UNIVERSITY OF CALGARY

Chaotic Spread Spectrum with Application to Digital Image Watermarking

by

Siyue Chen

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

CALGARY, ALBERTA

AUGUST, 2001

ABSTRACT

Digital watermarking is an enabling technology to prove ownership on copyrighted material, detect originators of illegally copies, and to monitor the usage of the copyrighted multimedia data. Currently, the most popular approach for digital image watermarking is the correlation-based spread spectrum technique. In this thesis, several novel techniques for image watermarking based on chaotic spread spectrum system are proposed. First, the chaotic spreading sequences are employed to spread out information signal instead of classical spreading sequences. Second, the chaotic parameter modulation is proposed to generate watermarks. The copyright information is embedded into the parameters of a chaotic dynamical system. The retrieval of copyright information is then a problem of parameter estimation from the extracted and possibly corrupted watermark signal. Different estimation methods are designed and the performances of them are analysed. It is shown that the chaotic parameter modulation based on mean value detection is not only superior to all the other estimation approaches, but it also has a better performance than the conventional correlation-based spread spectrum technique in terms of robustness, security and payload.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. H. Leung sincerely, for his guidance, advice, patience and encouragement throughout the course of this work, and for his constructive criticism offered during the writing of this thesis.

The author also wants to thank the persons working in the Multimedia Signal Processing Laboratory. The author has a wonderful collaborating experience with H. Yu, N. Xie, and V. Varadan during the whole research process. The assistances of K. Murali, W. Li and Y. Xia are also greatly appreciated.

Finally, the author wishes to thank all the staffs within the Department of Electrical and Computer Engineering for their support and assistances.

To

My dear mother and father

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

AC	Alternative current
A/D	Analog-digital conversion
BER	Bit error rate
CPM	Chaotic parameter modulation
D/A	Digital-analog conversion
DC	Direct current
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
DP	Dynamical programming
DSSS	Direct-sequence spread spectrum
DVD	Digital versatile disk
EKF	Extended Kalman filter
HM	Halving method
LMS	Least mean square
LSB	Least significant bit
MSE	Mean square error
MVD	Mean value detection
PN	Pseudo-noise
PWAM	Piecewise-affine Markov
RLS	Recursive least square
SNR	Signal-to-noise ratio
SS	Spread spectrum

CHAPTER 1

INTRODUCTION

In the past decade, there has been an explosion in the use and distribution of digital multimedia data. PCs with Internet connections have taken homes by storm and have made the distribution of multimedia data and applications much easier and faster. Electronic commerce applications and online services are rapidly being developed. Even the analog audio and video equipments in the home are in the process of being replaced by digital successors. As a result, we can see the digital mass recording devices for multimedia data entering the consumer market of today.

1.1 Need for Watermarking

Although digital data has many advantages over analog data, service providers are reluctant to offer services in digital form because they fear unrestricted duplication and dissemination of copyrighted materials. Because of possible copyright issues the intellectual property of digitally recorded materials must be protected [1]. The lack of such adequate protection systems for copyrighted contents was the reason for the delayed introduction of the digital versatile disk (DVD) [2]. Several media companies initially refused to provide DVD materials until the copy protection problem had been addressed [3, 4]. Representatives of the consumer, electronics industry and the motion picture industry have agreed to seek legislation concerning digital video recording devices. Recommendations describing ways that would protect both intellectual property and consumers' rights have been submitted to the U.S. Congress [4] and resulted in the Digital Millennium Copyright Act [5], which was signed by President Clinton on October 28, 1998. The European Union is also preparing such intellectual property rights protection for digital multimedia products including CDs and DVDs [6].

