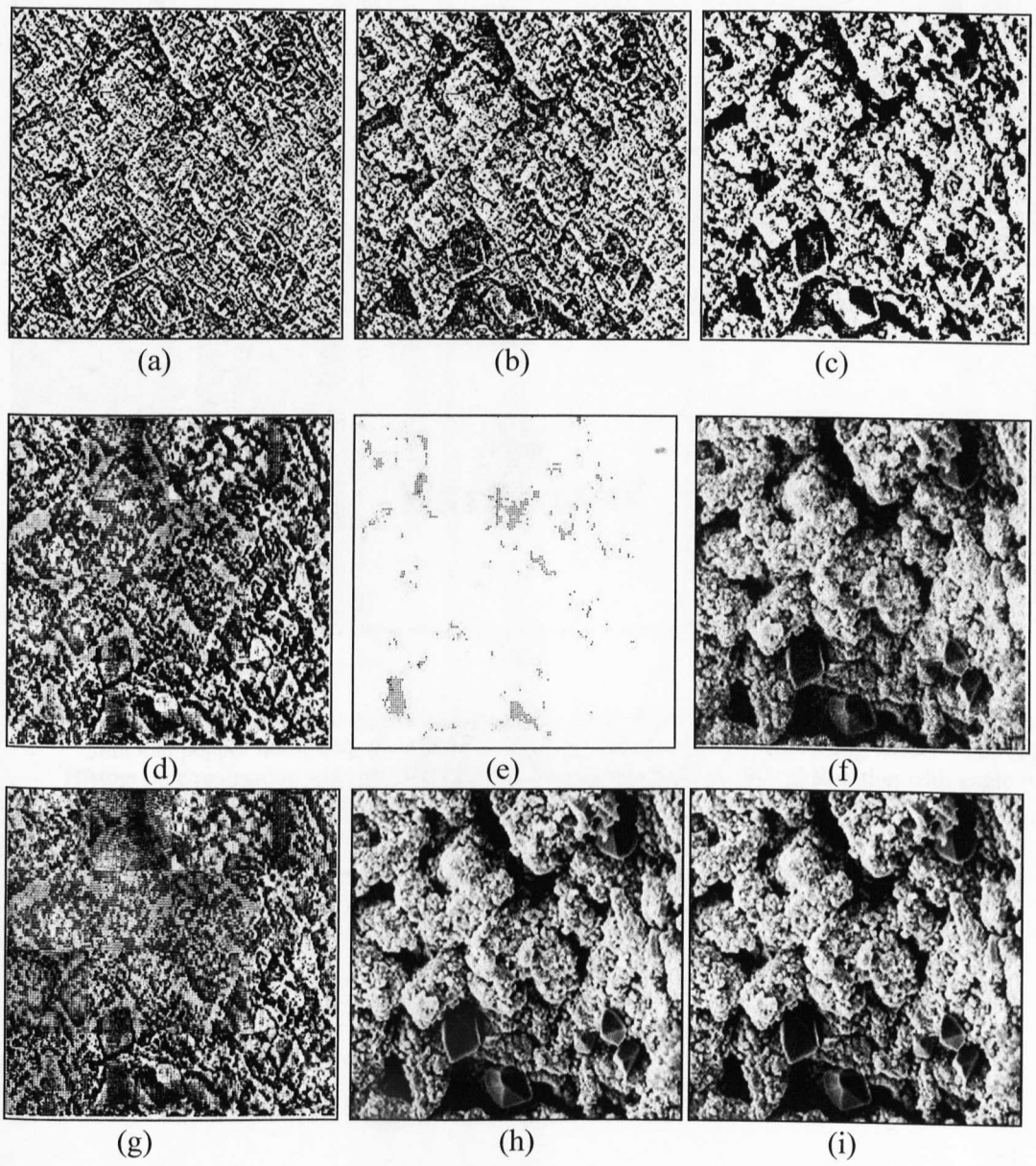


Figure 5.4Extracted HF watermark image of the LL sub-band (a) Gaussian noise with $\sigma = 0.005$ (b) Salt and pepper noise with $\sigma = 0.005$ (c) Sharpen with $\sigma = 1.0$ (d) Smoothing with $\sigma = 200$ (e) histogram equalisation with $\eta = 10$ (f) JPEG compression with $\sigma = 90$ (g) Rotation with angle = 30 (h) Translation with movement = 10 (i) Scale

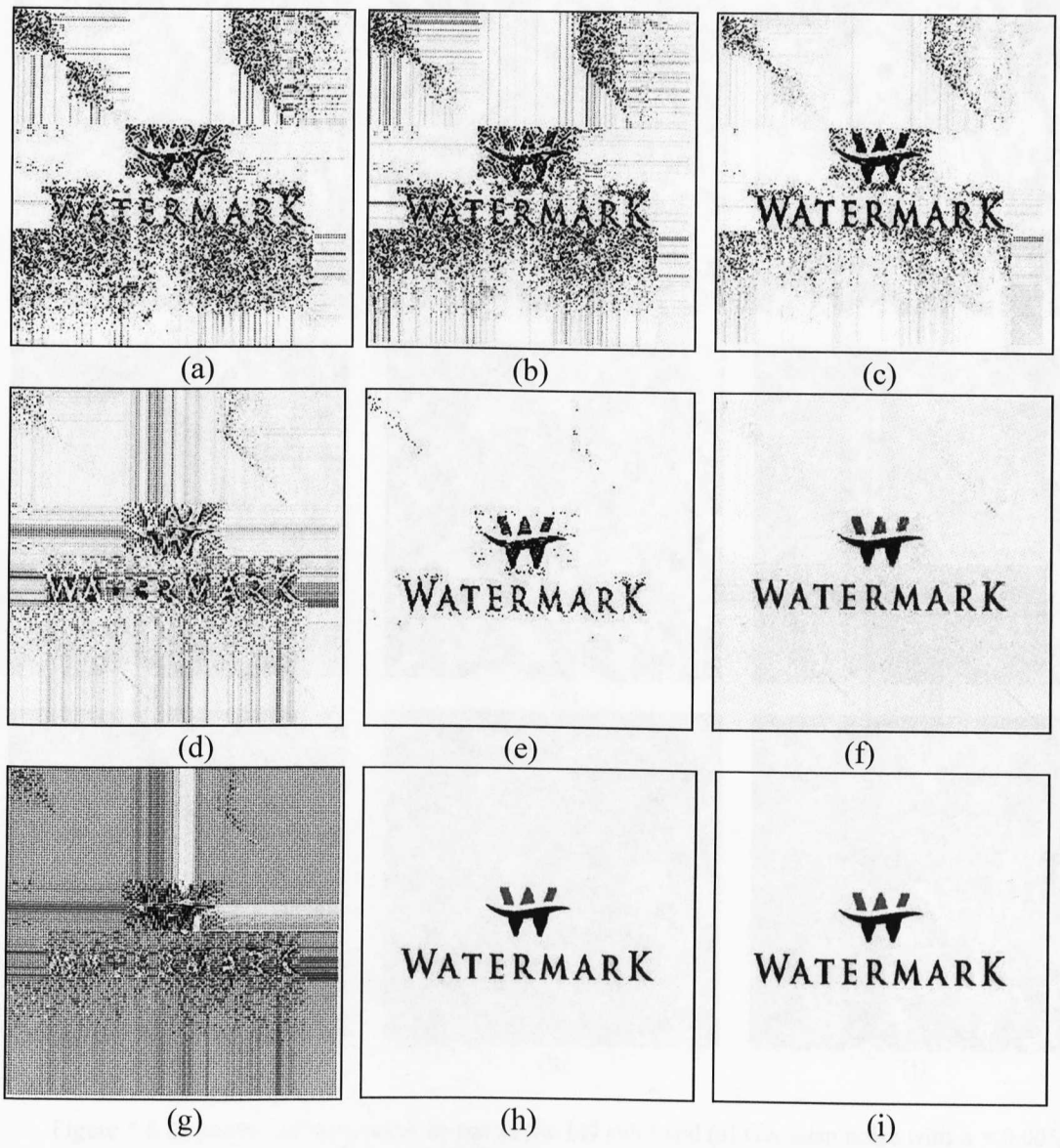


Figure 5.5 Extracted LF watermark image of the LL sub-band (a) Gaussian noise with $\sigma = 0.005$ (b) Salt and pepper noise with $\sigma = 0.005$ (c) Sharpen with $\sigma = 1.0$ (d) Smoothing with $\sigma = 200$ (e) Histogram equalisation with $\eta = 10$ (f) JPEG compression with $\sigma = 90$ (g) Rotation with angle = 30 (h) Translation with movement = 10 (i) Scale

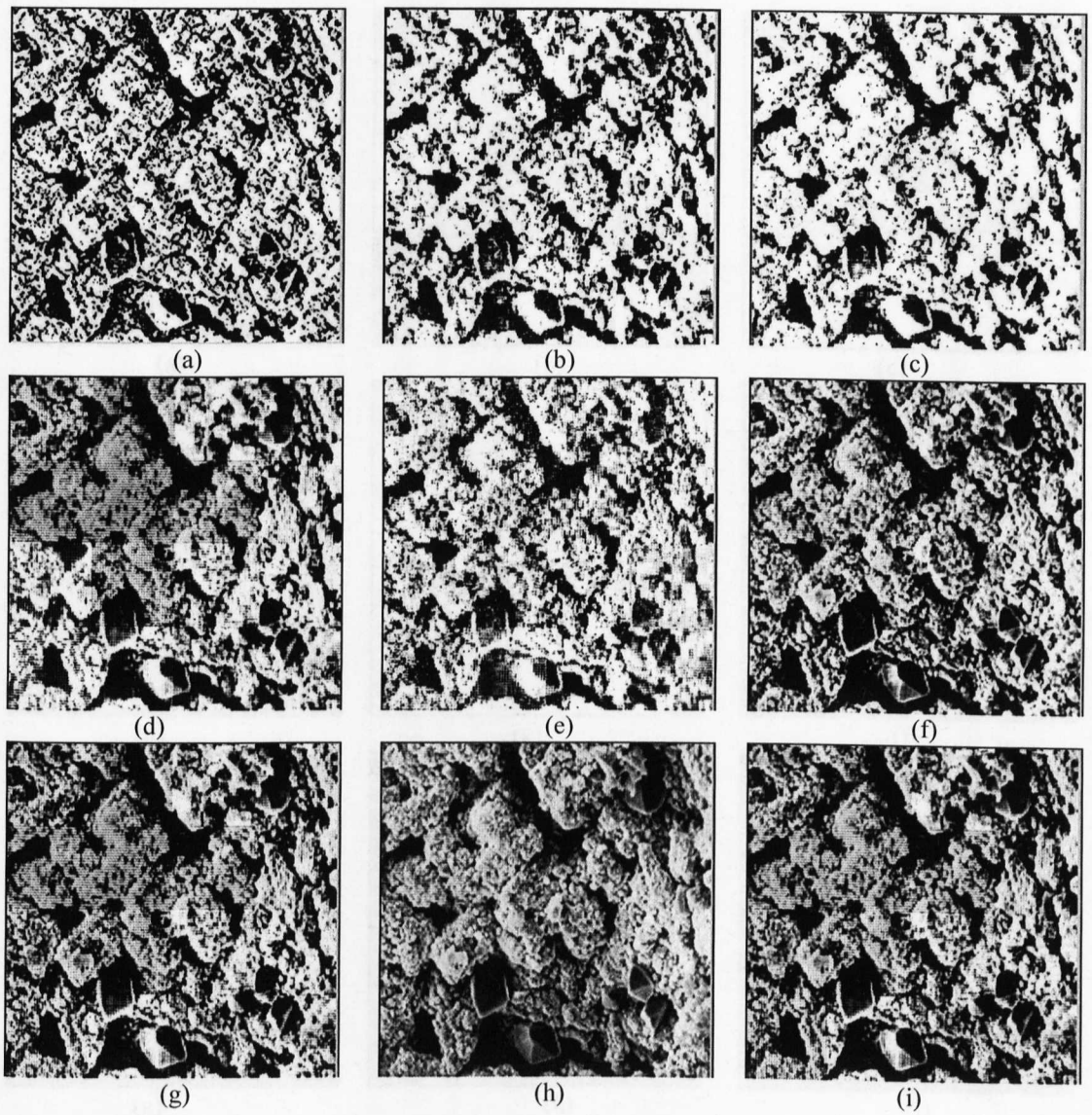


Figure 5.6 Extracted HF watermark image of the LH sub-band (a) Gaussian noise with $\sigma = 0.005$ (b) Salt and pepper noise with $\sigma = 0.005$ (c) Sharpen with $\sigma = 1.0$ (d) Smoothing with $\sigma = 200$ (e) Histogram equalisation with $\eta = 10$ (f) JPEG compression with $\sigma = 90$ (g) Rotation with angle = 30 (h) Translation with movement = 10 (i) Scale

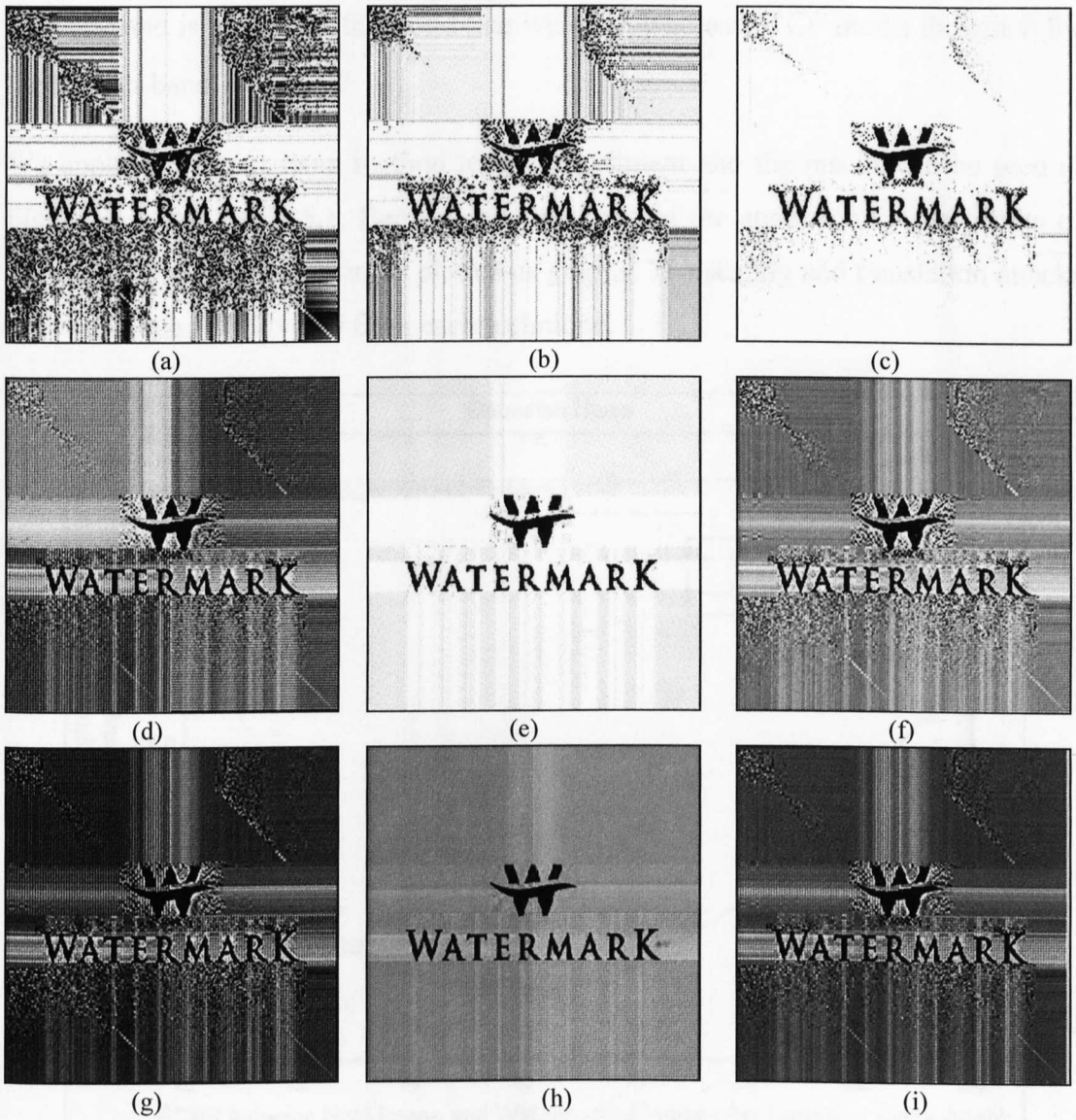


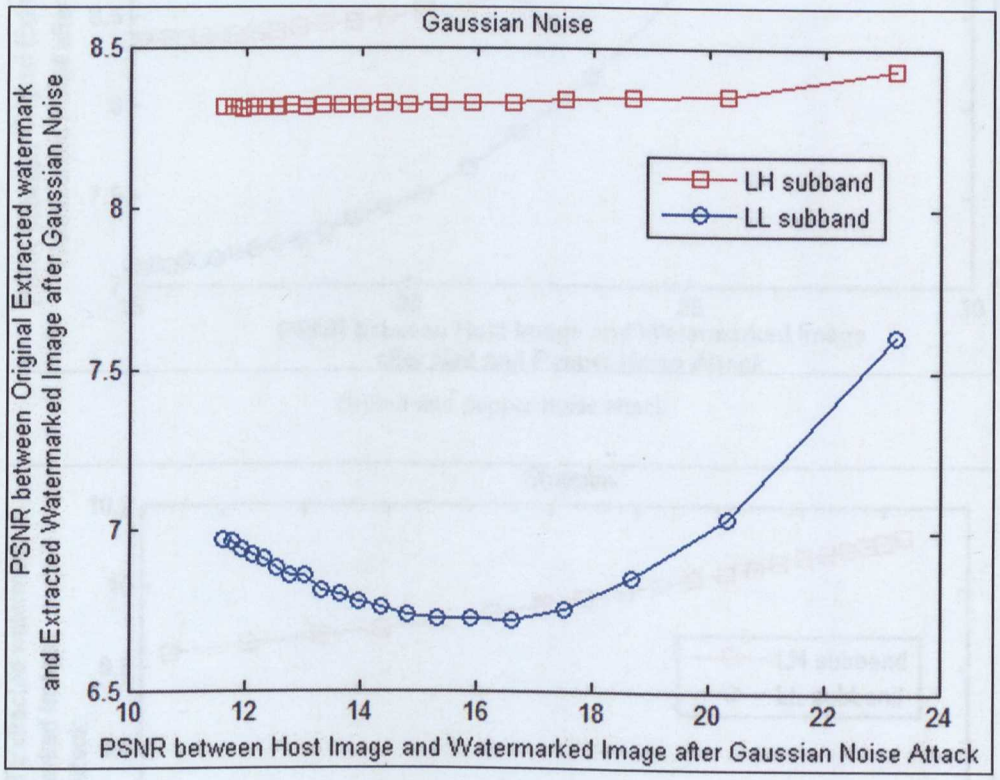
Figure 5.7 Extracted LF watermark image of the LH sub-band (a) Gaussian noise with $\sigma = 0.005$ (b) Salt and pepper noise with $\sigma = 0.005$ (c) Sharpen with $\sigma = 1.0$ (d) Smoothing with $\sigma = 200$ (e) Histogram equalisation with $\eta = 10$ (f) JPEG compression with $\sigma = 90$ (g) Rotation with angle = 30 (h) Translation with movement = 10 (i) Scale

To compare robustness between the LL and LH sub-bands, we conducted the experiment using 50 different host images. The parameters of the attack algorithms used in the experiment are the same as those described in Chapter 3.

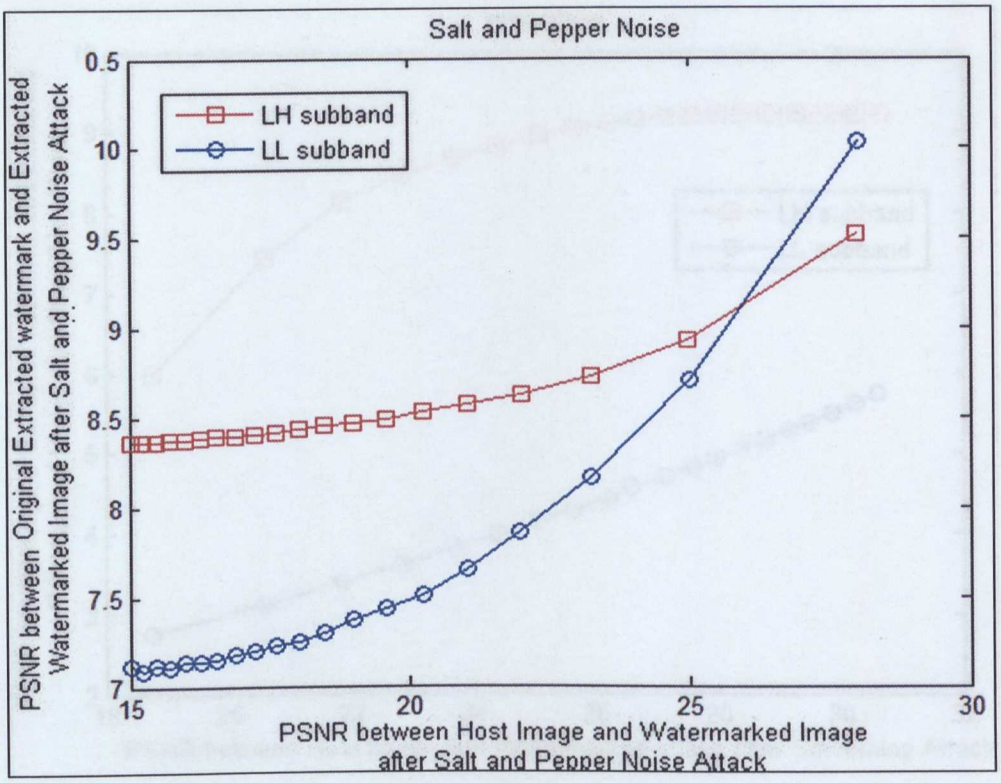
We start by plotting the 2D graphs of the PSNR values of the two sub-bands. Higher PSNR values indicate that the watermarking technique is more robust than other watermarking algorithms. The x direction is the PSNR value between the host image and the original watermarked image, and the y direction is the PSNR value between the watermark images and the extracted original watermark images. The graph for the

LH sub-band is marked with a solid line with '□', whereas '○' marks the graph for the LL sub-band.

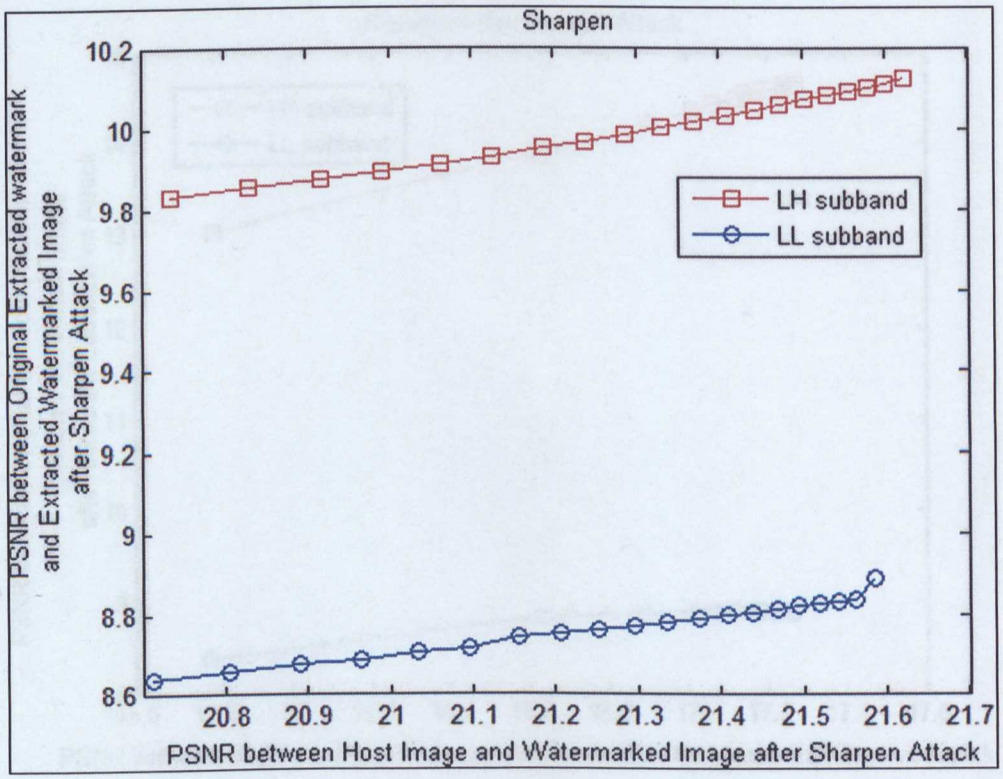
We apply this comparison method to our experiment and the results can be seen in Figures 5.8 and Table 5.1. Because of the nature of the attacks, only the results of seven out of nine attacks can be plotted as graphs. The scaling and translation attacks only yield one PSNR value from each technique.



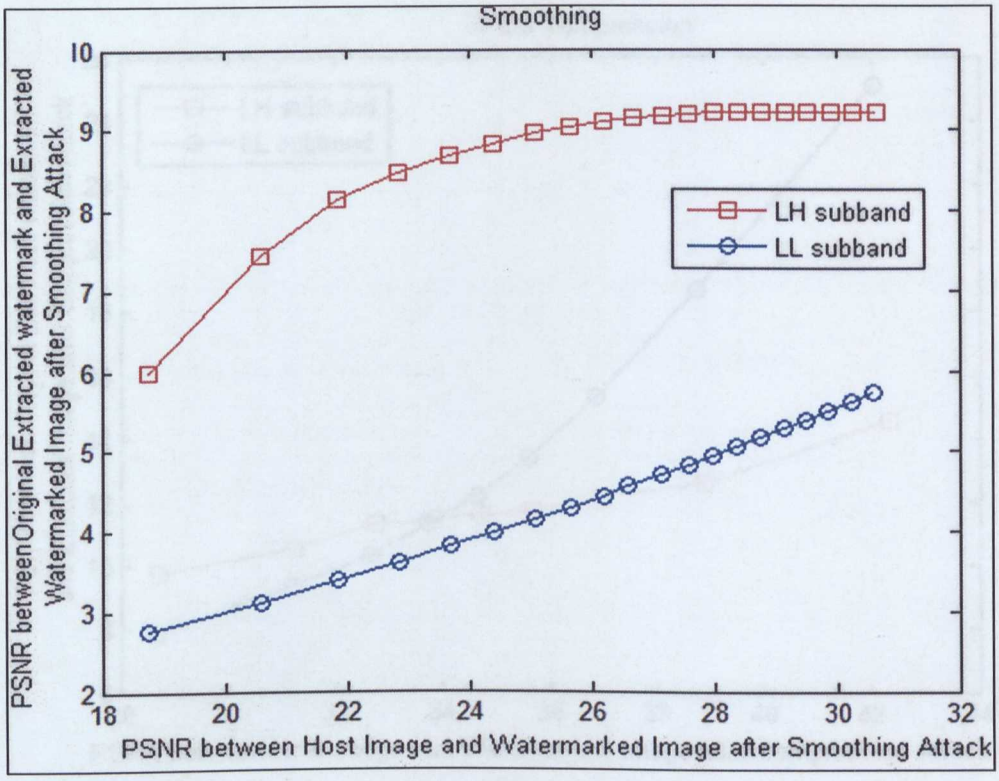
(a)Gaussian noise attack



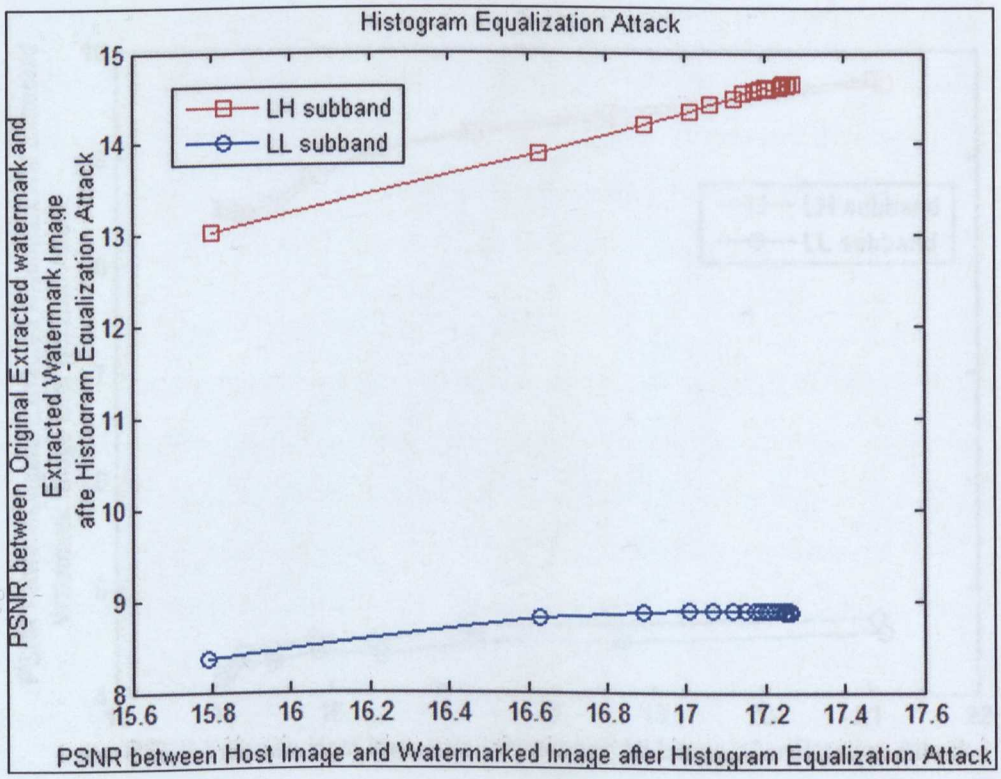
(b)Salt and pepper noise attack



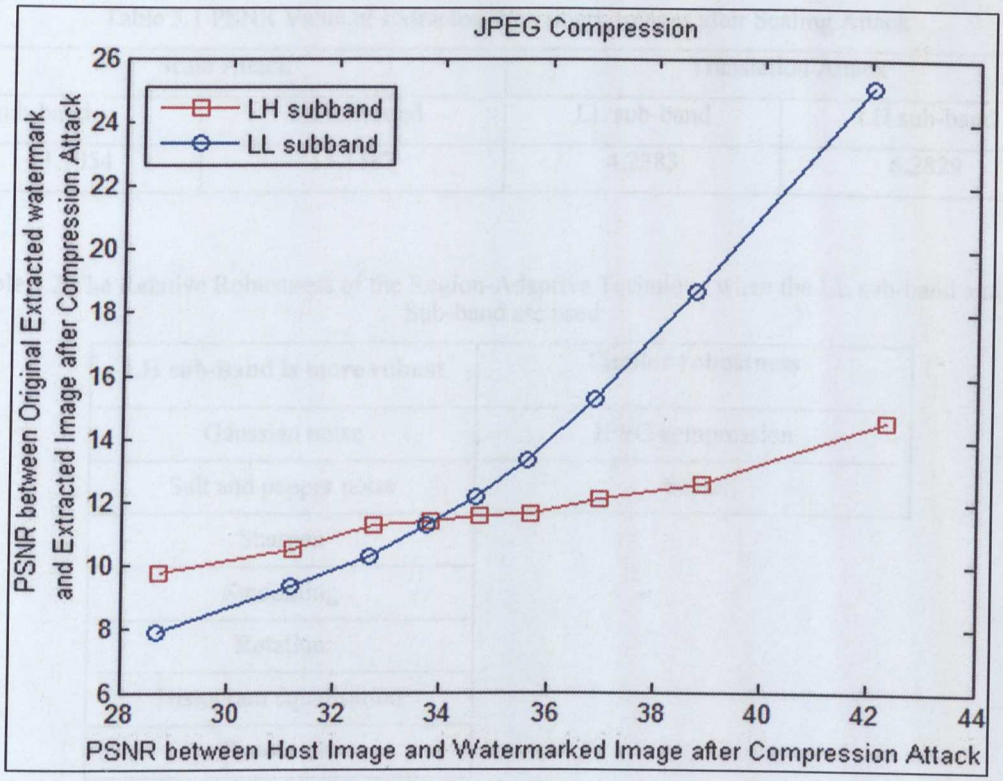
(c)Sharpen attack



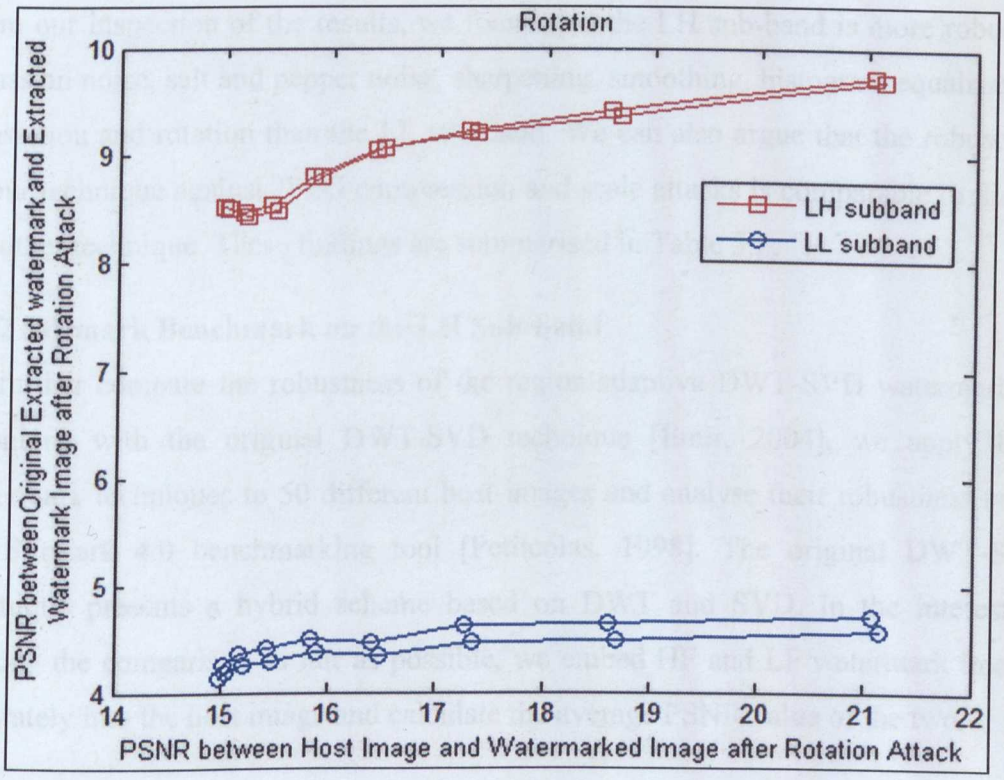
(d)Smoothing attack



(e)Histogram equalisation



(f)JPEG compression



(g)Rotation

Figure 5.8 Robustness comparisons between the LL sub-band and the LH sub-band

Table 5.1 PSNR Value of Extracted Watermark Images after Scaling Attack

Scale Attack		Translation Attack	
LL sub-band	LH sub-band	LL sub-band	LH sub-band
13.7054	13.1587	4.2383	6.2829

Table 5.2 The Relative Robustness of the Region-Adaptive Technique when the LL sub-band and LH Sub-band are used

LH sub-band is more robust	**Similar robustness**
Gaussian noise	JPEG compression
Salt and pepper noise	Scale
Sharpen	
Smoothing	
Rotation	
Histogram equalisation	
Translation	

From our inspection of the results, we found that the LH sub-band is more robust to Gaussian noise, salt and pepper noise, sharpening, smoothing, histogram equalisation, translation and rotation than the LL sub-band. We can also argue that the robustness of our technique against JPEG compression and scale attacks is comparable to that of the other technique. These findings are summarised in Table 5.2.

