ADAPTIVE MULTIMEDIA DATA HIDING AND WATERMARKING

by

KAN LI
B.Eng
Heifei, Anhui, P.R.China, 1998

A thesis

presented to Ryerson University

in partial fulfillment of the

requirement for the degree of

Master of Applied Science

in the Program of

Electrical and Computer Engineering

Toronto, Ontario, Canada, 2004

AUTHOR'S DECLARATION

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

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

Signature:

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

Signature:

Borrower's Page

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

Name	Signature	Address	Date

Adaptive Multimedia Data Hiding and Watermarking

Master of Applied Science 2004

KAN LI

Electrical and Computer Engineering

Ryerson University

Abstract

Watermarking is a technique of hiding a message about a work of media within that work itself in the purpose of protecting the digital information against illegal duplication and manipulation.

The objectives of this study are to analyze the robustness and distortion performance of watermarking system and to explore watermarking schemes which balance the robustness-distortion tradeoff optimally.

In this thesis, We present a detector algorithm to adaptively extract spread spectrum watermark by filtering the watermarked images with Wiener filter. Two optimization algorithms for quantization watermarking are proposed. First one optimizes uniform quantization based look-up table embedding which minimizes watermarking distortion. Secondly, we analyze the robustness-distortion tradeoff and formulate the robustness-distortion tradeoff into a Lagrangian function. Hence optimal quantizers for watermarking subject to given robustness or fidelity constraint are achieved.

Acknowledgment

First I would like to express my sincere gratitude to my advisor, Prof. Xiao-ping Zhang, for his guidance, nurturing, encouragement, and support in every stage of my graduate study. His knowledge, kindness, patience, openmindedness, and vision have provided me with lifetime benefits.

I am grateful to Profs. Mehmet Zeytinoglu, Sridhar Krishnan and Songnian Li in my thesis committee for their valuable comments and suggestions on the thesis drafts.

Conversations with Dr. Ling Guan were very helpful and furthered my understanding of multimedia signal processing.

I would like to thank the Electrical and Computer Engineering department at Ryerson University for the financial support of my study and research.

Special thanks go to the people in Communications and Signal Processing Applications Laboratory (CASPAL) for their friendship and advice.

I am very much indebted to my parents for their love, encouragement and complete support through my study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Digital Watermark

With the emergence of high-capacity digital recording devices, coupled with the recent growth of networked multimedia systems, the protection of ownership and prevention of unauthorized tampering of multimedia data become important concerns. Unlike analog media such as audio and VHS video tapes, multimedia data in digital form can be copied without degradation and distributed freely. Therefore, a major concern, with respect to protecting intellectual property rights, has arisen. One approach to addressing this problem is the embedding of an invisible digital watermark into multimedia data to "mark" the ownership. The embedded digital watermark may be copyright or authentication code, or an imperceptible "signature" of the originator, or recipient of the host data. In general, if it is useful to associate some additional information with a multimedia work (image/video/audio), this metadata can be embedded as a watermark [1]. Of course, there are other ways to associate information with a work, such as placing it in the header of a digital file, encoding it in a visible bar code on an image. Watermarking is distinguished from other techniques in three important ways. First watermarks are imperceptible [1][2][3]. Unlike bar codes, watermarks do not damage the art value of an image. Second, watermarks are inseparable from the work in which they are embedded. Unlike header fields, they are not removed when the image/video are displayed or converted to other file formats. Finally, watermarks undergo the same transformations. This means that it is possible to learn something about those transformations by looking at the resulting watermarks [1][2][3][4]. It is these three attributes that make watermarking invaluable for certain multimedia applications.

A typical data hiding framework is illustrated in Figure 1.1. Starting with the multimedia host data, or its transformed format, an embedding module puts the watermark and an optional public or secret key. The watermark often consists of a binary data sequence, representing a number, text, or even an image. The public or secret key is used to enforce security. The watermark sequence is embedded in the host data by making imperceptible modification to its content. The output of the watermark embedding algorithm is the modified, i.e. watermarked data.

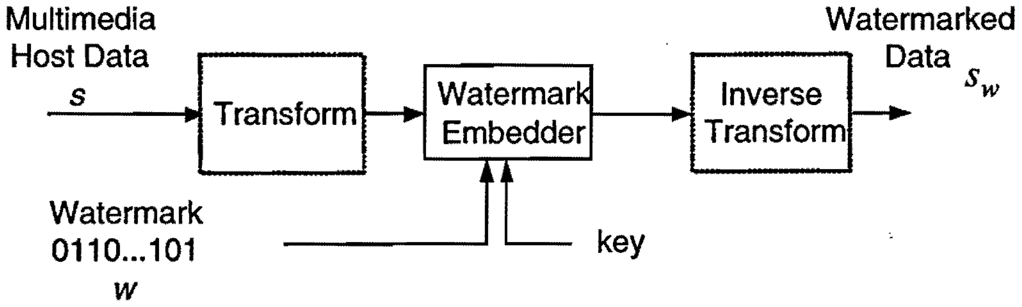


Figure 1.1: Watermark Embedding Process.

The general watermark extraction process is depicted in Figure 1.2. With or without the use of the key, the estimate of the original watermark is extracted from the watermarked multimedia data. In robust watermarking applications, the watermark must be recoverable even when the watermarked data undergo a reasonable level of distortion. According to whether or not the original host data is exploited during the watermark detection process, the existing schemes can be placed under two categories: blind watermark and non-blind watermark. Methods reported in [5] require the host signal for detection, whereas the scheme in [6][7] does not.

To protect copyright successfully, there are several fundamental requirements for watermarking.

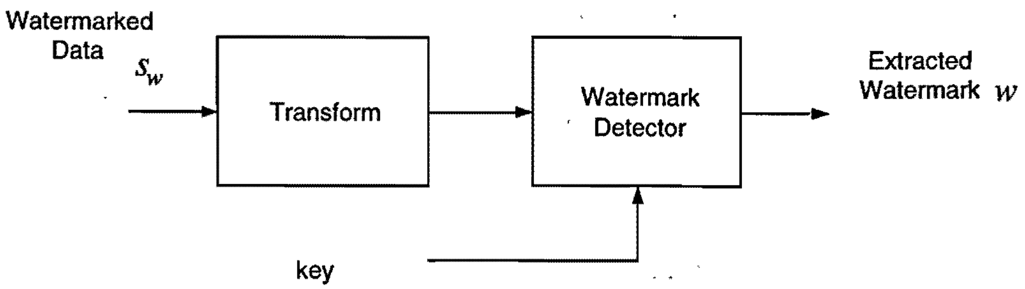


Figure 1.2: Watermark Extraction Process.

- Imperceptibility: The watermark should not be perceptible when embedded in the host data. In other words, the watermark embedding process should not introduce any perceptible artifacts into the host data. The commercial or art value of the host should not be affected.

- Robustness: The watermark should remain intact in the host data regardless of any change that may occur to the host data, including all possible signal processing, and possible malicious attacks that unauthorized parties may attempt. Robustness against all possible attacks may be impossible to achieve. Thus, the practical requirement is that the embedded watermark is computationally impossible to be removed without severely damaging the commercial or art value of the host data.

- Accuracy: The detection should be accurate, i.e. the probability of false alarm and miss detection should be as low as possible.

- Embedding Capacity: The total embedding capacity, namely, the number of bits that can be embedded and extracted with small probability of error is also an important measurement. Fortunately, not all scenarios require a high embedding capacity.

- There might be other requirements, such as blind detection, for those applications where the access to the original host data is impossible.

Unfortunately, the first two basic requirements are contradictory. For imperceptibility, the watermark embedding process should not introduce any perceptible artifacts into the host data. On the other hand, for high robustness, it is desirable that the watermark amplitude be as high as possible. Thus the design of watermarking methods always involves a tradeoff between imperceptibility and robustness.

Watermarking is a promising solution that can protect the copyright of multimedia data. Unlike encryption methods, digital watermarking does not restrict access to the multimedia work to prevent illicit acts. Instead, it provides evidence of a wrong-doing after it has taken place. Digital watermarking has the potential to provide protection even after the data is decrypted.

Also a reasonable expectation of applying watermarking techniques for copyright protection is to consider specific application scenarios, because the distortion behavior involved in these cases (quantization, compression and geometric distortions) could be predictable.

While the most prominent application of watermarking techniques is copyright protection, watermarking is also an attractive tool for any application where it is desirable to attach permanently hidden information to a multimedia signal, such as data monitoring and tracking, content labelling, multilingual captioning, usage control, and general covert communications [8]. An example for data monitoring is the automatic monitoring of broadcasted radio programs such that royalties are automatically paid to the copyright owners of the broadcast data. In transaction tracking applications, the owner or producer of the work would place a unique watermark in each copy; if the work were subsequently misused (leaked to the press or redistributed illegally), the owner could find out who was responsible [1]. In usage control applications, a digital watermark can be inserted to indicate the number of copies permitted. An example is digital video disc (DVD) [2].

In recent years, a number of practical data hiding systems have been proposed for image, audio or video watermarking. Most of recently reported schemes cast watermark into the transform domain [1][5][9], due to the fact that the transform domain watermarking schemes tend to achieve both perceptual transparency and robustness better than spatial/time domain schemes. However embedding watermark by directly modifying image pixels or audio samples is always simpler and faster [10][11].

In terms of different robustness requirements, watermarking schemes can be divided into robust watermark, fragile watermark and semi-fragile watermark. In a robust watermark system, the embedded data can survive common signal processing operations, whereas fragile watermark becomes undetectable after even minor modifications. Semi-fragile watermark is a hybrid of the two of above, only distortions that exceed a user-specified threshold will break the watermark.

All existing robust watermarking schemes can also be placed under two categories based on embedding mechanism: *coherent embedding* or *non-coherent embedding* . In non-coherent embedding algorithms, the embedded data have no relationship to the host data. Additive spread spectrum algorithm [5] is a representative of this category. In the second category, data hiding is achieved by enforcing a relationship between the bits to be embedded and the marked values. Quantization based watermarking schemes are in this category. Besides, Chen et. al. divided the existing embedding methods into two classes [12][13]: *host-interference non-rejecting methods* and *host-interference rejecting methods.* Generally, *host-interference rejecting embedding methods* correspond to *coherent embedding*, whereas *host-interference non-rejecting embedding methods* correspond to *non-coherent embedding.* In *host-interference rejecting embedding methods*, due to the enforced relationship between the watermark and the marked signal, the host signal is often not necessary during detection. On the contrary *host-interference non-rejecting methods* are primarily used where either the host signal is available at the detector or the host interference is small enough.

Much of the work on robust digital watermarking is based on spread spectrum principles [14][15]. Spread spectrum watermarking schemes borrow ideas from spread spectrum communications. In these schemes, a watermark is embedded into the host signal by adding a low energy pseudo-randomly generated white noise sequence. This specific pseudonoise sequence is detected by correlating the original watermark sequence with either the extracted watermark or the watermark signal itself (if the host data is not available for extraction). Spread spectrum watermarking has demonstrated great robustness and invisibility when the original host signal is available in detection [5]. However, in blind detection, the watermark experiences interference from the host data even when there is no noise from processing and intentional attack.

In another typical class of watermarking techniques, quantization based schemes [12][13][16], the watermark, often a binary sequence, is embedded into the host data by quantize-replace strategies that replace a quantized host signal with another quantization value. A simple example belonging to the class is the so called odd-even embedding: the host signal is replaced by the nearest even integer if to embed a "0" and the nearest odd integer if to embed a "1" [2]. This class of watermarking schemes are free from the interference from host data.

1.2 Optimal Watermarking and Data Hiding

In the design of any watermarking scheme, robustness against data distortion through signal processing or intentional attacks and the similarity between signal before and after watermarking are two major requirements. For some watermarking applications, watermarks are designed to survive normal processing and to resist any attempt by an adversary to thwart their intended purpose. In designing a robust watermark it is important to identify the specific processes that are likely to occur between embedding and detection. Examples of processes a watermark might need to survive include lossy compression, digital-to-analog-to-digital conversion, analog recording, printing and scanning, format conversion, and so on. For example, a video watermark designed for monitoring television advertisements [1] will need to survive the various processes involved in broadcasting — digital-to-analog conversion, lossy compression, and so on — but need not survive other processes, such as rotation or halftoning.

For other watermarking applications, fidelity is the primary perceptual measure. In these cases, the watermarked work must be indistinguishable from the original. In medical image applications, people may require this property of a watermark [1].

Clearly, various robustness and fidelity requirements involved in watermarking scheme design. In the applications where surviving the common signal processing operations is the primary concern, the robustness requirement should be satisfied first, we then maximize the fidelity. In other applications, the prerequisite is the perceptual similarity between the unwatermarked and watermarked signals, the robustness needs to be maximized. We define a watermarking system as *optimal watermarking* when it achieves a maximum robustness subject to a given fidelity criterion or a minimum distortion subject to a given robustness requirement.

Our objective in this thesis is to theoretically formulate this robustness-distortion tradeoff. The theoretical result may be applied directly to previously proposed as well as future robust watermarking algorithms in order to enhance their performance.

Although *optimal watermarking* is first introduced in this thesis, several papers related to this topic have been reported. Chen et. al. [12][13][17] introduced quantization index modulation (QIM) and theoretically proved that QIM achieves better robustness-distortion tradeoff than the current popular spread-spectrum methods. Wu [2][7] indicated that through a look-up table of nontrivial run, the probability of detection error can be considerably smaller than the traditional odd-even embedding. Optimal nonuniform quantization embedding is also studied by Wu et. al. [18]. They proposed algorithms for designing the optimal uniform quantization encoding scheme and optimal nonuniform quantization encoding scheme. In this thesis, we reformulate the robustness-distortion tradeoff and proposed an optimal watermarking system design method by addressing the tradeoff.

1.3 Organization of this Thesis

The remainder of the thesis is organized as follows. Chapter 2 is a brief review of background material that we employ later on. Then algorithms about robust watermarking scheme design are presented in the subsequent chapters, in which either the knowledge of reference watermarks or the knowledge of host data are employed in watermarking system. The following chapters are organized to highlight four principal contributions:

1. In Chapter 3, we consider a spread spectrum watermarking scheme where Wiener filter is employed into the design of the adaptive watermark detector. Two new local noise (watermark) variance estimation methods are employed in Wiener filtering.

2. Chapter 4 provides a look-up table (LUT) based watermark embedding scheme which achieves near minimum distortion. The LUT is determined by the probabilities that the feature to be embedded watermark "0" and "1" falls into each quantization cell.

3. Chapter 5 focuses on optimal watermarking. The robustness-distortion tradeoff is theoretically formulated as a robustness-distortion function. Lagrange's method is used to solve the robustness-distortion constrained optimization problem, hence an optimal watermarking scheme is obtained.

4. In Chapter 6, a novel image labelling system integrated with the cutting-edge still image compression standard–JPEG2000 is proposed.

Conclusions and suggestions for future research are discussed in Chapter 7. The proof of a useful theorem is provided in appendix.

Chapter 2

Preliminaries of Information Hiding

In this chapter, we present the background materials which will be employed later in the thesis. Channel capacity and performance indices of a watermarking scheme are briefly covered here.

2.1 Channel Capacity

The channel capacity of a memoryless channel is theoretically defined as the maximum mutual information between the channel input and output over all possible input distributions [19],

$$C = \max_{p(x)} I(X;Y). \tag{2.1}$$

The "operational" definition of channel capacity is the highest rate in bits per channel use at which information can be sent with arbitrarily low probability of error. Shannon's second theorem [19], called the channel coding theorem, establishes that the "information" channel capacity is equal to the "operational" channel capacity. Channel capacity serves as a good measure of the transmission potentiality of a channel.

Two simple examples of channel capacity will be employed in subsequent chapters. The first is a discrete memoryless channel, the binary symmetric channel (BSC); and the other is a continuous channel, the additive white Gaussian noise channel (AWGN).

2.1.1 Binary Symmetric Channel (BSC)

The binary symmetric channel with probability of bit error p is illustrated in Figure 2.1. The channel input X and output Y are binary signals [20]. When an error occurs, a "0" is received as a "1" is sent and visa versa. When $p \neq 0$, some of the received bits may be in error.

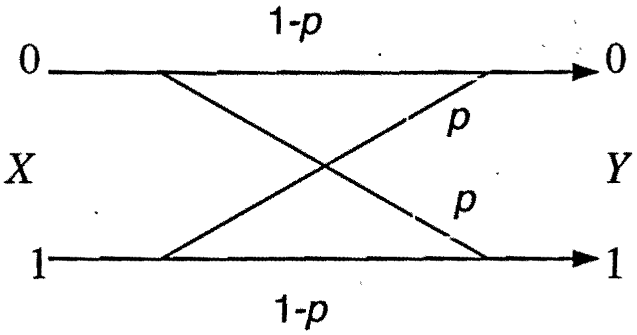


Figure 2.1: Binary symmetric channel.

However, by employing the channel coding theorem, we observe that we can still use such a communication channel to send information at a non-zero rate with an arbitrarily low probability of error. The capacity of a binary symmetric channel with probability of bit error p is given by

$$C = \max_{p(x)} I(X;Y) = 1 + plog_2 p + (1-p)log_2(1-p) \text{ bits per channel.} \tag{2.2}$$

where the maximum is taken over all possible input distributions $p(X)$.

The channel capacity of a BSC is achieved when the input X is equi-probable binary distribution, i.e. $P(X=0) = P(X=1) = \frac{1}{2}$.

2.1.2 Additive White Gaussian Noise (AWGN) Channel

Figure 2.2 illustrates the additive white Gaussian noise channel. In this channel, each element of the additive random noise vector is drawn independently from a Gaussian distribution. The capacity of the Gaussian channel with power constraint P is given by

$$C = \max_{p(x):E\{X^2\} \leq P} I(X;Y) = \frac{1}{2} log_2(1 + \frac{P}{N}) \text{ bits per channel.} \tag{2.3}$$

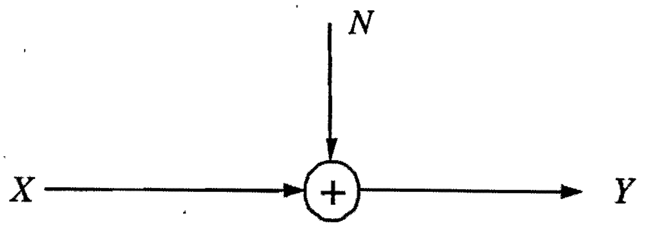


Figure 2.2: Additive white gaussian channel.

where the maximum is taken over all possible input distributions $p(X)$ satisfying the power constraint [21].

The capacity is achieved when X follows Gaussian distribution with zero mean and variance P, $X \sim \mathcal{N}(\prime, \mathcal{P})$.

2.2 Performance Indices

As mentioned in the introduction, an effective watermarking algorithm involves an appropriate tradeoff between imperceptibility and robustness. In this section, we present three quantitative measures to highlight this compromise.

1. Peak Signal-to-Noise Ratio (PSNR)

   The Peak Signal-to-Noise Ratio is defined as

   $$PSNR(f, f_w) = 10log_{10}\frac{MN\max_{m,n} f^2(m,n)}{\sum_{m=1}^{M}\sum_{n=1}^{N}(f_w(m,n) - f(m,n))^2}. \tag{2.4}$$

   in units of dB, where f is the original image and f_w is the watermarked image. $M \times N$ is the size of the image. Although this measure is generally not very accurate, the PSNR metric serves as a good rule of thumb measure of imperceptibility to assess the distortion introduced to the image as a result of embedding the watermark [22]. The larger the PSNR is, the better will be the performance of a watermarking scheme.

2. Probabilities of False Positive or False Negative

   The probability of false positive and the probability of false negative are two measures to objectively evaluate the robustness of watermarking schemes [2]. In security applications one must often detect the presence of a watermark or discover its unlawful modification by comparing the embedded and extracted marks. It is possible, however, that the presence of a watermark is not properly detected or that tampering is not identified. The chance of this occurring is called *the probability of false negative.* Similarly, the possibility of an incorrect watermark being detected in an image or the likelihood of false detection of tampering is termed *the probability of false positive.*

3. Bit Correct Ratio (BCR) or Bit Error Ratio (BER)

   Each time after watermark detection, the extracted watermark is compared with the original watermark to evaluate the robustness of the algorithm. The number of correct and erroneous bits divided by the total number of bits embedded are *the bit correct ratio* and *the bit error ratio*, respectively.

2.3 Common Signal Processing for Watermark Attack

Because many data hiding applications operate in a competitive environment where an adversary has the incentive to obliterate the embedded data, testing the systems' robustness and security against attacks is important. To evaluate the robustness of a scheme, a number of attacks against data hiding system are applied on the watermarked work before detection.

It is well accepted that no watermarking scheme can survive all attack methods, especially if the adversary has part or full knowledge of the watermarking algorithm. Several attacks as well as some countermeasures have been reported in the literature. Forging a fake "original" image for ownership claims can be thwarted by imposing invertibility requirement on watermarks. Collusion attack involves the averaging of multiple copies of the same original but having different markings. It is possible to systematically learn about the watermarks from the input-output relationship of a detector using many manipulated versions of watermarked images. Watermarks can also be attacked by geometric distortion, including rotation, scale, translation, warping, line dropping/adding, or in conjunction with moderate low-pass filtering and interpolation, but may not be always effective when the original image is available to perform registration.

In this thesis, one of our tasks is to search the optimal watermarking algorithm which is robust to common signal processing.

In the following sections, we examine the effects of three major types of common distortion on watermark detection : additive noise, low-pass filtering and lossy compression.

2.3.1 Additive Noise

Some processes that might be applied to a multimedia work have the effect of adding a random noise. That is

$$x = s + n \tag{2.5}$$

where s is the host data and n is a random vector chosen from some distortion, independently of s. For example, audio broadcast over a radio channel might be corrupted by white noise. In this case, the noise is independent of the multimedia work. Such noise process is a case of *additive noise.* Because of its simplicity, analysis of most watermarking algorithms assumes that the watermarked works are transmitted over an additive noise channel. Consequently, we will discuss watermarking system's robustness against additive noise in Chapter 3 through 6.

2.3.2 Filtering

Another common type of signal processing that may change multimedia signal in normal operations is filtering. That is,

$$x = s * f, \tag{2.6}$$

where s is the host data, f is a filter, and $*$ denotes convolution. Many normal operations on images and audio are explicitly implemented with filters. The blurring and sharping effects in image editing programs apply simple filtering operations. In addition, many lossy processes, although not explicitly implemented with filters, can be modelled as filtering.

2.3.3 Lossy Compression

Multimedia data contains redundancy with respect to what is needed for human perception, so losing information to a certain extent can be acceptable.