To provide copy protection and copyright protection for digital audio and video data, two complementary techniques are being developed: encryption and watermarking [7]. Encryption techniques can be used to protect digital data during the transmission from the sender to the receiver [8]. After the receiver has received and decrypted the data,

however, the data is identical to the original data and no longer protected. Watermarking techniques can complement encryption by embedding secret imperceptible signals, a watermark, directly into the original data in such a way that it always remains present. Such a watermark, for instance, can be used for the following purpose:

- Copyright protection: For the protection of intellectual property, the data owner can embed a watermark representing copyright information in his data. This watermark can prove his ownership in court when someone has infringed on his copyrights.
- Fingerprinting: To trace the source of illegal copies, the owner can use a fingerprinting technique. In this case, the owner can embed different watermarks in the copies of the data that are supplied to different customers. Fingerprinting can be compared to embedding a serial number that is related to the customer's identity in the data. It enables the intellectual property owner to identify customers who have broken their license agreement by supplying the data to third parties.
- Copy protection: The information stored in a watermark can directly control digital recording devices for copy protection purpose [9]. In this case, the watermark represents a copy-prohibit bit and watermark detectors in the recorder determine whether the data offered to the recorder may be stored or not.
- Broadcast monitoring: By embedding watermarks in commercial advertisements, an automated monitoring system can verify whether advertisements are broadcasted as contracted [10]. Not only commercials but also valuable TV products can be protected by broadcast monitoring [11]. News items can have a value of over US$100,000 per hour, which makes them very vulnerable to intellectual property rights violation. A broadcast surveillance system can check all broadcast channels and charge the TV stations according to their findings.
- Data authentication: Fragile watermarks [12] can be used to check the authenticity of the data. A fragile watermark indicates whether the data has been altered and supplies localization information as to where the data was altered.

Watermarking techniques are not only used for protection purposes. Other applications include:

- Indexing: Indexing of video mails, where comments can be embedded in the video content; indexing of movies and news items, where markers and comments that can be used by search engines are inserted.
- Medical safety: Embedding the patient's name and other information in medical images could be a useful safety measure.
- Data hiding: Watermarking techniques can be used for the transmission of secret private messages. Since various governments restrict the use of encryption services, people may hide their messages in other data.

1.2 Requirements for Watermarking

Each watermarking application has its own specific requirements. Therefore, there is no set of requirements to be met by all watermarking techniques. Nevertheless, some general directions can be given for most of the applications mentioned above:

- Robustness: A watermark should stay in the host data regardless of whatever happens to the host data, including all possible signal processing that may occur, and including all hostile attacks that unauthorized parties may attempt. It is a key requirement for copyright protection or conditional access applications, but less important for the applications of using fragile watermarks, since failure to detect the watermark proves that the host data has been modified and is no longer authentic. The possible signal processing that a watermark should be resistant to includes all data manipulations and modifications that the data might undergo in the transmission and storage, such as lossy compression, filtering, re-sampling, digital-analog (D/A) and analog-digital (A/D) conversion. "Attack" denotes the data manipulation with the purpose of impairing, destroying, or removing the embedded watermarks, such as noise addition, overmarking, and conspiracy. In general, there should be no way in which the watermark can be removed or altered without sufficient degradation of the perceptual quality of the host data so as to render it unusable.
- Imperceptibility: Another main requirement for watermarking is the perceptual transparency. The data embedding process should not introduce any perceptible artifact into the host data. On the other hand, for high robustness, it is desirable that the watermark amplitude is as high as possible. Thus, the design of a watermarking method

always involves a tradeoff between imperceptibility and robustness. It would be optimal to embed a watermark just below the threshold of perception. However, this threshold is difficult to determine for real-world images, video and audio signals. Several measures to determine objectively perceived distortion and the threshold of perception have been proposed for the mentioned media [13]. However, most of them are still not perfect enough to replace human viewers or listeners who judge the visual or audio fidelity through blind tests. Thus, a watermark-embedding procedure is truly imperceptible if humans cannot distinguish the original data from the data with the inserted watermark.

- Security: The security of watermarking techniques can be interpreted in the same way as the security of encryption techniques. Kerckhoff's assumption states that one should assume that the method used to encrypt the data is known to an unauthorized party and that the security must lie in the choice of a security key [14]. For example, in many schemes, pseudorandom signals are embedded as watermarks. In this case, the description and the seed of a pseudorandom number generator may be used as the security keys.
- Payload: Depending on the applications, some information such as transaction dates and serial numbers, etc., are all needed to be inserted in a watermark signal. It is therefore desirable to store as much as possible information in a watermark.
- Oblivious and nonoblivious watermarking [13]: In some applications, like copyright protection and data monitoring, watermark extraction algorithms can use the original unwatermarked data to find watermarks. This is called nonoblivious watermarking. In most other applications, e.g., copy protection and indexing, the watermark extraction algorithms do not have access to the original unwatermarked data. This renders the watermark extraction more difficult. Watermarking algorithms of this kind are referred to as oblivious watermarking. Watermark recovery is usually more robust if the original, unwatermarked data are available. The availability of the original data in the recovery process allows the detection and inversion of distortions, which change the data geometry. This helps, for example, if a watermarked image has been rotated by an attacker.