5.4.2 Stirmark Benchmark on the LH Sub-band

To further compare the robustness of the region-adaptive DWT-SVD watermarking technique with the original DWT-SVD technique [Emir, 2004], we apply both watermark techniques to 50 different host images and analyse their robustness using the Stirmark 4.0 benchmarking tool [Petitcolas, 1998]. The original DWT-SVD technique presents a hybrid scheme based on DWT and SVD. In the interest of making the comparison as fair as possible, we embed HF and LF watermark images separately into the host image and calculate the average PSNR value of the two.

In this experiment, it is possible that, due to the nature of the attacks, the extracted watermark does not bear any resemblance to the original watermark. When this is the case, we flag the result as an unsuccessful result. This is performed manually by observing the quality of the extracted watermark. Figure 5.9 shows an example of the

watermark images extracted successfully after watermark attack. Furthermore, Figure 5.10 shows an example of the failed extraction of watermark images after watermark attack. The success rate of a technique in this experiment therefore indicates how often the watermark can be extracted successfully.

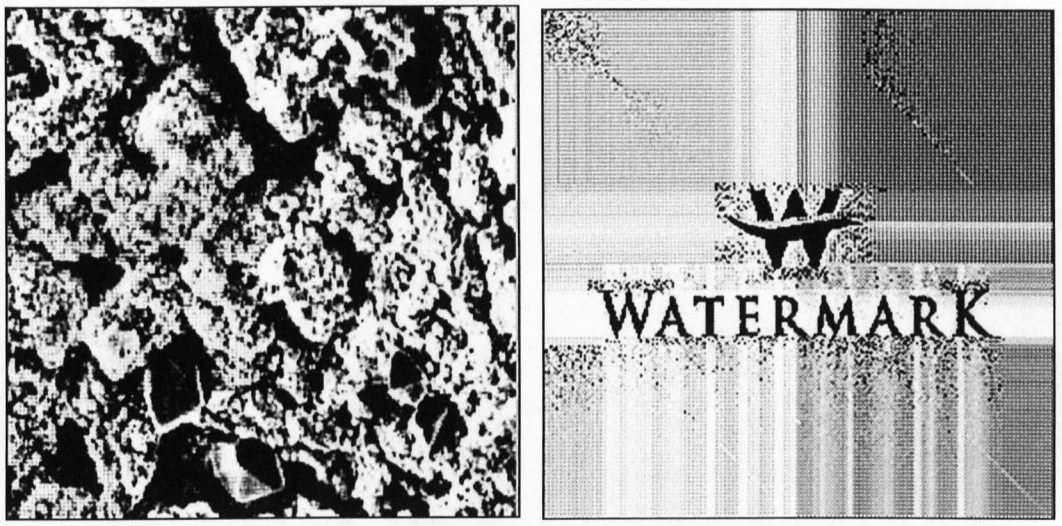


Figure 5.9 An example of watermark images extracted successfully

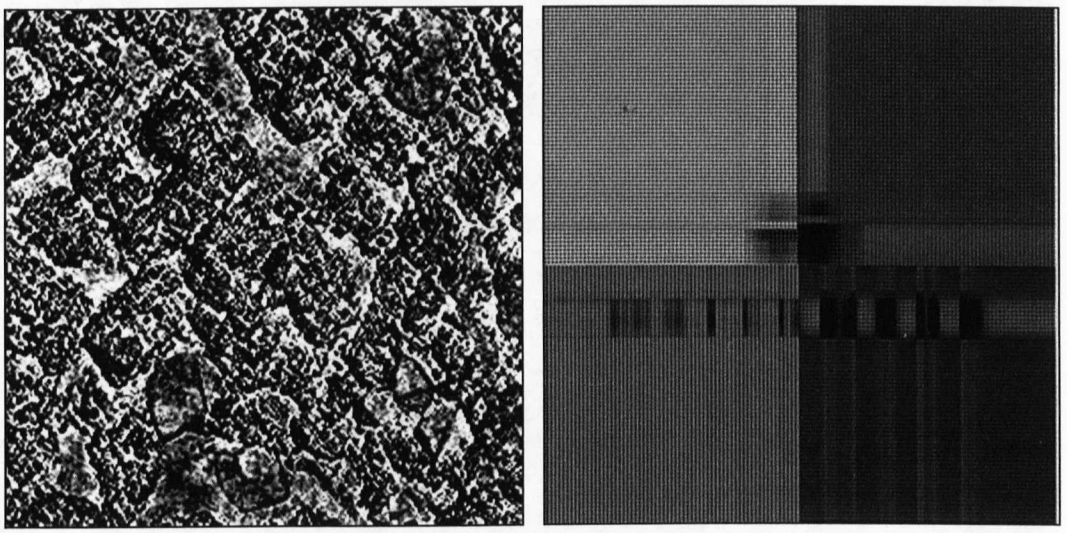


Figure 5.10 An example of watermark images where extraction failed

We used 13 watermark attacks as included in the Strimark benchmark, namely Affine, Conv, JPEG, LATESTRNNDIST, Median, Noise, PSNR, Resc, RML, RNDDIST, Rotation, RotCrop and RotScale.

5.4.2.1 Affine

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will placed under affine attack in this section. There are eight

different attack strengths, ranging from affine_1 to affine_8. Table 5.3 indicates the success rate of the extracted watermark images after affine attack on both watermarking algorithms. Furthermore, Figure 5.11 shows the representative results of extracted watermark images after affine attack (affine_1) on both techniques.

Table 5.3 The Success Rate of Extracted Watermark Image after Affine Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
AFFINE	98.39%	30.23%

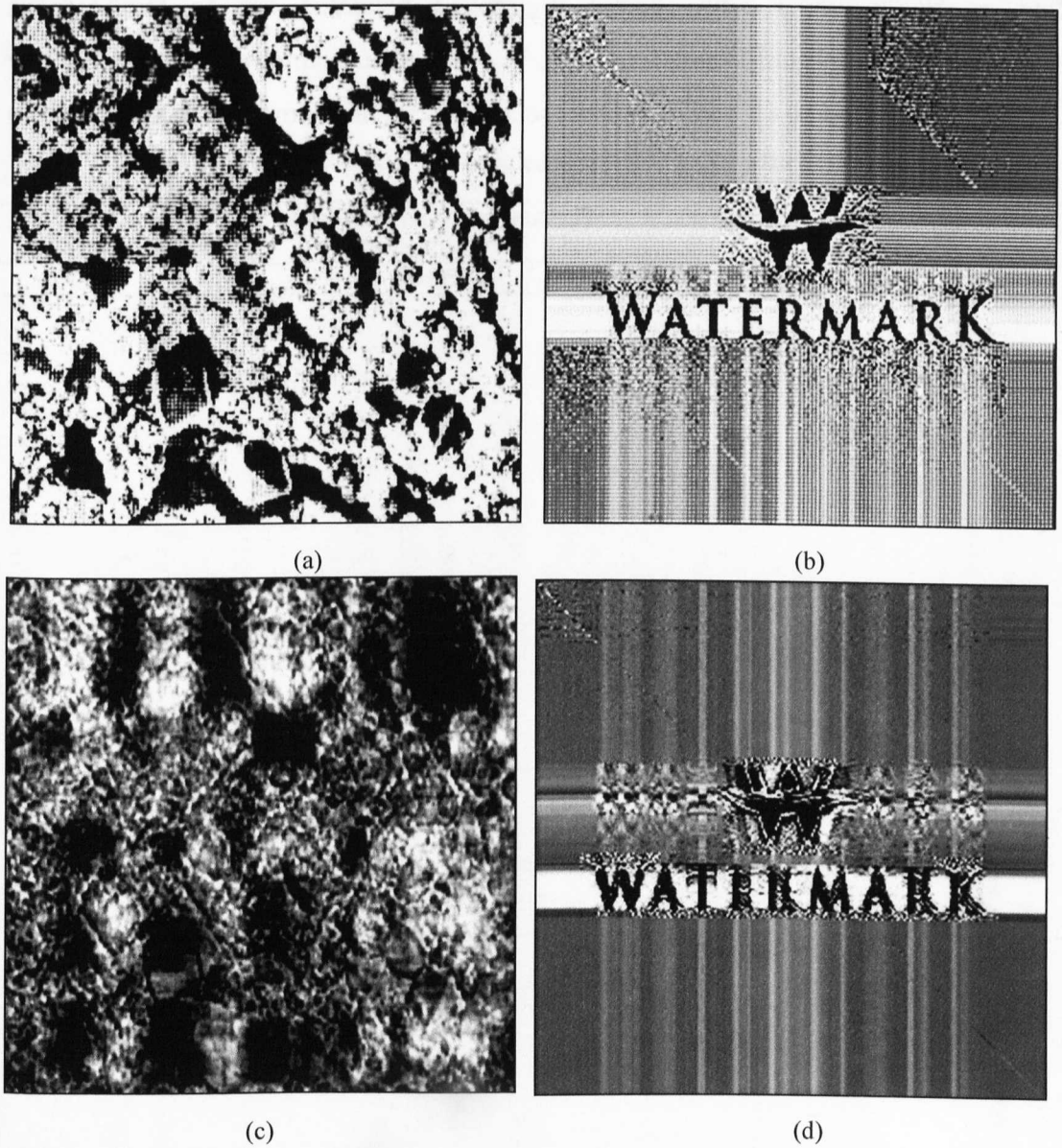


Figure 5.11 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after affine attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after affine attack

The results from Table 5.3 and Figure 5.11 indicate that the region-adaptive DWT-SVD watermarking algorithm is more robust than the original DWT-SVD watermarking algorithm after affine attack.

5.4.2.2 Conv

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm are put under Conv attack in this section. There are two different attack strengths, ranging from Conv_1 to Conv_2. Table 5.4 indicates the success rate of the extracted watermark images after Conv attack on both watermarking algorithms. Furthermore, Figure 5.12 shows the representative results of the extracted watermark images after Conv attack (Conv_1) on both techniques.

Table 5.4 The Success Rate of Extracted Watermark Image after CONV Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
CONV	96.49%	52.50%

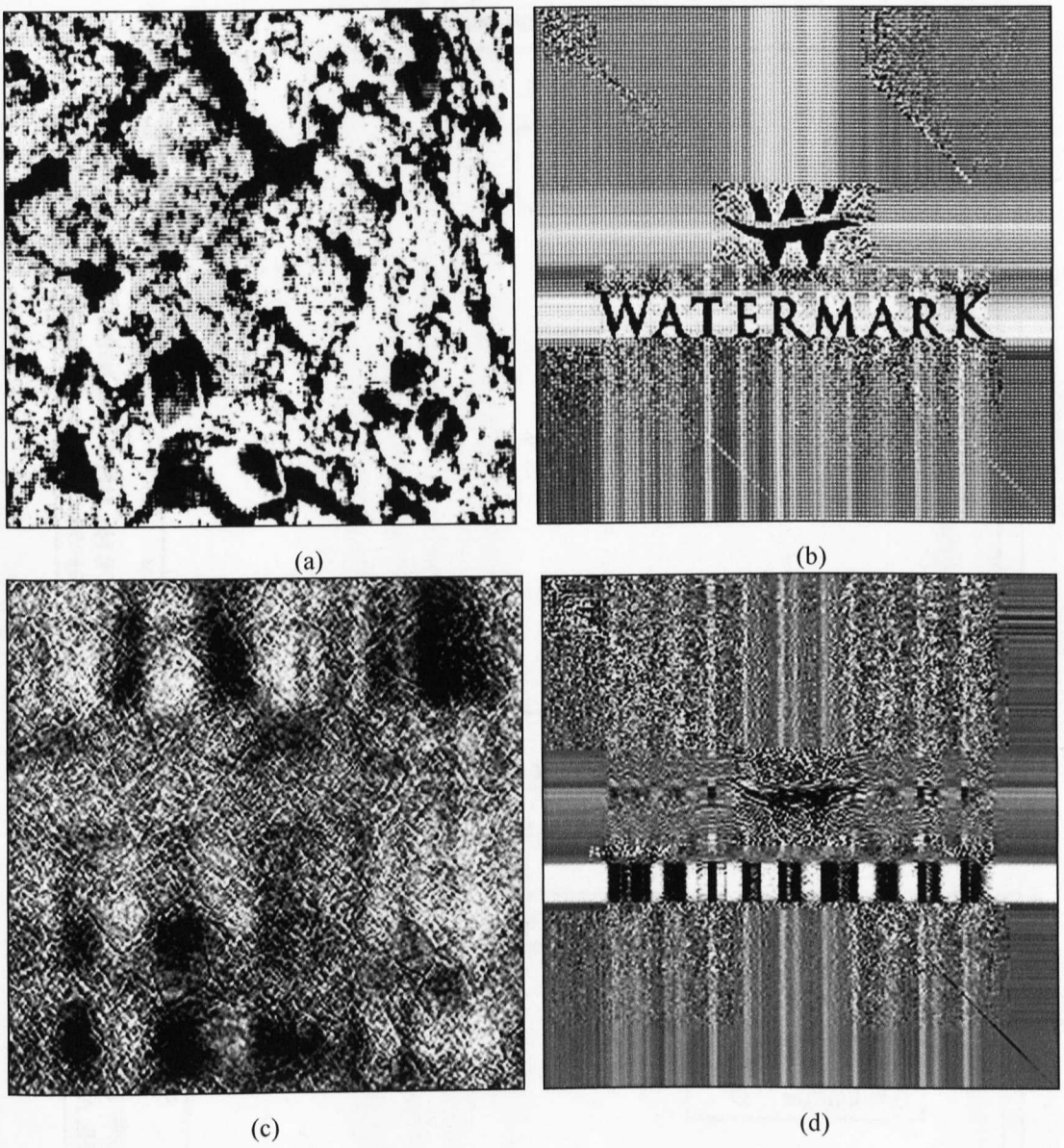
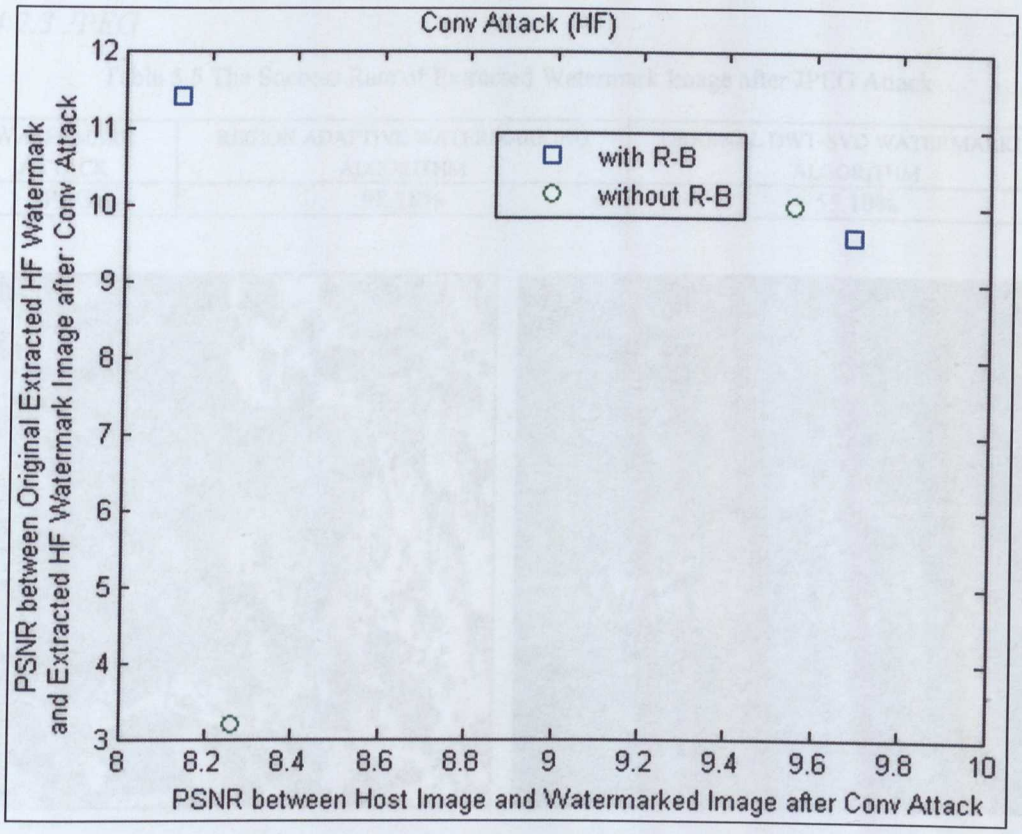
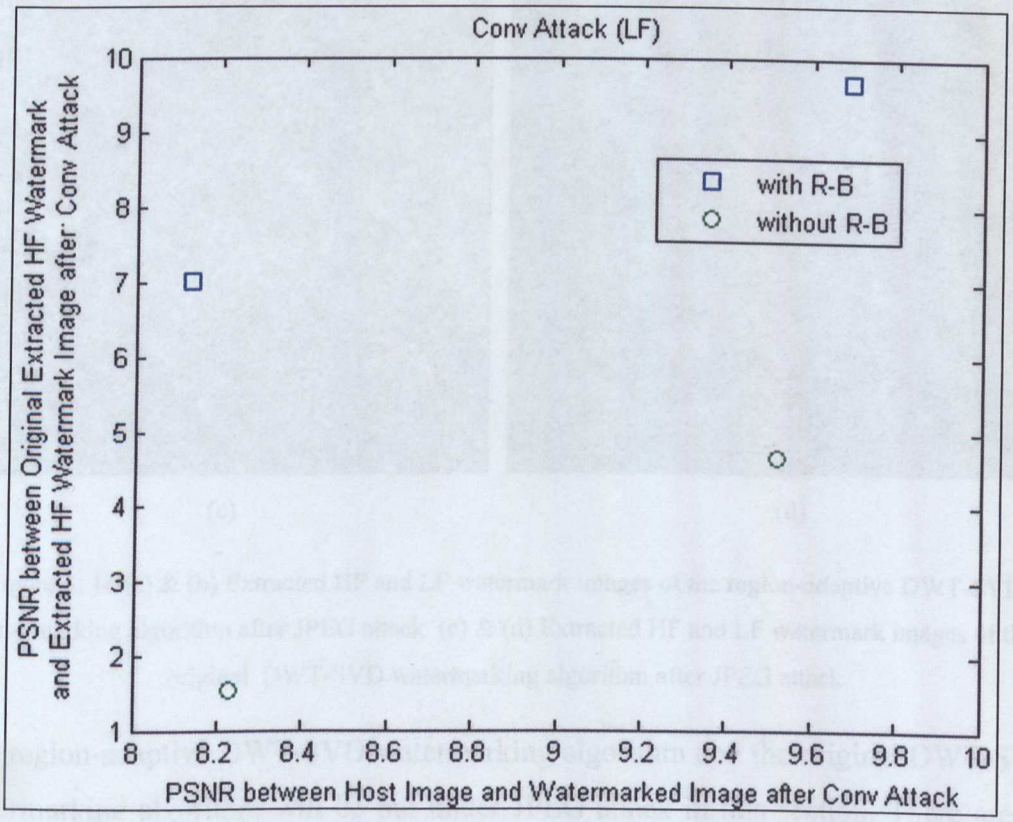


Figure 5.12 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after Conv attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after Conv attack

Furthermore, Figure 5.13 shows the performance comparison of the region-adaptive watermark and the original DWT-SVD technique after Conv attack, because both success rates are over 50%. The figures show the region-adaptive DWT-SVD watermarking algorithm has a higher watermark PSNR value for a given watermarked PSNR. The results of the region-adaptive DWT-SVD watermarking algorithm are marked with '□', whereas those of the original DWT-SVD algorithm are marked with '○'.



(a)



(b)

Figure 5.13 Performance of the proposed watermarking algorithm and the original DWT-SVD algorithm after Conv attack (a) HF watermark image, (b) LF watermark image

5.4.2.3 JPEG

Table 5.5 The Success Rate of Extracted Watermark Image after JPEG Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
JPEG	98.18%	55.10%

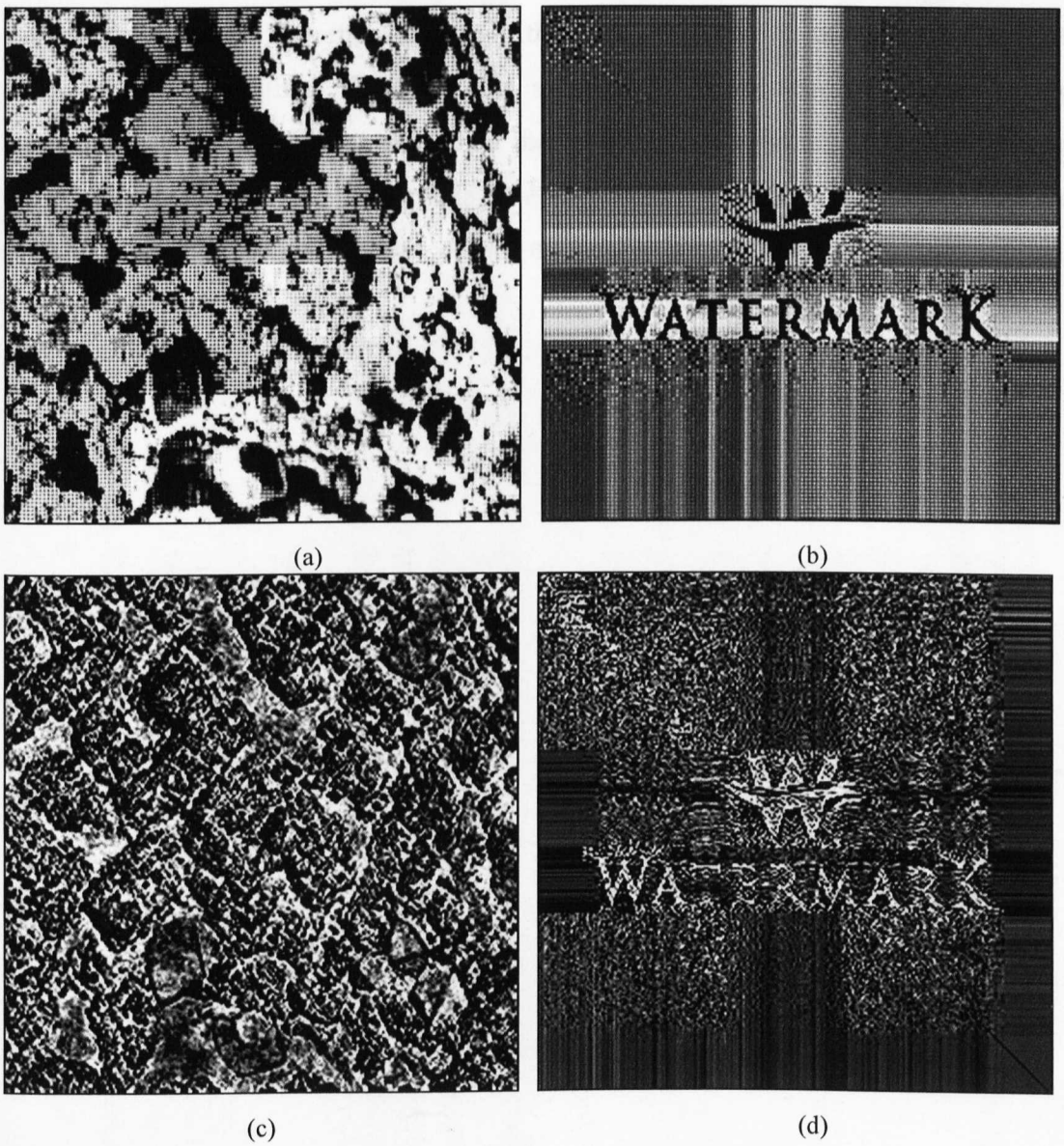
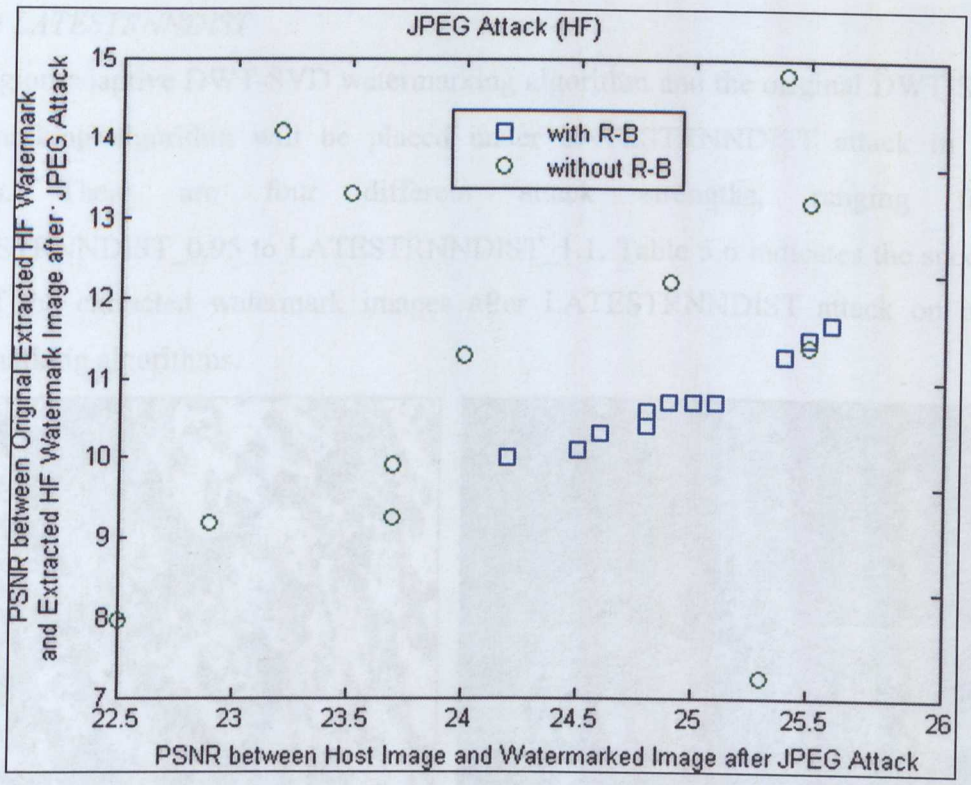


Figure 5. 14 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after JPEG attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after JPEG attack

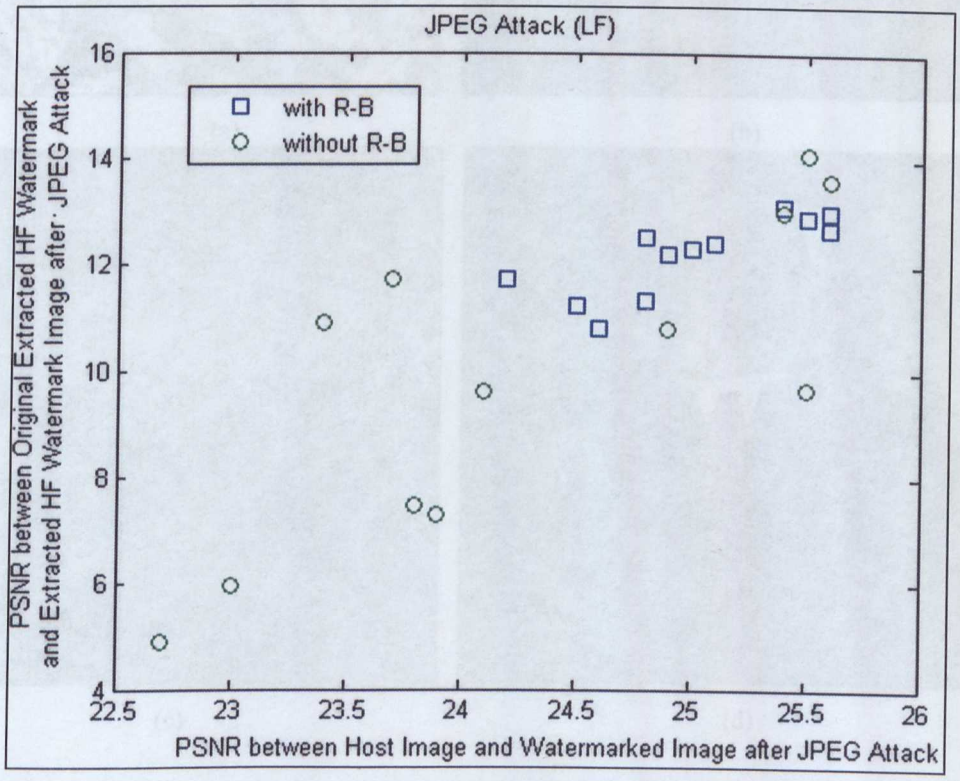
The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be put under JPEG attack in this section. There are 12 different attack strengths, ranging from JPEG_15 to JPEG_100. Table 5.5 indicates the success rate of the extracted watermark images after JPEG attack on both

watermarking algorithms. Furthermore, Figure 5.14 shows the representative results of the extracted watermark images after JPEG attack (JPEG_15) on both techniques.