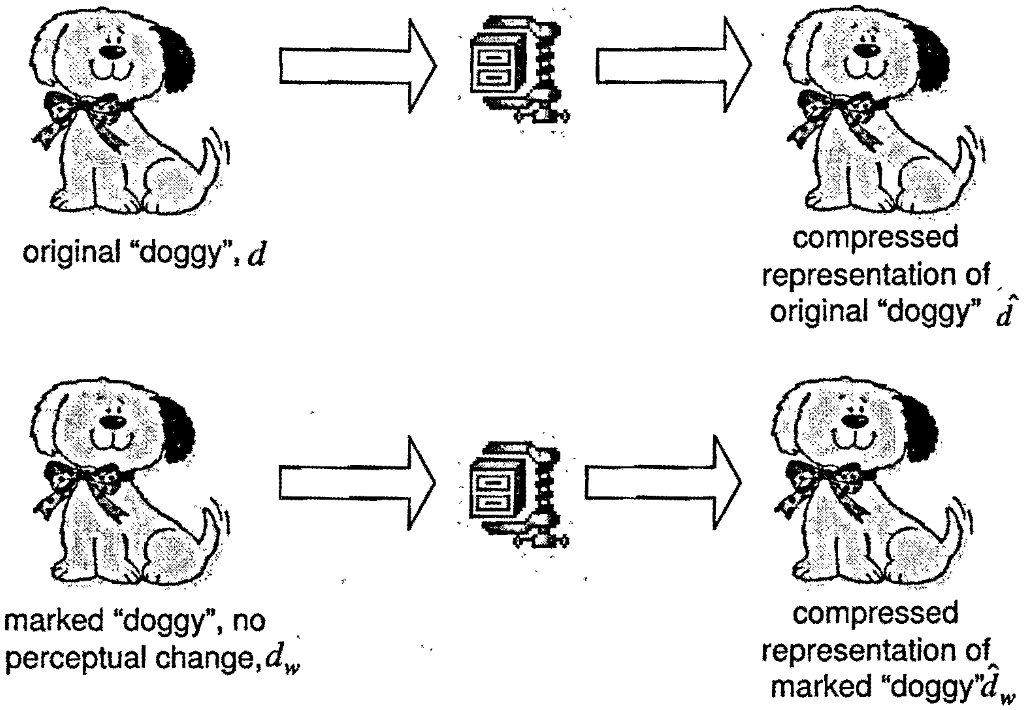

original "doggy", d
compressed representation of original "doggy" $\hat{d}$
marked "doggy", no perceptual change, d_w
compressed representation of marked "doggy" $\hat{d}_w$

Figure 2.3: Difference should exist between the watermarked compressed image and the original compressed image.

It is pointed out that a fundamental conflict exists between watermarking and lossy compression [1]. With an ideal lossy compressor, there should be a single compressed representation for all perceptually equivalent works. As illustrated in Figure 2.3, $\hat{d}$ should be equal to $\hat{d_w}$.

If there are two compressed representations resulting from perceptually equivalent works, then the lossy compressor does not remove all of the redundancy in the work. However, from a watermark embedder's viewpoint, to survive lossy compression, the compressed versions of the original and the watermarked data must be different, i.e., $\hat{d}$ should not be equal to $\hat{d_w}$.

Fortunately, lossy compression algorithms are far from ideal in practice, and there are still redundancies for a watermarking algorithm to survive lossy compression while maintaining excellent fidelity.

People also noticed that lossy compression and watermarking share some common characteristics [23]. Significant frequency coefficients must be achieved first for encoding in compression and for watermark casting in watermarking. Hence by integrating frequency domain watermarking with compression processes, the expensive transform computation can be saved. On the other hand, combining coding and watermarking is highly desired in some classic applications, such as copyright protection, copy and access control and annotation, where compression and watermarking are performed before spreading abroad. In Chapter 6, we will introduce a reliable image watermarking scheme integrated with state-of-art image compression standard — JPEG2000.

Chapter 3

Spread Spectrum Watermarking and Adaptive Filter Based Detector

3.1 Introduction

In this chapter, we focus on the spatial-domain spread spectrum watermarking scheme and the detector design where an adaptive filter technique is exploited. Our intention is to design the watermark detector that improves the detection response. We designed two detector algorithms: one uses the estimated local variance of the watermarked image, and the other uses not only the local variance of the received image but also the local variance of the reference watermark. The experimental results verify that a detector based on adaptive Wiener filter has better performance than high-pass filters.

3.2 Spread Spectrum Watermarking

Much of the work on robust digital watermarking is based on spread spectrum principles [5][9][14][24]. Spread spectrum watermarking schemes borrow ideas from spread spectrum communications. In spread spectrum communication, a narrow-band signal is spread across a wide band of frequencies. This can be accomplished by modulating the narrow-band signal (the watermark information) with a wide-band signal, such as white Gaussian noise.

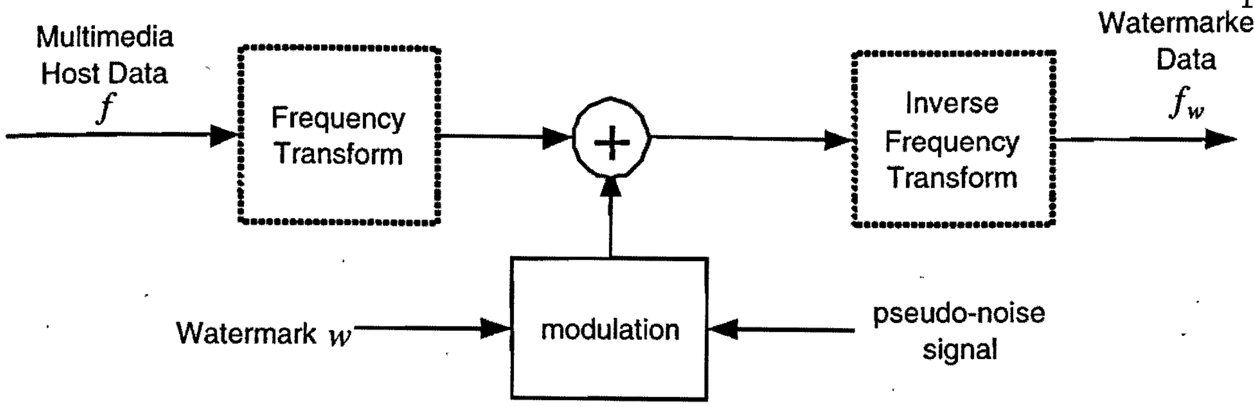


Figure 3.1: Typical block diagram of spread spectrum watermark embedding.

Therefore, the signal energy present in any single frequency is undetectable. Similarly, in spread spectrum watermarking system, the watermark is spread over many frequency bins so that the energy in any one bin is small and undetectable. Nevertheless, because the watermark verification process knows the location and content of the watermark, it is possible to concentrate these many weak signals into a single output with high signal-to-noise ratio. Destroying such a watermark would require noise of high amplitude to be added to all frequency bins. Thus, the commercial value of the watermarked multimedia work will also be destroyed. Spreading the watermark throughout the spectrum of an image/audio ensures a large measure of security against unintentional or intentional attacks

In real spread spectrum schemes, a watermark is embedded into the host signal by adding a low energy pseudo-random noise sequence which is often modulated by the intended message.

Figure 3.1 displays the block diagram of a typical spread spectrum watermarking process. The watermark embedding process can occur in either a spatial domain or a frequency domain. For frequency-domain techniques, an orthogonal transformation, such as discrete cosine transform (DCT) or discrete wavelet transform (DWT) is applied to the host data f. The transformation decomposes the host data s into coefficients to which the watermark is embedded.

Let $\mathrm{s} = [s_1, s_2, ..., s_N]$ be the coefficients in watermark domain. The watermark consists of a sequence of numbers, $\mathbf{w} = [w_1, w_2, ..., w_N]$ with a given statistical distribution, such as a normal distribution $N(0,1)$ with zero mean and unit variance. The watermark sequence is embedded into the coefficients s according to the relationship,

$$\mathbf{x} = \mathbf{s} + \alpha\mathbf{w}, \tag{3.1}$$

where α is a scaling parameter which determines the extent to which one can alter $\mathbf{s}$ without changing the fidelity of the multimedia work, $\mathbf{x}$ is the watermarked coefficient.

Taking the inverse transform on x produces the watermarked data which should be perceptually identical to the original data.

To detect the existence of the watermark, the receiver transforms the watermarked data into watermark domain and obtains the extracted signal $\mathbf{x}$ which contains both the watermark signal and the original signal in watermark domain. In order to suppress the interference from the host signal and to obtain detection result with small probability of error, people often subtract the original signal from $\mathbf{x}$ before correlation-based watermark verification operation. The existence of the original watermark $\mathbf{w}$ within the watermarked signal is detected by calculating the similarity between the original watermark $\mathbf{w}$ and the extracted signal $\hat{\mathbf{w}}$. The similarity measure is given by the correlation coefficient as follows:

$$\rho(\mathbf{w}, \hat{\mathbf{w}}) = \frac{\sum_{i=1}^{N} w_i \hat{w}_i}{\sqrt{\sum_{i=1}^{N} w_i^2}\sqrt{\sum_{i=1}^{N} \hat{w}_i^2}}. \tag{3.2}$$

If the correlation coefficient is above a given threshold, the watermark is considered to be present; otherwise, the watermark is considered not to be present in the received signal. Figure 3.2 presents the diagram of spread spectrum watermark detection procedure.

As discussed in the introduction, we are interested in blind watermarking for which the original multimedia work is not available. Thus the original signal can be regarded as a major noise source in detection. Hartung, et al. [15] proposed a spread spectrum blind image watermarking system in which subtraction of the original data is replaced by the pre-filtering. The high-pass filtered watermarked image is then demodulated using exactly the same pseudo-noise signal previously used for watermark embedding. In this way, the filtered

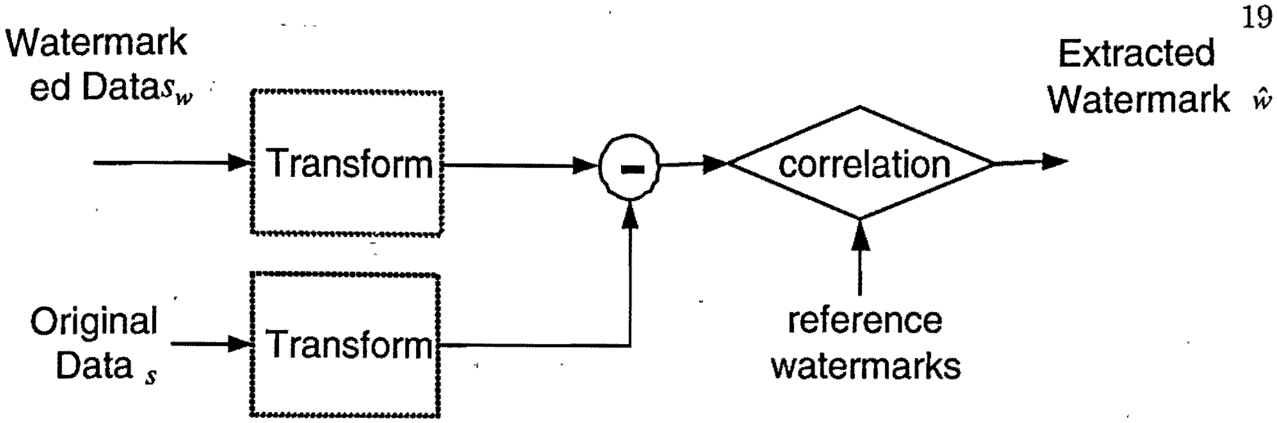


Figure 3.2: Typical block diagram of spread spectrum watermark extraction.

image is treated as the extracted watermark. The filtering based watermarking system is illustrated in Figure 3.3.

High-pass filtering can effectively suppress the original image's interference due to the fact that this interference is mainly contributed by low-frequency components, while the power spectrum of the original image at high-frequency is relatively small.

Wiener filter is a classic linear noise reduction filter. It is often used for image denoising applications. Exploiting Wiener filter in watermark extraction algorithm, as illustrated in Figure 3.4, we use the error signal at the output of the filter $v(x,y) = g(x,y) - f(x,y)$ as the extracted watermark.

3.3 Wiener Adaptive Filter Based Watermarking System

3.3.1 Wiener Filter

Wiener filter is the mean-square-error optimal stationary linear filter and based on the assumption that the power spectra of the ideal source and the noise are known. The goal of Wiener filtering is to obtain an estimate of the original signal from a degraded version of the signal. The degraded image $g(m,n)$ can be represented by

$$g(m,n) = f(m,n) + v(m,n), \tag{3.3}$$

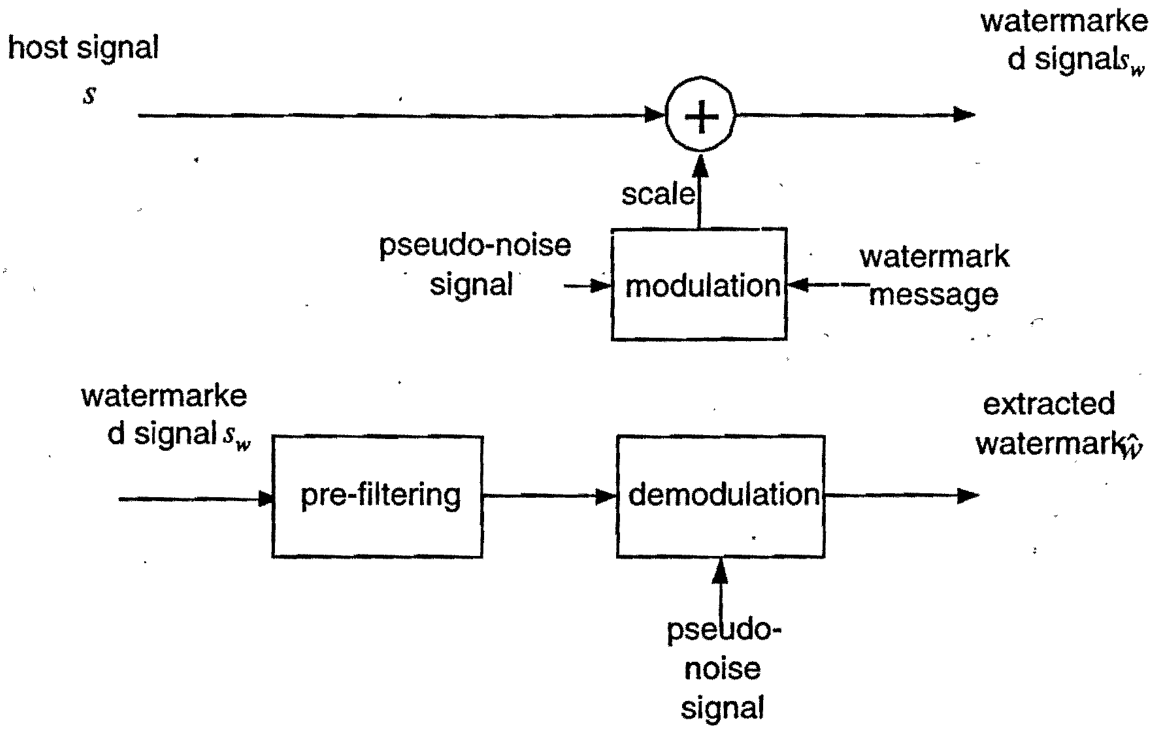


Figure 3.3: Filter based spread spectrum watermarking system.

where $f(m,n)$ is the nice, undegraded signal and $v(m,n)$ is the noise. Given the degraded signal $g(m,n)$ and some knowledge of the nature of $f(m,n)$ and $v(m,n)$, we want to come up with a function $h(m,n)$ that will output a good estimate of $f(m,n)$. This estimate is $p(m,n)$, and is defined by the following:

$$p(m,n) = g(m,n) * h(m,n), \tag{3.4}$$

$$P(\omega_1,\omega_2) = G(\omega_1,\omega_2)H(\omega_1,\omega_2). \tag{3.5}$$

where $P(\omega_1,\omega_2)$, $G(\omega_1,\omega_2)$ and $H(\omega_1,\omega_2)$ are the power spectra of $p(m,n)$, $g(m,n)$ and $h(m,n)$ respectively.

The Wiener filter generates an $h(x,y)$ that minimizes the mean square error, which is defined by:

$$E\{e^2(m,n)\} = E\{(g(m,n) - f(m,n))^2\}. \tag{3.6}$$

According to the orthogonality principle, the error, $e(m,n) = g(m,n)-f(m,n)$, is minimized

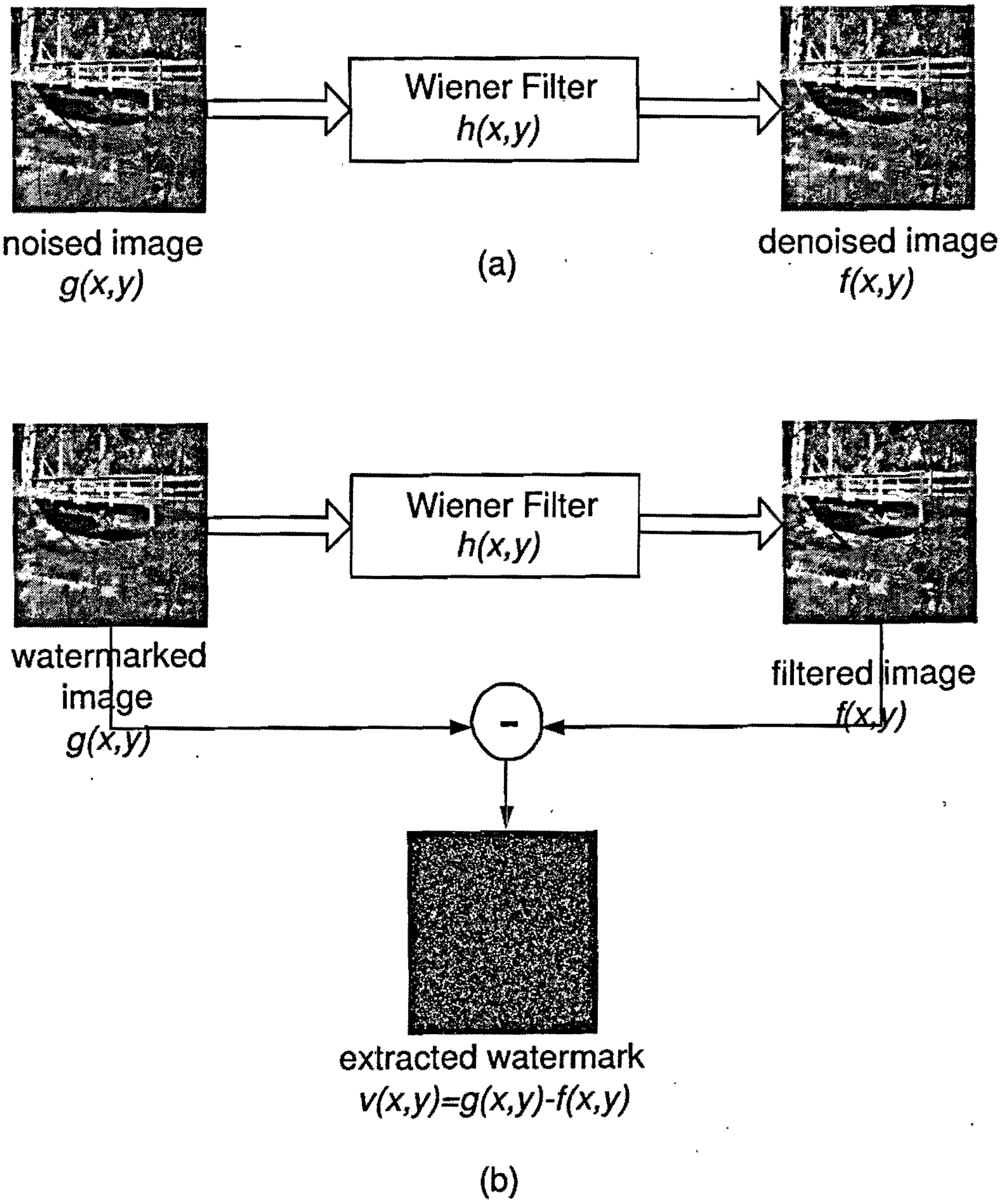


Figure 3.4: (a) Wiener filter for image denoising. (b) Wiener filter for watermark extraction.

by requiring that $e(m,n)$ be uncorrelated with any random variable of $g(m,n)$,

$$E\{e(m,n)g(m,n)\} = 0, \text{ for all } e(x,y) \text{ and } g(m,n)\ . \tag{3.7}$$

Then we have

$$\begin{aligned} E\{f(m,n)g(m,n)\} &= E\{(e(m,n)+p(m,n))g(m,n)\} \\ &= E\{p(m,n)g(m,n)\} \\ &= E\{(g(m,n)*h(m,n))g(m,n)\} \\ &= \textstyle\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} h(k_1,k_2)E\{g(x-k_1,y-k_2)g(m,n)\} \\ &= \textstyle\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} h(k_1,k_2)R_g(x-k_1-m,y-k_2-n\}, \end{aligned} \tag{3.8}$$

where $R_g(x,y)$ is the autocorrelation function of $g(x,y)$. So

$$R_{fg}(x,y) = h(x,y) * R_g(x,y), \tag{3.9}$$

and

$$H(\omega_1,\omega_2) = \frac{P_{fg}(\omega_1,\omega_2)}{P_g(\omega_1,\omega_2)}. \tag{3.10}$$

Suppose $f(x,y)$ is uncorrelated with $v(x,y)$,

$$R_{fg}(x,y) = R_f(x,y), \tag{3.11}$$

$$R_g(x,y) = R_f(x,y) + R_v(x,y), \tag{3.12}$$

and,

$$P_{fg}(\omega_1,\omega_2) = P_f(\omega_1,\omega_2), \tag{3.13}$$

$$P_g(\omega_1,\omega_2) = P_f(\omega_1,\omega_2) + P_v(\omega_1,\omega_2). \tag{3.14}$$

So,

$$H(\omega_1,\omega_2) = \frac{P_f(\omega_1,\omega_2)}{P_f(\omega_1,\omega_2) + P_v(\omega_1,\omega_2)}. \tag{3.15}$$

Since the power spectra $P_f(\omega_1,\omega_2)$ and $P_v(\omega_1,\omega_2)$ are real and nonnegative, $H(\omega_1,\omega_2)$ is also real and nonnegative. Therefore, the Wiener filter affects the spectral magnitude but not the phase.

3.3.2 2D Wiener Adaptive Filter

To obtain an accurate estimate of the power spectrum, an ensemble of many samples of the ideal image is required. However, in practical applications, it is unlikely that there will be an ensemble of ideal image samples available for estimation. In most applications, only the image to be restored is available and all prior knowledge about the ideal image signal has to be estimated from it. Hence, the power spectrum estimated from this single copy of the degraded image is far from the true power spectrum of the ideal image. For this reason, it is expected that the restoration filter is no longer optimal because of the lack of accurate prior information.