1.3 Overview of Previous Works on Image Watermarking

Most watermarking research and publications are focused on images. The reason might be that there is a large demand for image watermarking products due to the fact that there are so many images available at no cost on the World Wide Web, which need to be protected. Moreover, the method for image watermarking can also be extended and applied to video and audio data, etc..

The year of 1993 can be considered the beginning of the digital image watermarking era, although other publications from the early 1990's, such as Tanaka *et al* [15], already introduced the idea of tagging images to secretly hide information and assure ownership rights. Caronni [16] describes an overall system to track unauthorized image distributions. He proposes to mark images using spatial signal modulation and calls the process tagging. A tag is a block. All possible locations in an image where a tag could possibly be placed are identified by calculating the local region variance in the image and comparing it to the empirically identified upper and lower limits. Only locations with minimal variances are used for tagging. He also suggests to use the correlation coefficient between the original and the tagged image as a measure for the image degradation due to the tagging process. While the idea of hiding a spatial watermark in an image is fundamentally sound, this scheme may be susceptible to the attacks of filtering and redigitization. The fainter such watermarks are, the more susceptible they are to such attacks. And geometric shapes provide only a limited alphabet with which to encode information.

In the same year, approaches and ideas for digital image watermarking were proposed by Tirkel *et al.* [17] in their 1993 publication entitled "Electronic Water Mark". In this early publication on digital watermarking, the authors already recognized the importance of digital watermarking and proposed possible applications for image tagging, copyright enforcement, counterfeit protection, and controlled access to image data. Two methods were proposed for greyscale images. In the first approach, the watermark in form of an *m*-sequence-derived pseudo-noise (PN) code is embedded in the least significant bit (LSB) plane of image data. This method is actually an extension to simple LSB coding schemes in which the LSBs are replaced by the coding information. The watermark decoding is relatively straightforward since the LSB plane carries the

watermark without any distortion. In the second approach, the watermark, again in form of an *m*-sequence-derived code, is added to the LSB plane. The decoding process makes use of the unique and optimal autocorrelation function of *m*-sequences [18]. A modified version of the paper was published in 1994 [19] titled "A Digital Watermark", and being the first publication explicitly mentioning, and hence defining, the term digital watermarking. In 1995 [20], the idea of using *m*-sequences and the LSB addition was extended and improved by the authors through the use of two-dimensional *m*-sequences, which resulted in more robust watermarks. There are several drawbacks to these schemes. The most important is that they are susceptible to signal processing, especially requantization, and geometric attacks such as cropping. Furthermore, they degrade an image in the way that predictive coding and dithering can.

Since the above-mentioned publications, the interest and research activities on watermarking have largely increased. Even though the number of image watermarking methods is very large, most techniques share some common principles. The watermark signal is typically a pseudorandom signal with low amplitude, compared to the image amplitude, and usually with a spatial distribution of one information bit over many pixels. In order to avoid visibility of the embedded watermark, an implicit or explicit spatial or spectral [21, 22, 23] shaping is often applied with the goal to attenuate the watermark in areas of an image where it would otherwise become visible. The resulting watermark signal is sometimes sparse and leaves image pixels unchanged [24, 25], but usually it is dense and alters all pixels of the image to be watermarked. The signal embedding is done by addition [26, 27], or signal adaptive addition [28], mostly to the luminance channel alone. The addition can take place in the spatial domain, or in transform domains such as the discrete Fourier transform (DFT) domain [29], the full-image discrete cosine transform (DCT) domain [30, 31], the block-wise DCT domain [19, 32] and other domains [33, 34, 35]. The watermark recovery is usually accomplished by some sort of correlation method. Since a pseudorandom signal usually has a high auto-correlation and a low cross-correlation, a watermark can be detected when a high correlation value is observed.