Furthermore, Figure 5.15 shows a performance comparison after the JPEG compression attack, because both of the success rates are over 50%. Figure 5.15a shows the proposed watermarking algorithm has good PSNR values in x direction around 25. In addition, Figure 5.15b shows the region-adaptive watermarking algorithm has higher PSNR values than the original watermarking algorithm in most cases. The results of the region-adaptive DWT-SVD watermarking algorithm are marked with '□', whereas those of the original DWT-SVD algorithm are marked with '○'.



(a)

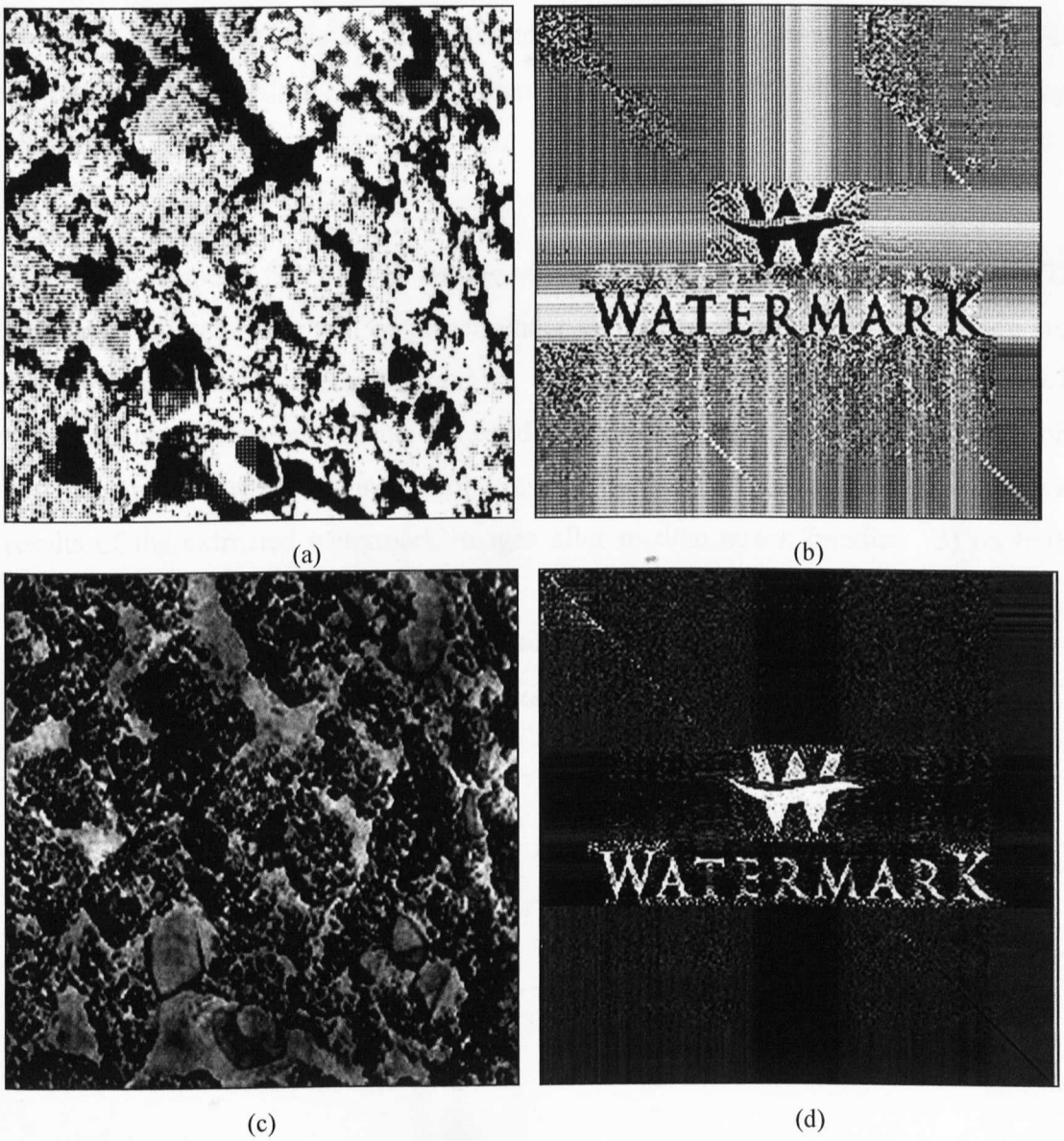


(b)

Figure 5.15 Performance of the region-adaptive watermarking algorithm and the original DWT-SVD algorithm after JPEG attack (a) HF watermark image, (b) LF watermark image

5.4.2.4 LATESTRNNDIST

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be placed under LATESTRNNDIST attack in this section. There are four different attack strengths, ranging from LATESTRNNDIST_0.95 to LATESTRNNDIST_1.1. Table 5.6 indicates the success rate of the extracted watermark images after LATESTRNNDIST attack on both watermarking algorithms.



(a) (b)

(c) (d)

Figure 5.16 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after LATESTRNNDIST attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after LATESTRNNDIST attack

Table 5.6 The Success Rate of Extracted Watermark Image after LATESTRNNDIST Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
LATESTRNNDIST	34%	58.24%

In conclusion, although the visual results of extracted HF & LF watermark images of the region-adaptive DWT-SVD watermarking algorithm look better than the original DWT-SVD watermarking algorithm after LATESTRNNDIST in this case, which are shown in Figure 5.16, the success rate indicates that the region-adaptive DWT-SVD watermarking algorithm is worse than the original DWT-SVD watermarking algorithm overall.

5.4.2.5 Median

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be placed under median attack in this section. There are four different attack strengths, ranging from median_3 to median_9. Table 5.7 indicates the success rate of the extracted watermark images after median attack on both watermarking algorithms. Furthermore, Figure 5.17 shows the representative results of the extracted watermark images after median attack (median _3) on both techniques.

Table 5.7 The Success Rate of Extracted Watermark Image after Median Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
MEDIAN	97.26%	61.76%

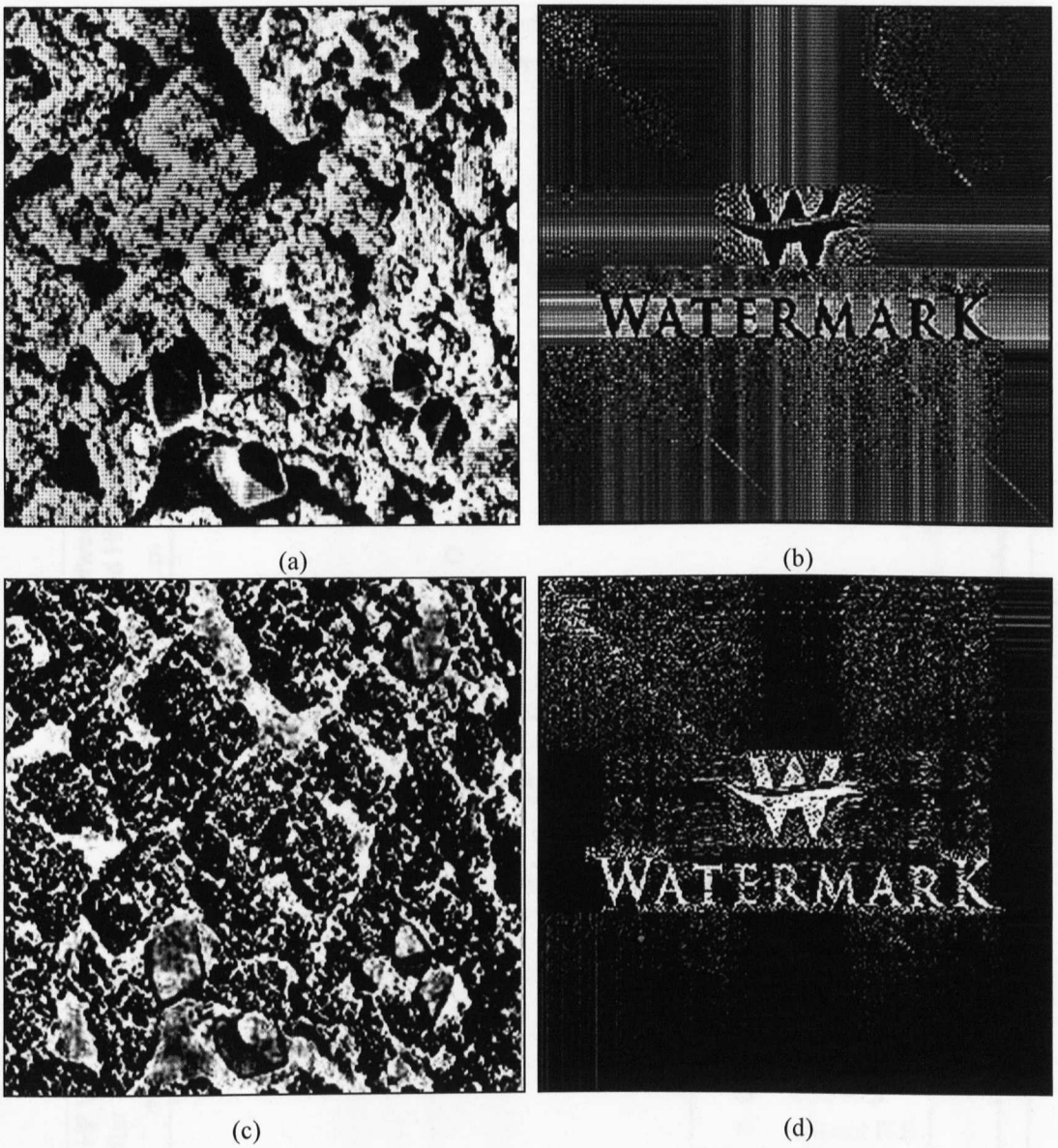
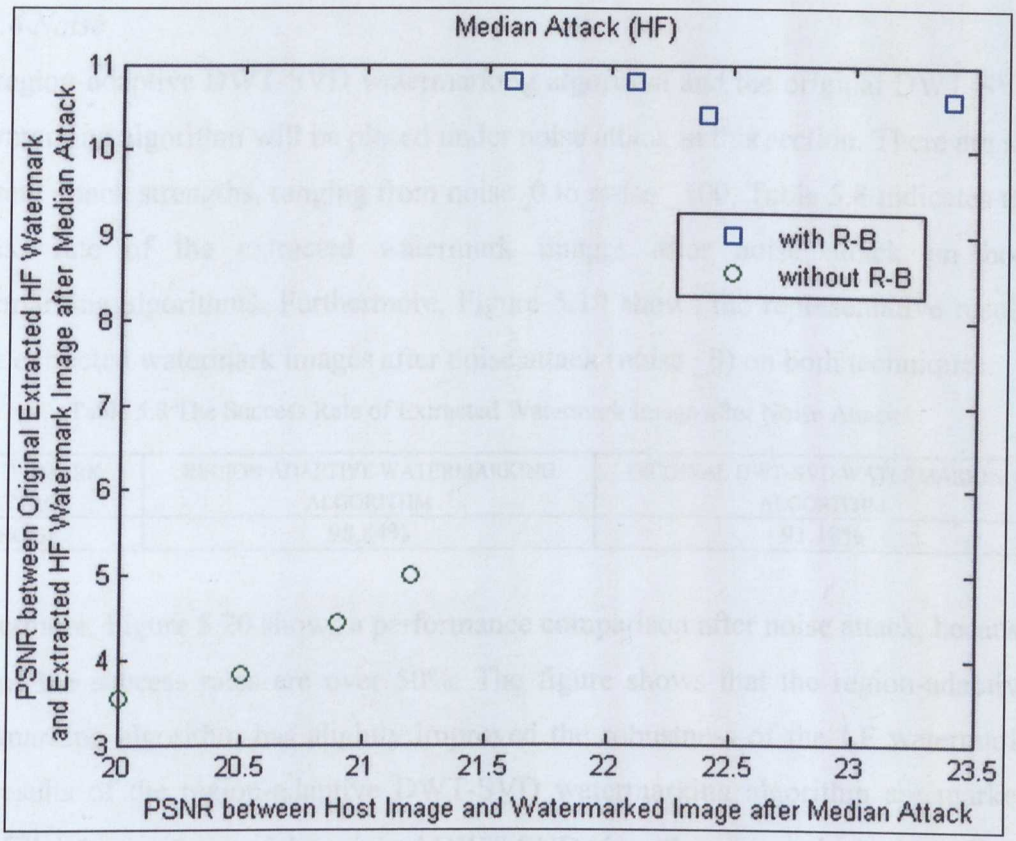
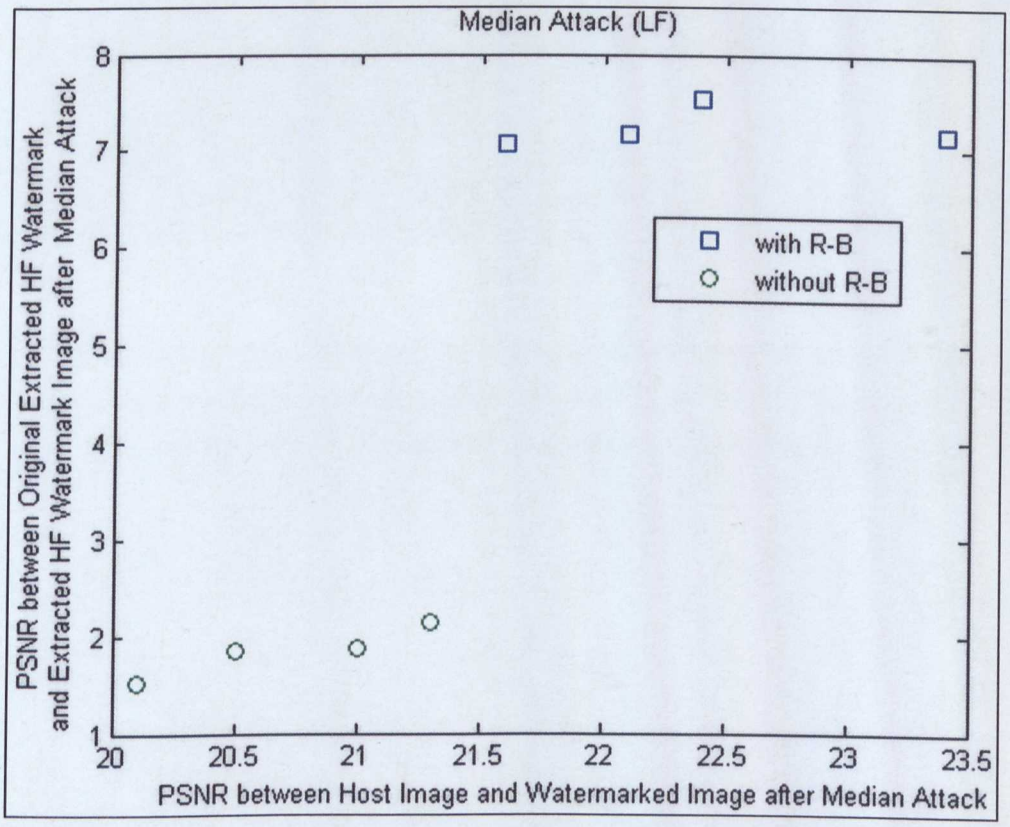


Figure 5.17 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after median attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after median attack

Furthermore, Figure 5.18 shows a performance comparison after median attack, because both of the success rates are over 50%. It indicates that the proposed region-adaptive approach significantly improves both the watermark and the watermarked PSNR values. The results of the region-adaptive DWT-SVD watermarking algorithm are marked with '□', whereas those of the original DWT-SVD algorithm are marked with '○'.



(a)



(b)

Figure 5.18 Performance of the region-adaptive watermarking algorithm and the original DWT-SVD algorithm after median attack (a) HF watermark image, (b) LF watermark image

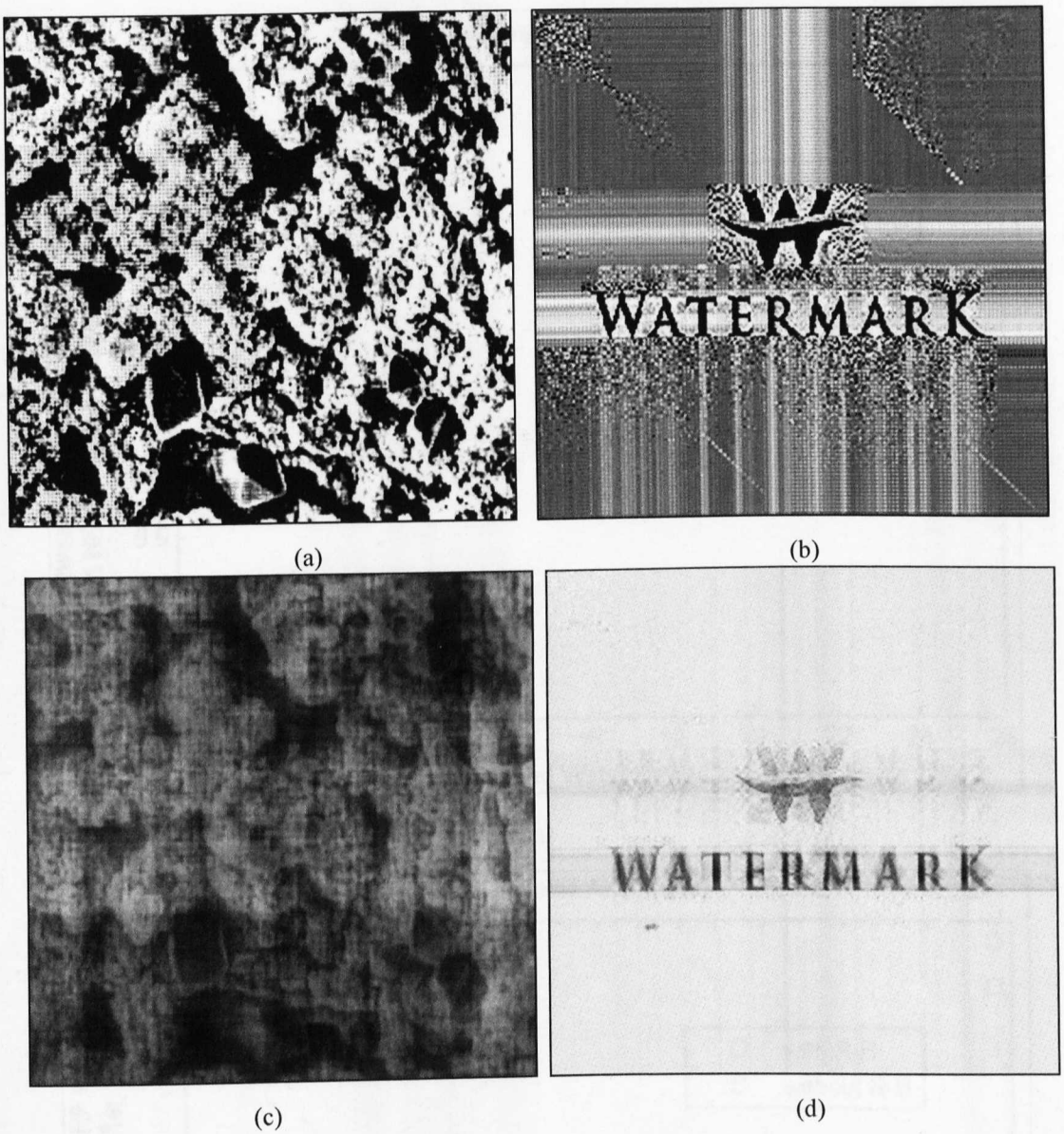
5.4.2.6 Noise

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be placed under noise attack in this section. There are six different attack strengths, ranging from noise_0 to noise _100. Table 5.8 indicates the success rate of the extracted watermark images after noise attack on both watermarking algorithms. Furthermore, Figure 5.19 shows the representative results of the extracted watermark images after noise attack (noise _0) on both techniques.

Table 5.8 The Success Rate of Extracted Watermark Image after Noise Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
NOISE	98.84%	91.18%

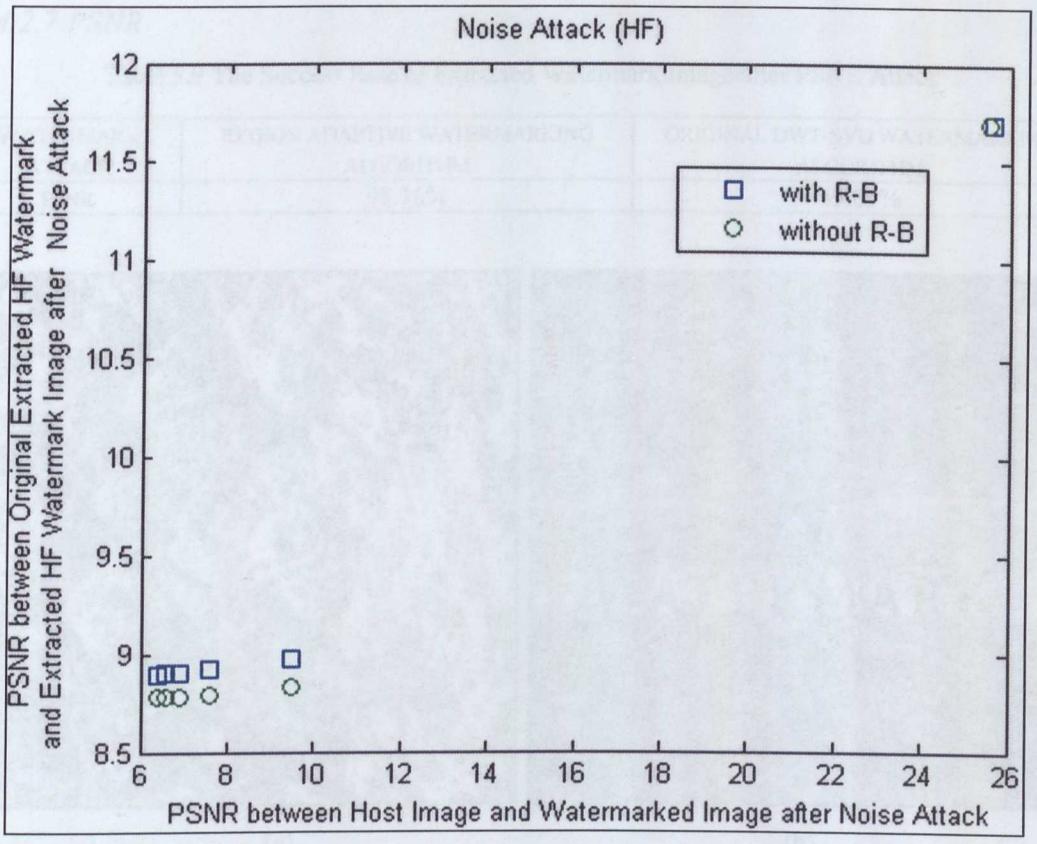
Furthermore, Figure 5.20 shows a performance comparison after noise attack, because both of the success rates are over 50%. The figure shows that the region-adaptive watermarking algorithm has slightly improved the robustness of the LF watermark. The results of the region-adaptive DWT-SVD watermarking algorithm are marked with '□', whereas those of the original DWT-SVD algorithm are marked with '○'.



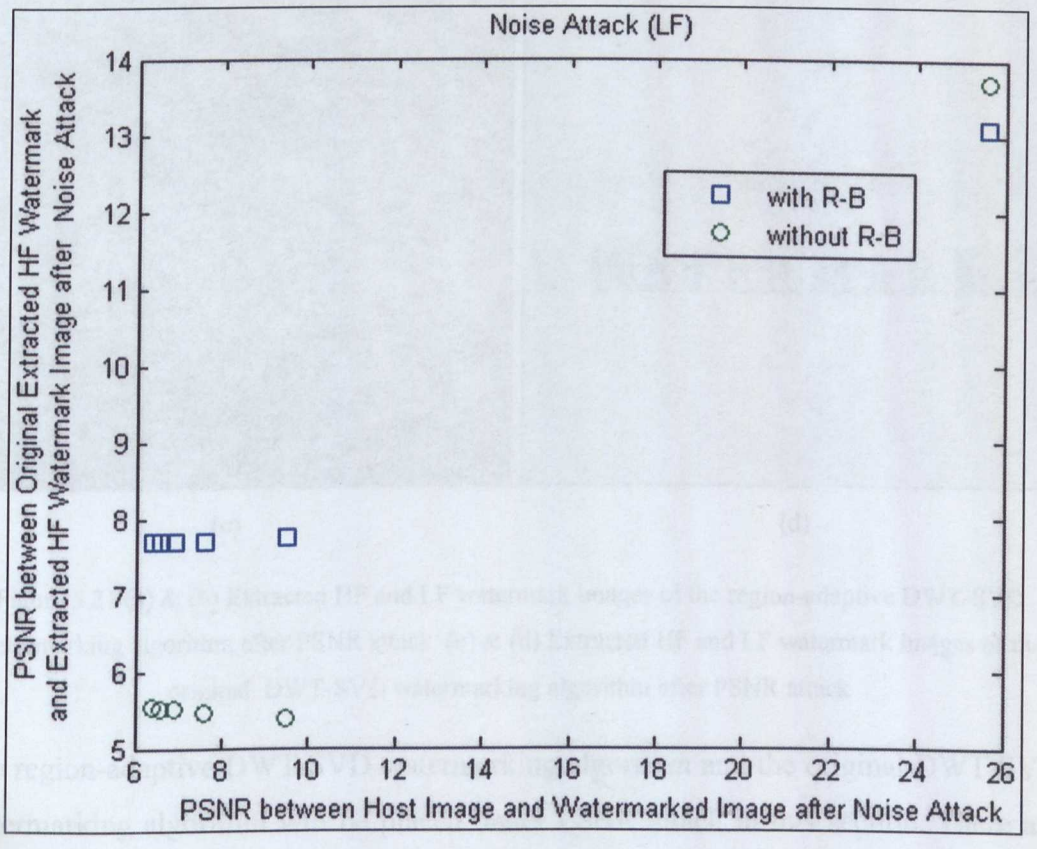
(a) (b)

(c) (d)

Figure 5.19 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after noise attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after noise attack



(a)



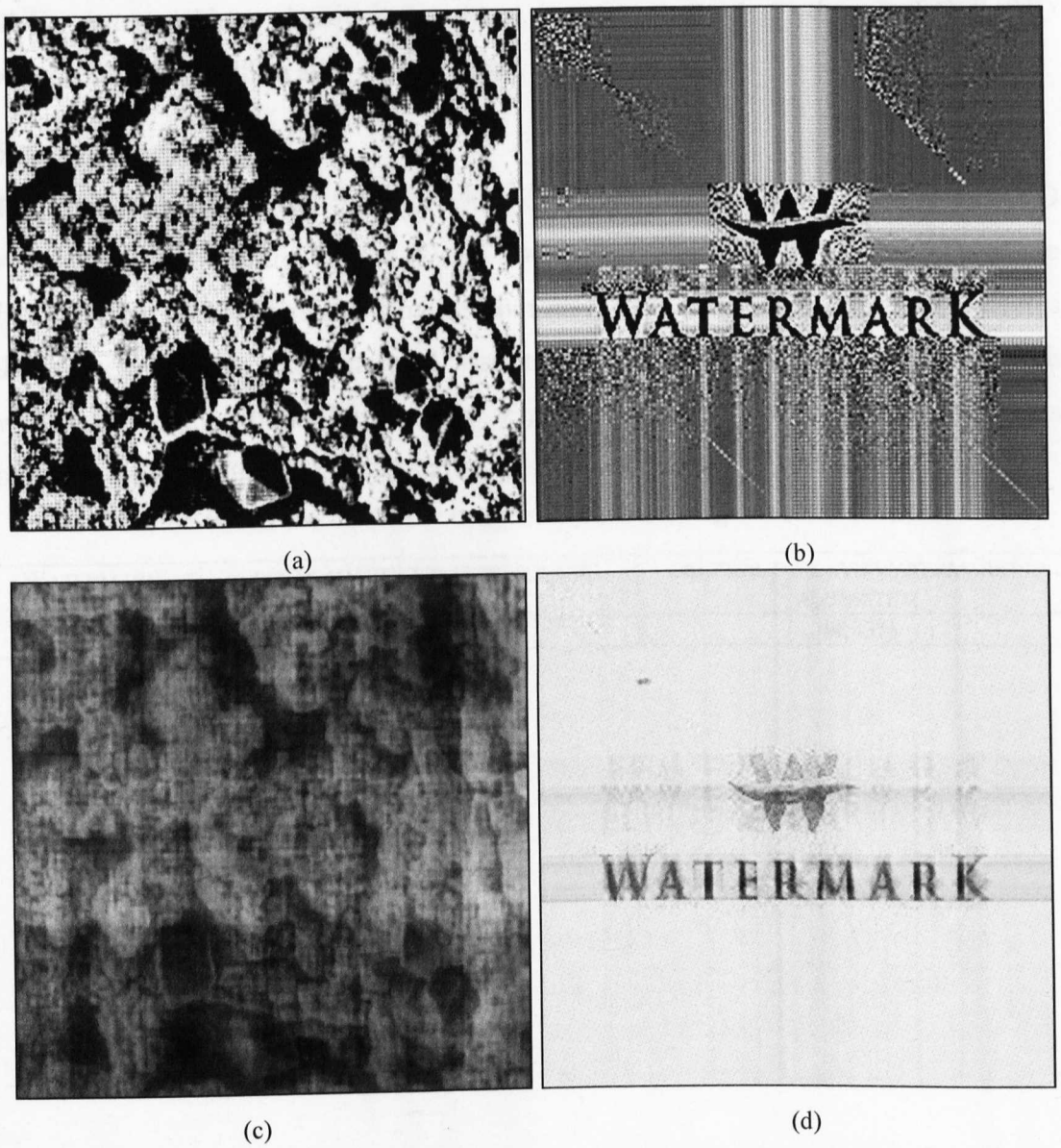
(b)

Figure 5. 20 Performance of the region-adaptive watermarking algorithm and the original DWT-SVD algorithm after noise attack (a) HF watermark image, (b) LF watermark image

5.4.2.7 PSNR

Table 5.9 The Success Rate of Extracted Watermark Image after PSNR Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
PSNR	98.36%	49.09%



(a) (b)

(c) (d)

Figure 5.21 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after PSNR attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after PSNR attack

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be placed under PSNR attack in this section. There are 11 different attack strengths, ranging from PSNR _0 to PSNR _100. Table 5.9 indicates the success rate of the extracted watermark images after PSNR attack on

both watermarking algorithms. Furthermore, Figure 5.21 shows the representative results of the extracted watermark images after PSNR attack (PSNR _0) on both techniques.