Although the Wiener filter is optimally derived, the success of Wiener filter in restoring real-world images depends on accurate estimation of the image power spectrum. In general, an image is modelled as an inhomogeneous random field. The Wiener filter requires estimating the signal mean μ_f, noise mean μ_v, signal power spectrum $P_f(\omega 1, \omega 2)$, and noise power spectrum $P_v(\omega 1, \omega 2)$. They can be estimated locally in adaptive Wiener filtering.

In [25], Lee proposed the whole calculation procedure. The additive noise $v(m,n)$ is assumed as zero mean and white with variance of σ_v^2. Its power spectrum $P_v(\omega_1, \omega_2)$ is then given by $P_v(\omega_1, \omega_2) = \sigma_v^2$. In a small local region the signal $f(x,y)$ is assumed homogeneous. Within the local region, the signal $f(x,y)$ is modelled by

$$f(m,n) = \mu_f + \sigma_f w(m,n), \tag{3.16}$$

where μ_f and σ_f are the local mean and standard deviation of $f(m,n)$, and $w(m,n)$ is zero-mean white noise with unit variance. Within the local region, the Wiener filter $H(\omega_1, \omega_2)$ and $h(m,n)$ are given by:

$$H(\omega_1, \omega_2) = \frac{P_f(\omega_1, \omega_2)}{P_f(\omega_1, \omega_2) + P_v(\omega_1, \omega_2)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_v^2}, \tag{3.17}$$

$$h(m,n)) = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_v^2} \delta(m,n). \tag{3.18}$$

Then, the restored image $p(m,n)$ within the local region can be expressed as

$$\begin{aligned} p(m,n) &= \mu_f + (g(m,n) - \mu_f) * \tfrac{\sigma_f^2}{\sigma_f^2+\sigma_v^2}\delta(m,n) \\ &= \mu_f + \tfrac{\sigma_f^2}{\sigma_f^2+\sigma_v^2}(g(m,n) - \mu_f), \end{aligned} \tag{3.19}$$

where μ_f and σ_f are assumed updated at each pixel,

$$p(m,n) = \mu_f(m,n) + \frac{\sigma_f^2(m,n)}{\sigma_f^2(m,n) + \sigma_v^2(m,n)}(g(m,n) - \mu_f(m,n)). \tag{3.20}$$

The new blind watermark detection technique is based on adaptive denoising filter. Given a received corrupted watermarked image X_c, pixel-wise adaptive denoising filtering is applied to X_c. This pixel-wise adaptive denoising filtering is based on statistics estimated from a local neighborhood of each pixel. The local image mean value μ and variance σ^2 are estimated using neighborhoods of size N-by-M:

$$\mu(m,n) = \frac{1}{NM} \sum_{m,n \in \eta} X_c(m,n), \tag{3.21}$$

$$\sigma^2(m,n) = \frac{1}{NM} \sum_{m,n \in \eta} X_c^2(m,n) - \mu^2, \tag{3.22}$$

where η is the N-by-M local neighborhood of each pixel in the image X_c. To extract watermark, an estimate of local noise (watermark) variance, ν^2 is necessary. Then a denoised image can be obtained by a pixel-wise adaptive Wiener denoising filtering according to [26]:

$$S_c(m,n) = \mu + \frac{\sigma^2 - \nu^2}{\sigma^2}(X_c(m,n) - \mu) = X_c(m,n) - \frac{\nu^2}{\sigma^2}(X_c(m,n) - \mu). \tag{3.23}$$

Then an estimation of the watermark W can be obtained by

$$\hat{W} = X_c - S_c = \frac{\nu^2}{\sigma^2}(X_c(m,n) - \mu). \tag{3.24}$$

Note that the estimate $\hat{W}$ may also contain other noises besides the desired spread spectrum watermark. However, as long as these noises are uncorrelated with the spread spectrum watermark, they will be eliminated by the subsequent correlation detector.

3.3.3 Wiener Adaptive Filter based Detector

From (3.22) and (3.23), it is clear that the local noise variance estimation is critical in extracting the watermark. Various methods have been proposed to estimate the local noise variance in digital image debluring and enhancement applications. Different to image enhancement applications, for image watermarking system, not only we have the watermarked image, the possible embedded watermarks and the watermark embedding strategy are also known to the detector. Hence watermark detection algorithm incorporates the reference watermarks may have further improved performance.

Here we propose two schemes to estimate the local noise variance ν^2. And therefore two types of blind detectors can be constructed.

Type I detector: The first scheme of the local noise (watermark) variance estimation is based on the method in [26]. It is assumed that the noise (watermark) variance ν^2 is uniform in the whole image and the original image has very low variation all over the image. Therefore, a global noise variance can be estimated by:

$$\nu^2 = E\{\sigma^2(m,n)\}. \tag{3.25}$$

Type II detector: In the second scheme of the local noise (watermark) variance estimation, we do not assume the uniform local noise (watermark) variance and the image flatness. Instead, the local noise (watermark) variance ν^2 is estimated adaptively according to the local statistics of the watermark:

$$\nu^2 = \frac{\sigma_w^2(m,n)}{E\{\sigma_w^2(m,n)\}} E\{\sigma^2(m,n)\}, \tag{3.26}$$

where σ_w^2 is the local variance of the watermark to be detected. It is calculated in the same fashion as in (3.22). If n watermarks may be present in the observed image, n different filtered images will be generated and used for detection. Only the real watermark is expected to have largest correlation coefficient with its corresponding extracted watermark. Figure 3.5 outlines the steps to embed and extract watermark.

The above-mentioned denoising filtering tailors itself to the local image variance. When the local variance is large, the filter performs little smoothing. When the variance is small,

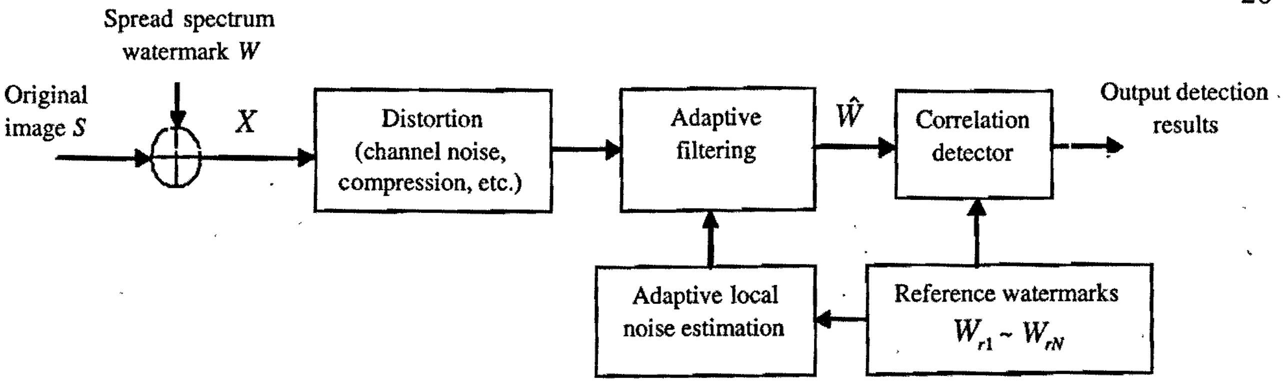


Figure 3.5: Block diagram for the Wiener denoising filter based watermarking scheme

the filter performs more smoothing. Note that theoretical Wiener filtering system assumes additive Gaussian noise. Previous work in the robust watermarking area has pointed out that the effect of distortions on the overall watermarked signal can be modelled as additive Gaussian noise [5]. So the extracted watermark is possibly corrupted by additive Gaussian noise. As we use spread spectrum watermark technique, additive Gaussian noise will not impact the final result of correlation-based detector.

3.4 Simulation

The 512×512 images of Lena and Bridge are used to demonstrate the robustness of the presented new blind watermarking detection method. Note that these two images are typical in that Lena image has rich grayscale information and the Bridge image is full of details and edges. Spread spectrum watermark is generated as 512×512 matrix, where each value is chosen independently according to $\mathcal{N}(/, \infty)$. The value of α is set to 5 to ensure that the change introduced by the watermark is perceptionally invisible. Figure 3.6 shows the results of digital watermarked images on Lena and Bridge.

First, type I detector is used and the noise variance is estimated according to (3.25). The size of neighborhood η is 3×3. By subtracting the filtered image S_c from X_c, the watermark is extracted. Then type II detector is used and the noise variance is estimated according to (3.26). All possible watermarks are examined. The similarity is evaluated with

Figure 3.6: Digital watermarking for Lena and Bridge.

respected to all the reference watermarks. The highest correlation coefficient indicates the real watermark. In the following demonstrations, the performance of the new watermark detectors are evaluated under three common image distortions: additive noise corruption, low-pass filtering and lossy compression. The results are analyzed and compared with the high-pass filter based detector proposed in [15]. Figure 3.7 shows the performance of three methods in additive white Gaussian noise, which is a simple simulation of channel noise. According to Figure 3.7, the detector's response value is highly related to the quality of the observed image. Given a watermarked image, the new locally adaptive denoising filter based detectors provide greater correlation coefficient than the high-pass filter based detector in [15]. Type II adaptive watermark detector also provides better performance for both images than type I detector does.

Figure 3.8 shows the results of the test for robustness against low-pass filtering distortion. By comparing the correlation coefficient values of the Wiener filter based method (0.6) and

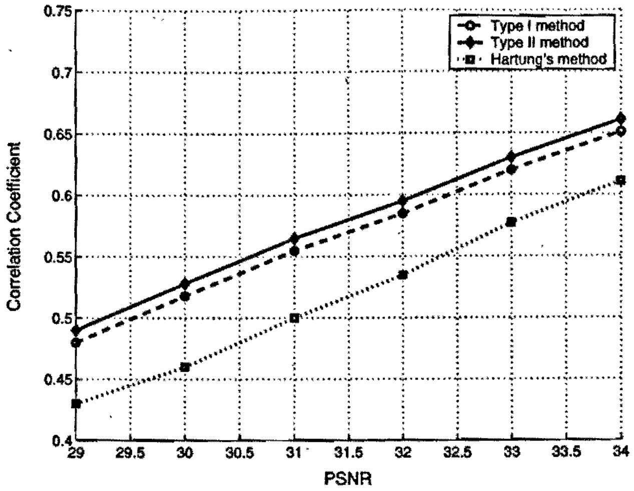


Figure 3.7: Performance of three different detectors subject to additive noise corruption.

the Hartung method (0.45), we can see that our method is more robust to low-pass filtering attacks.

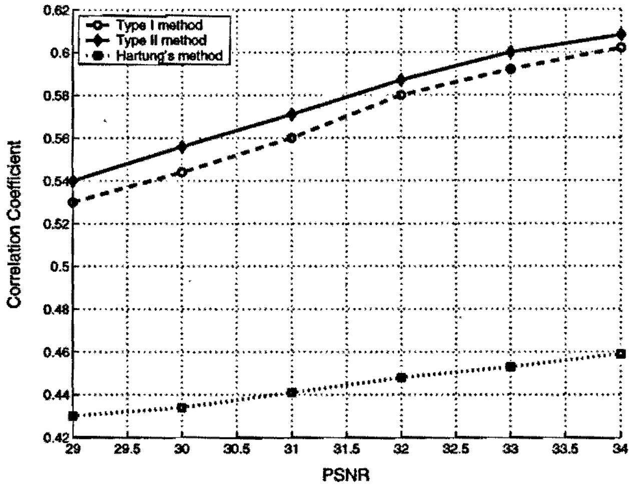


Figure 3.8: Performance of three different detectors subject to low-pass filtering corruption.

Figure 3.9 shows the detection response against JPEG compression with various compression degrees. It can be seen that, even after heavy compression, the adaptive filter based methods can still reliably detect the correct watermark. The correlation coefficient of the correct watermark is about 0.3, which is much higher than that of the high-pass filter based method in [15]. Again, type II adaptive watermark detector provides better detection performance for both images than type I detector.

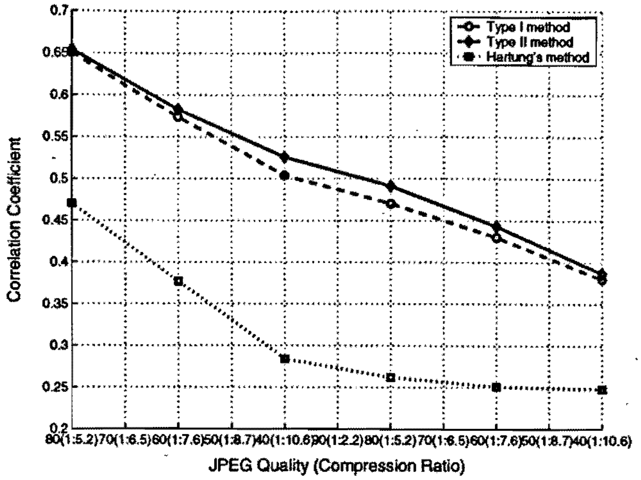


Figure 3.9: Performance of three different detectors subject to JPEG compression distortion.

Finally, Figure 3.10 shows the detection results against the state-of-art image compression method JPEG2000. When 0.6bpp (about 1:13.33) JPEG2000 compression is applied, the correlation coefficient achieved by Hartung's method becomes 0.2431, while of the same processing is applied to our method, the correlation coefficient value becomes around 0.5.

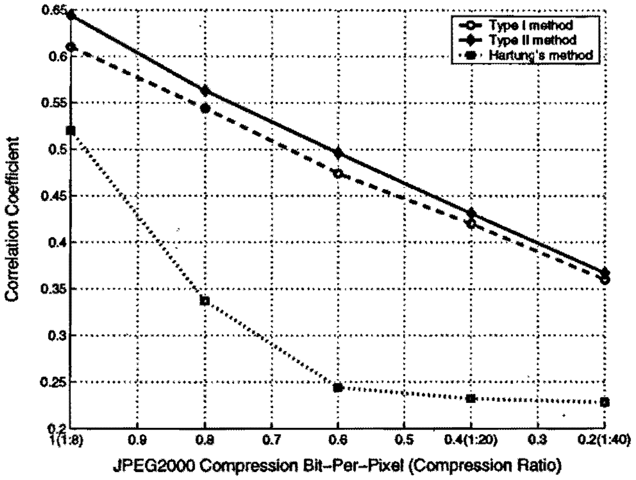


Figure 3.10: Performance of three different detectors subject to JPEG2000 compression distortion.

The experimental results demonstrate that even if the watermarked image has undergone severe distortion, the detector designed based on adaptive denoising filter can still detect the correct watermark.

3.5 Summary

This chapter focuses on developing blind watermark detection scheme. The presented scheme is developed by exploiting denoising technique, based on the principle that the additive spread spectrum watermark can be treated as uncorrelated noise with respect to the host data. A locally adaptive denoising filtering scheme is employed to construct the watermark detector. Adaptive local noise/watermark variance estimation schemes are presented to achieve best filtering performance. Two types of blind detectors are developed based on different local noise variance estimation schemes. Experimental results show that the presented methods are very effective and robust against most image processing attacks, such as lossy compression, noise addition and spatial filtering, etc., and they have superior detection performance compared to some conventional methods.

Chapter 4

Near Minimum Distortion LUT Embedding

4.1 Introduction

In this chapter, we focus on the LUT embedding algorithms. Look-up table (LUT) embedding is a simple embedding technique used to hide information into multimedia work for copyright protection, transaction tracking or content annotation. The LUT is often associated with a cryptographic key, thus provides security to embedding. This chapter studies the distortion introduced by LUT embedding where the maximum allowable run is limited to 2. Here run means the largest number of consecutive 0's or 1's in LUT. We find that designing LUT according to the distribution of the host data and the watermark data to be embedded can greatly reduce the distortion from LUT embedding. Hence a practical near-minimum-distortion look-up table design method is proposed. Meanwhile, security and robustness of the designed information hiding system are almost maintained. We apply this method into a wavelet domain image watermarking system. Because only significant wavelet coefficients can be selected to embed the watermark, an Expectation-Maximization (EM) algorithm based method is employed to model the statistical distribution.

4.2 Overview of the LUT Embedding

A LUT is a random sequence of 0's and 1's, with runs of 0's and 1's being limited in length. It also constitutes the key for the watermark extraction algorithm. Every possible value of the host data is quantized using a quantization function ($Q(\cdot)$) to a small set of values, equal in number to the size of the LUT. For example, a uniform quantizer with cell width q maps the original signal to $kq, k = 1, \cdots, K$, where K is the size of the LUT. The table then maps the quantized value to "0" or "1". To embed a "1" in a coefficient, the coefficient is unchanged if the entry of the table corresponding to that coefficient is also a "1". If the entry of the table is "0", then the coefficient is changed to its nearest neighboring value for which the entry is "1". The embedding of a "0" is similar. The look-up function ($Lookup(\cdot)$) simply returns a "0" or "1" depending upon the input index,

$$Lookup(s) = \text{value in Look-up table at index } s \tag{4.1}$$

The $LUT(\cdot)$ function takes the value of the original singal as the input and maps it to a "0" or "1" according to the LUT. Thus, the $LUT(\cdot)$ function is actually a simple composition of the lookup and the quantization functions:

$$LUT(s) = Lookup(Q(s)) \tag{4.2}$$

Figure 4.1 shows the general process of LUT embedding algorithm. An orthogonal transformation T_w decomposes the host data s into coefficients x in the watermark domain to which the watermark w is embedded. Then the coefficients are quantized and mapped according to LUT.

The entire process altering a coefficient can be abstracted into the following formula:

$$x = \begin{cases} s & \text{if } LUT(s) = b \\ s + d & \text{if } LUT(s) \neq b,\ d = \min_{|d|}(LUT(s+d) = b), \end{cases} \tag{4.3}$$

where s is the original coefficient, x is the marked one, b is the bit to be embedded.

For LUT embedding, once the LUT is known to the detector, the watermark can be extracted easily through a simple lookup from the LUT. The table is looked up as

$$\hat{b} = LUT(\hat{x}), \tag{4.4}$$

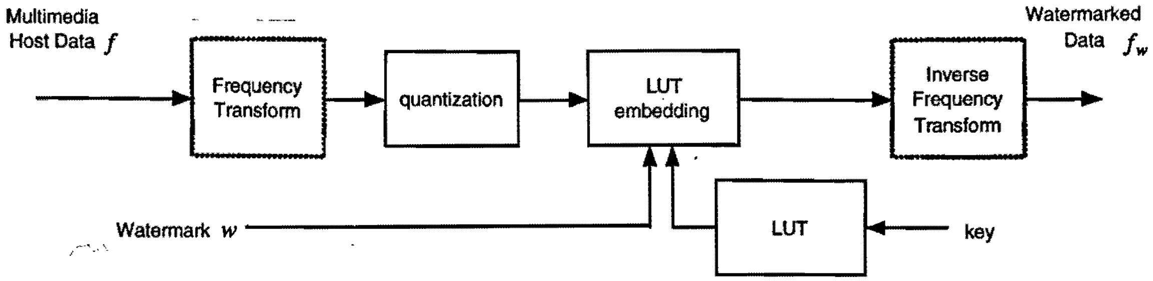


Figure 4.1: General LUT Embedding Algorithm

where $\hat{b}$ is the extracted bit and $\hat{x}$ is the watermark embedded, possibly corrupted signal.

A typical LUT embedding algorithm is the odd-even embedding. First, a uniform quantization function $Q(\cdot)$ is defined which partitions the signal space $\Re$ into subsets as illustrated in Figure 4.2. The host data is mapped to the nearest even numbered quantization point to embed a "0" and the nearest odd numbered quantization point to embed a "1". Thus a global relationship between the watermark bit and the marked signal is deterministically enforced.

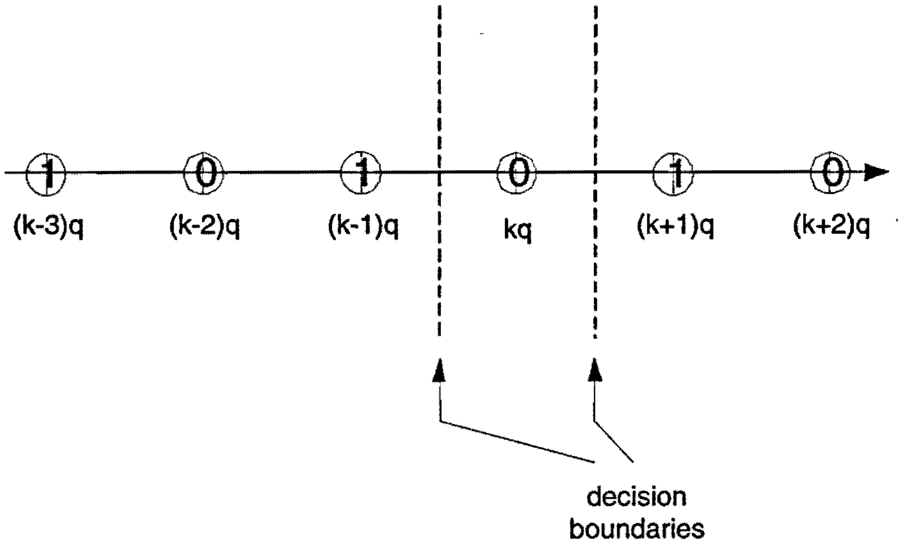


Figure 4.2: The odd-even embedding

The watermark bit is extracted by the following way,

$$\hat{b} = \begin{cases} 0 & \text{if } \frac{Q(\hat{x})}{q} \text{ is even} \\ 1 & \text{if } \frac{Q(\hat{x})}{q} \text{ is odd.} \end{cases} \tag{4.5}$$

In the odd-even embedding scheme, the table entries for embedding "1" and "0" are arranged in an interleaving order, $\{..., Lookup((k-1)q) = 0, Lookup(kq) = 1, Lookup((k+1)q) = 0, Lookup((k+2)q) = 1, ...\}$, which is also described as run=1 LUT embedding in [2]. It is pointed out that LUT embedding with larger run constraints introduces larger distortion but have smaller probability of detection error.

During LUT embedding, when $Lookup(Q(x))$ does not match the bit to be embedded b, we need to find a nearby entry in LUT that is associated with b. As such, the run of "1" and "0" entries of an LUT becomes a main concern which needs to be constrained to avoid excessive modification on the feature.