Despite the many efforts that are underway to develop and establish watermarking technology, watermarking is still not a fully mature and understood technology, and a lot

of questions are not yet answered. For instance, although some correlation-based watermarking systems [19, 29, 31] claim to be robust to various processing and attacks, their payloads are very small. That is, only one bit can be inserted into images. The invisible 10,000 bits per image watermark that resists all attacks whatsoever is an illusion for most schemes. The realistic numbers are approximately two orders of magnitude lower. Even when designed under realistic expectations, watermarks offer robustness against nonexperts but may still be vulnerable to attacks by experts. Therefore, we propose some new watermarking techniques based on chaotic spread spectrum, which can improve the robustness performance and increase the payload of watermarks at the same time.

The thesis is organized as follows:

Chapter 2 gives a review of the conventional correlation-based spread spectrum watermarking scheme. The detections with and without the original image available are both discussed. The performances of them are also analysed and tested.

In Chapter 3 we show the use of chaotic spreading sequences in the conventional correlation-based spread spectrum watermarking scheme. The chaotic spreading sequence is noise-like, wideband, and it possesses good auto-correlation and cross-correlation properties. Meanwhile, it is very easy to generate and store. Thus, the use of chaotic spreading sequences is superior to the use of classical spreading sequences, such as *m*-sequences and Gold sequences.

Chapter 4 presents a watermarking scheme based on chaotic parameter modulation. The copyright information will be stored in the parameter of a chaotic dynamical system and the detection of them becomes the problem of estimating the parameter. The proposed approach does not require the synchronization of a spreading sequence. Thus, the detection procedure is greatly simplified. Furthermore, since the chaotic parameter modulation is originally proposed for analog communications, it can modulate the information of numerical value directly so that the payload of watermarks can be increased. However, the robustness performance of this method is not so desirable.

Finally, in Chapter 5 we propose a novel watermarking scheme by using chaotic parameter modulation based on mean value detection. It solves all the problems encountered in the conventional spread spectrum technique and the chaotic parameter

modulation based on other estimation methods. It is shown that the new scheme is successfully resistant to all attacks while the payload is maintained high.

Conclusion remarks are given in Chapter 6.

CHAPTER 2

CONVENTIONAL CORRELATION-BASED SPREAD SPECTRUM WATERMARKING

2.1 Introduction

Currently, the most popular approach for digital watermarking is the correlation-based spread spectrum (SS) technique [27, 29, 36, 37], which follows the idea of the direct-sequence spread spectrum (DSSS) modulation for communications. Figure 2.1 illustrates a model for spread spectrum communications, in which a narrow band signal is transmitted over a much larger bandwidth by the use of a PN sequence. The signal energy present in any single frequency is so small that it has a low probability of being detected and intercepted.

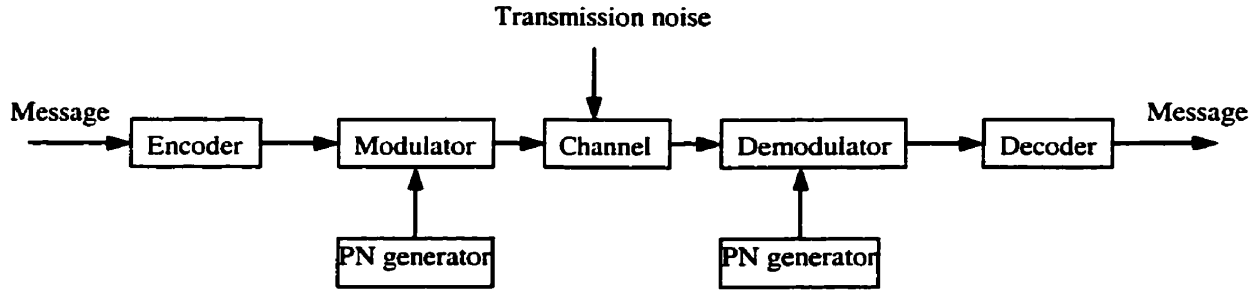

Figure 2.1 A model of a spread spectrum communication system

When the same rationale is applied to watermarking, images are viewed as communication channels. Correspondingly, the watermark is viewed as a signal that is transmitted through images. Image distortions, which are introduced by processing or attacks, are thus treated as noises that the immersed watermark must be immune to. Figure 2.2 shows the block diagram of a watermarking system as spread spectrum communications. The watermark is also spread over many frequency bins so that the energy in any one bin is very small and hence undetectable. Nevertheless, since the

location and content of the watermark is given in the extraction process, it is then possible to combine these many weak signals to form a single signal with a high signal-to-noise ratio (SNR). Another advantage of this approach is that to destroy such a watermark would require adding noises to all frequency bins. In other words, the quality of the original image will degrade greatly well before the watermark is lost.