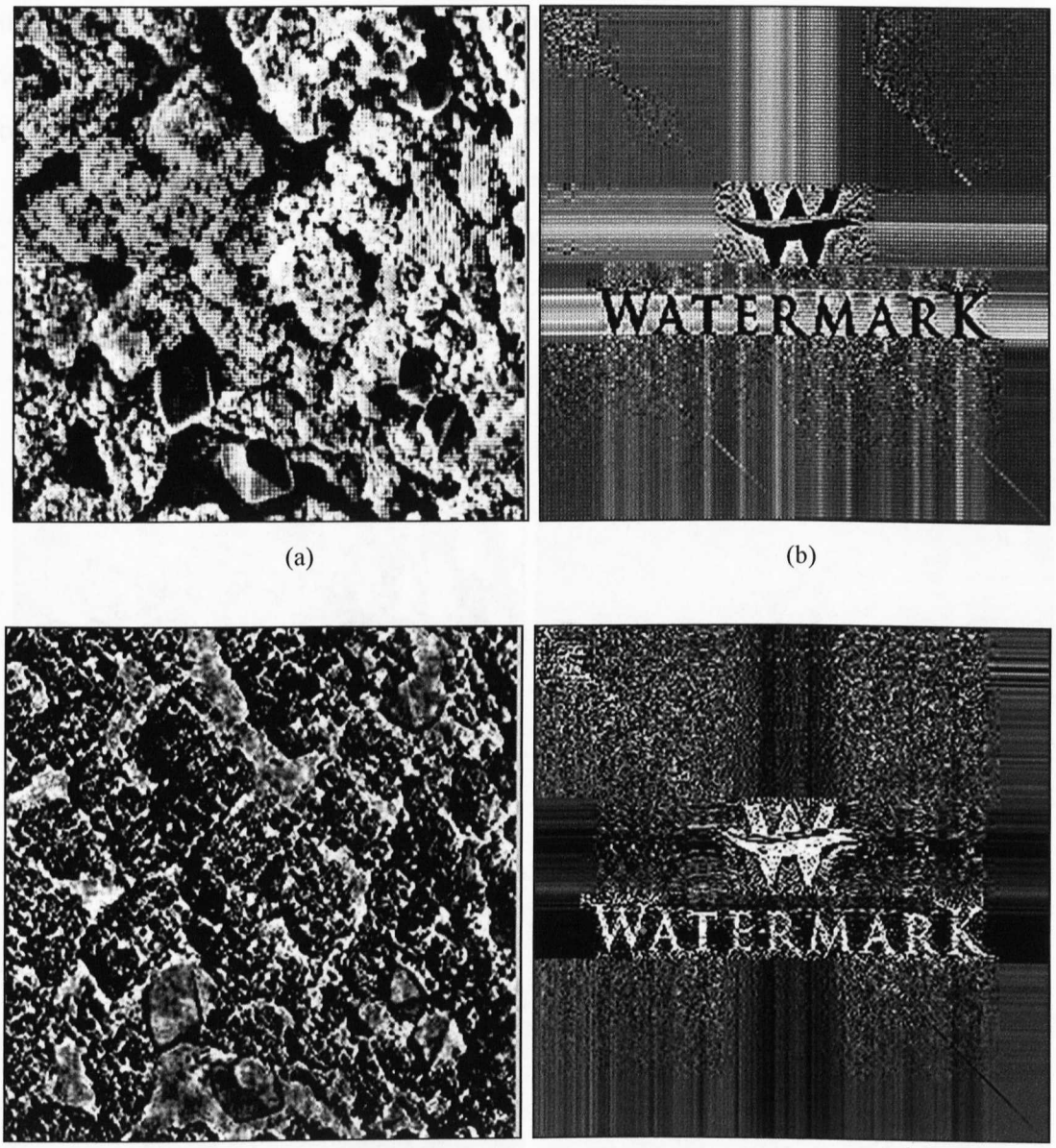
The results from Table 5.9 and Figure 5.21 indicate that the region-adaptive DWT-SVD watermarking algorithm is more robust than the original DWT-SVD watermarking algorithm after PSNR attack.

5.4.2.8 Resc

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be placed under Resc attack in this section. There are five different attack strengths, ranging from resc_75 to resc_200. Table 5.10 indicates the success rate of the extracted watermark images after Resc attack on both watermarking algorithms. Furthermore, Figure 5.22 shows the representative results of the extracted watermark images after Resc attack (Resc_150) on both techniques.

Table 5.10 The Success Rate of Extracted Watermark Image after RESC Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
RESC	81.48%	44.33%



(a) (b)

(c) (d)

Figure 5.22 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after Resc attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after Resc attack

The results from Table 5.10 and Figure 5.22 indicate that the region-adaptive DWT-SVD watermarking algorithm is more robust than the original DWT-SVD watermarking algorithm after Resc attack.

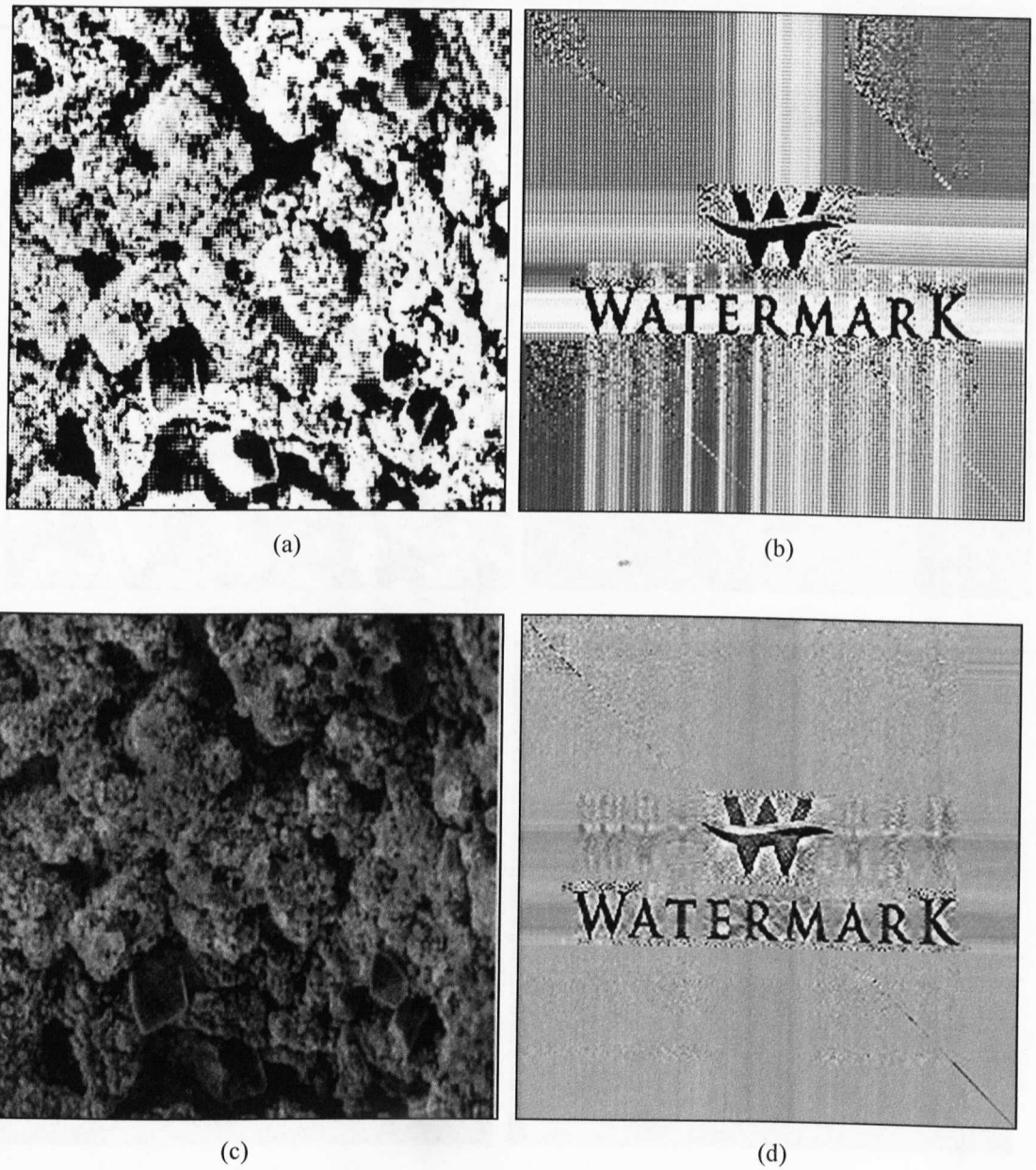
5.4.2.9 RML

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be placed under RML attack in this section. There are 10 different attack strengths, ranging from RML_10 to RML_100. Table 5.11 indicates

the success rate of the extracted watermark images after RML attack on both watermarking algorithms. Furthermore, Figure 5.23 shows the representative results of the extracted watermark images after RML attack (RML_50) on both techniques.

Table 5.11 The Success Rate of Extracted Watermark after RML Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
RML	97.22%	40.63%



(a) (b)

(c) (d)

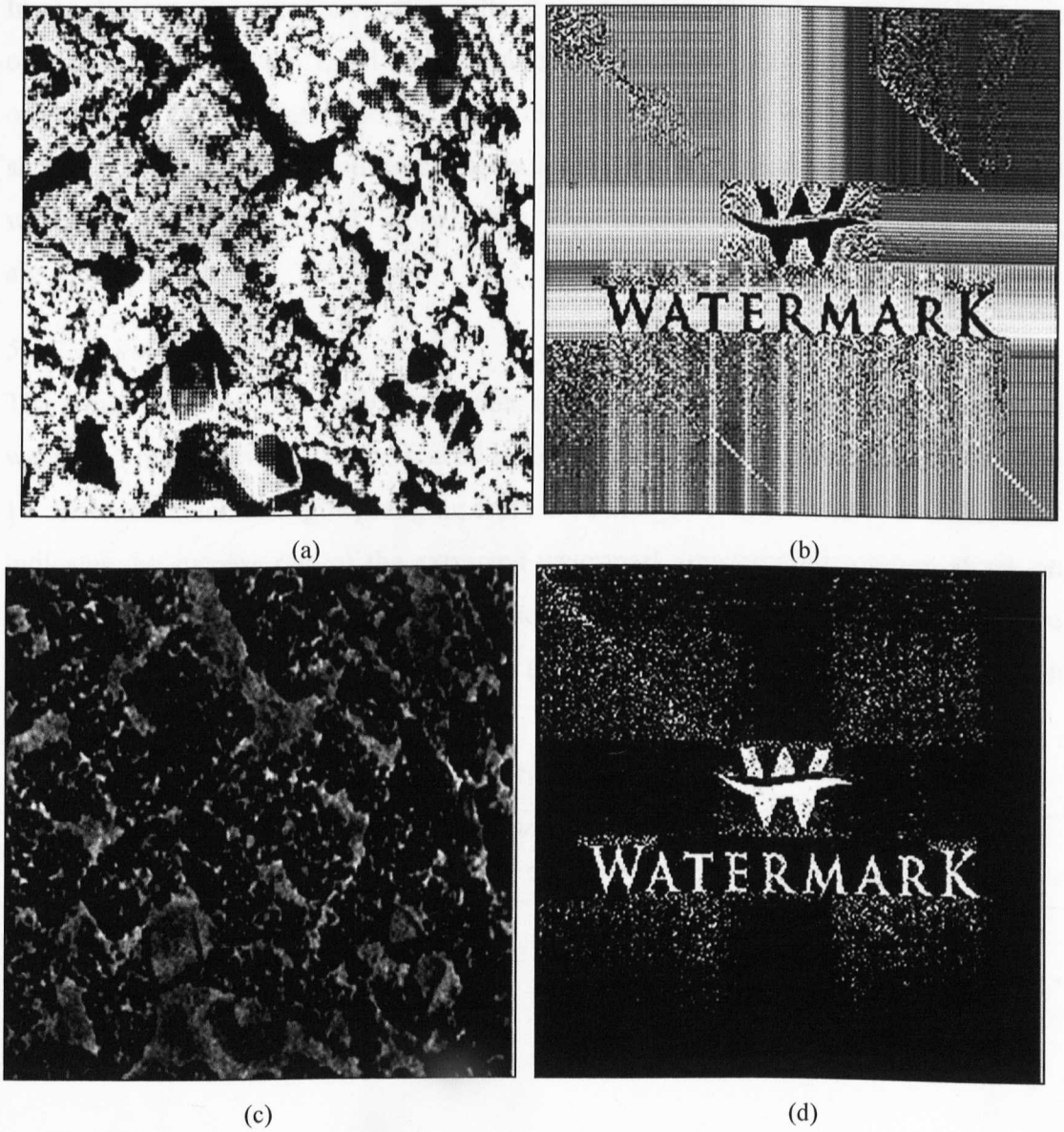
Figure 5.23 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after RML attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after RML attack

The results from Table 5.11 and Figure 5.23 indicate that the region-adaptive DWT-SVD watermarking algorithm is more robust than the original DWT-SVD watermarking algorithm after RML attack.

5.4.2.10 RNDDIST

Table 5.12 The Success Rate of Extracted Watermark Image after RNDDIST Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
RNDDIST	30.23%	50.75%



(a) (b)

(c) (d)

Figure 5.24 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after RNDDIST attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after RNDDIST attack

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be placed under RNDDIST attack in this section. There are four different attack strengths, ranging from RNDDIST _0.95 to RNDDIST _1.1. Table 5.12 indicates the success rate of the extracted watermark images after RNDDIST attack on both watermarking algorithms. Furthermore, Figure 5.24 shows the representative results of the extracted watermark images after RNDDIST attack (RNDDIST _0.95) on both techniques.

In conclusion, although the visual results of the extracted HF & LF watermark images of the region-adaptive DWT-SVD watermarking algorithm look better than the original DWT-SVD watermarking algorithm after RNDDIST in this case, which are shown in Figure 5.24, the success rate indicates that the region-adaptive DWT-SVD watermarking algorithm is worse than the original DWT-SVD watermarking algorithm overall.

5.4.2.11 Rotation

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be placed under rotation attack in this section. There are 16 different attack strengths, ranging from Rotation_0.25 to Rotation_-2. Table 5.13 indicates the success rate of the extracted watermark images after rotation attack on both watermarking algorithms. Furthermore, Figure 5.25 shows the representative results of the extracted watermark images after rotation attack (rotation_0.75) on both techniques.

Table 5.13 The Success Rate of Extracted Watermark Image after Rotation Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
ROTATION	89.29%	38.10%

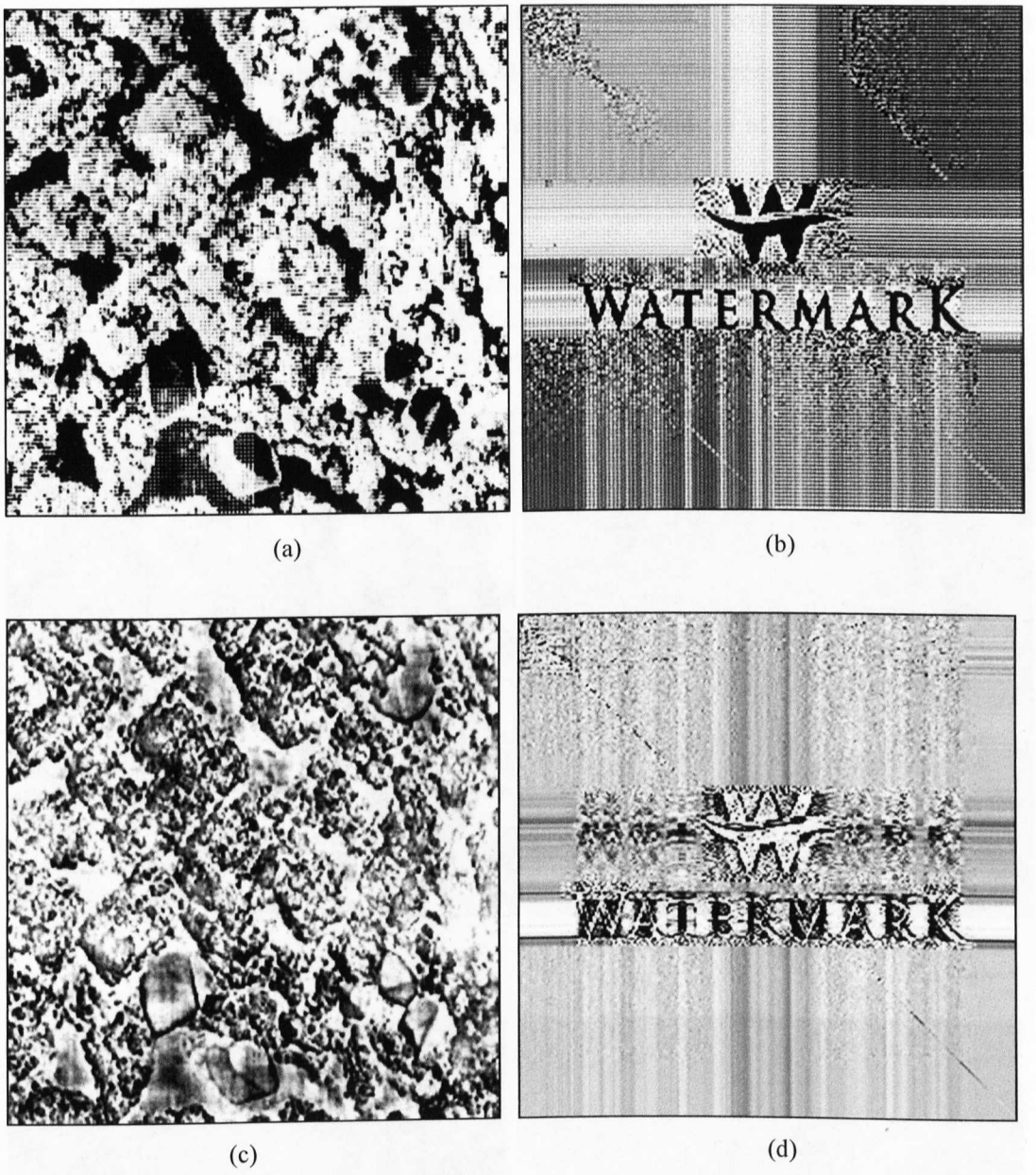


Figure 5.25 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after rotation attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after rotation attack

The results from Table 5.13 and Figure 5.25 indicate that the region-adaptive DWT-SVD watermarking algorithm is more robust than the original DWT-SVD watermarking algorithm after rotation attack.

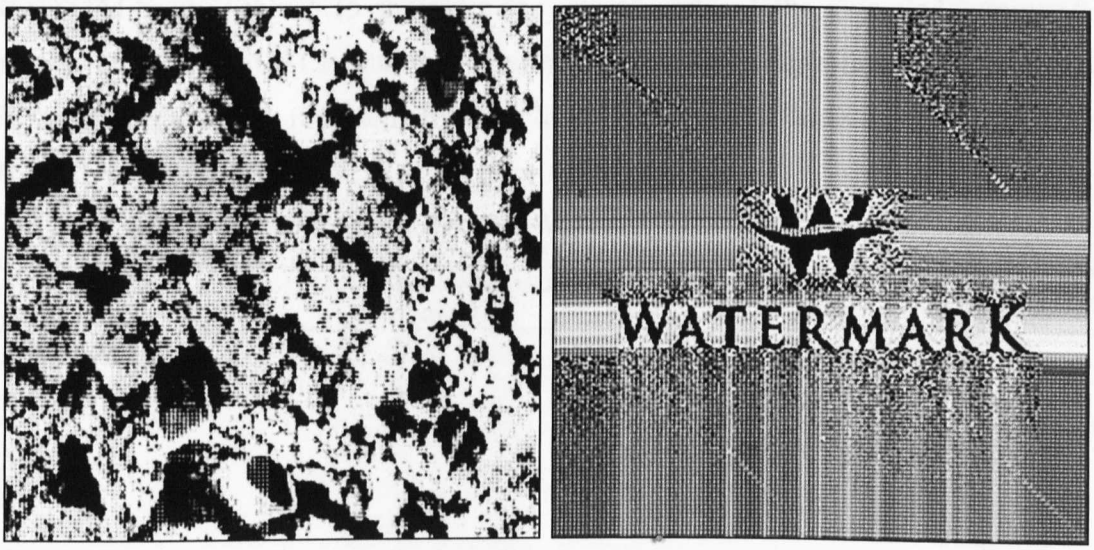
5.4.2.12 RotCrop

The region-adaptive DWT-SVD watermarking algorithm and original DWT-SVD watermarking algorithm will be placed under ROTCROP attack in this section. There are 10 different attack strengths, ranging from ROTCROP_0.25 to ROTCROP _-2.

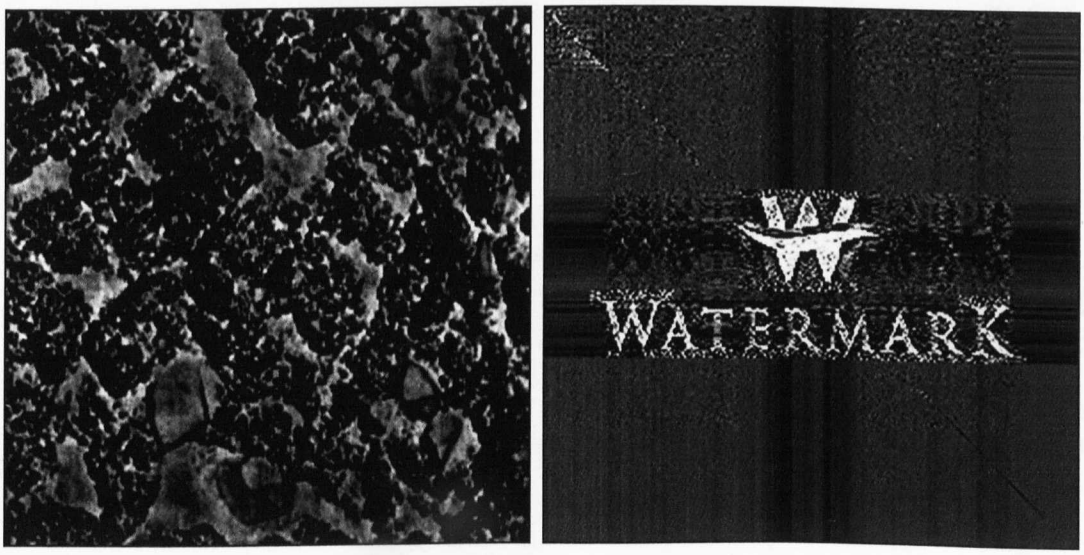
Table 5.14 indicates the success rate of the extracted watermark images after ROTCROP attack on both watermarking algorithms. Furthermore, Figure 5.26 shows the representative results of the extracted watermark images after ROTCROP attack (ROTCROP_0.75) on both techniques.

Table 5.14 The Success Rate of Extracted Watermark Image after RotCrop Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
ROTCROP	96.61%	47.92%



(a) (b)



(c) (d)

Figure 5.26 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after ROTCROP attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after ROTCROP attack

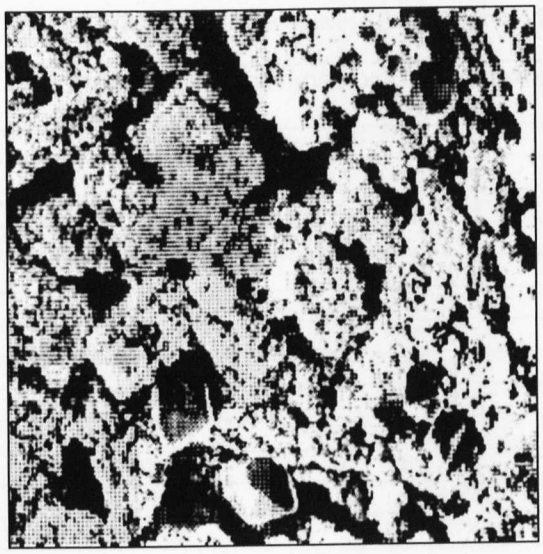
The results from Table 5.14 and Figure 5.26 indicate that the region-adaptive DWT-SVD watermarking algorithm is more robust than the original DWT-SVD watermarking algorithm after ROTCROP attack.

5.4.2.13 RotScale

The region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermarking algorithm will be placed under ROTSCALE attack in this section. There are 10 different attack strengths, ranging from ROTSCALE _0.25 to ROTSCALE_-2. Table 5.15 indicates the success rate of the extracted watermark images after ROTSCALE attack on both watermarking algorithms. Furthermore, Figure 5.27 shows the representative results of the extracted watermark images after ROTSCALE attack (ROTSCALE_0.75) on both techniques.

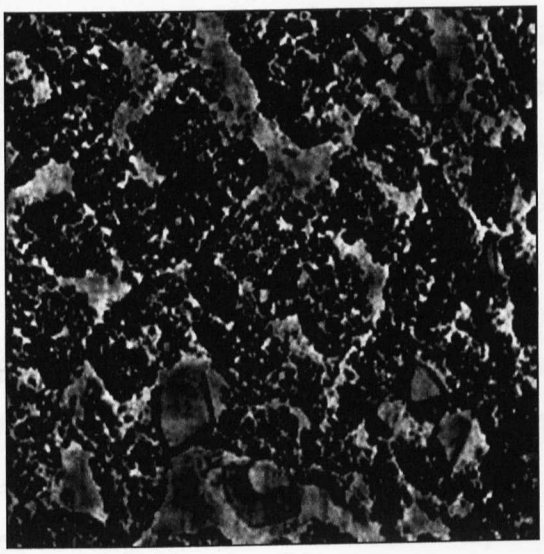
Table 5.15 The Success Rate of Extracted Watermark Image after ROTSCALE Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	ORIGINAL DWT-SVD WATERMARKING ALGORITHM
ROTSCALE	97.50%	47.19%



(a)

(b)



(c)

(d)

Figure 5.27 (a) & (b) Extracted HF and LF watermark images of the region-adaptive DWT-SVD watermarking algorithm after rotscale attack (c) & (d) Extracted HF and LF watermark images of the original DWT-SVD watermarking algorithm after rotscale attack

The results from Table 5.15 and Figure 5.27 indicate that the region-adaptive DWT-SVD watermarking algorithm is more robust than the original DWT-SVD watermarking algorithm after ROTSCROP attack.

To sum up, Table 5.16 shows the success rate of extracting watermark images using both techniques after each attack.

Table 5.16 The Success Rate of Extracting Watermark Image after Each Attack

WATERMARK ATTACK	REGION ADAPTIVE WATERMARKING ALGORITHM	NON-REGION ADAPTIVE WATERMARKING ALGORITHM
AFFINE	98.39%	30.23%
CONV	96.49%	52.50%
JPEG	98.18%	55.10%
LATESTRNNDIST	34%	58.24%
MEDIAN	97.26%	61.76%
NOISE	98.84%	91.18%
PSNR	98.36%	49.09%
RESC	81.48%	44.33%
RML	97.22%	40.63%
RNDDIST	30.23%	50.75%
ROTATION	89.29%	38.10%
ROTATECROP	96.61%	47.92%
ROTATESCALE	97.50%	47.19%

As can be seen from the results shown in the tables 5.16, the region-adaptive DWT-SVD watermarking algorithm produced higher PSNR values and a higher success rate in most attacks. Furthermore, as can be seen from the results shown in the figures, the region-adaptive DWT-SVD watermarking algorithm produced better visual results than the original DWT-SVD watermarking algorithm.

5.5. Chapter Summary

In this chapter, a novel region-adaptive watermark algorithm was developed based on findings from the previous chapter. This technique was derived from the other region-adaptive technique by using a singular value decomposition algorithm prior to embedding the watermark data. Comparing with the previous chapter, there are two main improvements: the first improvement is that this chapter applies the DWT-SVD embedding algorithm instead of the DWT algorithm in the previous chapter, and the second improvement is that this chapter applies both the LL and LH sub-bands as an embedding field. Therefore, this approach is expected to improve robustness in both removal and geometric attacks, and it is named the 'region-adaptive DWT-SVD watermarking algorithm'.

Through the experiment, we found that the region-adaptive DWT-SVD watermarking algorithm is capable of resisting various watermark attacks including both removal and geometric attacks in the LL and LH sub-bands. After that, there is a robustness comparison between the LL and LH sub-bands through nine different watermark attacks, the results of which in Figure 5.8 indicate that the LH sub-band is more

robust than its counterpart. Besides that, to further investigate the robustness of the LH sub-band, we applied Stirmark 4.0 in this chapter; the experiment shows that the LH sub-band can repel most watermark attacks. At the end of the experiment, we compared robustness between the region-adaptive DWT-SVD watermarking algorithm and the original DWT-SVD watermark technique. The results show that the region-adaptive DWT-SVD watermarking algorithm is far better than the original DWT-SVD watermark algorithm.

In the next chapter, we create a novel application by using the region-adaptive DWT-SVD watermarking algorithm to detect different watermark attacks.

CHAPTER 6

NOVEL APPLICATION OF THE REGION-ADAPTIVE WATERMARKING TECHNIQUE IN WATERMARK ATTACK DETECTION

So far, we have covered the design and development of a novel image watermarking technique using an adaptive region-adaptive approach. In this chapter, we will study how this new watermarking technology can be used in a unique way to detect the type of attacks which may have been applied to a watermark. The chapter is divided into five sections. In section 6.2, we will review image tamper and tamper detection. In section 6.3, we will describe watermark attack detection, and in section 6.4 we will show our experiment and its results. Furthermore, in section 6.5, we will discuss the advantages and disadvantages of this technique.

6.1. Introduction

Despite the availability of extremely powerful technologies in both generating and processing digital images, there is a severe lack of techniques and methodologies for validating their authenticity. In other words, it is becoming easier to tamper with digital images in ways that are difficult to detect. Figure 6.1 shows a modification made to an image using photo editing software such as Adobe Photoshop. In this

particular example, the original image is shown in Figure 6.1a and the modified image is shown in Figure 6.1b. If the tampered image was used as critical evidence in a legal case or police investigation, then this form of tampering might pose a serious problem [Ingemar J. Cox].

(b)

Figure 6.1 Modification made to an image using Adobe Photoshop

Image tampering and any problems that arise from it can be solved by watermarking. Digital watermarking technologies have been proposed for implementation in many applications such as image tamper detection. For example, digital signature technology has been used in digital cameras since the 1990s and was pioneered by Friedman [G.L.Friedman, 1996, G.L.Friedman, 1993], who suggested creating a "trustworthy camera" by computing a signature inside the camera. The digital signature would embed into the camera image directly by using watermarking. Epson offers such a system as an option on many of its digital cameras. We refer to such an embedded signature as an 'authentication mark' [Ingemar J. Cox, 2008], which is designed to become invalid after even the slightest modification/tamper. This is known as a 'fragile watermark'.

A fragile watermarking system as one of the main methods can detect tampering on digital images if they are modified. However, a robust watermarking system can also apply in this field. In this chapter, we will describe a novel tamper detection application by using the robust region-adaptive watermarking system.

6.2. Image Tamper & Tamper Detection

6.2.1 Image Tamper

Digital images can be modified, edited or tampered with in many possible ways, the most common of which are listed below [Granty, 2010].

6.2.1.1 Splicing

Image splicing is a technique that involves a composite of two or more images which are combined to create a fake image. Figure 6.2 presents an example of a spliced image. Figures 6.2a and 6.2b are images used in a spliced image and Figure 6.2c is the resultant image after splicing [Granty, 2010].

(c)

Figure 6.2 Example of image tampering on splicing (a) & (b) Original image (c) Spliced image

6.2.1.2 Image Retouching

Image retouching does not change an image significantly, but instead enhances or reduces certain features of the image. This technique is popular among magazine photo editors. It can be said that almost all magazine covers employ this technique to 'enhance' certain features of an image so that it is more attractive. A retouched image is shown in Figure 6.3. Figure 6.3a is the original image and 6.3b is the retouched image. [Granty, 2010].

Figure 6.3 Example of image tampering on retouching (a) Original image (b) Retouched image

6.2.1.3 Geometrical Transformation

Some images have a portion of the picture altered by some common geometric transformations such as translation, scaling and rotation. Copiers make a copy of the portion of the picture and make changes to it by geometrically modifying that portion of the image, which is shown in Figure 6.4. Figure 6.4a shows the original image of evidence showing a cartridge and Figure 6.4b is a fake image of evidence showing two cartridges; the second one is scaled and transformed [Granty, 2010].

Figure 6.4 Example of image tampering on geometrical transformation (a) Original image (b) Scaled and transformed image

6.2.1.4 Copy-Move Attack

A copy-move attack is more or less similar to image splicing in that both techniques modify certain image regions of a base image along with another image. However, instead of having an external image as the source, the copy-move attack uses a portion of the original base image as its source. In other words, the source and the destination of the modified image originate from the same image.

In a copy-move attack, parts of the original image are copied, moved to a desired location and pasted. This is usually done in order to conceal certain details or to duplicate certain aspects of an image. Figure 6.5 is an example of copy-move forgery. Figure 6.5(a) is an original image, Figure 6.5(b) is the resultant image of a copy-move attack and Figure 6.5(c) shows the copied portions [Granty, 2010].