4.3 Robustness Issue

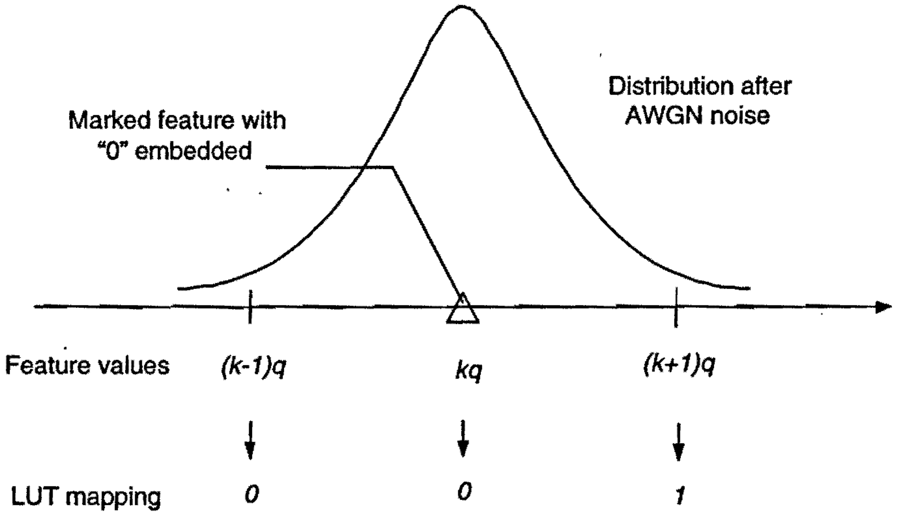


Figure 4.3: Illustration of reduced detection errors of LUT embedding as the maximum allowable run r increases.

To quantify the robustness in terms of the probability of detection error, we assume that the watermarked feature is at kq and the additive noise follows i.i.d. Gaussian distribution $\mathcal{N}(/, \sigma^{\in})$ with zero mean and variance σ^2. The probability of noise pushing a feature to other intervals that are far away from kq is small due to the fast decay of the tails of Gaussian distribution, so the probability of detection error can be approximated by considering only the nearby intervals around kq. An example is shown in Figure 4.3. When noise drags the watermarked feature away from kq to z, we will encounter detection error only when $LUT(z) \neq LUT(kq)$. For LUT embedding with a maximum allowable run of 2, there are three cases for the LUT entries of $(k-1)q$, kq and $(k+1)q$:

- Case 1: $\{Lookup(kq) = Lookup((k-1)q), Lookup(kq) \neq Lookup((k+1)q)\}$;
- Case 2: $\{Lookup(kq) \neq Lookup((k-1)q), Lookup(kq) = Lookup((k+1)q)\}$;
- Case 3: $\{Lookup(kq) \neq Lookup((k-1)q), Lookup(kq) \neq Lookup((k+1)q)\}$.

Table 4.1 shows all the possible combinations of the binary look-up table entries $(k-1)q$, kq and $(k+1)q$.

Table 4.1: All possible cases for LUT entries of $(k-1)q$, kq and $(k+1)q$ are listed where each of them can only be "0" or "1".

$(k-1)q$	kq	$(k+1)q$	
0	0	1	Case 1
0	1	0	Case 3
0	1	1	Case 2
1	0	0	Case 2
1	0	1	Case 3
1	1	0	Case 1

Suppose that each entry of the LUT has the equal probability to be "0" or "1".

$$P(Lookup(kq) = 0) = P(Lookup(kq) = 1) = \frac{1}{2}. \tag{4.6}$$

Using (4.6) and Table 4.1, we can find the probability of the first case as

$$P(Lookup(kq) = Lookup((k-1)q), Lookup(kq) \neq Lookup((k+1)q)) = \frac{1}{3}. \tag{4.7}$$

Similarly, the probabilities of the other two cases are

$$P(Lookup(kq) \neq Lookup((k-1)q), Lookup(kq) = Lookup((k+1)q)) = \frac{1}{3}, \tag{4.8}$$

$$P(Lookup(kq) \neq Lookup((k-1)q), Lookup(kq) \neq Lookup((k+1)q)) = \frac{1}{3}. \tag{4.9}$$

Thus the probability of detection error under Gaussian noise can be approximated by

$$\begin{aligned} P_e^{r=2} &\approx P(Lookup(kq) \neq Lookup((k-1)q), Lookup(kq) \neq Lookup((k+1)q)) \cdot 2 \cdot Q(\tfrac{q}{2\sigma}) \\ &+P(Lookup(kq) = Lookup((k-1)q), Lookup(kq) \neq Lookup((k+1)q)) \cdot Q(\tfrac{q}{2\sigma}) \\ &+P(Lookup(kq) \neq Lookup((k-1)q), Lookup(kq) = Lookup((k+1)q)) \cdot Q(\tfrac{q}{2\sigma}) \\ &= \tfrac{4}{3} \cdot Q(\tfrac{q}{2\sigma}) \end{aligned} \tag{4.10}$$

where the Q-function $Q(x)$ is the tail probability of a Gaussian random variable $N(0,1)$.

In contrast, for LUT with a maximum run of 1 (or equivalently, the odd-even embedding), detection error occurs as soon as the noise is strong enough to drag the watermarked feature to the quantization intervals next to the kq interval. The probability of detection error for this embedding is

$$P_e^{r=1} \approx 2 \cdot Q(\frac{q}{2\sigma}). \tag{4.11}$$

The above analytic approximations of the probability of detection error indicate that LUT embedding with maximum allowable run of 2 can potentially provide higher robustness than the commonly used quantization embedding with equivalent run 1.

4.4 Near Minimum Distortion LUT Embedding

4.4.1 Distortion Analysis

In LUT embedding, uniform quantization $Q(\cdot)$ divided the input signal space into K equi-spaced levels. Then if the k-th entry of LUT is b, the data samples of signal s in the quantization cell of $[(k-1/2)q, (k+1/2)q]$ to be embedded b is rounded to kq, the mean square distortion produced by this operation is calculated as

$$D_{kq}(s) = \int_{(k-1/2)q}^{(k+1/2)q} |s-kq|^2 f(s)ds, \tag{4.12}$$

where $f(s)$ is the Probability Density Function of s. However, if the desired bit for s is not b, the host data must be mapped to the nearest quantization points corresponding to the desired bit $(k \pm l)q, l > 0$. If the maximum allowable run is 2, the entry next to the k-th entry, i.e. either the $k-1$-th entry or $k+1$-th entry or both matches the desired bit. We subdivide the above situation into 3 cases as the previous section. In all the cases, further distortion will be introduced. If it is the first case, $\{Lookup(kq) = Lookup((k-1)q), Lookup(kq) \neq Lookup((k+1)q)\}$, the samples of s which are in the quantization cell of $[(k-1/2)q, (k+1/2)q]$ have to be rounded to $(k+1)q$. The distortion is

$$\begin{aligned} D_{Case1}(s) &= \int_{(k-1/2)q}^{(k+1/2)q} |s-(k+1)q|^2 f(s)ds \\ &= D_{kq} + q^2 \int_{(k-1/2)q}^{(k+1/2)q} f(s)ds - 2q \int_{(k-1/2)q}^{(k+1/2)q} (s-kq)f(s)ds. \end{aligned} \tag{4.13}$$

Similarly, in the second case, the samples in k-th cell have to be mapped to $(k-1)q$. we have

$$D_{Case2}(s) = D_{kq} + q^2 \int_{(k-1/2)q}^{(k+1/2)q} f(s)ds + 2q \int_{(k-1/2)q}^{(k+1/2)q} (s-kq)f(s)ds \tag{4.14}$$

In the third case, we have two nearest quantization points $(k+1)q$ and $(k-1)q$ corresponding to the desired bit simultaneously, then the original features in the range of $[(k-1/2)q, kq]$ will be rounded to $(k-1)q$, and the features in the other half interval $[kq, (k+1/2)q]$ will be mapped to $(k+1)q$. The distortion will be composed by two parts:

$$\begin{aligned} D_{Case3}(s) &= \int_{(k-1/2)q}^{kq} [s-(k-1)q]^2 f(s)ds + \int_{kq}^{(k+1/2)q} [s-(k+1)q]^2 f(s)ds \\ &= D_{kq} + q^2 \int_{(k-1/2)q}^{(k+1/2)q} f(s)ds + 2q \left[\int_{(k-1/2)q}^{kq} (s-kq)f(s)ds - \int_{kq}^{(k+1/2)q} (s-kq)f(s)ds\right] \end{aligned} \tag{4.15}$$

If the feature is approximately uniformly distributed within each cell or only fine quantizer is utilized, the last terms of (4.13) and (4.14) are close to 0, and the last term of (4.15), which is in the range of $[-q^2 \int_{(k-1/2)q}^{(k+1/2)q} f(s)ds, 0]$, approximates to $\frac{-q^2}{2} \int_{(k-1/2)q}^{(k+1/2)q} f(s)ds$.

For a binary data hiding system, we can divide the features into two categories: the features that are used to embed bit "0", denoted by s_0, and the features that are used to embed bit "1", denoted by s_1. The PDFs of s_0 and s_1 are $f_0(s_0)$ and $f_1(s_1)$, respectively. First we consider the overall mean squared distortion due to quantization only,

$$\begin{aligned} MSE_{quan} &= \sum_{k=1}^{K} \left[\int_{(k-1/2)q}^{(k+1/2)q} |s_0 - kq|^2 f_0(s_0)ds_0 + \int_{(k-1/2)q}^{(k+1/2)q} |s_1 - kq|^2 f_1(s_1)ds_1\right] \\ &= \sum_{k=1}^{K} \left[D_{kq}(s_0) + D_{kq}(s_1)\right]. \end{aligned} \tag{4.16}$$

Now consider that each of the K LUT entries is either "0" or "1". In all K quantization cells, either the data to be embedded "0" or the data to be embedded "1" are mapped to their local reconstruction points. Other data have to be mapped to neighboring reconstruction points where the above three cases appear. According to (4.13)-(4.16), the overall LUT embedding distortion can be formulated as

$$\begin{aligned} MSE_w &= MSE_{quan} + q^2 \textstyle\sum_{k=1}^{K} \alpha_{0,k} \int_{(k-1/2)q}^{(k+1/2)q} f_0(s_0)ds_0 + q^2 \sum_{k=1}^{K} \alpha_{1,k} \int_{(k-1/2)q}^{(k+1/2)q} f_1(s_1)ds_1 \\ &\quad - \tfrac{q^2}{2} \textstyle\sum_{k=1}^{K} \beta_{0,k} \int_{(k-1/2)q}^{(k+1/2)q} f_0(s_0)ds_0 - \tfrac{q^2}{2} \sum_{k=1}^{K} \beta_{1,k} \int_{(k-1/2)q}^{(k+1/2)q} f_1(s_1)ds_1 \end{aligned} \tag{4.17}$$

where α and β can be 0 or 1, $\alpha(0,k) = 1$ if the k-th reconstruction point is for "1" embedding, similarly $\alpha(1,k) = 1$ when the k-th LUT entry is "0", hence $\alpha(0,k) = mod(1 - \alpha(1,k), 2)$, only when Case 3 appears β is equal to 1, that is: if $\{Lookup((k-1)q) \neq Lookup(kq), Lookup(kq) \neq Lookup((k+1)q), Lookup(kq) = 1\}$, $\beta_{0,k} = 1$; if $\{Lookup((k-1)q) \neq Lookup(kq), Lookup(kq) \neq Lookup((k+1)q), Lookup(kq) = 0\}$, $\beta_{1,k} = 1$.

$\int_{(k-1/2)q}^{(k+1/2)q} f_0(s_0)ds_0$ and $\int_{(k-1/2)q}^{(k+1/2)q} f_1(s_1)ds_1$ represent the probability that s_0 and s_1 fall into the k-th quantization cell of $Q(\cdot)$, respectively. We denote them as $P_{0,k}$ and $P_{1,k}$, i.e.

$$P_{0,k} = \int_{(k-1/2)q}^{(k+1/2)q} f_0(s_0)ds_0, P_{1,k} = \int_{(k-1/2)q}^{(k+1/2)q} f_1(s_1)ds_1. \tag{4.18}$$

Thus we rewrite (4.17) as

$$MSE_w = MSE_{quan} + q^2 \sum_{k=1}^{K} \alpha_{0,k} P_{0,k} + q^2 \sum_{k=1}^{K} \alpha_{1,k} P_{1,k} - \frac{q^2}{2} \sum_{k=1}^{K} \beta_{0,k} P_{1,k} - \frac{q^2}{2} \sum_{k=1}^{K} \beta_{1,k} P_{1,k}. \tag{4.19}$$

If the original feature follows uniform distribution, the probabilities that the feature falls into each quantization cell will be exactly the same, then various LUTs have same overall distortion. Here we assume the distribution of the host signal is nonuniform. It is reasonable, as in real world most signals are not uniform. For example, the wavelet coefficients of a natural image do not follow uniform distortion. The probabilities that a nonuniform signal falls in each quantization cell are different to each other. Embedding watermark according to different LUT scheme can produce different distortion. From (4.19), we can see that the parameters $\alpha_{0,k}, \alpha_{1,k}, \beta_{0,k}$ and $\beta_{1,k}, k = 1, \cdots, K$ corresponding to each LUT scheme is unique. Therefore, various distortion can be obtained with various LUT schemes. Among them, the LUT which achieves near minimum distortion is our target.

4.4.2 Near Minimum Distortion LUT Embedding Algorithm

Figure 4.4 shows an example of wavelet coefficients which will be embedded binary watermark. The number of coefficients to be embedded "0" and "1" fall into each quantization cell is different to each other. We can design a variety of run of 2 LUTs, but only the one which achieves minimum distortion is what we want.

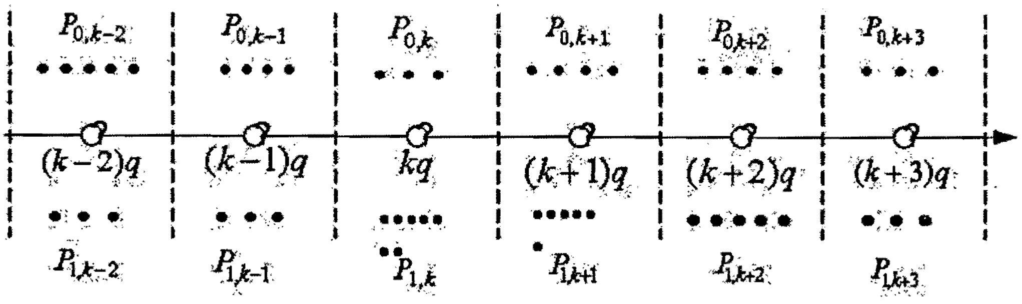


Figure 4.4: A minimum-distortion LUT needs to be generated according to the given wavelet coefficient distribution.

According to (4.19), the LUT embedding distortion is determined by the LUT (α and β) and the distribution of s_0 and s_1. We obtain the near-minimum distortion look-up table by looking at the probabilities that the coefficients to be embedded "0" and "1" falls into each cell, $P_{0,k}$ and $P_{1,k}$, $k = 1, 2...K$. First we sort $P_{0,k}$ and $P_{1,k}$ in descending order. The sorting result is a probability queue like $P_{0,k}, P_{1,k+1}, P_{0,k+2}, P_{1,k}, ...$ Then the look-up table is built in the way that the entry corresponding to the largest probability are set to its corresponding bit in priority. For example, if $P_{0,k}$ is the largest in the probability queue currently and the look-up table entry *Lookup*(kq) is still not determined yet, we set *Lookup*(kq) $= 0$. After each operation, we remove the largest probability value from the queue and move on the next. There is a rule we must keep in mind due to the maximum run constraint (r=2): the assignment should also satisfy the present maximum run number $r = 2$, i.e. the maximum run for "0" or "1" must be equal or less than 2 and at the border of the quantizer $Lookup(Q(\min s)) \neq Lookup(Q(\min s) + q), Lookup(Q(\max s)) \neq Lookup(Q(\max s) - q)$. The algorithm for run of 2 can be summarized in the following three steps:

STEP 1: Arrange the probabilities that the original feature associated with the desired bit falls into each quantization cell in descending order.

STEP 2: We find the largest probability $P_{j,k}$ from the above queue, where $j = 0$ or 1. If *Lookup*(kq) has been determined, go to **STEP 3**. Otherwise, among the determined look-up table entries, if any of the following situation occurs, $\{Lookup((k-2)q) = Lookup((k-1)q) = l\}$, $\{Lookup((k-1)q) = Lookup((k+1)q) = l\}$, $\{Lookup((k+1)q) = Lookup((k+2)q) = l\}$, $\{k = 2, Lookup(q) = l\}$, $\{k = L-1, Lookup(L) = l\}, l \in \{0,1\}$, *Lookup*($kq$) is set to the complement of l, $Lookup(kq) = mod(l+1, 2)$, no matter j value; otherwise $Lookup(kq) = j$, the original feature corresponds to $P_{j,k}$ is not shifted to other quantization points.

STEP 3: $P_{j,k}$ is removed from the queue. If the queue is not empty, go back to **STEP 2**.

To illustrate the above near minimum distortion look-up table generation algorithm, an example is provided in Figure 4.5.

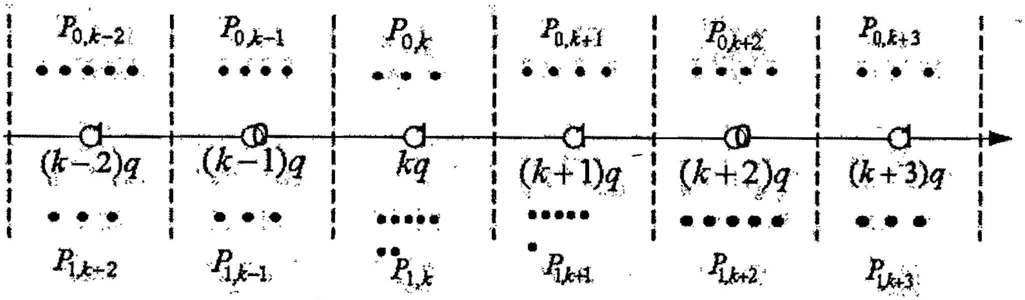


Figure 4.5: Example of a max-run=2 minimum distortion look-up table.

4.5 Significant Coefficient Selection Based on a Gaussian Mixture Model in the Wavelet Domain

In our scheme, only wavelet coefficients with large magnitude are selected to bear watermark. In general, these coefficients do not change significantly after image processing and compression attack. We propose a statistical method to pick the embeddable coefficients based on a Gaussian Mixture Model in a wavelet subspace by Expectation-Maximization (EM) algorithm [27]. The wavelet coefficients have a peaky, heavy-tailed marginal distribution, which record image texture and edge information at different scales [28]. Only a few significant coefficients take large values at the positions where edges occur, while most others take small values. This statistical characteristic can be expressed by using a two component Gaussian mixture:

$$p(w_i) = p_s \cdot g(w_i, 0, \sigma_s^2) + p_l \cdot g(w_i, 0, \sigma_l^2), \tag{4.20}$$

$$p_s + p_l = 1. \tag{4.21}$$

where the class of small coefficients is represented by subscript "s" and the class of large coefficients by subscript "l". The a priori probabilities of the two classes are represented by p_s and p_l, respectively. The Gaussian component corresponding to the small coefficients has a relatively small variance σ_s^2, capturing the peakiness around zero, while the component corresponding to the large state has a relatively large variance σ_s^2, capturing the heavy tails. An EM algorithm as in [28] can then be applied to find out the Gaussian mixture model by obtaining the model parameters $[p_s, p_l, \sigma_s^2, \sigma_l^2]$. The Gaussian mixture model is then used to find large coefficients for watermarking. The watermark is only embedded into the class of large coefficients because modifying coefficients which represent the edge information will introduce less perceptual degradation. We select significant coefficients by examining the coefficient magnitude that is larger than a threshold ρ determined by the Gaussian mixture model. That is, suppose that there are m coefficients in a detail subband s, the number of coefficients which is larger than ρ is approximately mp_l and the number of coefficients less than ρ is approximately mp_s. Coefficient $w_s(x, y)$ will be chosen for watermark embedding if $||w_s(x, y)|| \geq \rho$.

4.6 Simulation

The proposed watermarking scheme is tested on seven images of different types. We evaluate the quality of watermarked and attacked image by peak-signal-to-noise-ratio (PSNR), and the robustness under several intentional/unintentional attacks is denoted by bit correct ratio (BCR). First we inserted binary watermark into the images by applying the new method. Figure 4.6 shows one example (Lena). The modified significant coefficients are mainly at the edge of the image. The watermark robustness to common operations such as image compression is tested. The discrete cosine transform (DCT) based coding system, JPEG baseline, and the discrete wavelet transform (DWT) based coding system, JPEG 2000, are the two compression attacks in our tests. We evaluate the robustness by the average BCR for all test images. As shown in Figure 4.7, the BCR of the extracted watermark is larger than 75% until the compression quality factor is smaller than 60.

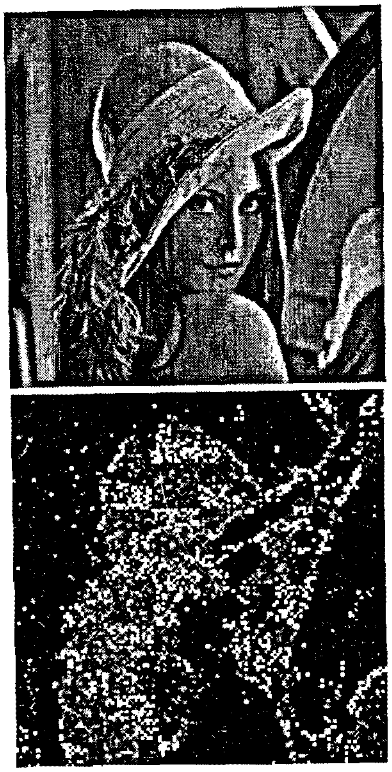


Figure 4.6: The watermarked image and the difference from the original image with black denoting zero difference.

Since the watermark is embedded in the wavelet domain, the presented algorithm has perfect robustness against DWT based JPEG 2000 compression attack. The results are shown in Figure 4.8. The decoded watermark can be 100% reconstructed after JPEG 2000 compression of 1bpp and is reliable until the compression bit-rate smaller than 0.2bpp (1:40). The embedded watermark and the extracted watermark in Figure 4.9 is an example of the watermark extraction with JPEG 2000 severe compression (1:40) and shows that the new scheme can survive JPEG 2000 compression very well. The comparison of watermark

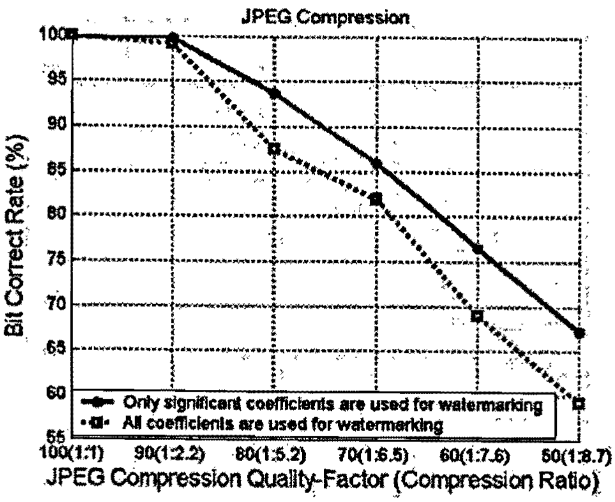


Figure 4.7: The robustness against JPEG compression.