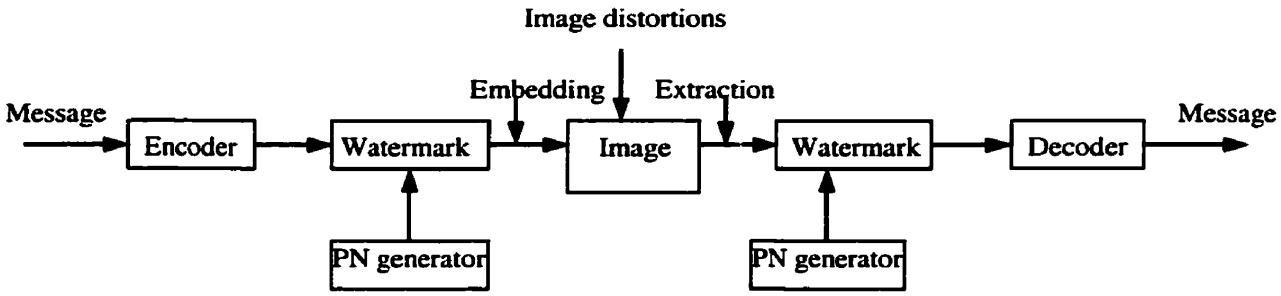

Figure 2.2 Watermarking as spread spectrum communications

As mentioned in Chapter 1, watermarking can be divided into two categories: nonoblivious watermarking and oblivious watermarking. In this Chapter, both of them are described and the performances of them are analysed.

2.2 Conventional Spread Spectrum Nonoblivious Watermarking

In a conventional SS watermarking system, the information signal that is to be inserted into watermarks has to be first encoded into a sequence of binary values denoted as $(b(0), b(1), \cdots b(i), \cdots b(L-1))$, where $b(i) \in \{1, -1\}$ and L is the length of the binary sequence. Each bit is modulated by multiplying with a PN spreading sequence $\mathbf{p} = (p_0, p_1, \cdots, p_{N-1})$, where N is the length of the spreading sequence. The modulated signal is then used as the watermark, that is, $\mathbf{x}(i) = b(i)\mathbf{p}$. To embed a watermark, a sequence of N values from the original image has to be extracted and to be added to the watermark, i.e., for $i \in \{0, 1, \cdots, L-1\}$

$$w_n = v_n + \alpha x_n(i) = v_n + \alpha b(i) p_n, \tag{2.1}$$

where $x_n(i)$ is the nth element of $\mathbf{x}(i)$. v_n is the nth element of the vector $\mathbf{v} = (v_0, v_1, \cdots, v_n, \cdots, v_{N-1})$ which is the N original image pixel values where the watermark signals are inserted. $\mathbf{w} = (w_0, w_1, \cdots, w_n, \cdots, w_{N-1})$ is the watermarked signal

and will be inserted back into the image in place of $\mathbf{v}$ to obtain the watermarked image for transmission. α is a scalar used to amplify or attenuate the power of the watermark signal. If a watermark is to be embedded into different images, the value of α will be adjusted to make sure that the watermark is invisibly embedded into them. There is a tradeoff existing in the selection of α since the energy of a watermark must be low to keep it perceptually invisible but is preferred to be high to increase the SNR for detection. Different images may exhibit more or less tolerance to modifications. Thus, we can view α as a relative measure of how much one must alter v_n to alter the perceptual quality of the image. A large α means that one can perceptually "get away" with altering the image pixel v_n by a large factor without degrading the image.