Figure 6.5 Example of image tampering on a copy-move attack (a) Original image (b) Resulting image, (c) Copied portions

6.2.2 Image Tamper Detection

Tamper detection can be classified broadly into two categories[Girija, 2010, Granty, 2010], namely active methods and passive methods. The active method requires that certain information is embedded inside an image during creation or before the image is disseminated to the public. This information can be used to either detect the source of an image or to detect possible modifications to an image. One of the techniques used under the active method is watermarking. On the other hand, the passive method does not require any pre-'image distribution' information to be inserted into the digital image. The method works purely by analysing binary information within the digital image, without any need for external information [Pai, 2006].

6.2.2.1 Active Methods

Active technology depends on watermarking and digital signatures to authenticate an image. A novel watermark algorithm is described in [Iwata, 2010], which can detect and recover JPEG images by using the Reed-Solomon code, an idea based on Minami's method [N.Minami, 2002]. The proposed method directly embeds the check symbols corresponding to an information-reduced original image into the LSBs of the quantised DCT coefficients in a JPEG image. For tamper detection, the watermark is firstly extracted from a suspicious JPEG image. Next, it is transformed into the check symbols by the inverse transform, while the information symbols are calculated from the suspicious JPEG image. In the experiment, the host image is 256×256, and the proposed method can theoretically recover a tamper part when the size of the tamper part is equal to 48×48. If the tamper part is larger than 48×48, such as 60×60, the modification in the tampered part is not recovered. Moreover, if the tamper part is less than 48× 8, such as 28×28, the modification in the tampered part is not recovered either.

In [Amira, 2010], another fragile watermarking is used. The proposed authentication scheme is based on inserting the logo into the least significant bits. The watermark using chaotic mixing and Eigen values is represented by 18 bits, which is inserted into a block of the original image using least significant bits. The original watermarked image in the experiment suffers from two tampers, namely geometrical transformation and splicing. The results show that this algorithm can detect tampered blocks in authenticated image accuracy.

6.2.2.2 Passive Methods

The past few years have seen the growth of research on passive digital image tampering detection, which can be categorised at three levels [Wei, 2009] :

- Low Level: Methods at this level use the statistical characteristics of digital image pixels or DCT/DWT coefficients. For example, a geometric attack or gamma correction during image acquiring processing will bring consistent correlations of adjacent pixels, whereas tampering will break this consistency. Investigating double JPEG compression for tampering detection is an example of using the statistical characteristics of DCT coefficients. Using a model of authentic images that tampered images do not satisfy for tampering detection also belongs to this level. In short, no semantic information is employed at this level [Mahdian, 2008, Mahdian, 2009, Micah, 2006].

- Middle Level: At this level, we detect traces of the tampering operation which has some simple semantic information, like splicing causing sharp edges, blur operation after splicing, inconsistencies in lighting direction, etc. [Dirik, 2009, Sutcu, 2007, Tao, 2010].

- High level: i.e. the semantic level. Actually, it is very hard for computers to use semantic information for tampering detection because the aim of tampering is to change the meaning of the image content it originally conveyed. Nevertheless, sometimes it still works. For example, it does not make sense to have an image in which George W. Bush is shaking hands with Osama bin Laden.

6.3. Novel Application of the Region-Adaptive Watermarking Technique in Watermark Attack Detection

6.3.1 Watermark Attack Detection Scheme

Chapter 5 showed that the region-adaptive DWT-SVD watermarking technique is more robust in removal and geometric attacks. However, this is not its only advantage. Tamper detection is another very important aspect in digital security. In this section, we introduce a novel application of the region-adaptive watermarking system, which detects tampers on a digital image. The low frequency area LL sub-band will be used in this algorithm.

Figure 6.6 shows the novel watermark attack detection scheme, which requires a certain threshold in addition to linear classification equations. The scheme starts by calculating PSNR between the original watermarked image and the tested watermarked image. If PSNR is higher than a certain threshold, it means that the original watermarked image and the test watermarked image are almost identical. However, if PSNR is lower than a certain threshold, it means that the test watermarked image may suffer from attack.

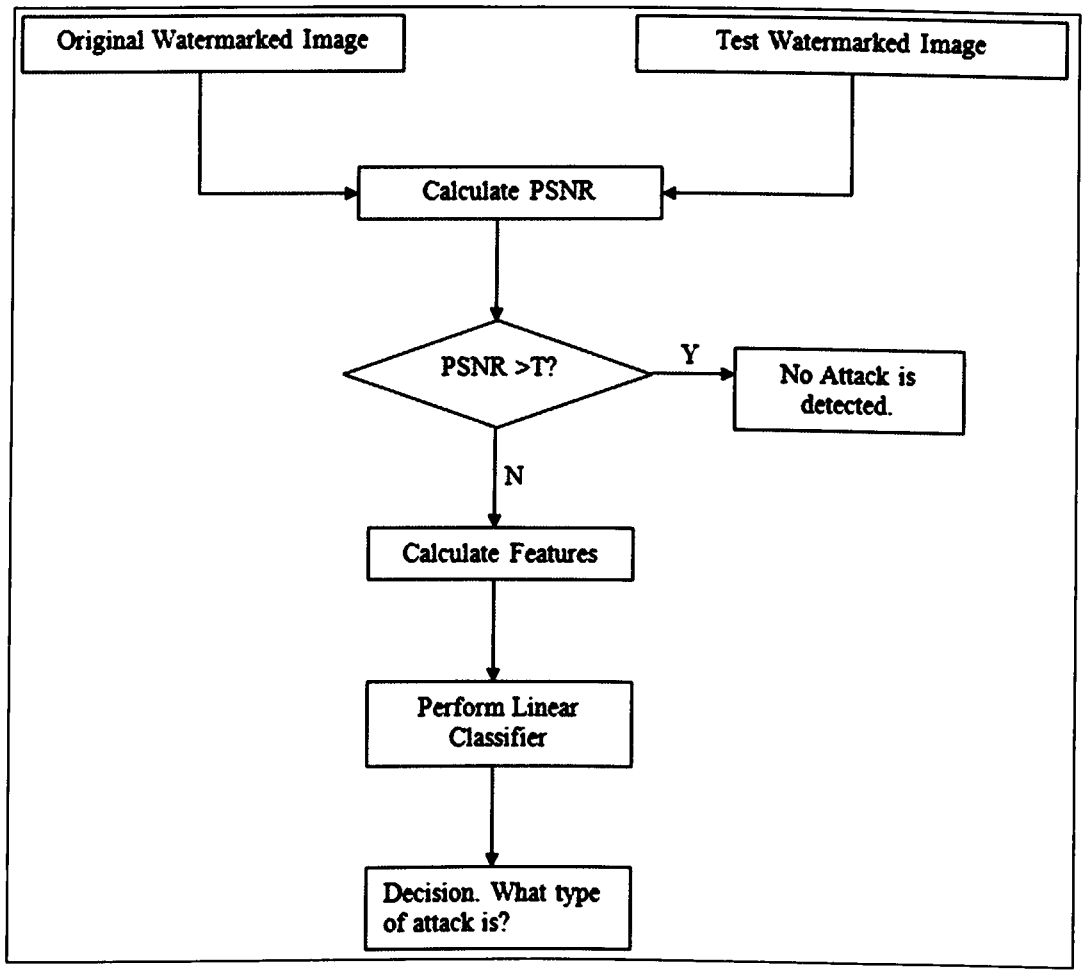


Figure 6.6 Watermark attack detection scheme

To explain the choice of threshold more clearly, an explanation of the experiments carried at to determine the threshold is given below. To determine the value of threshold, we conducted 2 experiments, the first experiment calculates the PSNR values between the host image and non-attacked watermarked image, these values are shown in table 6.1, for each different weight alpha values ranges from 0.2 to 0.005. The effect of using alpha values lower than 0.005 would be negated by the discretization of image pixel values. Hence, it will results in constant PSNR.

Table 6.1 PSNR values between host image and original watermarked image

Alpha	0.2	0.15	0.1	0.05	0.01	0.005
PSNR	24.7845	27.2813	30.664	36.2487	51.1495	54.9119

As we can find in table 6.1, when alpha value equals to 0.005, the PSNR value is just below 55, which point out the upper boundary of PSNR between the host image and non-attacked watermarked image. Therefore, the first experiment shows the threshold of PSNR value should be greater than 55 just to detect any changes in the host image.

The second experiment calculates the PSNR values between original watermarked image and attacked watermarked image. This experiment we just use compress attack. The compression ratio ranges from 5% to 1%. Any compression ratio which is lower than 1% would produce a negligible difference to the original image. Table 6.2 summaries the results.

Table 6.2 PSNR values between original watermarked image and compressed watermarked image

Compress ratio	5%	4%	3%	2%	1%
PSNR	49.2578	50.4686	52.0295	54.3672	59.1637

As can be seen from the second experiment result, the highest PSNR value is 59.1637 at 1% compression ratio. Using the he conclusion of both experiments we have decided to use 60 as the threshold.

To determine the type of attack a total of 10 coefficients were calculated. These coefficients are used as discriminating features in the linear classifier. They are:

A. Maximum difference between the Fourier spectrum of the HF watermark image and the Fourier spectrum of the host image.

B. Maximum difference between the Fourier spectrum of the LF watermark image and the Fourier spectrum of the host image.

C. PSNR value between the original watermarked image and the attacked watermarked image.

D. The percentage of the number of pixels having zero grey values in the attacked HF watermarked image.

E. The percentage of the number of pixels having zero grey values in the attacked LF watermarked image.

F. Percentage of dissimilar pixels between the original watermarked image and the tested watermarked image.

G. Min Fourier spectrum of the extracted HF watermark image: apply the Fourier spectrum to the extracted attacked HF watermark image, and then find out the minimum pixel value.

H. Max Fourier spectrum of the extracted HF watermark image: apply the Fourier spectrum to the extracted attacked HF watermark image, and then find out the maximum pixel value.

I. Max histogram of extracted HF watermark image: apply the histogram to the extracted attacked HF watermark image, and then find out the highest count in the histogram.

J. Max histogram of the extracted LF watermark image: apply the histogram to the extracted attacked LF watermark image, and then find out the highest count in the histogram.

Table 6.3 Criteria of Detection Coefficient

Watermark Attack	Coefficient	Explanations of coefficients selection
Gaussian smoothing	(I, J)	Histogram and extracted HF & LF watermark images are selected to affect to Gaussian smoothing attack
Sharpening	(G, H)	Fourier spectrum and extracted HF watermark image is selected to affect to Gaussian smoothing attack
Gaussian noise	C	PSNR value is selected to affect to Gaussian noise attack
Salt and pepper noise	F	Percentage of dissimilar pixels is selected to affect to salt and pepper noise
JPEG compression	(A, B)	Fourier spectrum, HF & LF watermark image and host image is selected to affect to JPEG compression attack
Histogram equalisation	(D, E)	The percentage of number of pixels is selected to affect to Histogram equalisation attack

Different watermark attacks have different coefficients to detect. Some of the attacks only require one coefficient, which include Gaussian noise and salt and pepper noise,

while the remainder require two factors. The criteria for detection are listed in Table 6.3.

6.3.2 Linear Classifier

Linear classification belongs to the field of statistical classification, and its goal is to use an object's characteristics to identify which class or group it belongs to. A linear classifier achieves this by making a classification decision based on the value of a linear combination of these characteristics. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. For example, a data point is viewed as a p-dimensional vector, and we want to know whether we can separate such points with a (p-1) dimensional hyperplane. There are many hyperplanes that might classify the data. One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximised. If such a hyperplane exists, it is known as the 'maximum-margin hyperplane' and the linear classifier it defines is known as a 'maximum margin classifier'.

Figure 6.7 indicates an example of linear classification. The solid circle and hollow circle represent two classes. H1, H2 and H3 represent three linear functions applied for separating the two classes. In addition, we found H3 does not separate the two classes. H1 does, by a small margin, and H2 does with maximum margin.

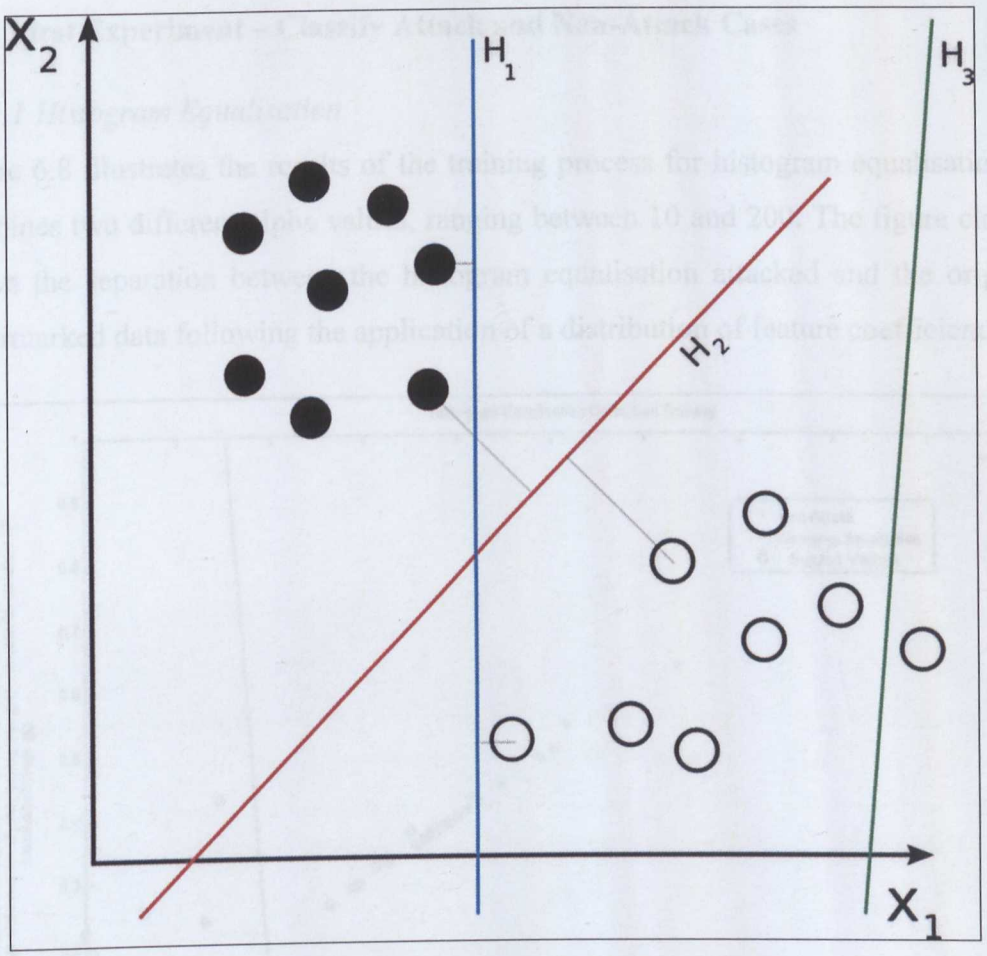


Figure 6.7 Examples of linear classification

6.4. Experiment and Results

To determine the performance of the proposed watermark attack detection scheme we conducted an experiment consisting of the training and testing of linear classifiers. In total, 100 images were used, 50 of which were the training images used to construct the linear classifiers. The resulting classifiers were then applied to the remaining 50 images to calculate how well they perform in correctly detecting watermark attacks.

Each experiment used two different attack parameters for each attack type. These parameters, denoted as alpha, were used to set the watermark attack strength. Implementation-wise, these parameters were coefficients that were used in the respective MATLAB function to apply the watermark attack. Furthermore, in our experiments, we conducted two types of experiments; the classified attack and non-attack cases and the other experiment classified one type of attack from other types of attack.

6.4.1 First Experiment – Classify Attack and Non-Attack Cases

6.4.1.1 Histogram Equalisation

Figure 6.8 illustrates the results of the training process for histogram equalisation. It combines two different alpha values, ranging between 10 and 200. The figure clearly shows the separation between the histogram equalisation attacked and the original watermarked data following the application of a distribution of feature coefficients.

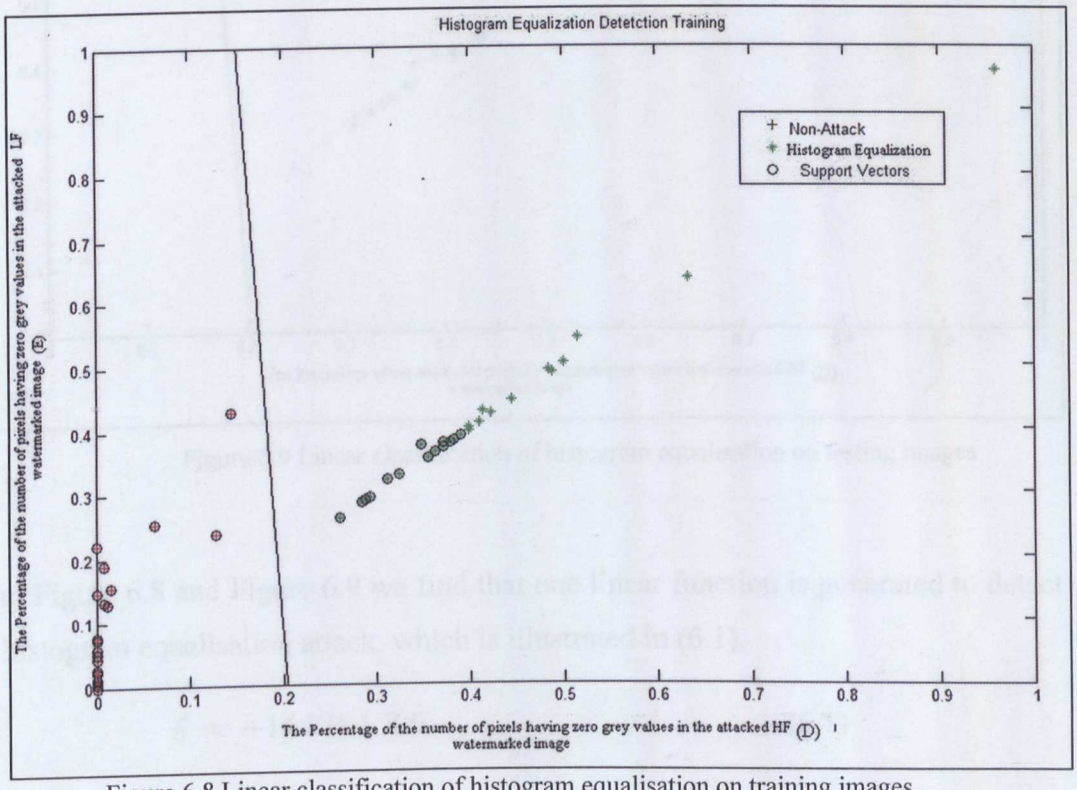


Figure 6.8 Linear classification of histogram equalisation on training images

Figure 6.9 shows the results of the testing process for histogram equalisation detection by calculating linear classifiers with the data. As also clearly indicated in the figure, the linear classifiers successfully separate the attacked and non-attacked cases using the test images. The results indicate that the correction ratio is 100%.

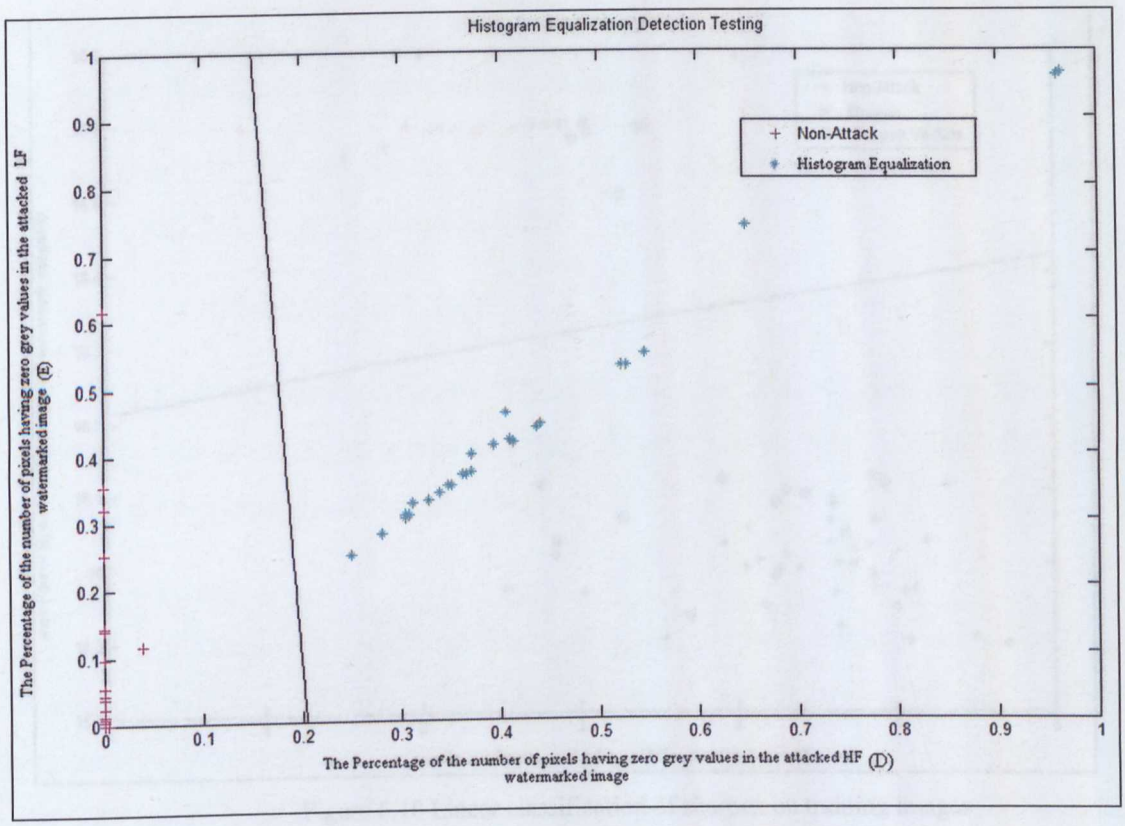


Figure 6.9 Linear classification of histogram equalisation on testing images

From Figure 6.8 and Figure 6.9 we find that one linear function is generated to detect the histogram equalisation attack, which is illustrated in (6.1).

$$E = -16.7D + 3.5 \qquad (6.1)$$

For a point (D, E) in the histogram equalisation detection scheme, if $D > (-E + 3.5)/16.7$, it represents that the histogram equalisation attack has been detected; otherwise, if $D < (-E + 3.5)/16.7$, it represents a non-attack detected.

6.4.1.2 Sharpen

The results of the training process for sharpen detection are illustrated in Figure 6.10, which shows the separation between a sharpen attack and the original watermarked image through the application of a distribution of feature coefficients. It can be seen that data is separated successfully by applying this algorithm, which combines two different alpha values, ranging between 0.05 and 1.0.

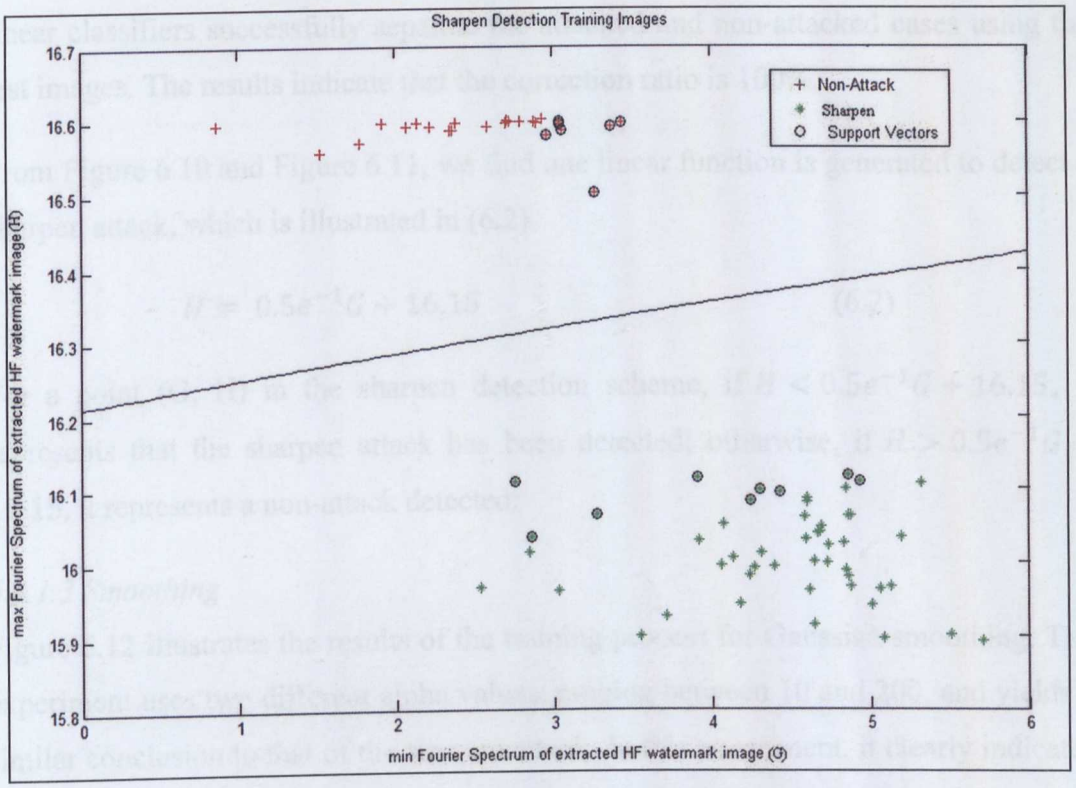


Figure 6.10 Linear classification of sharpen on training images

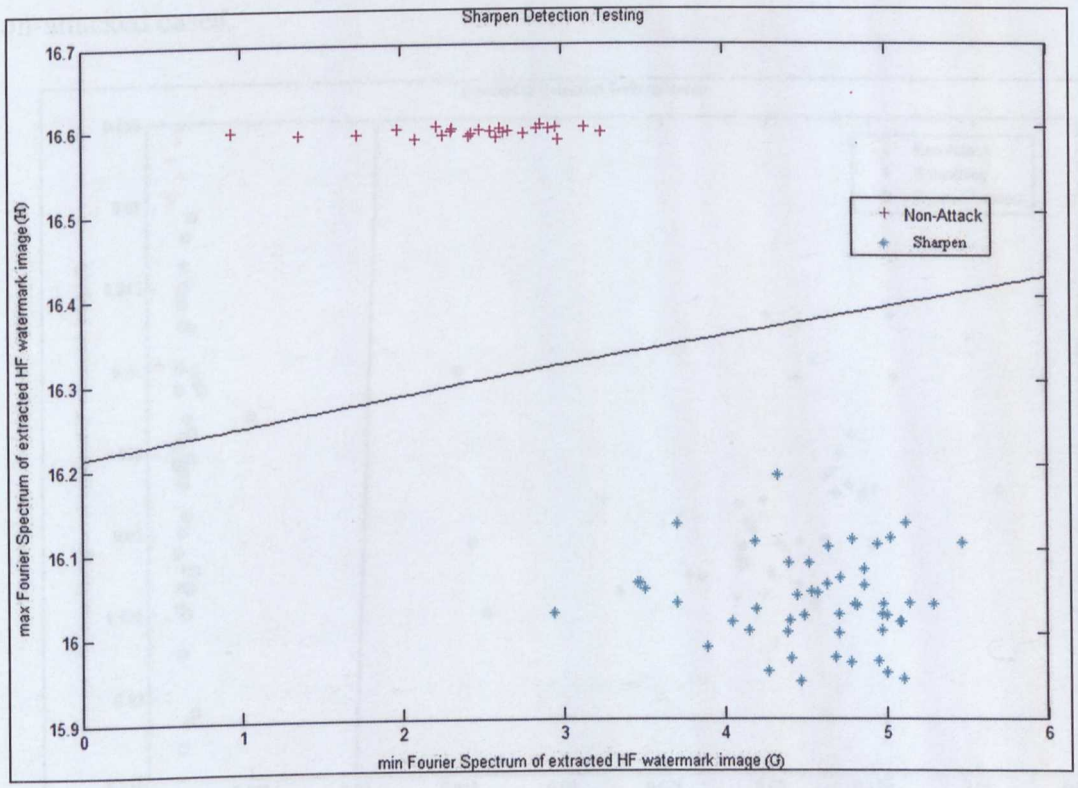


Figure 6.11 Linear classification of sharpen on testing images

Figure 6.11 shows the results of the testing process for sharpen detection by calculating linear classifiers with the data. As also clearly indicated in the figure, the

linear classifiers successfully separate the attacked and non-attacked cases using the test images. The results indicate that the correction ratio is 100%.

From Figure 6.10 and Figure 6.11, we find one linear function is generated to detect a sharpen attack, which is illustrated in (6.2).

$$H = 0.5e^{-1}G + 16.15 \qquad (6.2)$$

For a point (G, H) in the sharpen detection scheme, if $H < 0.5e^{-1}G + 16.15$, it represents that the sharpen attack has been detected; otherwise, if $H > 0.5e^{-1}G + 16.15$, it represents a non-attack detected.

6.4.1.3 Smoothing

Figure 6.12 illustrates the results of the training process for Gaussian smoothing. The experiment uses two different alpha values, ranging between 10 and 200, and yields a similar conclusion to that of the sharpen attack. In this experiment, it clearly indicates a clear separation in the distribution of the feature coefficients of the attacked and non-attacked cases.

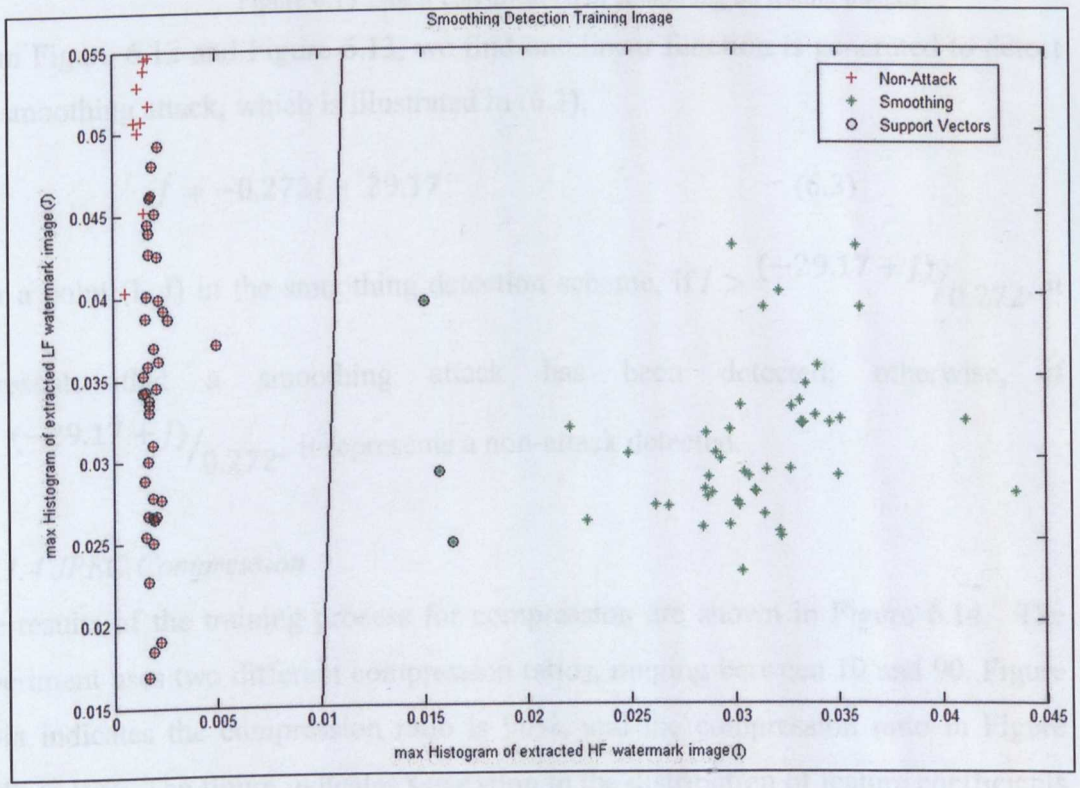


Figure 6.12 Linear classification of smoothing on training images

Figure 6.13 shows the results of the testing process for smoothing detection by calculating linear classifiers with the data. As also clearly indicated in the figure, the

linear classifiers successfully separate the attacked and non-attacked cases using the test images. The results indicate that the correction ratio is 100%.