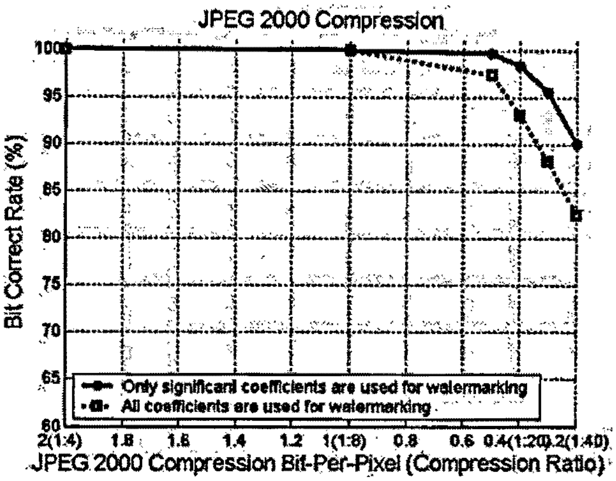


Figure 4.8: The robustness against JPEG 2000 compression.

embedding with and without the statistical model based significant coefficient selection is also shown in the Figure 4.7 and 4.8. The advantages of the new method with the coefficient selection are apparent.

Finally, we compared the distortion performance between the interleaving LUT (odd-even embedding, run=1), our distortion-minimized LUT and the average distortion of all LUTs with maximum allowable run of 2 in terms of various quantization level. According to Figure 4.10, the distortion of the new method is the minimum, though it provides better robustness.

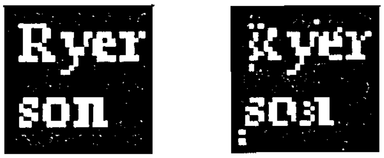


Figure 4.9: The embedded watermark (left) and the watermark (right) extracted after JPEG 2000 severely compressed (1:40) image.

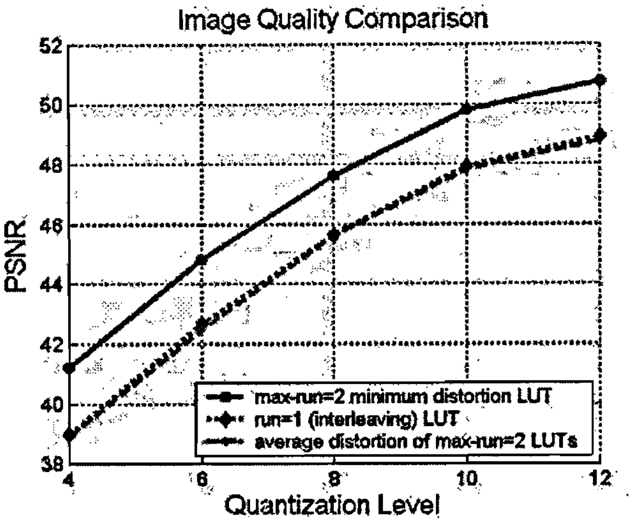


Figure 4.10: The image quality comparison between run=1 (interleaving) LUT, distortion-minimized run=2 LUT embedding and the average distortion of all maximum-run=2 LUTs.

4.7 Summary

We have analytically evaluated the distortion brought by LUT embedding with run constraint. Based on this analysis, we proposed a novel minimum-distortion algorithm to design LUT which can improve the watermarked signal quality. If security issue is taken into account, with a little bit change, the proposed approach can generate more than one near-minimum-distortion look-up tables. Thus our algorithm fits into some watermarking applications where joint-security-fidelity is required. For example, in transaction tracking applications, a unique watermark is embedded into each copy; if the multimedia work were subsequently leaked to the press or redistributed illegally, the owner could find out who was responsible. Experimental results show that the look-up table obtained with our method is superior to the odd-even(interleaving) embedding in terms of image quality.

Chapter 5

Optimal Quantization Based Watermarking Algorithm

5.1 Introduction

In this chapter, we focus on robust quantization based data hiding scheme. Robust data hiding techniques are required to achieve maximum robustness and fidelity simultaneously. However robustness and fidelity are always a pair of conflicting requirements. In this chapter, we consider the optimization of one given that the other is fixed. A Distortion-Robustness function, $D(R)$ and a Robustness-Distortion, $R(D)$ are formulated in the context of quantization based information hiding. Based on the theoretical analysis on robustness-distortion tradeoff, a new optimization strategy for data hiding given the embedding distortion or robustness constraint is proposed. This algorithm follows the general model of Quantization Index Modulation. The problem of designing the optimal nonuniform quantization encoder given fidelity or robustness criteria is formulated into a Lagrange function. Experimental results show that the optimal quantization watermarking scheme performs better than the existing schemes. The algorithm lends itself to applications where distortion or robustness is specifically requested.

This chapter studies optimal nonuniform quantization watermarking scheme. Starting with Lloyd-Max method based optimal quantization, the optimal quantization watermarking algorithm is designed by exploiting the statistics of the host data. Specifically, we assume the quantization level is fixed and consider two constrained optimization problems: (1) given robustness criterion, looking for quantization encoding scheme which minimizes distortion; (2) given fidelity constraint, looking for quantization encoding scheme which maximizes robustness.

5.2 Robust-Distortion function

5.2.1 Quantization Based Information Hiding

In quantization based information hiding schemes, the watermark information is conveyed in the choice of quantizer. The message symbol to be embeded is denoted by $m \in \{0, 1, \cdots, M-1\}$, which is also called M-ary watermark. In quantization based information hiding system shown in Figure 5.1, M quantizers are needed to embed M-ary watermark. To simplify the model, here we focus on binary watermark where two quantizers are needed for watermarking. We can divide the host data into two categories: the data samples that are used to embed message bit "0", denoted by s_0, and the data samples that are used to embed bit "1", denoted by s_1. Suppose two quantizers are denoted by Q_0 and Q_1, where Q_0 is used for s_0 and Q_1 is used for s_1.

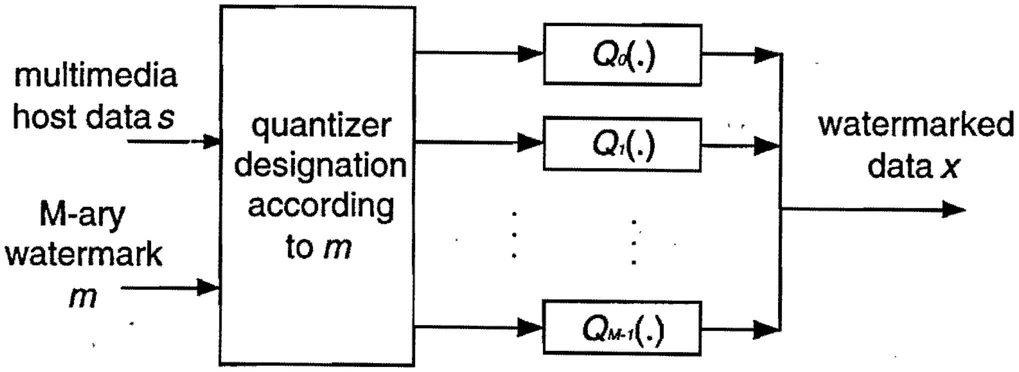


Figure 5.1: General diagram of embedder for quantization based data hiding system. The host data is quantized with the quantizer associated with the watermark bit.

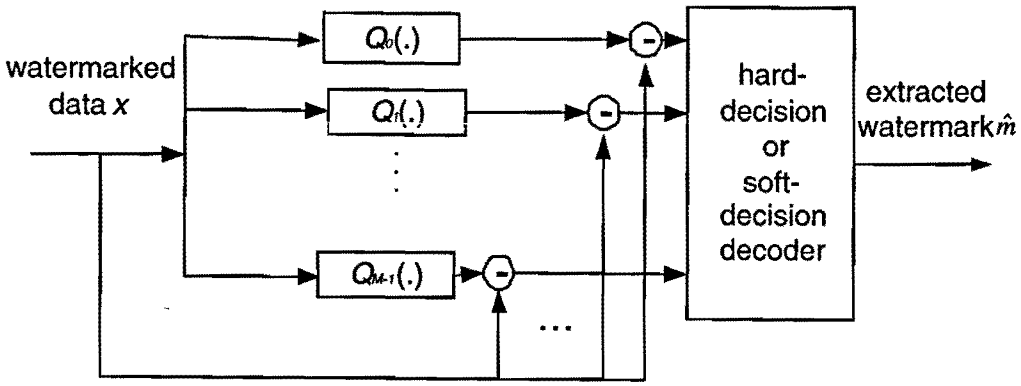

watermarked data x
$Q_0(.)$
$Q_1(.)$
$Q_{M-1}(.)$
-
-
-
hard-decision or soft-decision decoder
extracted watermark $\hat{m}$

Figure 5.2: General diagram of decoder for quantization based data hiding system. The distances between the received signal and the nearest quantizer reconstruction points are used for either soft-decision or hard-decision error correction decoding.

Denoting the k-th reconstruction point of Q_0 as $r_{0,k}, k = 1, \cdots, K$, we have

$$r_{0,k} = Q_0(s), s \in [d_{0,k-1}, d_{0,k}], \tag{5.1}$$

where $d_{0,k}, k = 0, \cdots, K$ is the decision level of the quantizer Q_0. Similarly, for quantizer Q_1,

$$r_{1,k} = Q_1(s), s \in [d_{1,k-1}, d_{1,k}], \tag{5.2}$$

where $r_{1,k}, k = 0, \cdots, K$ and $d_{1,k}, k = 0, \cdots, K$ represent reconstruction points and decision levels of the quantizer Q_1, respectively.

The message embedding procedure can be illustrated by Figure 5.1. For message bit $m = 0$, the host data is mapped to the nearest reconstruction point of the quantizer Q_0. For message bit $m = 1$, the host signal is mapped to the nearest reconstruction point of the quantizer Q_1. In another word, s_0 is quantized with Q_0 whereas s_1 is quantized with Q_1. An example of quantization ensembles for data embedding is illustrated in Figure 5.3, where $\triangle$ and $\triangledown$ represent quantization points for "0" and "1" embedding respectively.

A block diagram of the general decoding procedure is shown in Figure 5.2. The distance between the received signal y and the sets of reconstruction points of different quantizers are employed in hard-decision or soft-decision decoding algorithm.

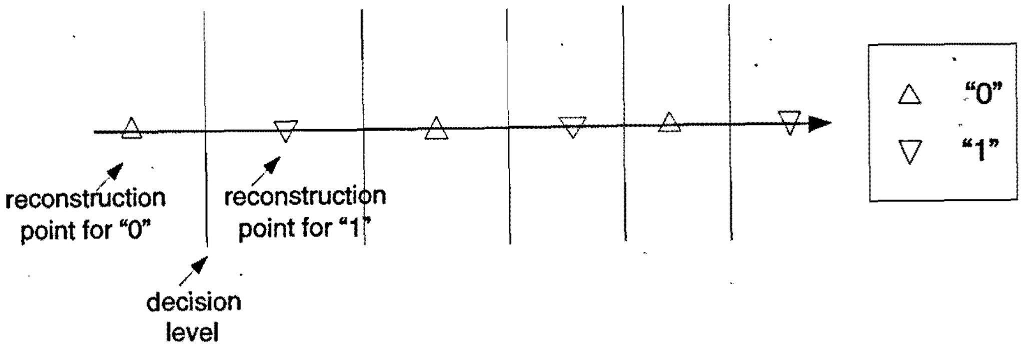


Figure 5.3: An example of signal constellation for quantization based data hiding.

Hard-decision Decoder

For hard-decision decoding, one can make decisions on each coded bit y [12][13],

$$\hat{m} = \arg\min_{m} ||y - Q_m(y)||^2,\ m \in \{0, 1\}, \tag{5.3}$$

where y is the received, maybe corrupted signal and $Q_m(y)$ is the reconstruction point of Q_0 or Q_1 that nearest to $\hat{x}$. Unless the channel noise is strong enough to drag the watermarked feature out of the enforced interval, the detection result is $\hat{m} = m$.

Soft-decision Decoder

Alternatively, for soft-decision decoding, the message to be embedded m is one of M binary sequences $m = m_i, i = 1, \cdots, M$ where each possible sequence is composed of L binary bits $m_i = \{m_{i1}, \cdots, m_{iL}\}$. The extracted watermark $\hat{m}$ is determined by evaluating the square-sum of distance between the received signal $y = \{y_1, \cdots, y_L\}$ and the nearest set of quantization ensembles.

$$\hat{m} = \arg\min_{m_i} \sum_{l=1}^{L} (y_l - Q_{m_{il}}(y_l))^2, m_{il} \in \{0, 1\}, i = 1, \cdots, M, l = 1, \cdots, L, \tag{5.4}$$

where $Q_{m_{il}}(y_l)$ is the reconstruction point of Q_0 or Q_1 which nearest to y_l.

5.2.2 Quantization Distortion

Apparently, through mapping the host signal to the nearest quantization value controlled by the watermark, distortion is introduced. The v-power difference distortion incurred by scalar quantization can be expressed as the sum of the distortions for each of the decision regions. The distortion produced by quantizing source s with a given L-levels quantizer Q is

$$\begin{aligned} D^{(v)} &= \textstyle\sum_{k=1}^{K} \left[\int_{d_{k-1}}^{d_k} |s - Q(s)|^v f(s) ds\right] \\ &= \textstyle\sum_{k=1}^{K} \left[\int_{d_{k-1}}^{d_k} |s - r_k|^v f(s) ds\right] \end{aligned} \tag{5.5}$$

where $d_k, k = 0, \cdots, K$ and $r_k, k = 1, \cdots, K$ are the decision levels and reconstruction points of the given quantizer Q respectively, $f(s)$ is the Probability Density Function (PDF) of s.

In our two quantizer case, the total distortion is the sum of Q_0 and Q_1 quantization distortion

$$D^{(v)} = D_0^{(v)} + D_1^{(v)} = \sum_{k=1}^{K} \left[\int_{d_{0,k-1}}^{d_{0,k}} |s_0 - r_{0,k}|^v f_0(s_0) ds_0\right] + \sum_{k=1}^{K} \left[\int_{d_{1,k-1}}^{d_{1,k}} |s_1 - r_{1,k}|^v f_1(s_1) ds_1\right] \tag{5.6}$$

where $f_0(s_0)$ and $f_1(s_1)$ are the PDFs of s_0 and s_1 respectively.

In the special case of $v = 2$, the difference distortion measure becomes the widely-used mean square error (MSE) criterion, and a quantizer which minimizes $D^{(2)}$ is termed MSE-optimal or minimum mean-square-error (MMSE) quantizer. In the rest of this chapter, we evaluate the quantization distortion by MSE, which is denoted as D,

$$D = \sum_{k=1}^{K} \left[\int_{d_{0,k-1}}^{d_{0,k}} |s_0 - r_{0,k}|^2 f_0(s_0) ds_0\right] + \sum_{k=1}^{K} \left[\int_{d_{1,k-1}}^{d_{1,k}} |s_1 - r_{1,k}|^2 f_1(s_1) ds_1\right]. \tag{5.7}$$

5.2.3 Robustness Measurement

In our information hiding system, we use a soft-decision decoder as (5.4). The square-sum of distance between the received data and the sets of reconstruction points of different quantizers are used to determine the embedded information. Therefore, we define the robustness measure as

$$R = \sum_{k=1}^{K} \left[||r_{0,k} - r_{1,k}||^2 \int_{d_{0,k-1}}^{d_{0,k}} f_0(s_0) ds_0\right] + \sum_{k=1}^{K} \left[||r_{0,k} - r_{1,k}||^2 \int_{d_{1,k-1}}^{d_{1,k}} f_1(s_1) ds_1\right]. \tag{5.8}$$

The overall mean squared distance between the nearest set of quantization ensembles are used to evaluate the robustness of quantization watermarking. $\int_{d_{0,k-1}}^{d_{0,k}} f_0(s_0) ds_0$ and $\int_{d_{1,k-1}}^{d_{1,k}} f_1(s_1) ds_1$

represent the probability that s_0 and s_1 fall into the k-th quantization cell of Q_0 and Q_1 respectively. We denote them as $P_{0,k}$ and $P_{1,k}$, i.e.

$$P_{0,k} = \int_{d_{0,k-1}}^{d_{0,k}} f_0(s_0)ds_0, P_{1,k} = \int_{d_{1,k-1}}^{d_{1,k}} f_1(s_1)ds_1. \tag{5.9}$$

(5.8) can be re-written as

$$R = \sum_{k=1}^{K} \left(P_{0,k}||r_{0,k} - r_{1,k}||^2\right) + \sum_{k=1}^{K} \left(P_{1,k}||r_{0,k} - r_{1,k}||^2\right) \tag{5.10}$$

5.2.4 Distortion-Robustness Function and Robustness-Distortion Function

When we design information hiding algorithm, we always face the tradeoff between fidelity and robustness requirements. Our purpose is to find a set of quantization ensembles which achieve the maximum robustness $R(D)$ subject to the given distortion D or achieve the minimum distortion $D(R)$ subject to the given robustness R. Then two constrained optimization problems are formulated as: (1)given robustness criterion, looking for quantization encoding scheme which minimizes distortion; (2)given fidelity constraint, looking for quantization encoding scheme which maximizes robustness.

The maximum R given that the distortion D is fixed can be represented by

$$R(D) = \max_{D}(R). \tag{5.11}$$

Similarly, we can define a $D(R)$ function to describe the minimum distortion subject to fixed robustness,

$$D(R) = \min_{R}(D). \tag{5.12}$$

The relationship between distortion and robustness is demonstrated in Figure 5.4. The curve shows the maximum robustness given distortion, therefore the region below the $R(D)$ curve is achievable and the region above it is unachievable.

Figure 5.5 illustrates the first constrained optimization problem, the region above the given robustness R_g and below $R(D)$ curve satisfies the robustness criterion. At the position

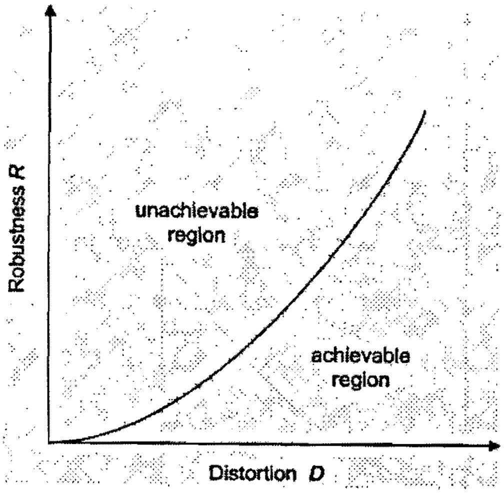


Figure 5.4: Distortion-Robustness function.

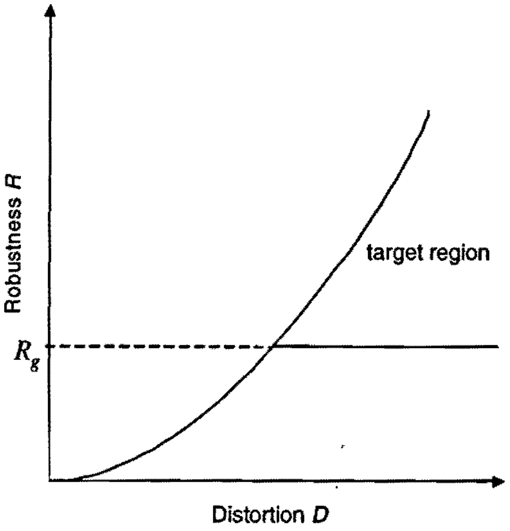


Figure 5.5: The region below $D-R$ curve above R_g is the target region.

where the horizontal line of R_g intersects the Robustness-Distortion curve, quantization encoding scheme minimizes distortion subject to R_g.

Quantization embedding scheme which achieves maximum robustness subject to given distortion is illustrated in Figure 5.6. Only the region right to the given distortion D_g and below $R(D)$ curve satisfies the specified distortion criterion. At the position where the line of $D = D_g$ intersects the curve, the robustness is the maximum.

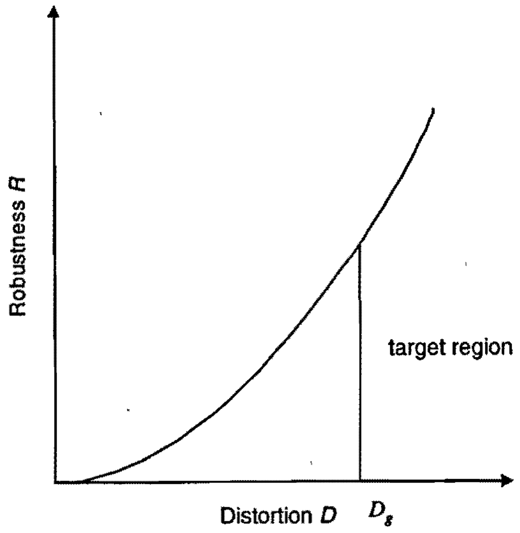


Figure 5.6: The region below $D-R$ curve right to D_g is the target region.

5.3 Properties of $R(D)$ and $D(R)$ Function

To clarify the approximate figure of $R(D)$ and $D(R)$ function, it is necessary to analyze the properties of $R(D)$ and $D(R)$ function in its defined region.

$R(D)$ and $D(R)$ functions are monotonous increasing functions

According to the definition of $R(D)$, among all possible quantization ensembles which generate equal or less than the given distortion D_g, we select the quantization ensembles which achieve the maximum robustness. When the allowable distortion D_g is enlarged, the set of choice quantization ensembles is widened which include all choice quantizers subject to previous D_g. Now searching the maximum robustness among this extended set of quantization ensemble, clearly the maximum robustness will increase, at most unaltered. So $R(D)$ is non-decreasing, with increased allowable distortion D_g, the maximum achievable robustness will increase. As $R(D)$'s inverse function, $D(R)$ function is also monotonous increasing.

The domain of definition of $R(D)$ and $D(R)$

For given host data s_0 and s_1, the minimum and maximum distortion, D_{min}, D_{max} and the achievable maximum robustness, R_{max} as well as the robustness corresponding to the minimum distortion $R(D_{min})$.

1. D_{min} and $R(D_{min})$

   First we discuss the minimum distortion where the host data and the quantization level are given. Without robustness concern, the desired quantization ensembles are simply the optimal quantizers in which the decision levels and reconstruction levels are selected that minimize the distortion subject to the quantization level constraint:

$$D_{min} = \min\left\{\sum_{k=1}^{K}\left[\int_{d_{k-1,0}}^{d_{k,0}} |s_0 - r_{k,0}|^2 f_0(s_0)ds_0\right] + \sum_{k=1}^{K}\left[\int_{d_{k-1,1}}^{d_{k,1}} |s_1 - r_{k,1}|^2 f_1(s_1)ds_1\right]\right\} \tag{5.13}$$

   Depending on the distribution of s_0 and s_1, the watermarking scheme based on the achieved optimal quantizers hase some extent of robustness. If the distribution of s_0 is similar to the distribution of s_1, the reconstruction points of the two optimal quantizers are also similar, then the watermarking system will have poor robustness performance. If the two quantizers are exactly same, then $R(D_{min}) = 0$.