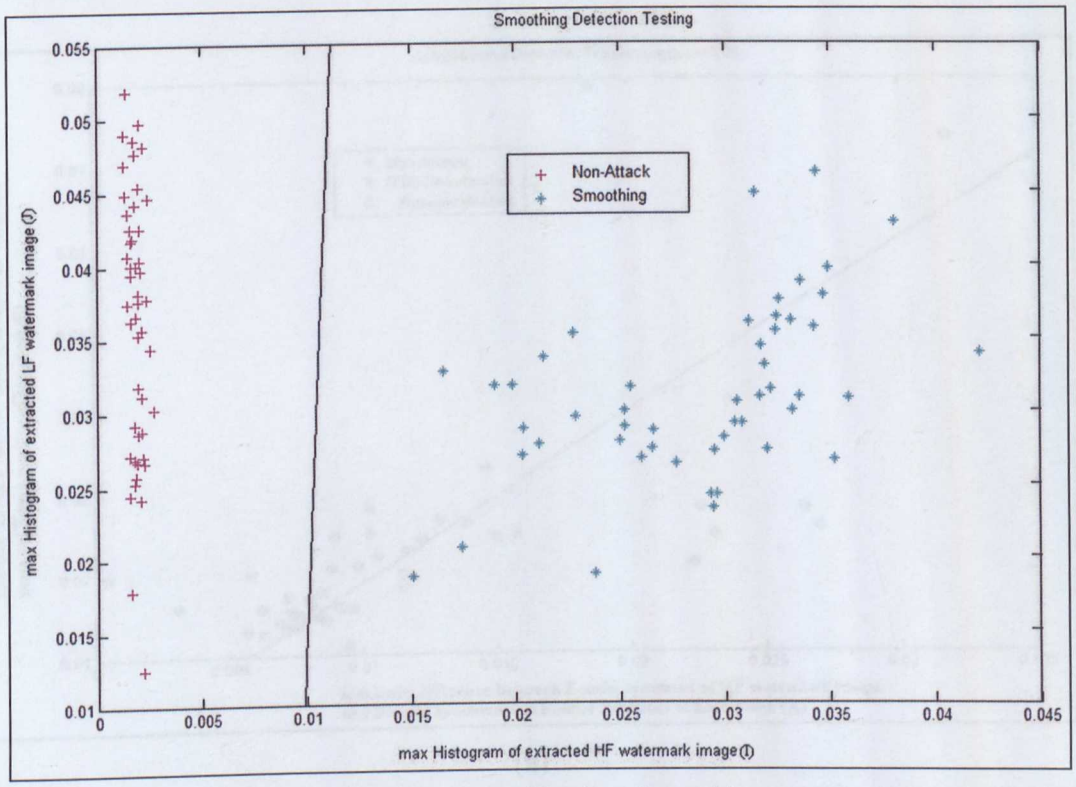


Figure 6.13 Linear classification of smoothing on testing images

From Figure 6.12 and Figure 6.13, we find one linear function is generated to detect the smoothing attack, which is illustrated in (6.3).

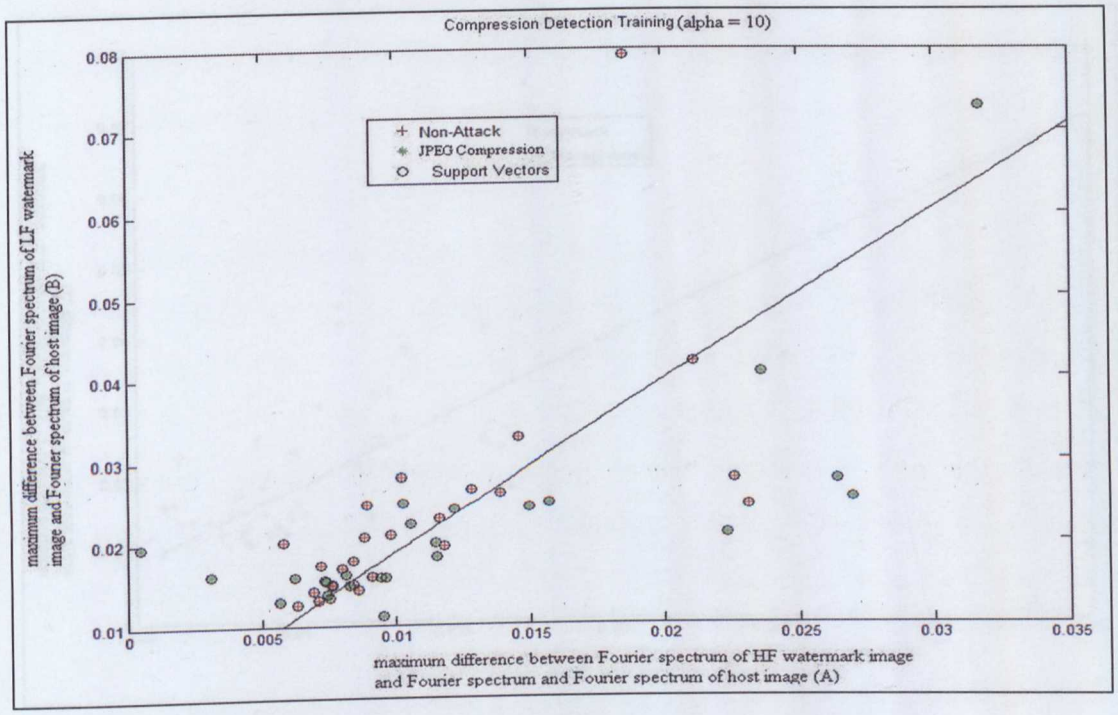
$$J = -0.272I + 29.17 \qquad (6.3)$$

For a point (I, J) in the smoothing detection scheme, if $I > {(-29.17 + J)}/{0.272}$, it represents that a smoothing attack has been detected; otherwise, if $I < {(-29.17 + J)}/{0.272}$, it represents a non-attack detected.

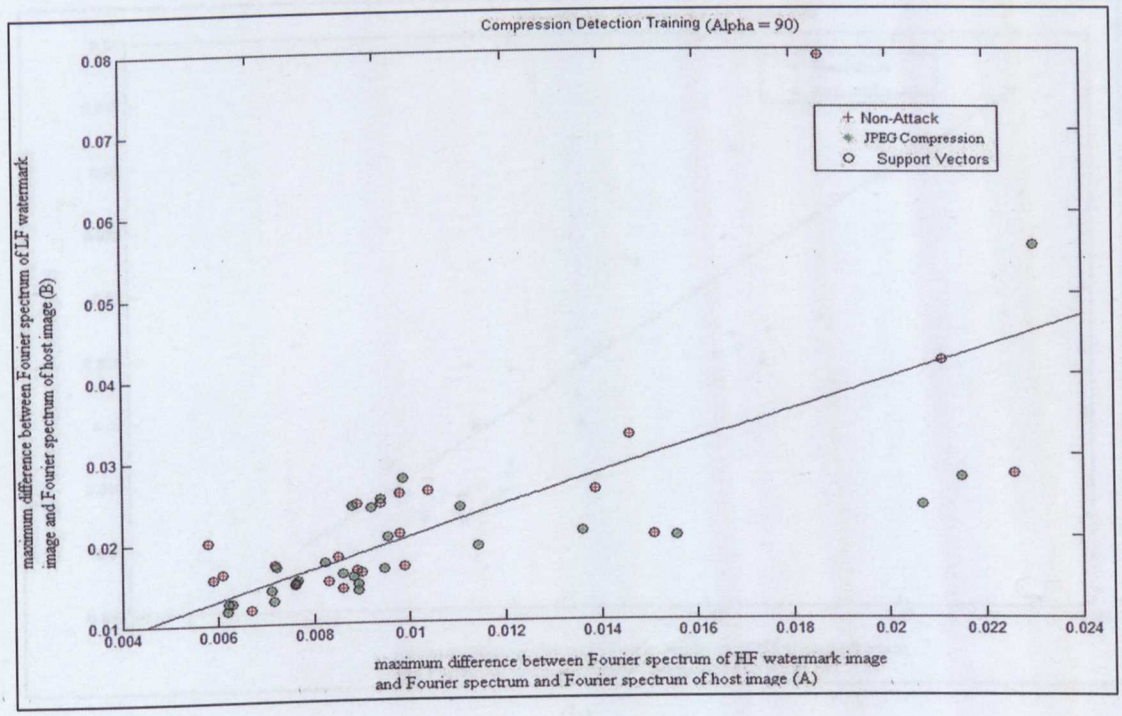
6.4.1.4 JPEG Compression

The results of the training process for compression are shown in Figure 6.14. The experiment uses two different compression ratios, ranging between 10 and 90. Figure 6.14a indicates the compression ratio is 90%, and the compression ratio in Figure 6.14b is 10%. The figure indicates separation in the distribution of feature coefficients of the JPEG compression attacked and non-JPEG compression attacked cases. It can

be seen that most of the data is separated successfully, and only a few of them are misclassified.



(a)

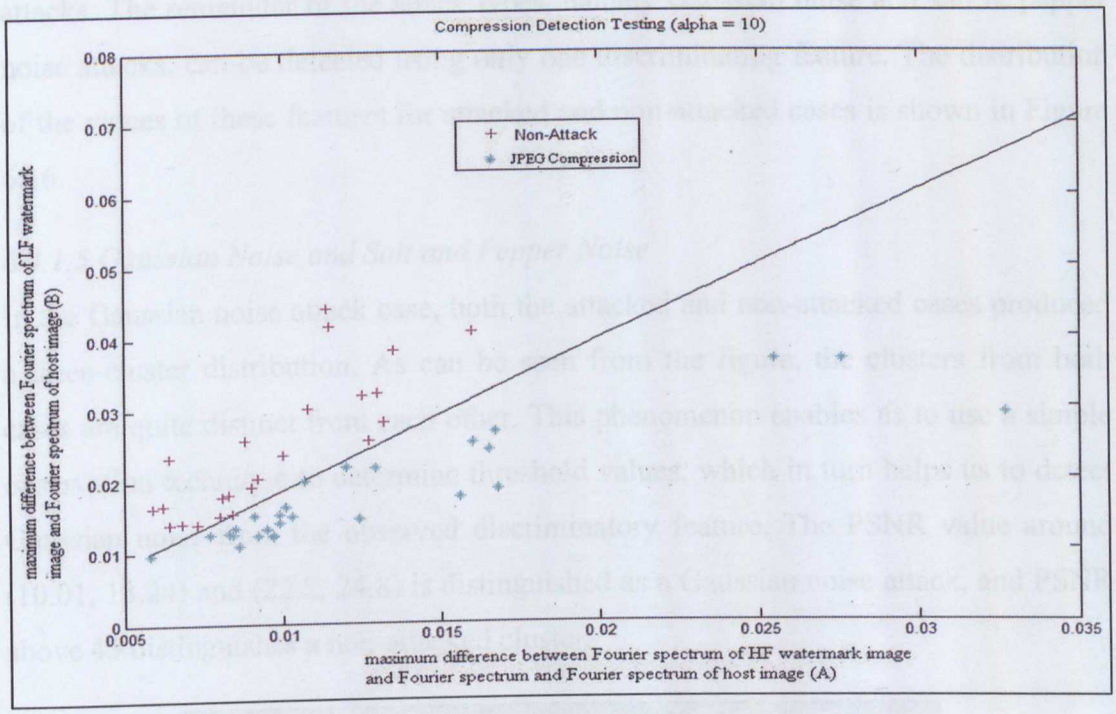


(b)

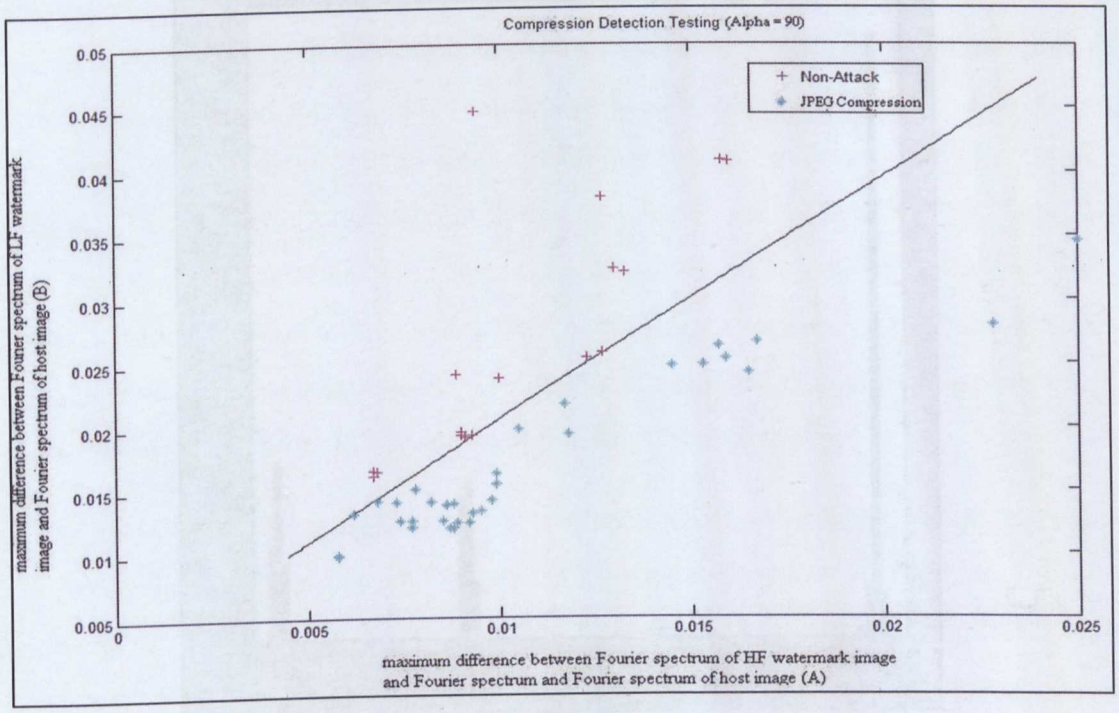
Figure 6.14 Linear classification of JPEG compression on training images

Furthermore, to show the results of the linear classifier in compression detection, we plot the calculated linear classifier with the data from the test images, the results of which are shown in Figure 6.15. As also clearly indicated in the figure, the linear

classifiers successfully separate most of the attacked and non-attacked cases using the test images.



(a)



(b)

Figure 6.15 Linear classification of jpeg compression on testing images

The results indicate that the overall correction ratio is 70%. It also shows that the attack detection scheme is not suitable for JPEG compression.

So far, we have shown the effectiveness of the watermark attack detection technique in detecting histogram equalisation, sharpen, smoothing and JPEG compression attacks. The remainder of the attack types, namely Gaussian noise and salt & pepper noise attacks, can be detected using only one discriminating feature. The distribution of the values of these features for attacked and non-attacked cases is shown in Figure 6.16.

6.4.1.5 Gaussian Noise and Salt and Pepper Noise

In the Gaussian noise attack case, both the attacked and non-attacked cases produced a three-cluster distribution. As can be seen from the figure, the clusters from both cases are quite distinct from each other. This phenomenon enables us to use a simple observation technique to determine threshold values, which in turn helps us to detect Gaussian noise from the observed discriminatory feature. The PSNR value around (10.01, 13.24) and (22.2, 24.8) is distinguished as a Gaussian noise attack, and PSNR above 45 distinguishes a non-attacked cluster.

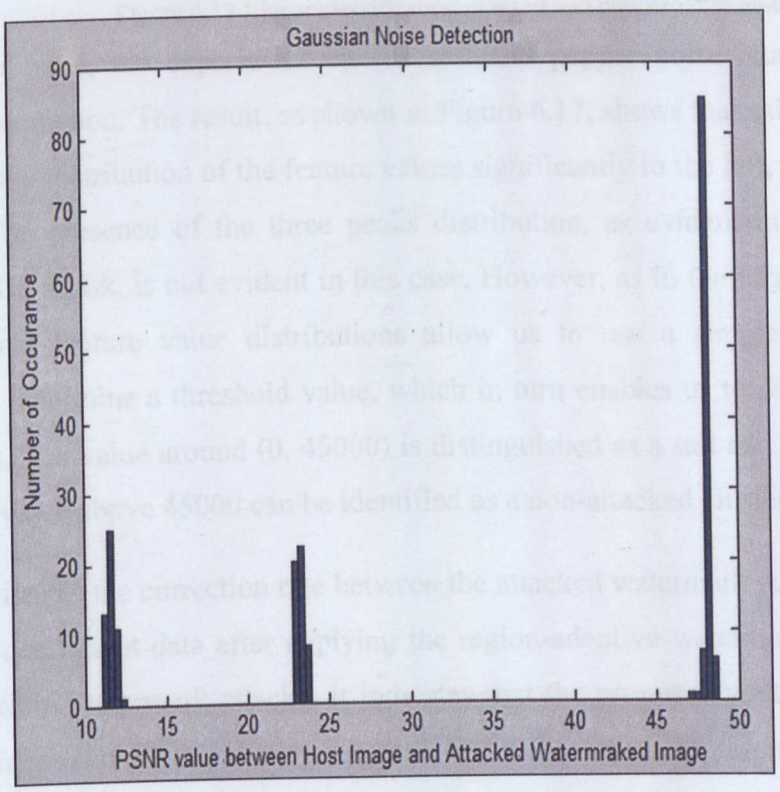


Figure 6.16 Linear classification of a Gaussian noise attack

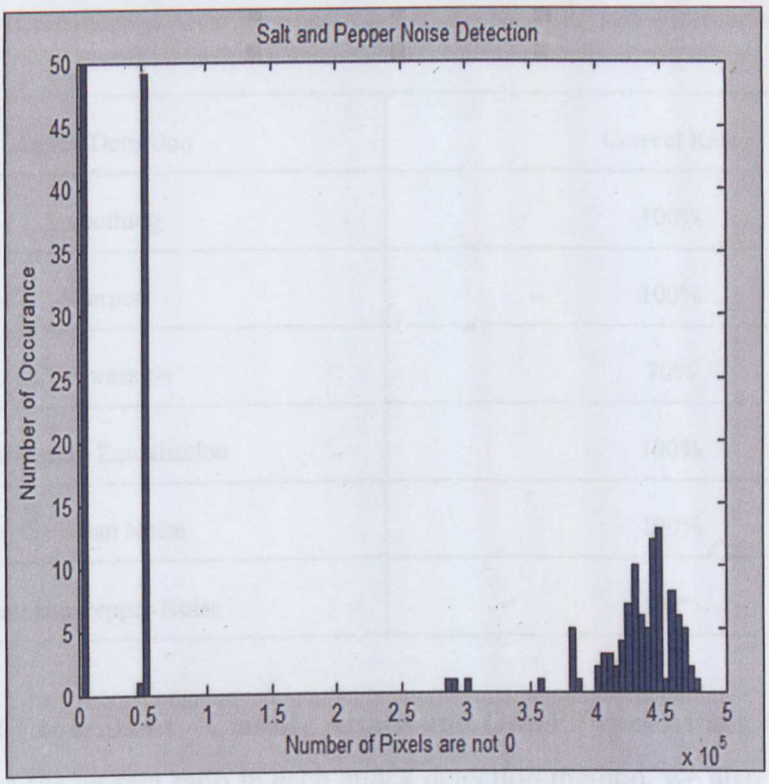


Figure 6.17 Linear classification of a salt and pepper noise attack

On the other hand, the experiment on the salt and pepper noise attack shows a different phenomenon. The result, as shown in Figure 6.17, shows that salt and pepper attacks shift the distribution of the feature values significantly to the left, i.e. reducing its values. The presence of the three peaks distribution, as evident in the case of Gaussian noise attack, is not evident in this case. However, as in the former case, the clearly distinct feature value distributions allow us to use a simple observation technique to determine a threshold value, which in turn enables us to detect salt and pepper noise. The value around (0, 45000) is distinguished as a salt and pepper noise attack and a value above 45000 can be identified as a non-attacked cluster.

Table 6.4 indicates the correction rate between the attacked watermark coefficient and the original coefficient data after applying the region-adaptive watermarking system to detect various watermark attacks. It indicates that the proposed watermark attack detection scheme can reliably distinguish most attacks, except compression.

Table 6.4 The Classification Accuracy when Attacked and Non-Attacked Watermarked Images are Used

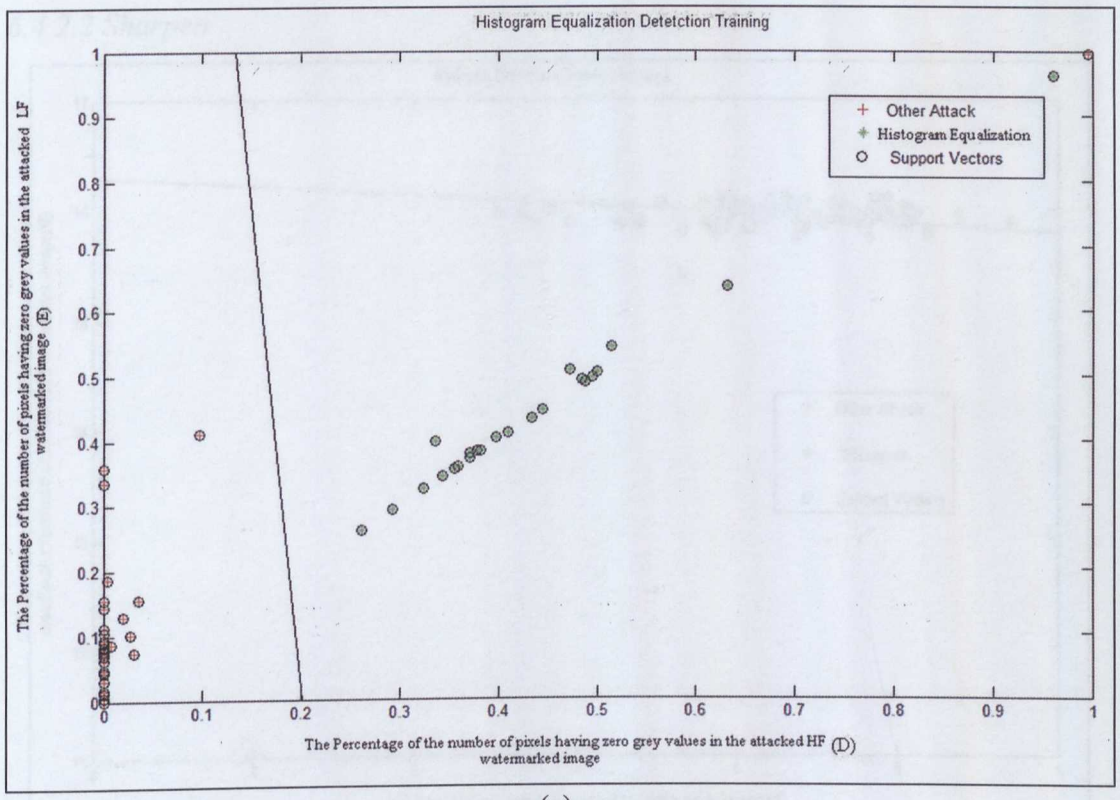
Attack Detection	Correct Rate
Smoothing	100%
Sharpen	100%
Compression	70%
Histogram Equalisation	100%
Gaussian Noise	100%
Salt and Pepper Noise	100%

6.4.2 Second Experiment – Classify Attack and Other Types Attack Cases

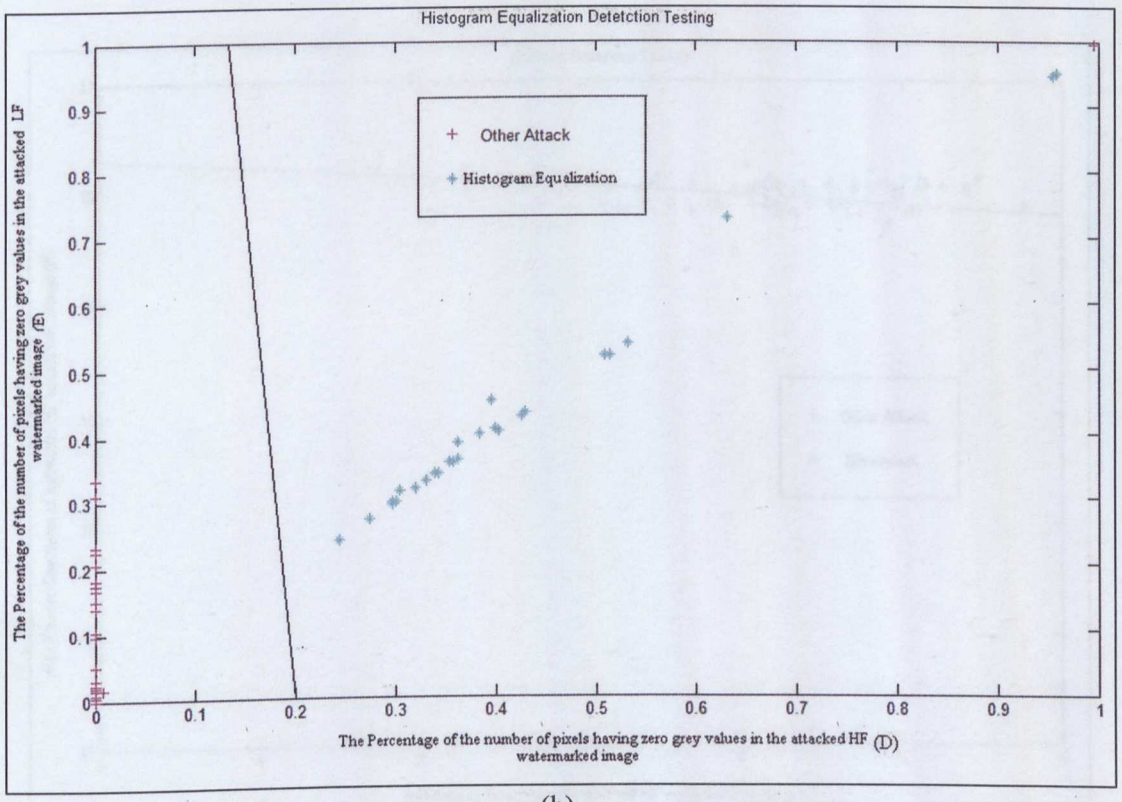
To further test the correct ratio in each attack detection method, we also need to apply the same coefficient to different attack to get different results. Figures 6.18 to 6.22 plot the results.

6.4.2.1 Histogram Equalisation

Figure 6.18 indicates the results of histogram equalisation. Figure 6.18a shows the results of the training process for histogram equalisation detection, and Figure 6.18b shows the results of test processing. It indicates that the data for histogram equalisation is clearly identified, except point (0.9961, 0.9961) which is located in the top-right corner of the figures.



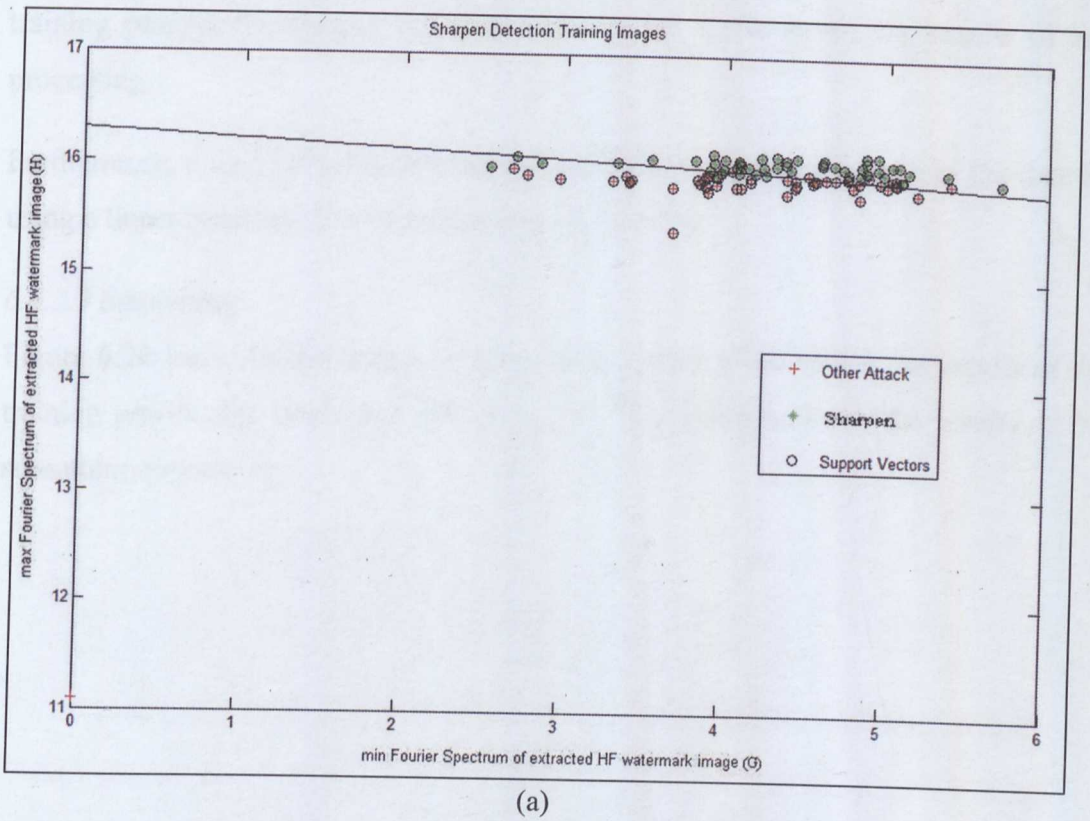
(a)



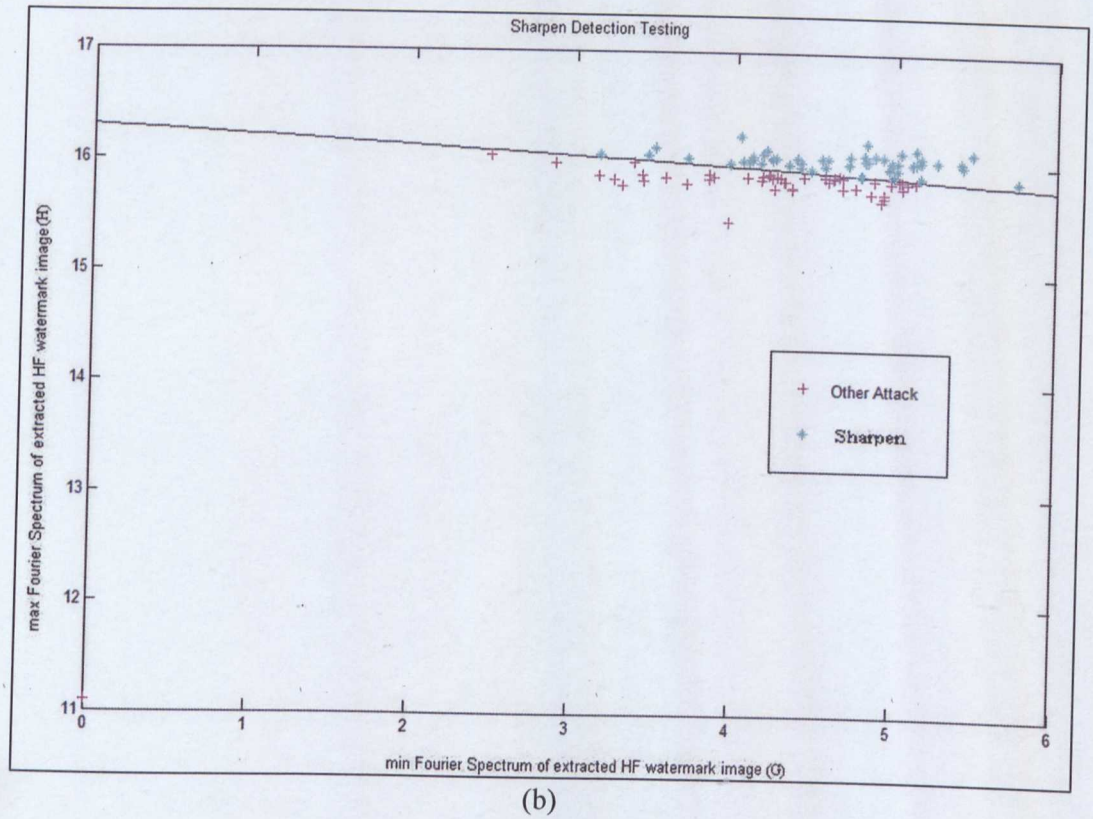
(b)

Figure 6.18 Linear classification of histogram equalisation on (a)Training images (b) Testing images between original results and other attack result in same coefficients

6.4.2.2 Sharpen



(a)



(b)

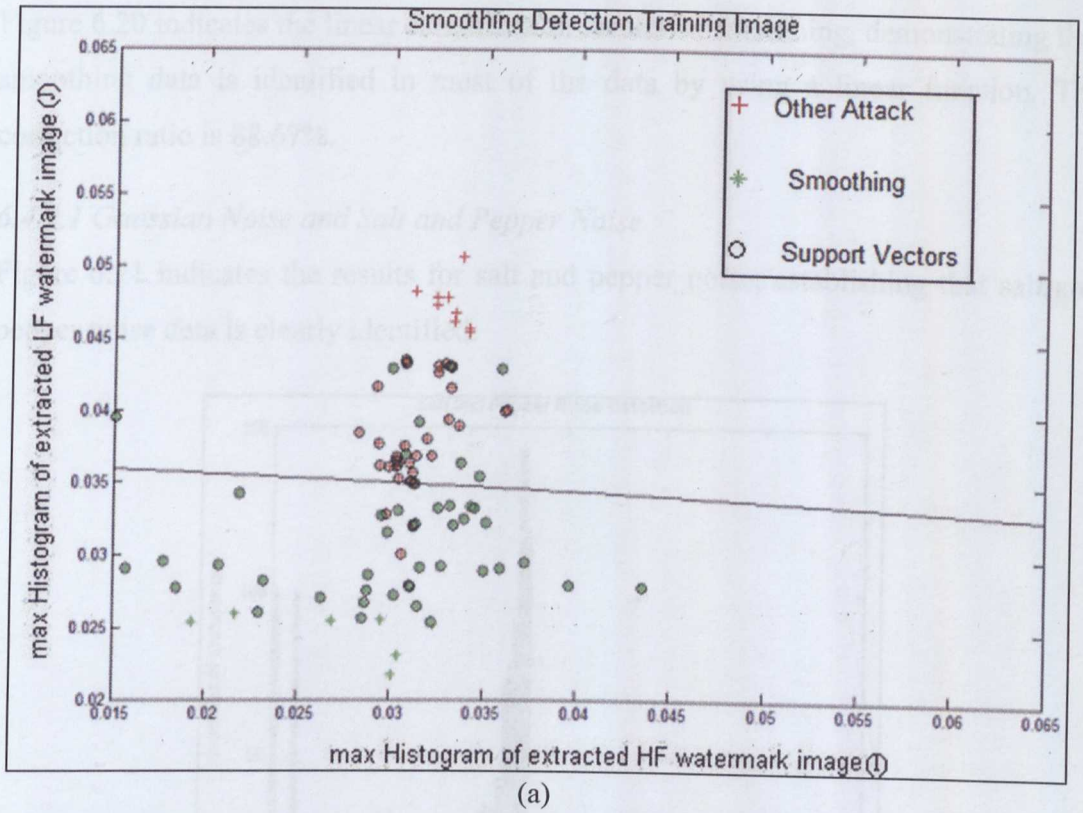
Figure 6.19 Linear classification sharpen on (a)Training images (b) Testing images between original results and other attack result in same coefficients

Figure 6.19 indicates the results in sharpen. Figure 6.19a shows the results of the training process for sharpen detection, and Figure 6.19b shows the results of test processing.

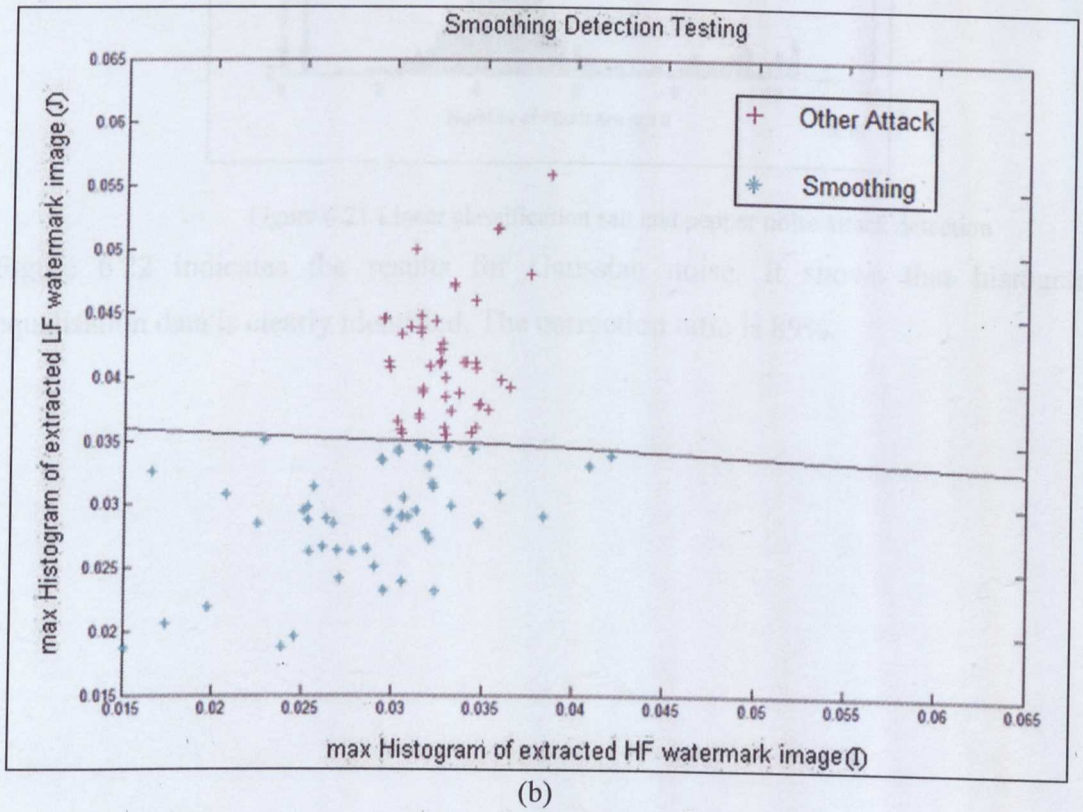
Furthermore, it also indicates that the sharpen data is identified in most of the data by using a linear function. The correction ratio is 94.67%.

6.4.2.3 Smoothing

Figure 6.20 indicates the results in smoothing. Figure 6.20a shows the results of the training process for smoothing detection, and Figure 6.20b shows the results of the smoothing processing.



(a)



(b)

Figure 6.20 Linear classification smoothing on (a)Training images (b) Testing images between original results and other attack result in same coefficients

Figure 6.20 indicates the linear classification results on smoothing, demonstrating that smoothing data is identified in most of the data by using a linear function. The correction ratio is 88.67%.

6.4.2.1 Gaussian Noise and Salt and Pepper Noise

Figure 6.21 indicates the results for salt and pepper noise, establishing that salt and pepper noise data is clearly identified.

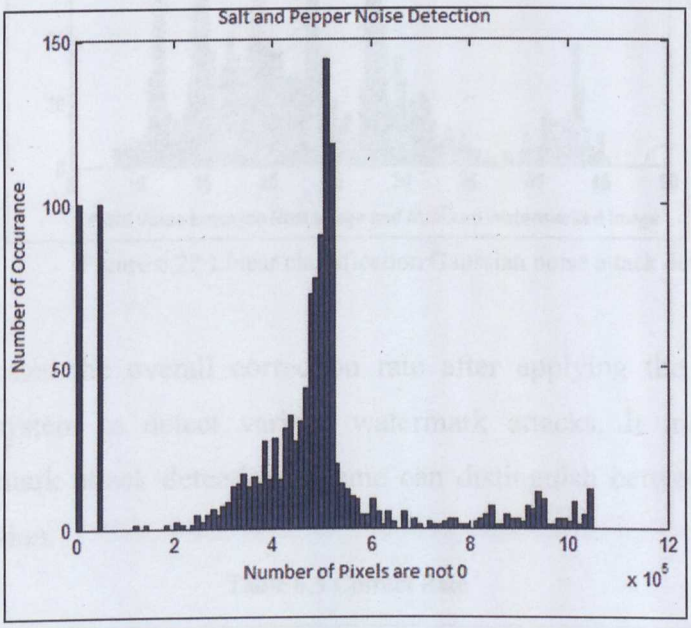


Figure 6.21 Linear classification salt and pepper noise attack detection

Figure 6.22 indicates the results for Gaussian noise. It shows that histogram equalisation data is clearly identified. The correction ratio is 89%.

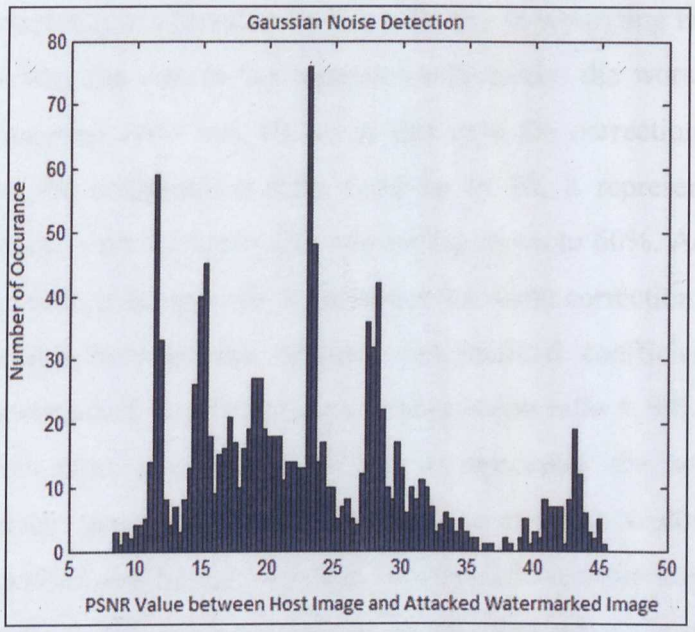


Figure 6.22 Linear classification Gaussian noise attack detection

Table 6.5 indicates the overall correction rate after applying the region-adaptive watermarking system to detect various watermark attacks. It indicates that the proposed watermark attack detection scheme can distinguish between most attacks, except compression.

Table 6.5 Correct Rate

ATTACK DETECTION	CLASSIFICATION ACCURACY WHEN ATTACKED AND NON-ATTACKED WATERMARKED IMAGES ARE USED	CLASSIFICATION ACCURACY WHEN ATTACKED AND OTHER TYPES OF ATTACKED WATERMARKED IMAGES ARE USED	OVERALL CORRECT
Smoothing	100%	88.67%	88.67%
Sharpen	100%	94.67%	94.67%
Histogram Equalisation	100%	100%	100%
Gaussian Noise	100%	89%	89%
Salt an Pepper Noise	100%	100%	100%
Overall			94.5%

As we can see from the experiment results and Table 6.5, we find that most attacks can be successfully detected, apart from when JPEG compression attack is used, in

which case the technique is found to be less effective in separating the data. There are several reasons why the data is not separated effectively: the worst attack occurred when the compression ratio was 10, so in this case the correction ratio was 80%. However, when the compression ratio went up to 90, it represented the weakest compression attack, with the correction rate falling down to 60%. Accordingly, when the compression ratio goes up to 90, it generates the worst correction rate because it is hard to distinguish between the original watermarked coefficient data and the compressed watermarked coefficient data (compression ratio = 90). However, when the compression ratio goes down to 10, it represents the worst compressed watermarked image generated, in other words it is easier to separate data between original watermarked coefficient data and compressed watermarked coefficient data (compression ratio = 10). Furthermore, we assume that when the compression ratio goes down, the correct ratio goes up.

6.5. Discussion

The advantages and disadvantages of using a region-adaptive DWT-SVD watermarking system for watermark attack detection will be discussed in this subsection.

6.5.1 Advantages

There are several advantages of using a watermarking attack detection system, including its unique detection method, correction ratio and computation cost.

- Unique detection method: In image tamper detection algorithms, most watermarking algorithms belong to fragile watermark systems, but this application uses a robust watermark system. Furthermore, most watermark algorithms can only detect one tamper[Al-Qershi, 2010, Amira, 2010] ; however, this algorithm can detect five different watermark attacks include Gaussian noise attack, salt and pepper noise attack, Gaussian smoothing attack, sharpen attack and histogram equalisation attack.

- Correction ratio: The overall correction ratio is 94.5% in total, except when detecting compression attack. Moreover, the correction ratio for histogram equalisation and salt and pepper noise attack is 100%.

- In this technique, the discriminating features have been chosen such that linear

classifier can be used to detect the attacks to a very high detection success rate without having to use higher order classifier. Therefore this technique has a low computational and algorithmic complexity [Bernhard, 1998] .

6.5.2 Disadvantages

There are also various disadvantages, which are listed below:

- Non-district detection of watermark attacks: This is a significant disadvantage, as it applies a watermark attack detection algorithm which can only detect attacks on the whole image but cannot detect tampering on a district of the digital image. For an example, see Figures 6.3, 6.4 and 6.5.
- Non-detection of geometric attack: The algorithm can detect removal attack only, but cannot detect a geometric attack such as rotation, scaling, etc.

6.6. Chapter Summary

Robust and fragile watermarks are suited to different purposes, given their contrasting characteristics. Robust watermarks are good for copyright protection due to their robustness against image distortions. On the other hand, fragile watermarks are sensitive to changes in image. However, in this chapter, we presented a novel application of robust watermarks which can detect changes in an image.

In this algorithm, we show a novel use for the region-adaptive watermarking technique, which is described in Chapter 4 and 5, as a means to detect whether certain types of attacks have occurred. This is a unique feature of our watermarking algorithm, which separates it from other state-of-the-art watermarking techniques. The watermark detection process uses coefficients derived from the region-adaptive watermarking algorithm in a linear classifier. The experiment conducted to validate this feature shows that, on average, 94.5% of all watermark attacks can be correctly detected and identified.

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

7.1. Thesis Summary

Digital image watermarking technology has been drawing the attention of researchers and practitioners as a new method of protecting copyright for digital images. It is realised by embedding data that is invisible to the human visual system, and is called watermark. So watermarking in digital images is the process by which a discrete data stream is hidden within an image, imposing imperceptible changes on the image.

A watermark attack is one of the major problems in a digital image watermarking system. They are often deliberately applied to a watermarked image in order to remove or destroy any watermark signals in the host data. The purpose of the attack is aimed at disabling the copyright protection system offered by watermarking technology.

The aims of the research in this thesis was to advance state-of-the-art digital image watermarking technology by means of improving its transparency and robustness to different watermark attacks, and at the same time provide a novel application of the solution in watermark attack detection.

In this thesis, we discussed the analysis and findings resulting from research into this subject. These research contributions can be summarised as:

1. Analysis and comparisons of the effects of different watermark attacks.
2. Theory on robust watermarking techniques using a region-adaptive approach.
3. A geometrically robust region-adaptive watermarking system with a DWT-SVD algorithm.
4. The application of the region-adaptive technique for watermark attack detection and identification.

7.1.1 Analysis and comparisons of the effects of different watermark attacks

The success of a watermarking technology used in copyright protection or digital rights management system relies heavily on its robustness against watermark attacks, which are aimed at removing or destroying any watermark signals in the host data. It is therefore important to understand how these attacks work, in order to design a good and robust watermarking technique.

In this contribution, we analysed the effect of six different watermark attacks. The analysis was carried out using two image analysis tools, namely an image histogram and Fourier transforms. The histogram of an image refers to the distribution of the pixel intensity values. This histogram is a graph showing the number of pixels in an image at each different intensity value found in that image. A Fourier transform convert image data into its frequency spectrum. As with many other signal transforms, the Fourier transform converts time or spatial domain signals into alternative representations that are more amenable for certain types of analysis. The low frequency is located at the centre of the image and high frequency around the edge.

In the experiment, our analysis indicates that watermark attacks can be divided into two general categories: high frequency and low frequency watermark attacks. The former category attacks the high frequency part of the signal while keeping most of the low frequency component relatively unchanged, while the latter category attacks vice versa. This finding suggests that the spectral distribution of the watermarked image plays an important role in the overall robustness of the watermark algorithm used for a particular attack. Based upon this finding, we hypothesised that in order to maximise the robustness of the watermarked data, the spectral distribution of the host

data and the watermark data should be similar. Furthermore, in order to counter both types of attacks, the watermark data should possess both strong high and low frequency parts.

7.1.2 Theory on robust watermarking techniques using a region-adaptive approach

This contribution comes from the findings from analysing different watermark attacks. It was conducted to analyse the effects of different attacks on watermarked data, and it provided us with the conclusion that each attack affects high and low frequency parts of the watermarked image spectrum differently. Furthermore, the findings also showed that the effect of the attack can be alleviated by using a watermark image with a similar frequency spectrum to that of the host image. The results of this experiment led to a hypothesis which would be proven by applying a watermark embedding technique which takes into account all of the above phenomena.

To test this hypothesis we developed a region-adaptive watermark technique. This technique employs two unique solutions. Firstly, it uses two watermark images, each with strong high frequency or low frequency components, and secondly it embeds different parts of the watermark images into the host image based on the differences between their spectral distributions. Furthermore, the DWT embedding and extraction processes are applied only to the LL sub-band.

The experiment results have provided sufficient evidence to confirm the initial hypothesis and prove our theory that a region-adaptive approach can increase the robustness of a watermarking technique against removal attacks.

7.1.3 A geometrically robust region-adaptive watermarking system with a DWT-SVD algorithm

The design of the region-adaptive watermarking technique developed previously was improved by incorporating SVD, in order to counter geometric attack including translation, rotation, scaling, etc. In addition, we investigated the difference between variances in performance when the watermark is inserted into the LL and LH sub-bands.

Our experiment showed that the region-adaptive DWT-SVD watermarking technique improves the robustness of the watermark data to image processing attacks as well as

geometric attacks, regardless of whether it is inserted in either the LL or LH sub-band of the DWT coefficients. However, we also noticed that the results are slightly better when the watermark is inserted into the LH sub-band as compared to the LL sub-band.

7.1.4 The application of the region-adaptive technique for watermark attack detection and identification

Traditionally, tamper detection could only be achieved using a fragile watermarking technique, so robust watermarking techniques were not used to detect image tampering. Therefore, in this contribution, we looked at a novel application of the region-adaptive watermarking algorithm to detect tampering with digital images.

The process works by using coefficients calculated during the watermarking process to detect and identify the presence of attacks in a watermark image. The detection and identification process can be carried out using a linear classifier technique while avoiding using more complex non-liner ones. This is made possible due to the choice of coefficients in the classifiers, which are derived from the watermarking process.

The experiment conducted to validate this feature showed that, on average, 94.5% of all watermark attacks can be correctly detected and identified.

7.2. Future Research

In this section, we give some possible research directions.

The region-adaptive digital image watermarking algorithm could apply directly to digital cameras. The watermarked image would be generated automatically after pressing the flash button. This strategy would not only protect the copyright, but also could detect watermark attacks if they occur. However, one aspect needs to be improved – updating the greyscale host image to a coloured host image – because most output camera images are coloured.

The region-adaptive digital image watermarking system developed in this thesis could be extended to biomedical image analysis. For biomedical images, regions that contain disease symptoms are of major concern to physicians and patients [Farooq, , Jayanthi, 2009] . For example, in retinal images the image viewer usually cares more about vasculature and lesions than the background. In biomedical images, the portions that attract more attention from image views are called the 'regions of interest' (ROI).

Furthermore, the finding in my thesis showed that the effect of an attack can be alleviated by using a watermark image with a similar frequency spectrum to that of the host image. Therefore, it is desirable to embed different watermarks in ROI to give them better protection.

The science of modelling 3D watermarking algorithms [Shrivastava, 2011, Yan-hong, 2010] has developed rapidly in recent years. One of the characteristics of 3D watermarking technology has to take in to account the fact that a model can be represented in different ways. For example, a collection of parametric curves (e.g. NURBS) can be represented by a set of parameters or by polygonal meshes. Furthermore, 3D watermark attacks such as *mesh smoothing, polygonal simplification and re-meshing* are more complex than 2D watermark attacks. In addition, we could attempt to use the region-adaptive watermark algorithm in 3D models. For example, a 3D mesh could partition into different sub-meshes based on criteria such as smoothing area on 3D models, and then the watermark embedding procedure could be applied to each sub-mesh.

List of Publications

1. C.Song, S. Sudirman and M.Merabti, "A Robust Region-Adaptive Dual Image Watermarking Technique", *Journal of Visual Communication and Image Representation* Vol.23, 549-568, April, 2012
2. C.Song, S. Sudirman and M.Merabti, "Region-Adaptive Watermarking System and its Application", *IEEE Conference in Developments in e-Systems Engineering,* Dec, 2011
3. C.Song, S. Sudirman and M.Merabti, "Region Adaptive Digital Image Watermarking System Using DWT-SVD Algorithm", *Proc 12th of PostGraduate Network Symposium*, June, 2011.
4. C.Song, S.Sudirman and M.Merabti, "Robust Digital Image Watermarking using Region Adaptive Embedding Technique", *IEEE conference on Progress in Informatics and Computing*, Dec, 2010.
5. C.Song, S.Sudirman and M.Merabti, "A Spatial and Frequency Domain Analysis of the Effect of Removal Attacks on Digital Image Watermarks", *Proc 11th of PostGraduate Network Symposium*, 119-124, June, 2010.
6. C.Song, S.Sudirman, M.Merabti and D.L.Jones, "Analysis of Digital Image Watermark Attacks", *6th IEEE International Workshop on Digital Rights Management*, 2010.
7. C.Song, S.Sudirman and M.Merabti, "Recent Advances and Classifications of Watermark techniques in Digital Images", *Proc 10th of PostGraduate Network Symposium*, 283-288, 2009.

REFERENCE

[Al-Qershi, 2010] O. M. Al-Qershi and K. Bee Ee, "ROI-based tamper detection and recovery for medical images using reversible watermarking technique," *2010 IEEE International Conference on Information Theory and Information Security (ICITIS)*, 17-19 Dec. 2010, pp. 151-155.

[Amira, 2010] H. Amira, R. Rhouma and S. Belghith, "An eigen value based watermarking scheme for tamper detection in gray level images," *2010 7th International Multi-Conference on Systems Signals and Devices (SSD)*, 27-30 June 2010, pp. 1-5.

[Ayangar, 2010] V. R. Ayangar and S. N. Talbar, "A novel DWT-SVD based watermarking scheme," *2010 International Conference on Multimedia Computing and Information Technology (MCIT)*, 2-4 March 2010, pp. 105-108.

[Becker, 2003] E. Becker, W. Buhse, D. G¨Unnewig and N. Rump, Digtial Rights Management: Technological, Economic, Legal and Policital Aspect, ed. Berlin, Germany Springer Lecture Notes in Computer Science 2770, 2003.

[Bernhard, 1998] S. Bernhard, Lkopf, S. Alexander, M. Klaus-Robert and Ller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299-1319, 1998.

[Bhatnagar, 2008] G. Bhatnagar, B. Raman and K. Swaminathan, "DWT-SVD based Dual Watermarking Scheme," *First International Conference on Applications of Digital Information and Web Technologies*, 2008. ICADIWT 2008. pp. 526-531.

[Carson, 2002] C. Carson, S. Belongie, H. Greenspan and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 8, pp. 1026-1038, 2002.

[Chad, 1999] C. Chad, T. Magan, B. Serge, J. H. and J. M, "Blobworld: A System for Region-Based Image Indexing and Retrieval," *Visual Information and Information Systems, Lecture Notes in Computer Science*, vol. 1614, 1999.

[Cheddad, 2008] A. Cheddad, J. Condell, K. Curran and P. Mckevitt, "Combating digital document forgery using new secure information hiding algorithm," *Third International Conference on Digital Information Management, 2008*. 13-16 Nov. 2008, pp. 922-924.

[Chih-Chin Lai, 2010] Chih-Chin Lai and C.-C. Tsai, "Digital Image Watermarking Using Discrete Wavelet Transform and Singular Value Decomposition," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 11, 2010.

[Chin-Chen] C. Chin-Chen, C. Kuo-Nan and H. Ming-Huang, "A Robust Public Watermarking Scheme Based on DWT," *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (IIH-MSP), 15-17 Oct. 2010, pp. 21-26.

[CL.Song, 2009] Cl.Song, S.Sudirman and M.Merabti, "Recent Advances and Classification of Watermark Techniques in Digital Images," *Proc 10th of PostGraduate Network Symposium,* 283-288, 2009.

[Congxu, 2006] Z. Congxu, C. Zhigang and X. Xu, "Image Dual Watermark Algorithm Based on Lifting Wavelet and Liu Chaotic System," *The Sixth World Congress on Intelligent Control and Automation*, 2006, pp. 10476-10480.

[D.Lahaner, 1989] D.Lahaner, C.Moler and S.Nash, "Numerical Mehthods and Software," 1989.

[Dey, 2007] S. Dey, A. Abraham and S. Sanyal, "An LSB Data Hiding Technique Using Natural Number Decomposition," *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2007, pp. 473-476.

[Di, 2006] W. Di, "Segmentation, registration and selective watermarking of retinal images" Texas A&M University, Doctoral 2006

[Dirik, 2009] A. E. Dirik and N. Memon, "Image tamper detection based on demosaicing artifacts," 2009 16th IEEE International Conference on Image Processing (ICIP), 7-10 Nov 2009, pp. 1497-1500.

[E.Bell, 1999] A. E.Bell, "The Dynamic Digital Disk," *IEEE Spectrum*, vol. 36, no. 10, pp. 28-35, 1999.

[E.Ganic, 2004] E.Ganic and A.Eskicioglu, "Robust DWT-SVD Domain Image Watermarking: Embedding Data in All Frequencies," *MM&Sec '04 Proceedings of the 2004 workshop on Multimedia and security,* 166-174, 2004.

[Ellinas, 2007] J. N. Ellinas, "A Robust Wavelet-based Watermarking Algorithm Using Edge Detection," *Proceedings of World Academy of Science, Engineering and Technology,* vol. 25 2007.

[Emir, 2004] G. Emir and M. E. Ahmet, "Robust DWT-SVD domain image watermarking: embedding data in all frequencies," *in Proceedings of the 2004 workshop on Multimedia and security*, Magdeburg, Germany: ACM, 2004

[Erdogan, 2001] C. Erdogan and W. Deliang, "Texture Segmentation Using Gaussian-Markov Random Field and Neural Oscillator Networks," *IEEE Transactions on Neural Networks*, vol. 12, 394-404, 2001.

[Eugene, 2007] L. Eugene, H. Ching-Tang and H. Kuo-Ming, "Blind watermarking technique based on integer discrete cosine transform and AC prediction," *in Proceedings of the 8th Conference on 8th WSEAS International Conference on Automation and Information* - Volume 8, Vancouver, British Columbia, Canada: World Scientific and Engineering Academy and Society (WSEAS), 2007.

[Fabien A. P. Petitcolas, 1999] Fabien A. P. Petitcolas, Ross J. Anderson and M. G. Kuh, "Information Hiding A Survey," *Proceedings of the IEEE, special issue on protection of multimedia content,*, vol. 87, no. 7, pp. 1062-1078, 1999.

[Farooq] O. Farooq, A. L. Vyas, S. Datta and D. Mulvaney, "Watermarking biomedical signal for authentication using integer wavelet transform," 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), 15-17 Oct. 2011, pp. 892-896.

[G.Depovere, 1999] G.Depovere, T.Kalker, J.Haitsm, M.Maes, L. D. Strycker, P.Termont, J.Vandewege, A.Langell, C.Alm, P.Norman, B.O'reilly, G.Howes, H.Vaanholt, R.Hintzen, P.Donnelly and A.Hudson, "The VIVA project: Digital watermarking for broadcast monitoring," *IEEE Internation Conference on Image Processing*, 1999, pp. 202-205.

[G.L.Friedman, 1996] G.L.Friedman, "Digital Camera with Apparatus for Authentication of Images Produced from an Image File," *United Sates Patent*, vol. 5, 1996.

[G.L.Friedman, 1993] G.L.Friedman, "The trustworth camera: Restoring Credibility to the Photographic Image," *IEEE trans on Consumer Electronic*, vol. 39, 905-910, 1993.

[Girija, 2010] C. Girija, "Blind and passive digital video tamper detection based on multimodal fusion," *in Proceedings of the 14th WSEAS international conference on Communications*, Ed.^Eds., ed. Corfu Island, Greece: World Scientific and Engineering Academy and Society (WSEAS), 2010.

[Gonzalez, 2008] R. C. Gonzalez, Digital Image processing, 2008.

[Granty, 2010] R. E. J. Granty, T. S. Aditya and S. S. Madhu, "Survey on passive methods of image tampering detection," *2010 International Conference on Communication and Computational Intelligence (INCOCCI)*, 27-29 Dec. 2010, pp. 431-436.

[H.C.Andrews, 1976] H.C.Andrews and C.L.Patterson, "Singular Value Decomposition (SVD) Image Coding," *IEEE Transactions on Communication*, vol. 24, 425-432, 1976.

[Hadjidemetriou, 2000] E. Hadjidemetriou, M. D. Grossberg and S. K. Nayar, "Histogram preserving image transformations," *IEEE Conference on Computer Vision and Pattern Recognition*, 2000. vol.1, pp. 410-416.

[Hartung, 1999] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc IEEE*, vol. 87, no. 7, pp. 1079-1107, 1999.

[Hu, 2011] Y. Hu, Z. Wang, H. Liu and G. Guo, "A Geometric Distortion Resilient Image Watermark Algorithm Based on DWT-DFT," *Journal of Software*, vol. 6, 1805-1812, 2011.

[Hubballi, 2009] N. Hubballi and K. D. P, "Novel DCT based watermarking scheme for digital image," *International Journal of Recent Trends in Engineering*, 1430-433, 2009.

[Hwei-Jen, 2010] L. Hwei-Jen, L. Che-Wei and C. Chiang-Ming, "DWT-based watermarking technique associated with embedding rule," *in Proceedings of the 10th WSEAS international conference on Signal processing, computational geometry and artificial vision*, Ed.^Eds., ed. Taipei, Taiwan: World Scientific and Engineering Academy and Society (WSEAS), 2010.

[Ingemar J. Cox, 2008] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, Jessica Fridrich and T. Kalker, Eds. (2008). Digital Watermarking and Steganography. Burlington, Morgan Kaufmann.

[Iwata, 2010] M. Iwata, T. Hori, A. Shiozaki and A. Ogihara, "Digital watermarking method for tamper detection and recovery of JPEG images," *2010 International Symposium on Information Theory and its Applications (ISITA)*, 17-20 Oct. 2010, pp. 309-314.

[Jain, 1990] A. K. Jain and F. Farrokhnia, "Unsupervised Texture Segmentation Using Gabor Filters," *Systems, Man and Cybernetics*, 14-19, Nov 1990.

[Jayanthi, 2009] V. E. Jayanthi, V. M. Selvalakshmi and V. Rajamani, "Digital watermarking robust to geometric distortions in biomedical images," *2009 International Conference on Control, Automation, Communication and Energy Conservation, 2009.*, 4-6 June 2009, pp. 1-6.

[Joseph, 1998] J. K. Joseph, Ruanaidh and P. Thierry, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Process.*, vol. 66, no. 3, pp. 303-317, 1998.

[K.F.Tsang, 2001] K.F.Tsang and O.C.Au, "A Review on Attacks, Problems and Weaknesses of Digital Watermarking and the Pixel Reallocation Attack," *SPIE Proceedings Series*, vol. 4314, 385-393, 2001.

[k.Zebbiche, 2008] K.Zebbiche and F.Khelifi, "Region-Based Watermarking of Biometric Images: Case Study in Fingerprint Images," *International Journal of Digital Multimedia Broadcasting*, 2008.