2. R_{max} and $D(R_{max})$

   The maximum robustness is obtained when the data to be marked "0" is mapped to the minimum value that the data can be changed to and the data to be marked "1" is mapped to the maximum value or vice versa. This time the distortion is

$$D(R_{max}) = \int_{\eta_{min}}^{\eta_{max}} (s_0 - \eta_{min})^2 f(s_0)ds_0 + \int_{\eta_{min}}^{\eta_{max}} (\eta_{max} - s_1)^2 f(s_1)ds_1 \tag{5.14}$$

   where η_{min} and η_{max} are the minimum and maximum of the host data.

$$\eta_{min} = \min(s);\ \eta_{max} = \max(s) \tag{5.15}$$

5.4 Optimal Watermarking Implementation

The performance of optimal nonuniform quantization encoding scheme is represented by a point on the curve of Figure 5.4 with given distortion D_g or given robustness R_g. If we can find the curve, i.e. $R(D)$ function, we can achieve the optimal watermarking scheme with any given constraint. Leave the robustness requirement aside, to achieve the minimum distortion due to the quantization operation, the Q_0 and Q_1 quantizers should be the minimum-distortion quantizers for feature to be embed "0" and "1" respectively.

The problem of minimum-distortion quantization design is to select decision levels and reconstruction levels that minimize distortion subject to a constraint on the number of levels K. This amounts to the simple requirement that the partial derivatives of with respect to the decision levels and reconstruction levels be zero [29][30][31]:

$$\frac{\partial D^2}{\partial d_k} = 0; k = 1, ..., K-1 \tag{5.16}$$

$$\frac{\partial D^2}{\partial r_k} = 0; k = 1, ..., K \tag{5.17}$$

Lloyd found the necessary and sufficient conditions for a fixed-rate quantizer to be locally optimal (minimum-distortion) [29]: the quantizer partition must be optimal for the set of reproduction levels, and the set of reproduction levels must be optimal for the partition. Solving equations (5.15) and (5.16), the decision level $d_k, k = 0, \cdots, K$ is the *average* of the surrounding quantization levels,

$$d_k = \frac{r_k + r_{k+1}}{2} \tag{5.18}$$

and the reproduction level $r_k, k = 1, \cdots, K$ corresponding to a given cell is the centroid of the source value given that it lies in the specific cell:

$$r_k = \frac{\int_{d_{k-1}}^{d_k} s f(s) ds}{\int_{d_{k-1}}^{d_k} f(s) ds} \tag{5.19}$$

After performing Lloyd-Max algorithm on s_0 and s_1 respectively, we can obtain the initial optimal quantizers Q_0 and Q_1.

However, if the binary watermark is distributed randomly, the probability that feature to be embedded "0" is close to the probability that it is embedded "1". It is found that the PDF of s_0 is similar to the PDF of s_1. The minimum-distortion quantizer is determined by the signal's statistical distribution. Because s_0 and s_1 have similar PDFs and the fixed quantization levels are same, the two optimal quantizers obtained by Lloyd-Max method are similar, that is for any i, $1 < i < L$,

$$r_{i,0} \simeq r_{i,1} \tag{5.20}$$

and for any j, $i \neq j$

$$\begin{aligned} Dist(r_{i,0}, r_{i,1}) &\ll Dist(r_{i,0}, r_{j,1}) \\ Dist(r_{i,0}, r_{i,1}) &\ll Dist(r_{i,0}, r_{j,1}) \\ Dist(r_{i,0}, r_{i,1}) &\ll Dist(r_{i,1}, r_{j,1}) \end{aligned} \tag{5.21}$$

where $Dist(\cdot)$ is the distance between two reconstruction points.

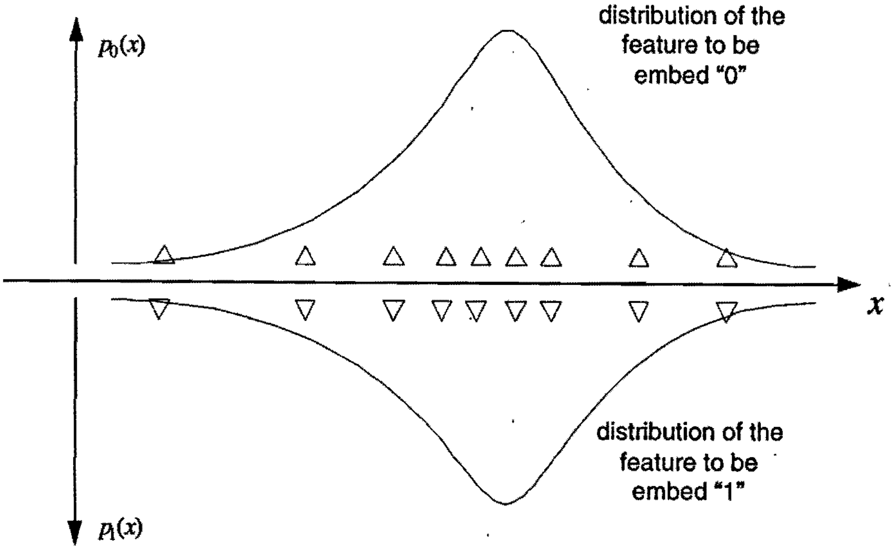

$p_0(x)$
distribution of the feature to be embed "0"
x
distribution of the feature to be embed "1"
$p_1(x)$

Figure 5.7: The signals that follow similar PDFs have similar optimal quantizers.

In Figure 5.7, the two PDFs are plotted and the Δ points and ∇ points represent the optimal quantizers for "0" embedding and "1" embedding respectively.

The distance between the sets of reconstruction points of different obtained optimal quantizers, i.e. $Dist(r_{i,0}, r_{i,1})$, is so small that the embedded watermark can be destructed even by a faint perturbation. Intuitively, we need to adjust the two quantizers by enlarging $Dist(r_{i,0}, r_{i,1})$ to improve the embedding scheme's robustness performance. For example, if the i-th reconstruction point of 0 quantizer, $r_{i,0}$ is close to but less than $r_{i,1}$, to decrease the error probability, we should reduce $r_{i,0}$ and enlarge $r_{i,1}$ as illustrated in Figure 5.8.

Before the adjustment for robustness improvement, the distortion is only contributed by quantization. The distortion produced by quantizing source s with K level optimal

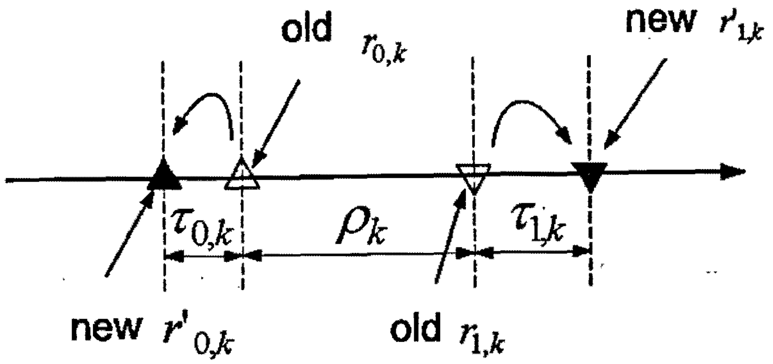


Figure 5.8: Reconstruction points need to be adjusted to achieve the optimal distortion-robustness tradeoff.

quantizers

$$D = \sum_{k=1}^{K} \left(\int_{d_{k-1}}^{d_k} |s - r_k|^2 f(s) ds \right) \tag{5.22}$$

where $d_k, k = 0, \cdots, K$ and $r_k, k = 1, \cdots, K$ are the decision levels and reconstruction points of source s's minimum-distortion quantizer. The reconstruction points are optimal in the MSE sense so that any change on them will introduce further distortion. Assume that the k-th optimal quantization point r_k is moved to r'_k, $k = 1, \cdots, K$, the distortion of new constructed quantizer is:

$$\begin{aligned} D_w &= \textstyle\sum_{k=1}^{K} \left[\int_{d_{k-1}}^{d_k} |s - r'_k|^2 f(s) ds \right] \\ &= \textstyle\sum_{k=1}^{K} \left[\int_{d_{k-1}}^{d_k} |s - r'_k + r_k - r_k|^2 f(s) ds \right] \\ &= \textstyle\sum_{k=1}^{K} \left[\int_{d_{k-1}}^{d_k} |s - r_k|^2 + |r_k - r'_k|^2 + 2|r'_k - r_k||r_k - s| f(s) ds \right]. \end{aligned} \tag{5.23}$$

Since $\int_{d_{k-1}}^{d_k} r_k f(s) ds = \int_{d_{k-1}}^{d_k} s f(s) ds$, we have

$$\begin{aligned} D_w &= \textstyle\sum_{k=1}^{K} \left[\int_{d_{k-1}}^{d_k} |s - r_k|^2 + |r_k - r'_k|^2 f(s) ds \right] \\ &= \textstyle\sum_{k=1}^{K} \left[\int_{d_{k-1}}^{d_k} |r_k - r'_k|^2 f(s) ds \right] + D \end{aligned} \tag{5.24}$$

In our bi-quantizer case, the total distortion is

$$\begin{aligned} D_w &= \textstyle\sum_{k=1}^{K} \left[\int_{d_{0,k-1}}^{d_{0,k}} |r_{0,k} - r'_{0,k}|^2 f_0(s_0) ds_0 \right] + D_0 \\ &+ \textstyle\sum_{k=1}^{K} \Big] \int_{d_{1,k-1}}^{d_{1,k}} |r_{1,k} - r'_{1,k}|^2 f_1(s_1) ds_1 \Big] + D_1 \\ &= \textstyle\sum_{k=1}^{K} \left[\int_{d_{0,k-1}}^{d_{0,k}} \tau_{0,k}^2 f_0(s_0) ds_0 \right] + D_0 \\ &+ \textstyle\sum_{k=1}^{K} \Big] \int_{d_{1,k-1}}^{d_{1,k}} \tau_{1,k}^2 f_1(s_1) ds_1 \Big] + D_1, \end{aligned} \tag{5.25}$$

where D_0 and D_1 are only due to the initial Q_0 and Q_1 optimal quantizers which is achieved prior to adjustment, and

$$\tau_{0,k} = r_{0,k} - r'_{0,k}, \tau_{1,k} = r_{1,k} - r'_{1,k}. \tag{5.26}$$

We argue that robustness of quantization watermarking scheme is not determined by the minimum distance between the sets of reconstruction points of different quantizers. Take the distributions of the host data s_0 and s_1 into account, according to (5.10), the statistical weighted square quantizer-distance robustness measurement after adjustment is as

$$\begin{aligned} R &= \sum_{k=1}^{K} \left[P_{0,k}|r'_{0,k} - r'_{1,k}|^2\right] + \sum_{k=1}^{K} \left[P_{k,1}|r'_{0,k} - r'_{1,k}|^2\right] \\ &= \sum_{k=1}^{K} [P_{0,k}(\tau_{0,k} + \rho_k + \tau_{1,k})^2] + \sum_{k=1}^{K} [P_{k,1}(\tau_{0,k} + \rho_k + \tau_{1,k})^2], \end{aligned} \tag{5.27}$$

where $\tau_{0,k}$ and $\tau_{1,k}$ are the magnitude that we adjust the reconstruction points $r_{0,k}$ and $r_{1,k}$, respectively and ρ_k is the distance between the reconstruction points of the initial Q_0 and Q_1 quantizers.

Then we can formulate designing the optimal quantization ensembles for watermarking into a Lagrangian function,

$$\begin{aligned} J(k,l) &= -R + \lambda D \\ &= -\sum_{k=1}^{K} [P_{0,k}(\tau_{0,k} + \rho_k + \tau_{1,k})^2] - \sum_{k=1}^{K} [P_{1,k}(\tau_{0,k} + \rho_k + \tau_{1,k})^2] \\ &+\lambda[D_0 + \sum_{k=1}^{K} \left(P_{0,k}\tau_{0,k}^2\right) + D_1 + \sum_{k=1}^{K} \left(P_{1,k}\tau_{1,k}^2\right)], \end{aligned} \tag{5.28}$$

which combines our proposed robustness measurement. The cost function measures the MSE between the original data and the watermarked data. Because the MMSE quantizers are achieved before this optimization procedure, D_0, D_1 are supposed to be known as well as $\rho_k, k = 1, 2, ..., K$, the initial distances between Q_0 and Q_1 quantizer points. A simple solution to the above equation is that the partial derivatives of J with respect to k_i and l_i be zero.

$$\begin{aligned} &\tfrac{\partial J}{\partial \tau_{0,k}} = -2P_{0,k}(\tau_{0,k} + \rho_k + \tau_{1,k}) - 2P_{1,k}(\tau_{0,k} + \rho_k + \tau_{1,k}) + 2\lambda P_{0,k}\tau_{0,k} = 0 \\ &\Rightarrow \tau_{0,k} = \tfrac{(P_{0,k}+P_{1,k})(\rho_k+\tau_{1,k})}{(\lambda-1)P_{0,k}-P_{1,k}} \end{aligned} \tag{5.29}$$

$$\begin{aligned} &\tfrac{\partial J}{\partial \tau_{1,k}} = -2P_{0,k}(\tau_{0,k} + \rho_k + \tau_{1,k}) - 2P_{1,k}(\tau_{0,k} + \rho_k + \tau_{1,k}) + 2\lambda P_{1,k}\tau_{1,k} = 0 \\ &\Rightarrow \tau_{1,k} = \tfrac{(P_{0,k}+P_{1,k})(\rho_k+\tau_{1,k})}{(\lambda-1)P_{1,k}-P_{0,k}} \end{aligned} \tag{5.30}$$

Thus we have $2K$ equations for $2K+1$ real variables ($\tau_{0,k}, \tau_{1,k} k = 1, \cdots, K$ and λ). This question is unsolvable. Fortunately there is a given criterion, it may be a robustness condition, $R = \sum_{k=1}^{K} [P_{0,k}(\tau_{0,k} + \rho_k + \tau_{0,k})^2] + \sum_{k=1}^{K} [P_{1,k}(\tau_{0,k} + \rho_k + \tau_{0,k})^2] \geq R_g$ or a fidelity condition, $D_w = D_0 + \sum_{k=1}^{K} P_{0,k}\tau_{0,k}^2 + D_1 + \sum_{k=1}^{K} P_{1,k}\tau_{0,k}^2 \leq D_g$. The additional equation makes the variables in k_k, l_k and λ achievable.

In Figure 5.9, we present all the steps involved in the watermark embedding process.

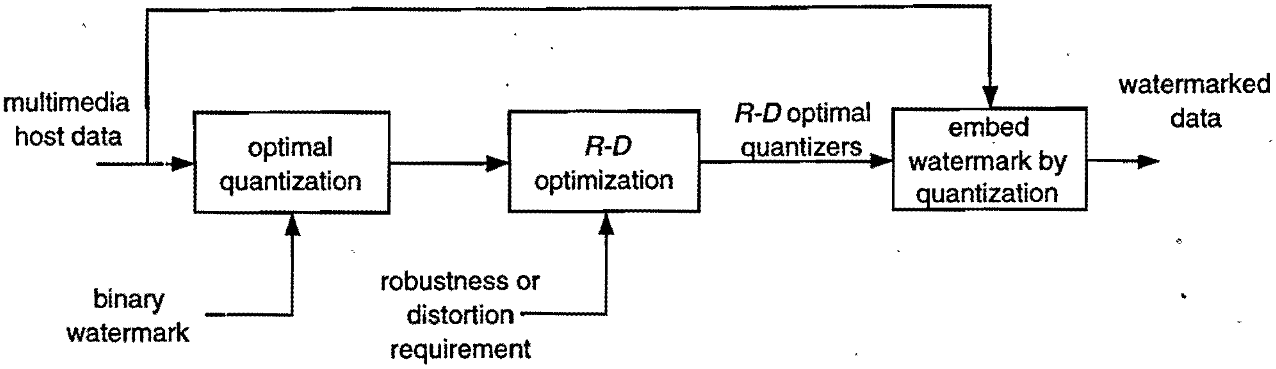


Figure 5.9: Watermark embedding algorithm.

5.5 Performance Evaluation

In this section we show some experimental results to demonstrate the performance of the proposed scheme on source which subject to various distributions. To demonstrate the necessity of the adjustments on the quantizers, the robustness of the data hiding schemes with and without quantizers adjustment are compared. Since uniform quantization based odd-even embedding [2] is widely used in information hiding systems, this algorithm is also compared in the experiments.

Multimedia data can be depicted by various distributions, for example, image signal is often modelled as Laplace distribution or generalized Gaussian distribution. For each distribution except uniform distribution, the obtained optimal quantizer is always totally different from and outperforms uniform quantizer. Meanwhile, uniformly distributed multimedia data is quite rare, so our proposed optimal quantization based embedding scheme is more effective than uniform scalar quantization based embedding scheme in solving multimedia information hiding problems.

First, we examine the robustness of three quantization strategies on uniform distributed source. Figure 5.10 shows that the performances of the uniform quantizer and the minimum-distortion quantizer are two single points in the Robustness-Distortion coordinate. That is because, given a source, there is only one fixed-level uniform quantizer and minimum-distortion quantizer. The quantizers adjustment in accordance with the given robustness or distortion criterion generates various sets of quantization ensembles, which robustness-distortion performance is represented by a $R-D$ curve in Figure 5.10. For uniform distribution, the achieved minimum-distortion quantizer is close to uniform, so the performance of the two quantizers are close to each other and are all around the border of $R-D$ curve. Nevertheless, both minimum-distortion quantizer and uniform quantizer do not adapt to specific robustness or fidelity requirement of various data hiding applications.

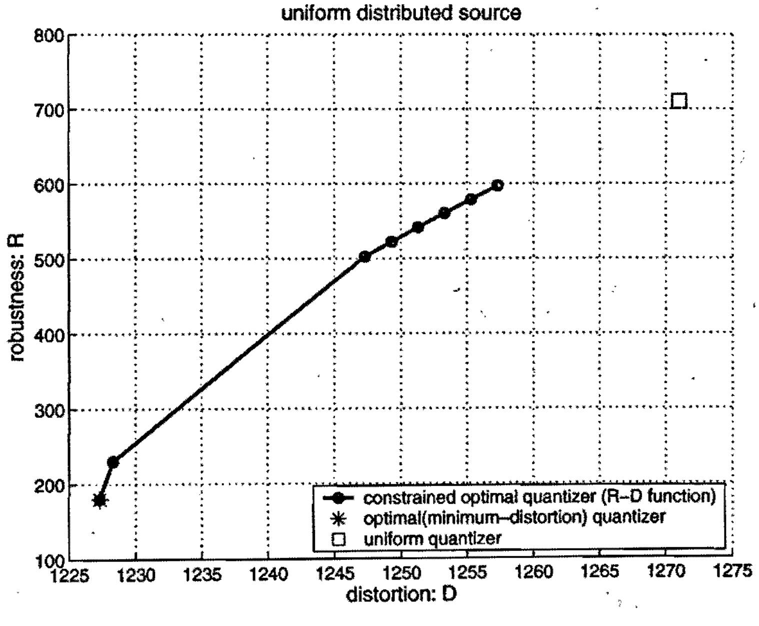


Figure 5.10: R-D performance of quantizers with uniform distributed source.

Second, Gaussian distributed source is exploited to evaluate the performance of the new scheme. We can see that the constrained Robustness-Distortion optimal quantizers have better performance than the uniform quantizer. An obvious proof is that the point which represents the performance of the uniform quantizer is below the $R-D$ curve. The quantization ensembles obtained with our method present about 16.4% higher robustness than the uniform quantization ensembles at the same expense of distortion. At the same robustness, the constrained optimal quantizers have 12.4% less distortion than the uniform quantizers.

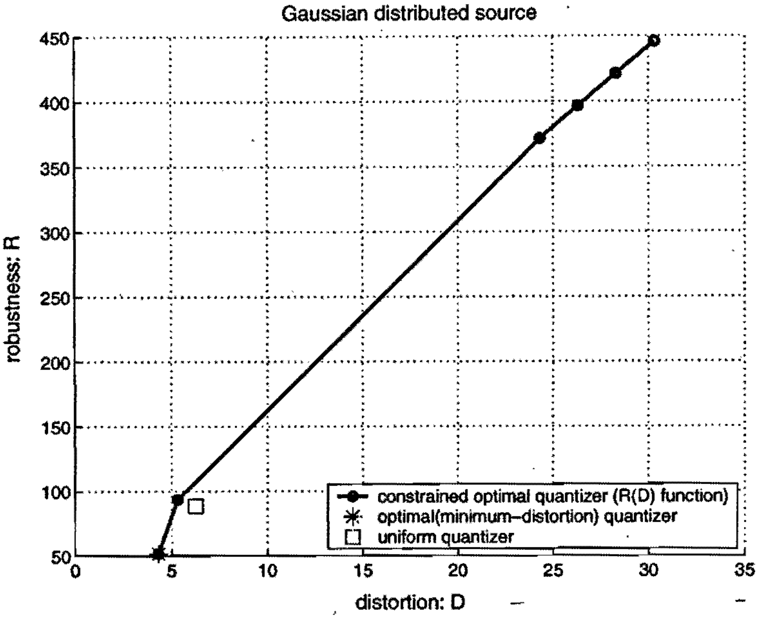


Figure 5.11: R-D performance of quantizers with Gaussian distributed source.

Figure 5.12 shows the performance of the method on the Laplace distributed source. The quantization ensembles generated with our method show about 20, relatively 18.2% higher robustness than the uniform quantization ensembles or 5.7% fidelity advantage over the uniform quantization strategy.

Finally, we evaluate the performance of new scheme on digital image "Lena". Figure 5.13 shows that we can achieve non-uniform quantizers which perform better than the uniform quantizer. The robustness of the non-uniform quantization can be 18 or relatively 23% higher than the uniform quantizer at the same expense of distortion. We can also find quantizers which show the same robustness to the uniform quantizer but 4% less distortion.