[Kargupta, 2004] H. Kargupta and B. H. Park, "A Fourier spectrum-based approach to represent decision trees for mining data streams in mobile environments," *IEEE Transactions on Knowledge and Data Engineering,* vol. 16, no. 2, pp. 216-229, 2004.

[Kipper, 2004] G. Kipper, Ed. (2004). Investigator's Guide to Steganography. New York, Auerbach.

[Kutter, 2000] M. Kutter, S. Voloshynovskiy and Herrigel, "The watermark copy attack," *Electronic Imaging 2000, Security and Watermarking of Multimedia Content II*, vol. 3971, 2000.

[Licks, 2005] V. Licks and R. Jordan, "Geometric attacks on image watermarking systems," *Multimedia, IEEE*, vol. 12, no. 3, pp. 68-78, 2005.

[M.Eskicioglu, 2004] E. G. A. A. M.Eskicioglu, "A DFT-Based Semi-Blind Multiple Watermarking Scheme for Images " *4th New York Metro Area Networking (NYMAN 2004) Workshop*, 2004.

[Mahdian, 2008] B. Mahdian and S. Saic, "Blind Authentication Using Periodic Properties of Interpolation," *IEEE Transactions on Information Forensics and Security,* vol. 3, no. 3, pp. 529-538, 2008.

[Mahdian, 2009] B. Mahdian and S. Saic, "Detection and description of geometrically transformed digital images," *Media Forensics and Security*, vol. 7254, 2009.

[Mark de Berg, 2000] Mark De Berg, Marc Van Kreveld, Mark Overmars and O. Schwarzkopf, Computational Geometry (2nd revised ed.), Springer-Verlag, 2000.

[Micah, 2006] K. J. Micah and F. Hany, "Exposing digital forgeries through chromatic aberration," in Proceedings of the 8th workshop on Multimedia and security, ed. Geneva, Switzerland: ACM, 2006, pp.

[N.Garguir, 1979] N.Garguir, "Comparative Performance of SVD and Adaptive Cosing Transform in Coding Images," *IEEE Transactions on Communication*, vol. 27, 1230-1234, 1979.

[N.Minami, 2002] N.Minami and M.Kasahara, "A New Decoding Method for Digital Watermark based on Error Correcting Codes and Cryptography," *IEICE Trans. Fundamentals (Japanese Edition)*, vol. J87-A, no. 7, pp. 965-975, 2002.

[Nasir, 2007] I. Nasir, Y. Weng and J. Jiang, "A New Robust Watermarking Scheme for Color Image in Spatial Domain," *Third International IEEE Conference on Signal-Image Technologies and Internet-Based System*, 2007. SITIS '07. 942-947, 2007.

[Ng, 2006] H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh and W. L. Nowinski, "Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm," *in Proceedings of the 2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, IEEE Computer Society, 2006, pp.

[Nikolaidis, 2001] A. Nikolaidis and I. Pitas, "Region-based image watermarking," *IEEE Transactions on Image Processing*, vol. 10, no. 11, pp. 1726-1740, 2001.

[Pai, 2006] P. Pai, N. Peng and D. S. Reeves, "On the secrecy of timing-based active watermarking trace-back techniques," 2006 IEEE Symposium on Security and Privacy, 21-24 May 2006.

[Parameswaran, 2008] D. L. Parameswaran and D. K. Anbumani, "Content-Based Watermarking for Image Authentication Using Independent Component Analysis," Informatica, vol. 32, 299-306, 2008.

[Pereira, 2000] S. Pereira and T. Pun, "Robust Template Matching for Affine Resistant Image Watermarks," *IEEE Transactions on Image Processing*, vol. 9, 1123-1129, 2000.

[Petitcolas, 1998] F. A. P. Petitcolas, R. J. Anderson and M. G. Kuhn, "Attacks on copyright marking systems," *Information Hiding, Second International Workshop*, IH'98, 1998.

[Petrovic, 2007] R. Petrovic, B. Tehranchi and J. M. Winograd, "Security of Copy-Control Watermarks," *8th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, 2007.* 26-28 Sept. 2007, pp. 117-126.

[Phadikar, 2009] A. Phadikar, S. P. Maity and H. Rahaman, "Region Specific Spatial Domain Image Watermnarking Scheme," *2009 IEEE International Advance Computing Conference*, 888-893, 2009.

[Porikli, 2005] F. Porikli, "Integral histogram: a fast way to extract histograms in Cartesian spaces," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. 20-25 June 2005, vol. 1. pp. 829-836

[Qian-Chuan, 2008] Z. Qian-Chuan, Z. Qing-Xin and Z. Ping-Li, "A Spatial Domain Color Watermarking Scheme based on Chaos," *International Conference on Apperceiving Computing and Intelligence Analysis*, 2008. 2008, pp. 137-142.

[Qieshi] Z. Qieshi, H. Inaba and S. Kamata, "Adaptive Histogram Analysis for Image Enhancement," *2010 Fourth Pacific-Rim Symposium on Image and Video Technology*, 14-17 Nov. 2010, pp. 408-413.

[R.Anderson, 1996] R.Anderson, "Information Hiding," *Lecture Notes in Computer Science, Springer-Verlag*, vol. 1174, 1996.

[R.Karkarala, 2001] R.Karkarala and P.O.Ogunbona, "Signal Analysis Using a Multiresolution Form of the Singular Value Decomposition," *IEEE Transactions on Image Processing*, 724-735, 2001.

[Ren, 2009] G. Ren, C. Zhang and X. Yang, "Blind Mesh Watermarking Based on the Featured Points in the Frequency Domain," *Second International Workshop on Knowledge Discovery and Data Mining*, 2009. 701-704, 2009.

[Ridzon, 2008] R. Ridzon and D. Levicky, "Robust digital watermarking in DFT and LPM domain," ELMAR, 2008. 50th International Symposium, vol. 2, 651-654, 2008.

[S.O.Aase, 1999] S.O.Aase, J.H.Husoy and P.Waldemar, "A Critique of SVD-Based Image Coding System," *IEEE International Symposium on Circuits and Systems VLSI*, vol. 4, 13-16, 1999.

[Sabbag, 2006] E. Sabbag and N. Merhav, "Optimal Watermark Embedding and Detection Strategies Under Limited Detection Resources," *2006 IEEE International Symposium on Information Theory*, 9-14 July 2006, pp. 173-177.

[Sharkas, 2005] M. Sharkas, D. Elshafie and N. Hamdy, "A Dual Digital-image Watermarking Technique," *Proceedings of World Academy of Science, Engineering and Technology*, vol. 5, 2005.

[Shrivastava, 2011] S. Shrivastava and S. Choubey, "A Secure Image Based Watermark for 3D Images," *2011 International Conference on Communication Systems and Network Technologies (CSNT)*, 3-5 June 2011, pp. 559-562.

[Skodras, 2006] A. N. Skodras and T. Ebrahimi, "JPEG2000 image coding system theory and applications," *2006 IEEE International Symposium on Circuits and Systems*, 2006. 0-0 0 2006, pp. 4 pp.-3869.

[Sudirman, 2009] S. Sudirman, "High order polynomial signature embedding in wavelet transform domain for robust digital image watermarking," *Proc. 5th IASTED European Conference on Internet and Multimedia Systems and Applications(EuroIMSA 2009)*,1-6, 2009.

[Sun, 2009] J. Sun, J. Li and Z. Li, "An Improved Algorithm of Digital Watermarking Based on Wavelet Transform," *2009 WRI World Congress on Computer Science and Information Engineering*, 2009, pp. 280-284.

[Sutcu, 2007] Y. Sutcu, B. Coskun, H. T. Sencar and N. Memon, "Tamper Detection Based on Regularity of Wavelet Transform Coefficients," IEEE International Conference on Image Processing, 2007. Sept. 16 2007-Oct. 19 2007, pp. I - 397-I - 400.

[Sverdlov, 2004] A. Sverdlov, S. Dexter and A. M. Eskicioglu, "Robust DCT-SVD Domain Image Watermarking for Copyright Protection: Embedding Data in all Frequencies," *International Multimedia Conference, Proceedings of the 2004 workshop on Multimedia and security*, 166-174, 2004.

[Tao, 2010] J. Tao, L. Xinghua and Z. Feifei, "Image tamper detection algorithm based on Radon and Fourier-Mellin transform," *2010 IEEE International Conference on Information Theory and Information Security (ICITIS)*, 17-19 Dec. 2010 2010, pp. 212-215.

[Tao, 2004] P. N. Tao and Eskicioglu, "A robust multiple watermarking scheme in the Discrete Wavelet Transform domain," *Internet Multimedia Management System Conference*, vol. 5601, 133-144, 2004.

[Tobias, 2002] O. J. Tobias and R. Seara, " IEEE Transactions on Image segmentation by histogram thresholding using fuzzy sets,", vol. 11, no. 12, pp. 1457-1465, 2002.

[Verma, 2009] A. Verma and S. Tapaswi, "A novel reversible visible watermarking technique for images using Noise Sensitive Region Based Watermark Embedding (NSRBWE) approach," EUROCON 2009, EUROCON '09. IEEE, 2009, pp. 1374-1377.

[Voloshynovskiy, 2001] S. Voloshynovskiy, S. Pereira, T. Pun, J. J. Eggers and J. K. Su, "Attacks on digital watermarks: classification, estimation based attacks, and benchmarks," *Communications Magazine, IEEE*, vol. 39, no. 8, pp. 118-126, 2001.

[Wei, 2009] W. Wei, D. Jing and T. Tieniu, "A Survey of Passive Image Tampering Detection," in Proceedings of the 8th International Workshop on Digital Watermarking, vol., Ed.^Eds., ed. Guildford, UK: Springer-Verlag, 2009.

[Wong, 1999] P. W. Wong and E. J.Delp, "Security and Watermarking of Multimedia Contents," Society of Photo-optical Instrumentation Engineers, vol. 3657, 1999.

[Wong, 2000] P. W. Wong and E. J.Delp, "Security and Watermarking of Multimedia Contents II," Society of Photo-optical Instrumentation Engineers, vol. 3971, 2000.

[Woo, 2007] C.-S. Woo, "Digital Image Watermarking Methods for Copyright Protection and Authentication" Queensland Univerisity of Technology, Doctor of Philosophy, 2007

[Xiaochuan, 2006] Z. H. Q. C. G. Xiaochuan, "Low Luminance Smooth Blocks Based Watermarking Scheme in DCT Domain," *2006 International Conference on Communications, Circuits and Systems Proceedings*, vol. 1, 2006.

[Xiaojun, 2008] Z. Xiaojun, X. Jiangtao, Y. Yanguang and J. Chichero, "The Fourier Spectrum Analysis of Optical Feedback Self-Mixing Signal under Weak and Moderate Feedback," *4th IEEE International Symposium on Electronic Design, Test and Applications*. 23-25 Jan. 2008, pp. 491-495.

[Xu J, 2004] Xu J, Sung, A.H, Shi P and Liu Q, "JPEG compression immune steganography using wavelet transform" *International Conference on Information Technology: Coding and Computing*, 2004, Vol. 2, 704-708.

[Yan-hong, 2010] Z. Yan-Hong, "Robust 3D Watermarking Technique for Authentication of 3D Polygonal Medel," *2010 Sixth International Conference on Semantics Knowledge and Grid (SKG)*, 1-3 Nov. 2010, pp. 325-329.

[Ying-Hua, 2007] L. Ying-Hua, Q. Jing, K. Jun and L. Si-Hui, "A robust digital watermarking scheme based on optimal coefficients selector about subimages," *International Conference on Wavelet Analysis and Pattern Recognition, 2007*. 2-4 Nov. 2007, pp. 1865-1869.

[Zaboli, 2007] S. Zaboli and M. S. Moin, "CEW: A Non-Blind Adaptive Image Watermarking Approach Based on Entropy in Contourlet Domain," *IEEE International Symposium on Industrial Electronics*, 4-7 June 2007, pp. 1687-1692.

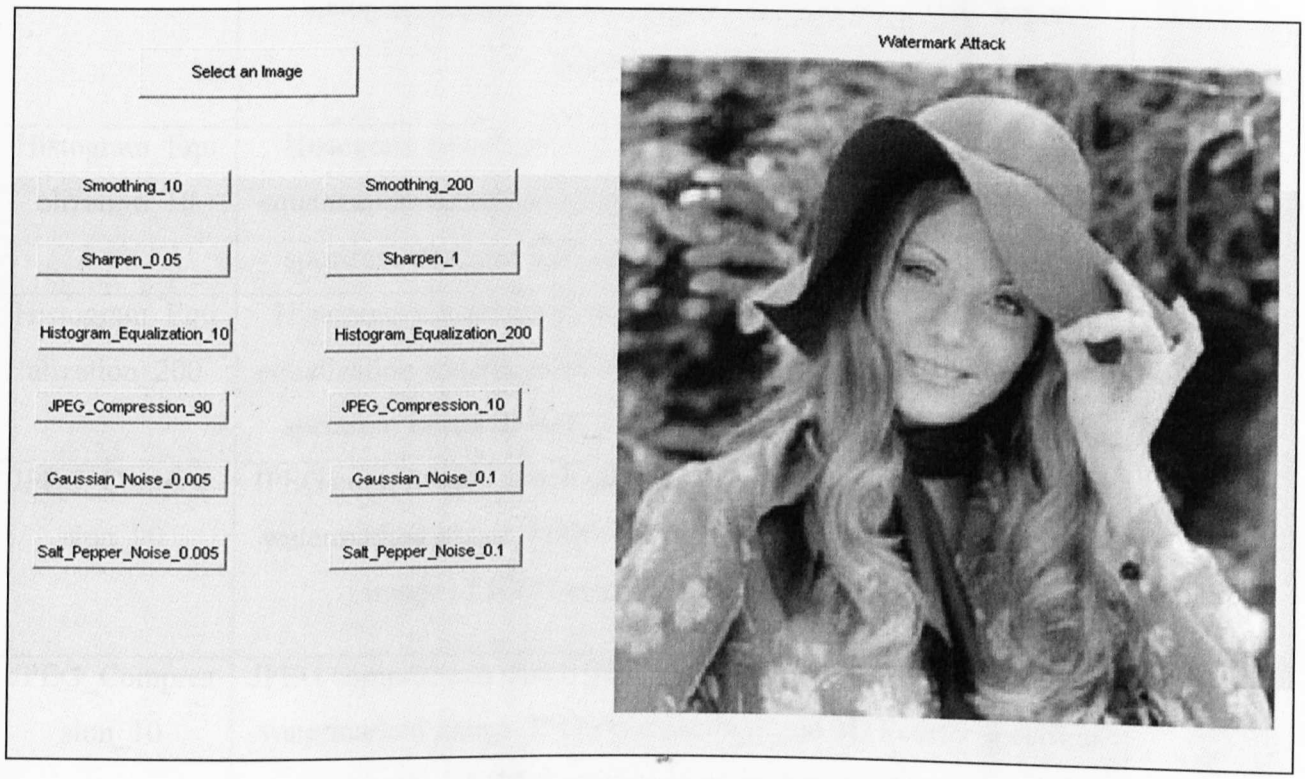
[Zhao, 2008] R.-M. Zhao, H. Lian, H.-W. Pang and B.-N. Hu, "A Blind Watermarking Algorithm Based on DCT," *Second International Symposium on Intelligent Information Technology Application*, 2008. 20-22 Dec. 2008 2008, pp. 821-824.

[Zhiguo, 2007] Q. Zhiguo and J. Cong, "Self-Restoring Dual Watermarking Technology," *IEEE International Conference on System of Systems Engineering*, 2007. 2007, pp. 1-5.

[Zhou, 2004] B. Zhou and J. Chen, "A Geometric Distortion Resilient Image Watermarking Algorithm Based on SVD," *Chinese Journal of Image and Graphics*, vol. 9506-512. 2004.

Appendix A: Graphic User Interface Design

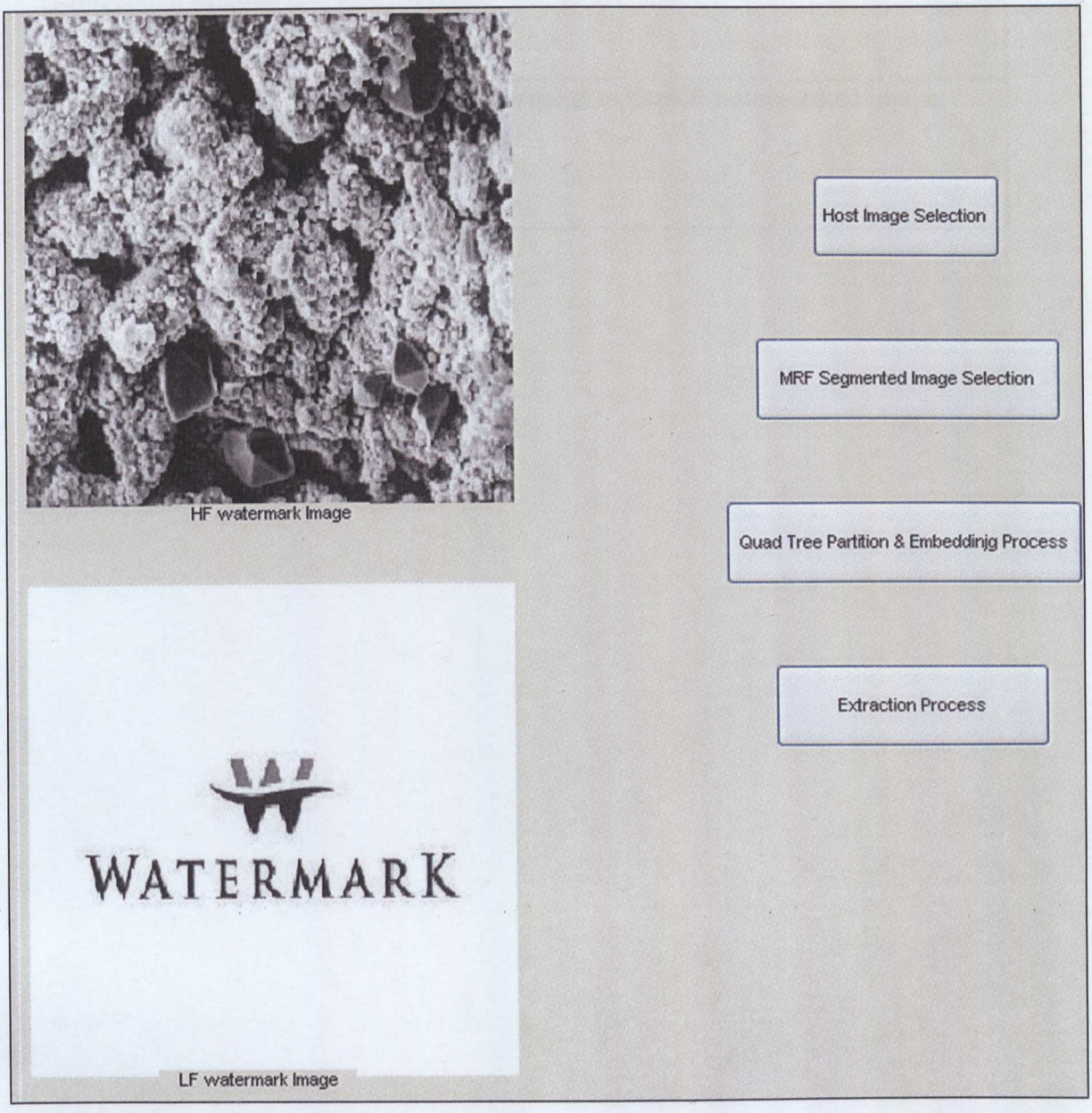
Subsection 1: Analysis Different Watermark Attacks



Button	Function
Select an Image	Select host image manually and generate watermarked image by DWT
Smoothing_10	Smoothing attack (alpha = 10), produce smoothing watermarked image, histogram image and 3D Fourier spectrum image of smoothing watermarked image.
Smoothing_200	Smoothing attack (alpha = 200), produce smoothing watermarked image, histogram image and 3D Fourier spectrum image of smoothing watermarked image.
Sharpen_0.05	Sharpen attack (alpha = 0.05), produce sharpen watermarked

	image, histogram image and 3D Fourier spectrum image of sharpen watermarked image.
Sharpen_1	Sharpen attack (alpha = 1), produce sharpen watermarked image, histogram image and 3D Fourier spectrum image of sharpen watermarked image.
Histogram_Equalization_10	Histogram equalization attack (alpha = 10), produce histogram equalization watermarked image, histogram image and 3D Fourier spectrum image of histogram equalization watermarked image
Histogram_Equalization_200	Histogram equalization attack (alpha = 200), produce histogram equalization watermarked image, histogram image and 3D Fourier spectrum image of histogram equalization watermarked image
JPEG_Compression_90	JPEG compression attack (alpha = 90), produce JPEG compression watermarked image, JPEG compression and 3D Fourier spectrum image of JPEG compression watermarked image.
JPEG_Compression_10	JPEG compression attack (alpha = 10), produce JPEG compression watermarked image, JPEG compression and 3D Fourier spectrum image of JPEG compression watermarked image.
Gaussian_Noise_0.005	Gaussian noise attack (alpha = 0.005), produce Gaussian noise watermarked image, Gaussian noise and 3D Fourier spectrum image of Gaussian noise watermarked image.
Gaussian_Noise_0.1	Gaussian noise attack (alpha = 0.1), produce Gaussian noise watermarked image, Gaussian noise and 3D Fourier spectrum image of Gaussian noise watermarked image.
Salt_Pepper_Noise_0.005	Salt and Pepper noise attack (alpha = 0.005), produce Salt and Pepper noise watermarked image, Salt and Pepper noise and 3D Fourier spectrum image of Salt and Pepper noise watermarked image.
Salt_Pepper_Noise_0.1	Salt and Pepper noise attack (alpha = 0.1), produce Salt and Pepper noise watermarked image, Salt and Pepper noise and 3D Fourier spectrum image of Salt and Pepper noise watermarked image.

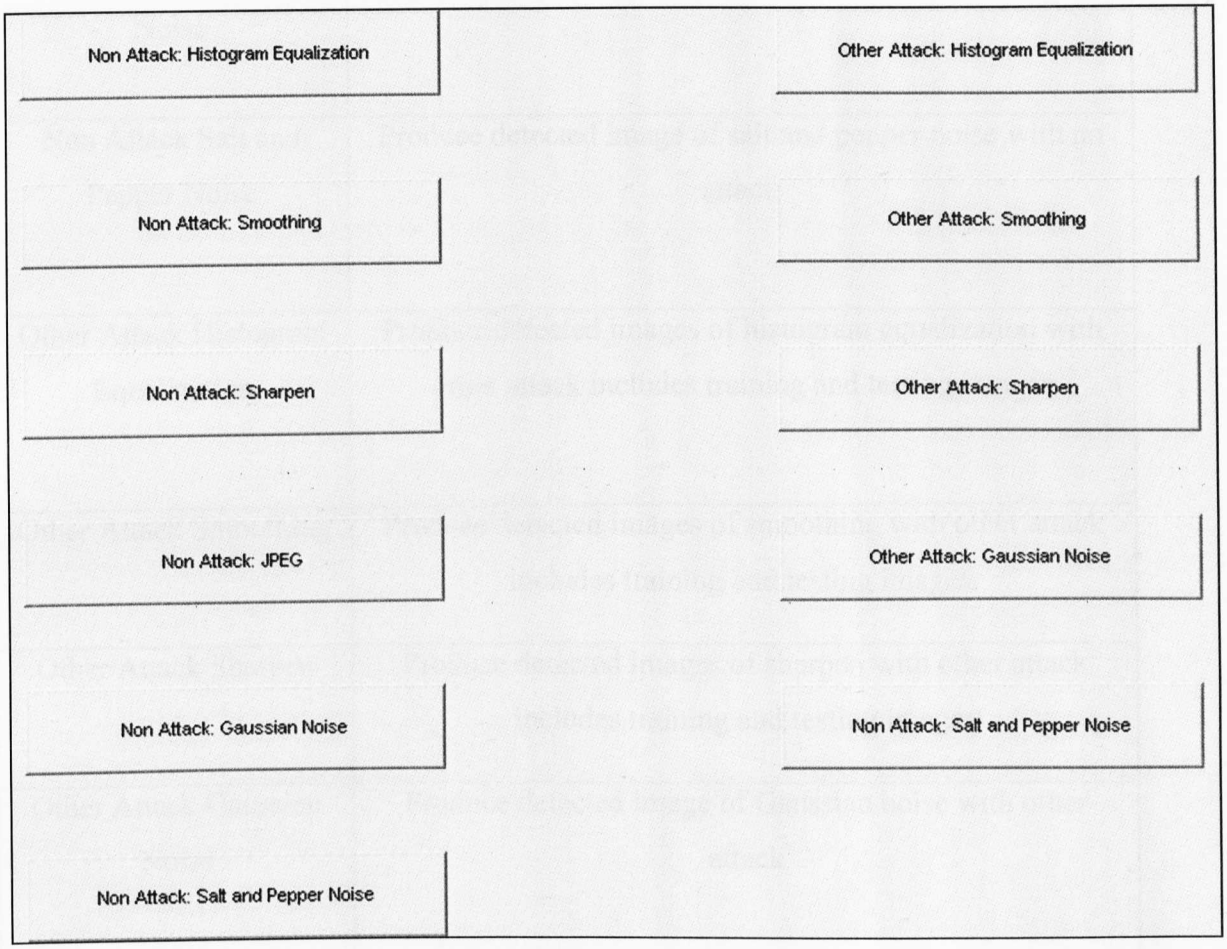
Subsection 2: Region-Adaptive Watermarking algorithm by DWT-SVD



Button	Function
Host Image Selection	Select host image manually, produce host image.
MRF Segmented Image Selection	Select corresponding MRF segmented image manually (After select host image manually *by Host Image button*), produce MRF segmented image.

Quad Tree Partition and Embedding Process	Produce quad tree image and watermarked image.
Extracted Process	Produce original extracted watermarked image.

Subsection 3: Watermark Attack Detection



Button	Function
Non Attack Histogram Equalisation	Produce detected images of histogram equalization with no attack includes training and testing images
Non Attack Smoothing	Produce detected images of smoothing with no attack includes training and testing images.
Non Attack Sharpen	Produce detected images of sharpen with no attack includes training and testing images
Non Attack JPEG	Produce detected images of JPEG compression with no attack includes training and testing images

Non Attack Gaussian Noise	Produce detected image of Gaussian noise with no attack
Non Attack Salt and Pepper Noise	Produce detected image of salt and pepper noise with no attack
Other Attack Histogram Equalization	Produce detected images of histogram equalization with other attack includes training and testing images
Other Attack Smoothing	Produce detected images of smoothing with other attack includes training and testing images.
Other Attack Sharpen	Produce detected images of sharpen with other attack includes training and testing images
Other Attack Gaussian Noise	Produce detected image of Gaussian noise with other attack
Other Attack Salt and Pepper Noise	Produce detected image of salt and pepper noise with other attack

Appendix B: Image Segmentation by using Markov Random Field

Markov Random Fields

Markov random field theory holds the promise of providing a systematic approach to the analysis of images in the framework of Bayesian probability theory. Markov random fields (MRFs) model the statistical properties of images. This allows a host of statistical tools and approaches to be turned to solving so called *ill-posed problems* in which the measured data does not specify a unique solution.

Sites and Labels

A Markov random field is defined on a set of sites. The sites may be regularly spaced on a lattice or irregularly spaced. Regularly spaced sites are suitable for modelling pixel intensity levels in images and will be used throughout this work. Irregularly spaced sites are useful for high level vision problems in which features have been extracted from the image. Irregularly spaced sites are usually referred to in the statistical literature as point processes rather than Markov random fields. Let _ be a set of m discrete sites

$$S = \{1,\ldots,m\} \quad (1)$$

in which 1,...,m are indices. A set of sites on a square n * n lattice can also be written as $S = \{(i,j)|1 \leq i,j \leq n\}$

Each site has a label associated with it. The set of possible labels may be continuous or discrete. The adoption of either a continuous or a discrete label set is one of the first decisions that need to be made as this determines the nature of the solution space. If the label set is continuous, the probability distribution used to model the problem must also be continuous in which case it is known as a probability density function. If the label set is discrete, the probability distribution used to model the problem must also be discrete and is called a probability mass function. For now, a set _ of _ discrete labels will be adopted such that

$$L = \{l_1,\ldots,l_M\} \quad (2)$$

The labelling for a set of sites, S _, will be denoted by

$$f = \{f_1,\ldots,f_M\} \quad (3)$$

where f_i _ is the label at site i. The set of all possible configurations is called F. The size of the configuration space F is given by M^m where M is the number of candidate labels for each site and m is the number of sites on the lattice. Many problems in machine vision can be cast into this form where the problem is to estimate the best labelling for a set of sites.

The Markov Property

The defining characteristic of MRFs is that the interaction between labels is limited to a local region. This region is called the neighbourhood of a site. The sites of a Markov random field on a lattice S are related to each other via a neighbourhood system, N, such that

$$N = \{N_i|\forall i \epsilon S\} \quad (4)$$

Where N_i _ is the set of sites neighbouring site i. A site cannot be a neighbour to itself.

A random process is said to be Markov if the following condition holds. The conditional probability function for the label at a site i_ given the labels of all other sites on S is equal to the conditional probability for that label given only the labels in the neighbourhood of site. Following the notation can be written as

$$P(f_i|f_{S-\{i\}}) = P(f_i|f_{Ni}) \quad (5)$$

Equation 1.5 does not mean that the labels of sites not in each other's neighbourhood are independent, but rather that all information about the distribution at a site is given by its neighbours and no more information can be gained by considering sites outside of that sites neighbourhood. In other words, correlations may extend far beyond the local neighbourhood of a site. The conditional distribution of a site gives the probability of possible labels at that site given the labels at neighbouring sites. It is

difficult to specify a Markov random field by its conditional probability structure as there are highly restrictive consistency conditions. Fortunately Gibbs distributions provide a way to specify a Markov random field by its joint probability distribution. The joint probability assigns a probability to each possible configuration on the lattice S. It is the joint probability that is required for the maximum *a posteriori* estimation algorithm.

The Gibbs Distribution

Markov random fields and Gibbs distributions are equivalent. The Gibbs distribution of a Markov random field is just the joint probability of that Markov random field.

Let P(f) be a Gibbs distribution on a lattice S. Then P(f) has a form given by

$$P(f) = Z^{-1} * e^{(-1/T)U(f)} \quad (6)$$

Where

$$Z = \sum_{f \epsilon F} e^{\left(-\frac{1}{T}\right)U(f)} \quad (7)$$

is a normalizing constant called the partition function. Calculating the partition function exactly involves normalizing over all possible configurations which is computationally prohibitive for even moderately sized images as the number of possible configurations is given by M^m where M is the number of labels for each site and m is the number of sites on the lattice. The term Z is sometimes called the free energy of the system.

The energy function U(f) in Equation 6 is the sum of clique potential functions, $V_c(f)$, over all cliques C on the lattice S as given by Equation 8. Configurations with higher energy have less probability of occurring.

$$U(f) = \sum_{c \epsilon C} Vc(f) \quad (8)$$

The energy U(f) and the clique potential functions $V_c(f)$ should be positive for all possible label configurations, to enable correct normalization of Equation 6. This positivity constraint can be enforced on clique potential functions by subtracting the

minimum value of the potential function over the domain L from the potential function as shown in Equation 9. This is done for all clique potential functions except for the uniform prior defined in Equation 9 which is defined to be positive for all possible label

$$\text{Vc(f)} \leftarrow Vc(f) - minVc(li) \quad (9)$$

The order of a clique is given by the number of sites in the clique. A first order clique potential is thus a function of the label at one site. A second order clique potential is a function of the labels at two sites and is also the lowest order clique potential to convey contextual information or to model dependence between the labels at neighbouring sites.

The term T in Equation 6 is a scalar that represents temperature in physical systems and will be referred to as the temperature here. As the value of T is increased the distribution approaches a uniform distribution, for which each configuration has the same probability. Similarly, as the temperature is lowered the distribution becomes more peaked with the probability mass concentrating at the most likely configurations. The temperature term T is prominent in the simulated annealing optimization algorithm where the search strategy involves sampling the same distribution at different temperatures.