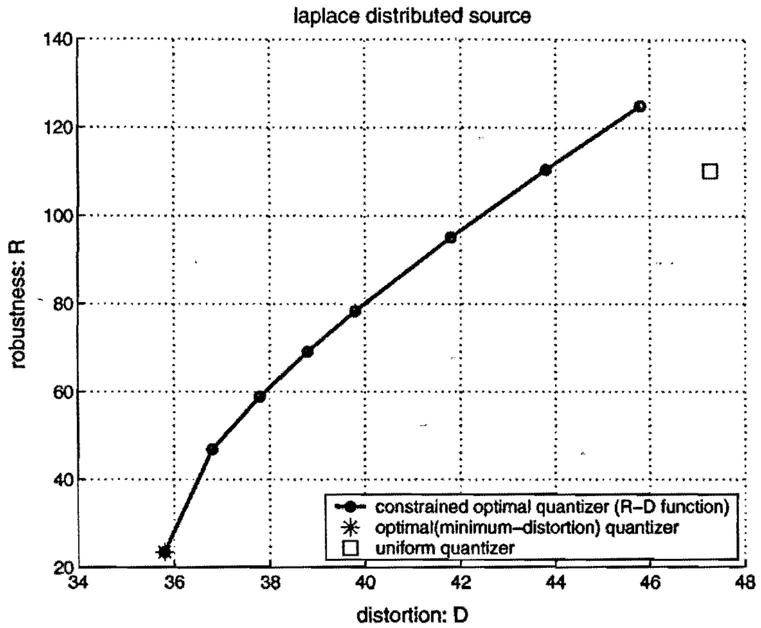


Figure 5.12: R-D performance of quantizers with Laplace distributed source.

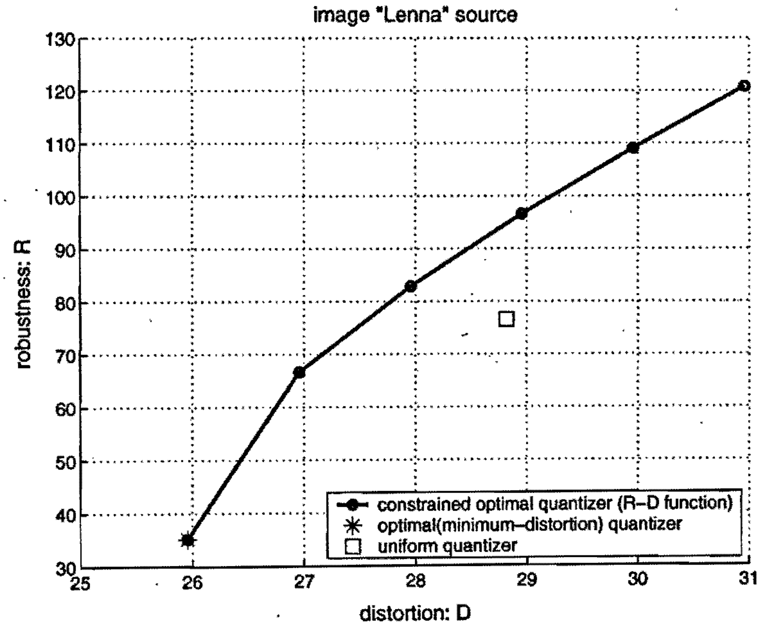


Figure 5.13: R-D performance of quantizers with image "Lena".

5.6 Summary

In summary, this chapter studies designing fixed-level non-uniform quantizers for robust information hiding. The robustness-distortion tradeoff is formulated into a Distortion-

Robustness function $D(R)$ and a Robustness-Distortion function $R(D)$. Then the properties of $R(D)$ and $D(R)$ function are analyzed. We quantify the robustness of non-uniform quantization watermarking scheme in terms of the distances between the sets of reconstruction points of different quantization ensembles. The distortion can also be formulated as a function of the adjustment magnitude on the reconstruction points of all involved quantizers. Based on the formulated $R(D)$ function, plus a given robustness or distortion constraint, a Lagrangian function is established. By solving it, we achieve the robustness or distortion constrained optimal quantizers. A series of experiments have been made, in which various distribution sources are used. Results show that our proposed nonuniform quantization embedding method performs better than existing scalar quantizer based watermarking schemes and can satisfy users' various robustness or fidelity requirements.

Chapter 6

Image Labelling System Based On JPEG2000 Compression Standard

6.1 Introduction

Watermarking has now come as a widely accepted approach for copyright protection and ownership identification. A lot of efforts have been dedicated to the development of robust watermarking schemes to achieve these goals. In this chapter, we consider identifying the ownership and distribution of image in digital network environment.

There are many practical requirements for successful ownership and distribution identification. In order to be effective and workable in a multimedia environment, the copyright label must be difficult to remove and survive processing which does not seriously reduce the value of the image. This encompasses a wide range of possibilities including format conversions, data compression, and low-pass filtering. Besides these well known robustness requirements, a copyright labelling system should also satisfy the following basic functional requirements to be a reliable identification tool:

1. The image must contain a label or code, which makes it as property of the copyright holder.
2. The image data must contain a user code, which verifies the user is in legal possession

of the data.

3. The image data is labelled in a manner which allows its distribution to be tracked.

First, a digital terminal, for example "Cell-phone A", sends an image request to the appropriate image server, "Image-server A" which belongs to "Vendor A". "Image-server A" replies by sending an image to the cell-phone. For copyright protection purpose, the image is watermarked. To reliably identify the ownership and distribution, the watermark message should includes but not limited to some key words like "Vendor A", "Image-Server A", "Cell-phone A", etc. Thus even if the watermarked image is passed to "PC B" by "Cell-phone A", the extracted watermark can still clearly demonstrate the owner, the origin and the distributor of the image. In network environment, transmitting a complete image from server to client is not only time-consuming, the precious network resource is also taken. So the multimedia data(image/audio/video) is often in its compression format in network applications. Clearly in image labelling system, the watermark embedding method should be integrated with image compression.

It is worthwhile to point out that image compression and frequency-domain watermarking share some common characteristics. In image compression, we encode significant frequency coefficients first because these coefficients convey more fundamental visual information about the image. In watermarking, we choose significant coefficients (coefficients with large amplitude) for watermark casting to enhance its robustness since these coefficients often remain stable after the attack. If they do change substantially, the reconstructed image will be perceptually different from the original one, and the value of protecting the intellectual property right of such a seriously degraded image becomes low. With this similarity, efficiency can be achieved by integrating frequency-domain watermarking procedures with compression processes, since the most expensive computation related to the image transform has already been computed as one part of compression and decompression algorithms.

In this chapter, we propose a watermarking scheme integrated with the new ISO/ITU-T still image coding standard, JPEG2000. This scheme satisfies the design criterion for image labelling on network.

6.2 Brief Review of JPEG2000 and Discrete Wavelet Transform

The image compression scheme, on which our watermarking approach is based, is the latest still image compression standard, JPEG2000.

JPEG2000 adopted a discrete wavelet transform (DWT) based technology in its compression scheme [32]. This means that the first step in the algorithm is to decompose the input image into a set of subbands via a discrete wavelet transform.

6.2.1 Discrete Wavelet Transform

The basic idea in the DWT for a one dimensional signal is the following. A signal is split by a pair of low-pass and high-pass filters into two parts, high frequencies and low frequencies. The edge, texture and detail components of the signal are largely confined to the high frequency part. Conversely, the low-pass filter preserves the low frequencies of a signal while attenuating or eliminating the high frequencies, thus resulting in a blurred version of the original signal. The low-pass and high-pass filter pair is known as *analysis filter-bank* [32]. The low frequency part is split again into two parts of high and low frequencies. This process is continued until the signal has been entirely decomposed or stopped before by the application at hand. For compression applications, generally no more than five decomposition steps are computed. Furthermore, from these DWT coefficients, the original signal can be reconstructed with another pair of low-pass and high-pass filters, known as the *synthesis filter-bank* [32]. This reconstruction process is called the inverse DWT. The DWT and IDWT for one dimensional signal $f(n)$ is best understood as successive applications of analysis and synthesis filter-banks, as illustrated in Figure 6.1.

The 1-D DWT can be easily extended to two dimensions (2-D) by applying the filter-bank in a separable manner. At each level of the wavelet decomposition, each row of a 2-D image is first transformed using a 1-D horizontal analysis filter-bank. The same filter-bank is then applied vertically to each column of the filtered and subsampled data. The result of a one-level wavelet decomposition is four filtered and subsampled images, referred to as

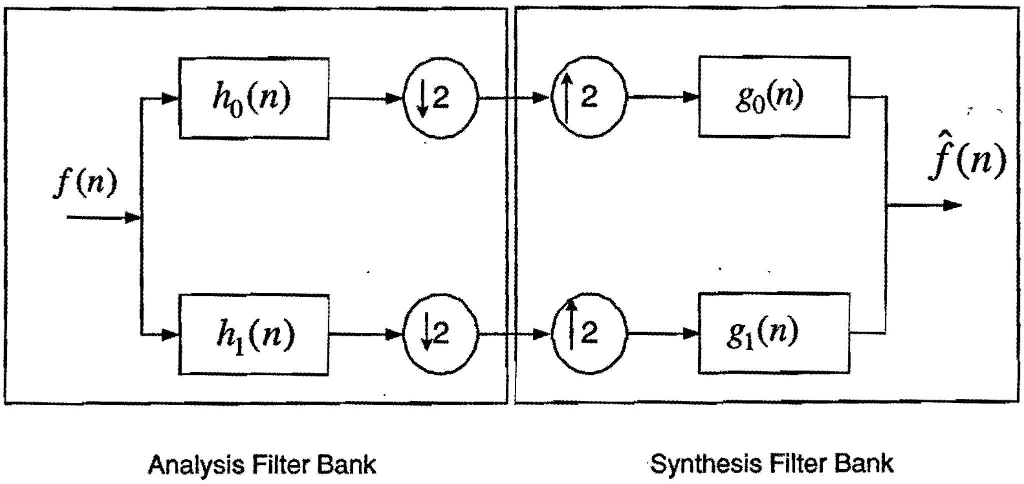


Figure 6.1: 1-D, 2-band wavelet analysis and synthesis filter-bank.

subbands. In a 2-D dyadic decomposition, the lowest frequency subband (denoted as the LL band) is further decomposed into four smaller subbands, and this process may be continued until no tangible gains in compression efficiency can be achieved. Figure 6.2 shows a 3-level, 2-D dyadic decomposition and the corresponding labelling for each subband. For example, the subband label kHL indicates that a horizontal high-pass(H) filter has been applied to the rows, followed by a vertical low-pass(L) filter applied to the columns during the kth level of the DWT decomposition. Figure 6.3 shows a 3-level, 2-D DWT decomposition of the Lena image.

6.2.2 JPEG2000

Each subband will then contain different frequency components of the information in the original image with the appropriate subsampling. The used wavelet transform can be either a floating- or a fixed-point wavelet, which implies lossy coding due to limited precision, or a reversible integer wavelet, which enables lossless coding.

The JPEG2000 image coding standard is based on a scheme originally proposed by Taubman and known as EBCOT ("Embedded Block Coding With Optimized Truncation"). The major difference between previously proposed wavelet-based image compression algorithms such as EZW ("Embedded Zerotree Wavelet ") [33] or SPIHT ("Set Partitioning in Hier-

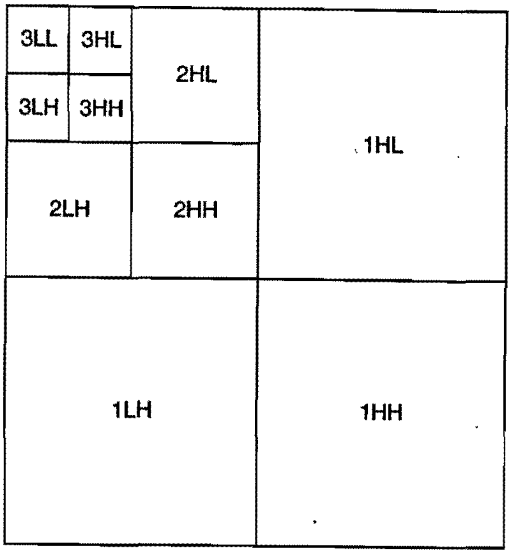


Figure 6.2: 2-D, 3-level wavelet decomposition.

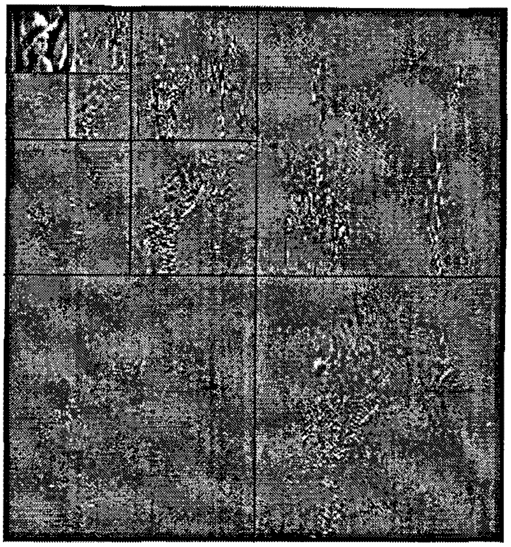


Figure 6.3: 2-D, 3-level wavelet decomposition of Lena.

archical Trees") [34] is that EBCOT as well as JPEG2000 operate on independent, non-overlapping blocks which are coded in several bit layers to create an embedded, scalable bitstream. Instead of zerotrees, the JPEG2000 scheme depends on a per-block quad-tree structure since the strictly independent block coding strategy precludes structures across subbands or even code-blocks. These independent code-blocks are passed down the "coding pipeline" shown in Figure 6.4. and generate separate bitstreams. Transmitting each bit layer corresponds to a certain distortion level. The partitioning of the available bit budget between the code blocks and layers, i.e. "truncation points" is determined using a sophisticated optimization strategy for optimal rate/distortion performance.

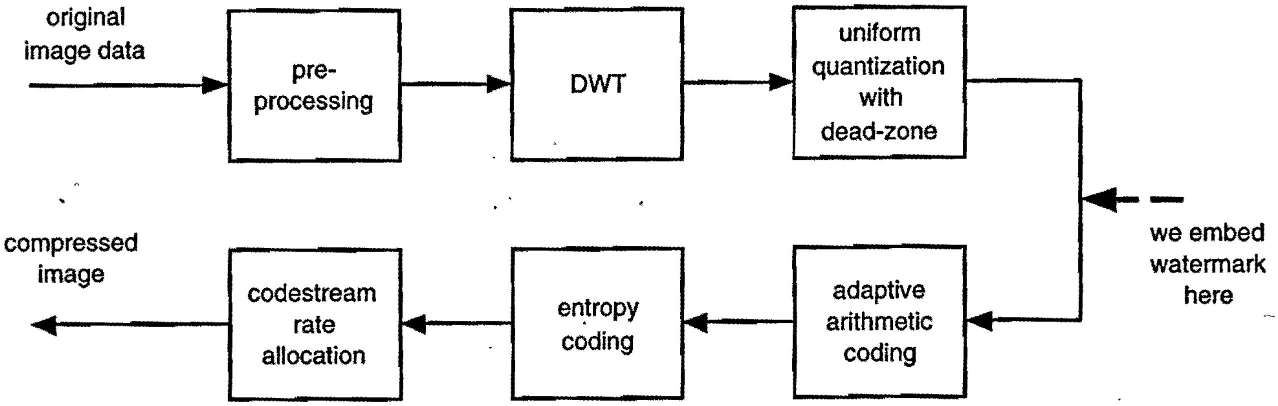


Figure 6.4: JPEG 2000 compression standard fundamental building blocks.

6.3 Previous Work

Several attempts to introduce image watermarking technique into JPEG2000 system have been reported in the recent literatures [23][35][36]. All of them embed watermark into the wavelet domain. In Su's scheme [23], a random noise sequence is generated as watermark and in each code-block, wavelet coefficients, which larger than a certain threshold value δ, are selected to bear watermark. To detect the embedded watermark, correlation detection is performed before dequantization to identify watermark. Hence reference watermarks are absolutely necessary in the watermarking system. In [35], Meelward exploited quantization index modulation (QIM) [13] to embed and detect watermark. Although watermark can be decoded directly in this way, the amount of data that can be embedded is quite limited. Chen et al. proposed a watermarking scheme, in which the watermark is scattered, embedded by bit-plane modification, followed by distortion reduction operation.

Both the above three methods cast watermark just after the stage of quantization. Although they fit into JPEG 2000 coding pipeline, the rate allocation procedure is not seriously considered. In JPEG2000, to generate an optimal image for a target file size (bit-rate), a rate control process is performed after entropy coding. In [37], an efficient rate control method is proposed that achieves a desired rate based on post-compression R-D optimization. The rate control algorithm finds the optimal bit allocation for all code-blocks, such that the total distortion is minimized subject to the target bit-rate.

Clearly the distortion brought by the rate control threatens the existence of watermark. The strength of watermark should be related to the compression ratio or bit-rate. For mild compression, the rate distortion is small and the energy of watermark should be low to maintain the image quality as good as possible. On the contrary, for higher compression ratio, the energy of watermark should be high enough to survive strong distortion caused by rate control. Furthermore, with declined image quality caused by compression, even a strong watermark becomes imperceptible. In our new watermarking algorithm, wavelet coefficients are modified depending on the compression ratio (bit-rate). A weak watermark is embedded into a mildly compressed image, while a strong watermark is applied to heavily compressed image. It is shown in this chapter that the new adaptive method integrates with JPEG2000 standard very well. The watermark is detected in a simple and fast way without assistance from either the original image or the reference watermarks. Meanwhile, by taking into account the compression ratio, the tradeoff between imperceptibility and robustness is balanced.

6.4 Watermarking System Integrated with JPEG2000

6.4.1 Watermark Embedding

In the proposed watermarking method, watermark is embedded into the detail sub-bands of middle resolution after stages of quantization and region of interests (ROI) scaling. During the stage of quantization, a wavelet coefficient $s_b(u,v)$ in subband b is mapped to a quantized index value $q_b(u,v)$. It is normalized as the most significant bit (MSB) carries the sign bit and the remaining bits represent the absolute magnitude of the coefficient. In this work, we assume that 8 bits are utilized to represent the integer part of $q_b(u,v)$. Thus the values of wavelet coefficients fall into [-255,255].

For wavelet coefficient, the operation of finding the optimal truncation point is nothing but performing an optimal non-uniform scalar quantization. Table 6.1 records the value of some wavelet coefficients before and after rate allocation with different target bit-rate φ. We find that the degree of quantization highly depends on φ. This point is quite easy to understand as the aim of this truncation operation is to achieve φ. Since the optimal truncation point is determined in the way that the whole image is represented best, the quantization step of each individually processed code-block differs from each other and not only determined by φ, factors like the significance of this code-block among all code-blocks are also taken into account. Fortunately, it is not necessary to estimate the accurate quantization step; a coarse estimated quantization interval Q is enough for watermarking. Actually we prefer Q a little bit larger than the true interval, especially when the compression ratio is low, in order to reliably decode the watermark. The basic idea is to make the image distortion caused by the watermarking conforming to the distortion caused by the entropy coding such that the watermark embedding capacity is maximized.

Table 6.1: Example wavelet coefficient values before and after rate control with different compression degrees(coefficients are selected randomly from 3HH subband of image "Baboon" with Jasper).

original coefficient	$\varphi = 1bpp$	$\varphi = 0.625bpp$	$\varphi = 0.5bpp$	$\varphi = 0.25bpp$	$\varphi = 0.1bpp$	$\varphi = 0.08bpp$
2	2	3	0	0	0	0
12	12	14	12	0	0	0
23	23	22	20	24	0	0
-32	-32	-34	-36	-48	-48	0
45	45	46	44	48	48	0
-66	-66	-66	-68	-80	-96	-64
104	104	106	108	112	96	128

The binary watermark w is embedded into selected code-blocks as follows:

- **Step 1.** Estimate Q with the provided target bit-rate according to Table 6.2.

Table 6.2: Estimated quantization intervals for different bit-rate value.

bit-rate φ	estimated quantization interval Q
> 1	2
$1 - 0.8$	4
$0.8 - 0.5$	8
$0.5 - 0.25$	16
$0.25 - 0.2$	32
$0.2 - 0.1$	48
< 0.1	64

- **Step 2.** The watermark bits to be embedded are each repeated M times: M is also determined by φ, for watermark bits embedded in mild compressed image are more fragile to common image processing, hence a large M is need to improve its robustness. Note that the spread watermark is still binary.

- **Step 3.** Coefficients belong to $[-Q/2, Q/2]$ are excluded to bear watermark.

- **Step 4.** Positive coefficients are mapped to the nearest even multiples of Q except 0 to embed "0" and the nearest odd multiples of Q to embed "1", while negative coefficients are mapped to the nearest even multiples of Q except 0 to embed "1" and the nearest odd multiples of Q to embed "0". The above operation of encoding bit "1" and "0" can be formulated as (6.1) and (6.2),

$$y = \begin{cases} \left[\left(\lfloor \frac{x}{2Q} \rfloor + 0.5\right) \times 2Q\right] & x > 2Q \\ \left(\lceil \frac{x-Q}{2Q} \rceil \times 2Q\right) & x < -Q \\ Q & Q/2 \leq x \leq 2Q \\ -2Q & -Q \leq x \leq -Q/2, \end{cases} \tag{6.1}$$

$$y = \begin{cases} \left(\lfloor \frac{x+Q}{2Q} \rfloor \times 2Q\right) & x > Q \\ \left[\left(\lceil \frac{x}{2Q} \rceil - 0.5\right) \times 2Q\right] & x < -2Q \\ 2Q & Q/2 \leq x \leq Q \\ -Q & -2Q \leq x \leq -Q/2, \end{cases} \tag{6.2}$$

where x is the original data and y is the modified data.

6.4.2 Watermark Retrieval

Watermark is decoded before dequantization during image decompression. In terms of the bit-rate of compressed image, we can achieve the same quantization interval Q as the one for watermark casting. The received coefficient $\hat{y}$ is mapped to the nearest multiple of Q, d. Then the watermark sequence is recovered by

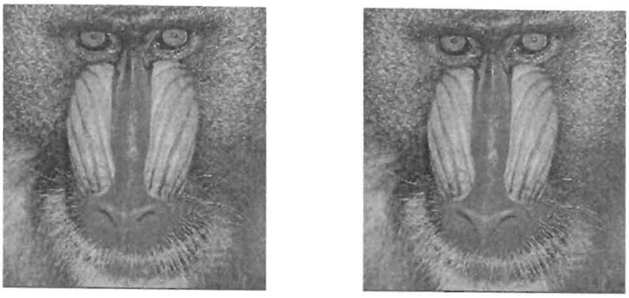
$$\hat{b} = \begin{cases} 1 & d = (2n-1)Q \text{ or } d = (-2n)Q \\ 0 & d = (2n)Q \text{ or } d = (-2n+1)Q; \end{cases} \quad n = 1, 2, 3 \cdots . \tag{6.3}$$

Finally the M consecutive decoded watermark bits are summed and a threshold decision yields the output bits. Thus, the results of the watermark decoder are the same watermark bits that have been embedded.

6.5 Simulation Results

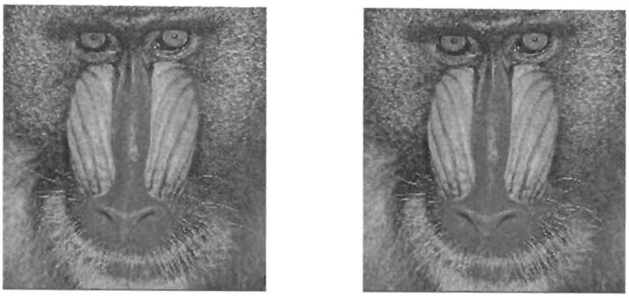
In this section, the performance of the proposed watermarking scheme to various distortions is demonstrated by experiments on grayscale image. The JPEG2000 codec that is used to test the new watermarking system is Jasper, an implementation of the JPEG2000 encoder/decoder [6]. The objective quality of watermarked image is indicated by PSNR. The robustness under several intentional/unintentional attacks is represented by bit correct ratio (BCR). The grayscale image "baboon" is for demonstration here. The testing results for other images are similar.

In the experiment, the watermark is embedded into the quantized coefficients of 3HH subband (5 decomposition levels are default for Jasper Codec). The original and watermarked images are shown in Figure 6.5. In Table 6.3, the PSNR (Peak-Signal-Noise-Ratio) of compressed image with and without watermark as well as watermark embedding capacity (i.e. the maximal number of watermark bits can be embedded) are shown in terms of various compression degrees. The results show that, the new bit-rate adaptive approach is superior to the watermark-strength fixed scheme in that the new method takes advantage of the compression to improve the watermark embedding capacity while minimizes the image distortion on top of the compressed images with various compression bit-rates.



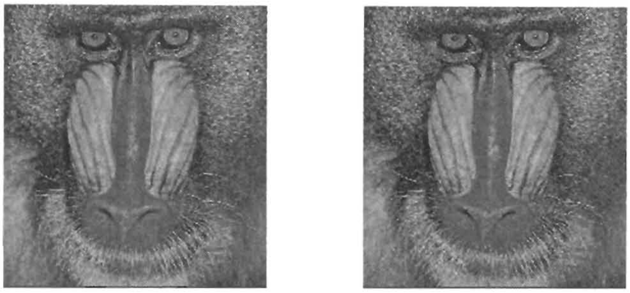
(a) bit-rate=1bpp

(b) bit-rate=1bpp



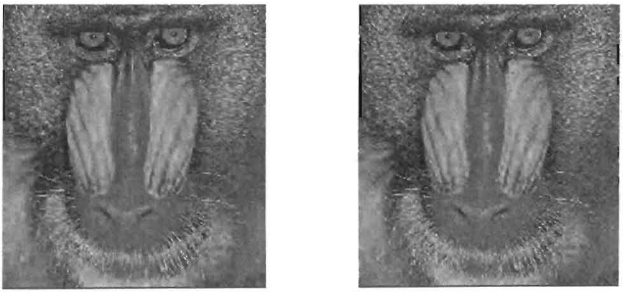
(c) bit-rate=0.5bpp

(d) bit-rate=0.5bpp



(e) bit-rate=0.25bpp

(f) bit-rate=0.25bpp



(g) bit-rate=0.1bpp

(h) bit-rate=0.1bpp

Figure 6.5: (a)(c)(e) and (g) are JPEG2000 compressed images with bit-rate=1, 0.5, 0.25, 0.1bpp respectively; (b)(d)(f) and(h) are their watermark embedded counterparts respectively. The details are presented in Table 3.

Table 6.3: Watermark detector results, measured in BCR, with various compression degrees.

φ	PSNR without watermark	PSNR with watermark	capacity (bits)	BCR
1	60.89dB	50.38dB	220	100%
0.625	45.63dB	43.58dB	220	100%
0.5	40.83dB	38.07dB	373	100%
0.25	30.59dB	28.83dB	373	100%
0.1	23.91dB	23.41dB	251	100%

Robustness is tested on four conditions: JPEG baseline, JPEG2000 compression, additive noise and low-pass (noise-removal) filtering.

JPEG is widely used for image compression. Figure 6.6 shows the results of the test for robustness against JPEG compression. When JPEG compression quality factor is between 50 and 100, the BCR is 100% for $\varphi = 0.1$ and nearly 100% for $\varphi = 0.25$ and 0.5. For $\varphi = 1$, the watermark scheme is reliable until the quality factor is smaller than 60, as shown in .

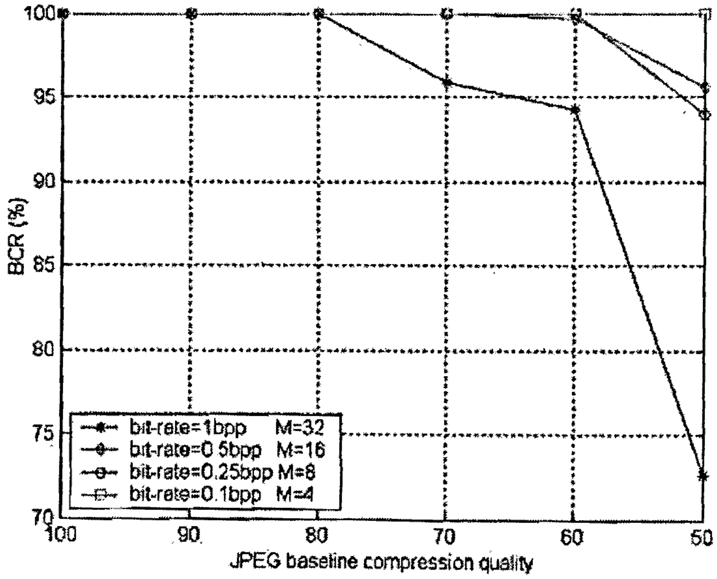


Figure 6.6: The robustness of watermarked images against JPEG baseline compression with four different bit-rates.

Also we test the watermarked image with JPEG 2000 compression with different bit rate φ. The results are shown in Figure 6.7. For $\varphi = 1$ and $\varphi = 0.5$, the decoded watermark is reliable until the compression bit-rate smaller than 0.5bpp.

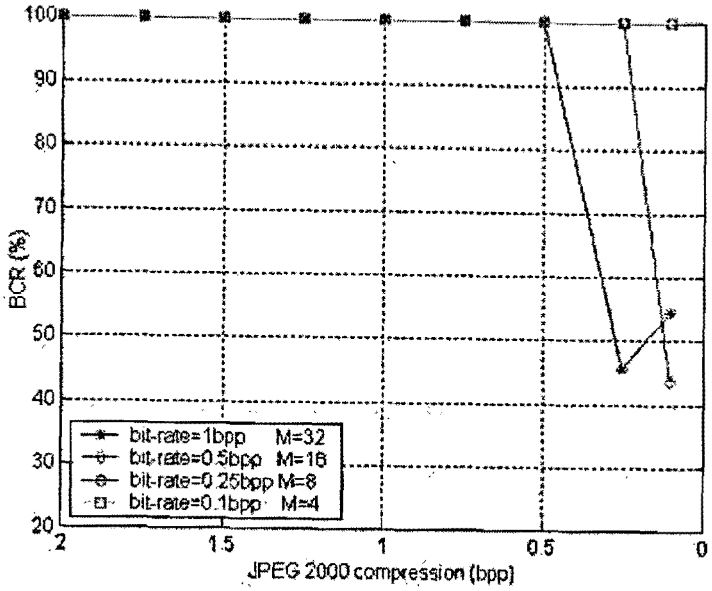


Figure 6.7: The robustness of watermarked images against JPEG2000 compression with four different bit-rates.

Noise is the most common distortion in image processing and transmission. In the experiments, Gaussian noise with variance from 0.01 to 0.05 is added into the watermarked image. As shown in Figure 6.8, The BCR is about 100% for $\varphi = 0.1$ and 0.25. For $\varphi = 1$ and 0.5, the detected watermark is not reliable when the noise variance is larger than 0.02.

Noise-removal filter is another common attack to the watermarked image. In the experiments, Wiener filter is used to filter the watermarked image with estimated noise variance of 0.01, 0.02, 0.03, 0.04 and 0.05. Refer to Figure 6.9, the BCR is 100% for $\varphi = 0.1$, around 95% for $\varphi = 0.5$ and 0.25, and 90% for $\varphi = 0.1$.

It can be seen from the experimental result, the new compression adaptive watermarking algorithm is robust to these common attacks and distortions while keeping an acceptable visual quality of the image.

6.6 Summary

In this chapter, we presented a compression degree adaptive watermarking method integrated with JPEG2000 image compression standard. Binary watermark is embedded into

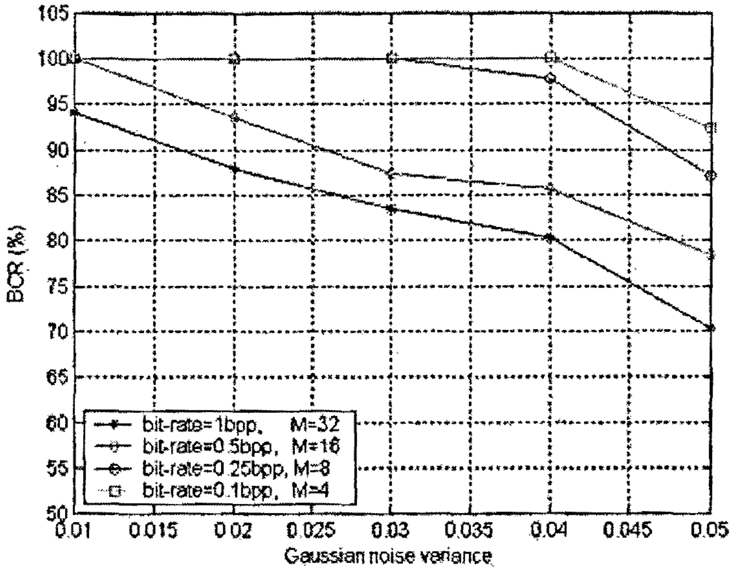


Figure 6.8: The robustness of watermarked images against additive noise with four different bit-rates.

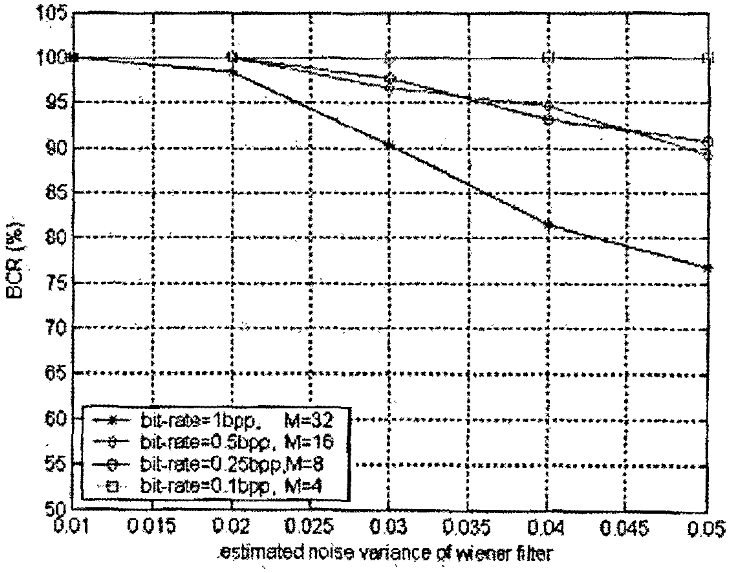


Figure 6.9: The robustness of watermarked images against noise-removal filtering with four different bit-rates.

middle frequency wavelet coefficients after quantization. During image decompression, the watermark is decoded without assistance from either the original image or the reference watermarks. The interval of quantization is designed based on the target bit rate, hence the strength of watermark is proportional to the compression ratio. We point out that, in this way, not only watermark can survive rate distortion, the tradeoff between visual quality and robustness is also balanced. The experiments show that the new algorithm has a good performance in terms of both robustness and fidelity.

Chapter 7

Conclusions and Future Work

Imperceptibility, robustness against moderate compression and processing are the basic but rather contradictory requirements for watermarking applications. The design of a successful watermarking scheme always involves a tradeoff between imperceptibility and robustness. This thesis focuses on the situation in which the watermarked signal undergoes common signal processing: additive noise, filtering, lossy compression. Distortion brought by watermarking is also considered as keeping a multimedia work's commercial value is a prerequisite for all data hiding algorithms.

In this work, two common types of watermarking algorithms are considered: spread spectrum and quantization based watermarking algorithm. For the spread spectrum watermarking schemes, we present an adaptive Wiener denoising filter based watermark detector and the experimental results show that it has better performance than Hartung's low-pass filter based detector.

For quantization based watermarking algorithms, the problem of designing embedding algorithm is transformed to designing signal constellations to which the host data is mapped to embed watermarking. Here both the look-up table for LUT embedding and the quantization ensembles for quantization watermarking are deemed as signal constellations. The signal constellation determines the property of the watermarking scheme: robustness, distortion, etc. That is, depending on LUT or the position of quantizer points, a quantization based watermarking algorithm presents robustness to attacks and distorts the host data. But no scheme can achieve both maximum robustness and minimum distortion at the same time. In this thesis, This problem is changed to look for an optimal watermarking strategy with respect to the embedding distortion given robustness constraint or with respect to robustness given fidelity criterion. A solution to optimizing quantization watermarking schemes is provided based on information theory, besides, robustness-distortion function $R(D)$ and distortion-robustness function $D(R)$ are developed. Experimental results show that the generated watermarking scheme is superior in terms of robustness and fidelity.

Proposed methods in this thesis can be applied directly in most applications where robustness and fidelity are major concern, or applied to some previously proposed and future robust watermarking algorithms to enhance performance. It is important to notice that our work in this thesis does not cover all aspects of multimedia data hiding. This field is so wide that various disciplines such as image/audio/video signal processing, computer security, human perception and business are involved. Therefore, studying various aspects of data hiding continues to be necessary.

A few possible future research directions are:

1. We have noticed that changes in different coefficients may have different perceptual sensitivity on human eyes. Thus, human perceptual models are often theoretically and experimentally derived to determine the changes on a signal which remain imperceptible. One of these is the Just-Noticeable-Difference (JND) model. The JND threshold is such that changes in the frequency content in the image/audio/video in the particular frequency hand below the threshold are not noticeable. It would be interesting to incorporate JND model into our analysis.

2. The watermark embedding and extraction are treated as a watermark communication channel. The capacity associated with the watermark channel is used to evaluate the efficiency of watermark scheme. Channel capacity is a theoretical upper bound of how many bits of information can be reliably transmitted through the channel with arbitrarily low probability of bit error. Channel coding theorem states that all

rates below capacity are achievable. However, for the real-world scenarios in today's data hiding research, there exists a discrepancy between the theoretical capacity and practically achievable watermark embedding capacity. A potential further research problem is how to encode and decode information to approach the channel capacity.

3. In a spread spectrum algorithm, or quantization based algorithm using repetition code, each watermark bit is transmitted through parallel channels simultaneously. Each channel has its own noise characteristics. How to spread and embed watermark into frequency coefficients to survive various expectable channel noise and how to extract watermark from individual extracted watermark bits are challenging topics. Channel coding which involves the addition of redundancy to allow robustness to a noisy transmission environment is a promising solution to this problem.

4. More emphasis should be placed on applications. We can see more and more real-time multimedia services are delivered through internet to a mix of users. A possible solution is source coding (quantization and compression) methods combined with transmission schemes providing different grades of services. Watermarking schemes integrated with this **joint source and channel coding** is yet to be studied in detail and optimized. As an example, to hide information in video stream over internet is needed to defend pirates, track transaction and access control. Then more factors, such as the property of the network, video CODEC, etc., have to be taken into account before a successful data hiding algorithm can be achieved.

Appendix A

Constrained Optimization: Lagrange's Method

The Lagrange method is a method used for constraint optimization. Suppose we want to maximize (or minimize) a function of n variables:

$$f(x) = f(x_1, x_2, ..., x_n) \text{ for } \mathbf{x} = (x_1, x_2, ..., x_n) \tag{A.1}$$

subject to p constraints

$$g_1(\mathbf{x}) = c_1, \quad g_2(\mathbf{x}) = c_2, ..., \text{ and } g_p(\mathbf{x}) = c_p \tag{A.2}$$

The first step of Lagrange's solution is to introduce p new parameters and write down the Lagrangian function:

$$L = f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \lambda_2 g_2(\mathbf{x}) + ... + \lambda_p g_p(\mathbf{x}), \tag{A.3}$$

the new parameters λ is called the Lagrange multiplier. L has became the function we want to maximize. Therefore we take partial derivatives of L and set them equal to zero. Hence the constrained optimization problem is formulated and solve by the following theorem.

Theorem (Lagrange) Assuming appropriate smoothness conditions, minimum or maximum of $f(\mathbf{x})$, subject to the constraints (A.2), that is not on the boundary of the region where $f(\mathbf{x})$ and $g_j(\mathbf{x})$ are defined can be found by introducing p new parameters $\lambda_1, \lambda_2, ..., \lambda_p$ and solving the system

$$\frac{\partial}{\partial x_i}(f(\mathbf{x}) + \sum_{j=1}^{p} \lambda_j g_j(\mathbf{x})) = 0, \qquad 1 \leq i \leq n \tag{A.4}$$

$$g_j(\mathbf{x}) = c_j, \qquad 1 \leq j \leq p \tag{A.5}$$

This amounts to solving $n + p$ equations for the $n + p$ real variables in $\mathbf{x}$ and λ.

Bibliography

[1] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Morgan Kaufmann, 2001.

[2] M. Wu, *Multimedia Data Hiding*, Ph.D. thesis, Princeton University, 2000.

[3] C.-Y. Lin, *Watermarking and Digital Signature Techniques for Multimedia Authentication and Copyright Protection*, Ph.D. thesis, Columbia University, 2000.

[4] D. Kundur, *Multiresolution Digital Watermarking: algorithm and implication for multimedia signals*, Ph.D. thesis, University of Toronto, 1999.

[5] J. Cox, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing*, vol. 6, pp. 1673–1687, 1997.

[6] F. Hartung and M. Kutter, "Multimedia watermarking technique," *Proc. of IEEE*, vol. 87, July 1999.

[7] M. Wu, "Joint security and robustness enhancement for quantization based data embedding," *IEEE Trans. on Circuits and Systems for Video Tech*, vol. 13, no. 8, pp. 831–841, Aug. 2003.

[8] J. Zhao, E. Koch, and C.Luo, "In buisness today and tomorrow," *Communications of the ACM*, vol. 41, no. 7, pp. 67–72, July 1998.

[9] D. Kundur and D. Hatzinakos, "A robust digital image watermarking method using wavelet-based fusion," *Proc. Int. Conf. Image Processing*, vol. 1, pp. 544–547, 26-29 Oct. 1997.

[10] R. B. Wolfgang and E. J. Delp, "A watermark for digital images," *Proc. Int. Conf. Image Processing*, vol. 3, pp. 219-222, 1996.

[11] J. J. K. ÓRuanaidh, W. J. Dowling, and F. M. Boland, "Watermarking digital images for copyright protection," *IEE Proc. Vision, Image and Signal Processing*, vol. 143, pp. 250–256, Aug. 2000.

[12] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1423–1443, May 2001.

[13] B. Chen and G. W. Wornell, "An information-theoretic approach to the design of robust digital watermarking systems," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 4, pp. 2061–2064, 1999.

[14] X.-G. Xia, C. G. Boncelet, and G. R. Arce, "A multiresolution watermark for digital images," *Proc. Int. Conf. Image Processing*, vol. III, pp. 548–551, Santa Barbara, CA, Oct. 26-29 1997.

[15] F. Hartung, J. K. Su, and B. Girod, "Spread spectrum watermarking: Malicious attacks and counterattacks," *Proc. SPIE Security and Watermarking of Multimeida Contents*, vol. 3657, Jan. 1999.

[16] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proc. of IEEE*, vol. 87, no. 7, July 1999.

[17] B. Chen, *Deisgn and Analysis of Digital Watermarking, Information Embedding, and Data Hiding Systems*, Ph.D. thesis, MIT, 2000.

[18] G. Wu, E.-H. Yang, and W. Sun, "Optimization strategies for quantization watermarking with application to image authentication," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp. 672–675, 2003.

[19] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, pp. 289–293, 1958.

[20] J.G. Proakis, *Digital Communication*, McGraw-Hill Science/Engineering/Math, 4th edition, August 15 2000.

[21] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley *&* Sons, Inc, 1991.

[22] M. Kutter and F. Petitcolas, "A fair benchmark for image watermarking systems," *Proc. SPIE IS&T/SPIE's 11th Annu. Symp. Electrionic Imaging '99: Security and Watermarking of Multimedia Contents*, vol. 3657, Jan. 1999.

[23] P. Su, H. M. Wang, and C. C. J. Kuo, "An integrated approach to image watermarking and jpeg-2000 compression," *Journal of VLSI Signal Processing*, vol. 27, pp. 35–53, 2001.

[24] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, "The effect of matching watermark and compression transforms in compressed color images," *Proc. Int. Conf. Image Processing*, vol. 1, 1998.

[25] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, pp. 165–168, March 1980.

[26] J. S. Lim, *Two-Dimensional Signal and Image Processing*, Englewood Cliffs, 1990.

[27] H. Yuan and X.-P. Zhang, "Fragile watermark based on the gaussian mixture model in wavelet domain for image authentication," *Proc. Int. Conf. Image Processing*, vol. 1, pp. 505–508, 14-17, Sep. 2003.

[28] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," *IEEE Trans. on Image Processing*, vol. 10, no. 7, July 2001.

[29] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

[30] J. Max, "Quantizing for minimum distortion," *IRE Transaction on Information Theory*, vol. IT-6, pp. 7–12, Mar. 1960.

[31] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 1–63, Oct. 1998.

[32] M. Rabbani and R. Joshi, "An overview of the jpeg 2000 still image compression standard," *Signal Processing: Image Communication*, 2000.

[33] J. M. Shapiro, "An embedded hierarchical image coder using zerotrees of wavelet coefficients," *Data Compression Conference*, pp. 214–223, 1993.

[34] A. Said and W. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuit and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.

[35] P. Meerwald, "Quantization watermarking in the jpeg2000 coding pipeline," *Communications and Multimedia Security Issues of The New Century, IFIP TC6/TC11 Fifth Joint Working Conference on Communications and MultimediaSecurity*, pp. 69–79, May 2001.

[36] T.-S. Chen and J.-G. Chen, "A simple and effective watermarking technique based on jpeg2000 codec," 2002, http://www.bohr.idv.tw/pdf/F022.pdf (in chinese).

[37] D. Taubman, "High performance scalable image compression with ebcot," *IEEE Trans. on Image Processing*, vol. 9, no. 7, pp. 1158–1170, July 2000.