Since the original image is available at the receiver end, the watermark can be extracted by simply reversing the insertion procedure. That is,

$$\mathbf{y}(i) = \frac{1}{\alpha}(\mathbf{w}' - \mathbf{v}), \tag{2.2}$$

where $\mathbf{y}(i)$ is the retrieved watermark and $\mathbf{w}'$ is the watermarked image signal which may be corrupted by image processing and attacks. If we let $\mathbf{d}$ be the distortion, we have

$$\begin{aligned} \mathbf{y}(i) &= \frac{1}{\alpha}(\mathbf{w} + \mathbf{d} - \mathbf{v}) \\ &= \mathbf{x}(i) + \frac{1}{\alpha}\mathbf{d} \\ &= b(i)\mathbf{p} + \frac{1}{\alpha}\mathbf{d}. \end{aligned} \tag{2.3}$$

To detect the information bit $b(i)$, the correlator given below is applied

$$\begin{aligned} \hat{b}(i) &= sign(\frac{1}{N}\mathbf{y}(i)\cdot\mathbf{p}) \\ &= sign(\frac{1}{N}\sum_{n=0}^{N-1} y_n(i)p_n) \\ &= sign\left(\frac{1}{N}\sum_{n=0}^{N-1}[(b(i)p_n + \frac{1}{\alpha}d_n)p_n]\right) \\ &= sign\left(\frac{1}{N}(\sum_{n=0}^{N-1} b(i){p_n}^2 + \sum_{n=0}^{N-1}\frac{1}{\alpha}d_n p_n)\right) \\ &= sign\left(\frac{1}{N}(\Sigma_1 + \Sigma_2)\right), \end{aligned} \tag{2.4}$$

where the two summations $\Sigma_1 = \sum_{n=0}^{N-1} b(i) p_n^2$ and $\Sigma_2 = \sum_{n=0}^{N-1} \frac{1}{\alpha} d_n p_n$ in (2.4) correspond to the correlation sums from the watermark and the distortion respectively. When there is no distortion, we have $d_n = 0$, and so $\Sigma_2 = 0$. Since p_n^2 is equal to 1, the sign of Σ_1 will be the same as that of *b(i)*. Thus, the information bit can always be detected correctly if there is no distortion introduced to the watermarked image.

However, when distortions exist, i.e., $d_n \neq 0$, the correlation sum between the distortion and PN sequence, Σ_2, will have an impact on the detection results. And now, there is no guarantee that we can always retrieve *b(i)* correctly. A bit error occurs if $\hat{b}(i)$ obtained in (2.4) is not equal to $sign(\Sigma_1)$. That is, $\Sigma_1 < -\Sigma_2$ when the transmission bit is 1; or $\Sigma_1 > -\Sigma_2$ when the transmission bit is –1. Therefore, the probability of detection error can be written as

$$\begin{aligned} P_e &= P(|\Sigma_2| \geq \Sigma_1) \\ &= 2\frac{1}{\sqrt{2\pi}\sigma_{\Sigma_2}} \sum_{\Sigma_2=\Sigma_1}^{\infty} \exp\left(-\frac{\Sigma_2^2}{2\sigma_{\Sigma_2}^2}\right) \\ &= 2\frac{1}{\sqrt{2\pi}\sigma_{\Sigma_2}} \sqrt{\pi/2}\sigma_{\Sigma_2} erfc\left(\frac{\Sigma_1}{\sqrt{2}\sigma_{\Sigma_2}}\right) \\ &= erfc\left(\frac{\Sigma_1}{\sqrt{2}\sigma_{\Sigma_2}}\right). \end{aligned} \tag{2.5}$$

Assume that the distortion is white Gaussian noise with zero mean and variance σ_d^2, and that the PN sequence has a mean of μ_p and variance of σ_p^2. Since the distortion d_n and the PN sequence p_n are uncorrelated, the variance of Σ_2 can be obtained as

$$\begin{aligned} \sigma_{\Sigma_2}^2 &= E[\Sigma_2^2] = E\left[\left(\sum_{n=0}^{N-1}\left(\frac{1}{\alpha} d_n p_n\right)\right)^2\right] \\ &= \frac{1}{\alpha^2} E\left[\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} d_n d_m p_n p_m\right] \\ &= \frac{1}{\alpha^2} N(\mu_p^2 + \sigma_p^2)\sigma_d^2. \end{aligned} \tag{2.6}$$

Σ_1 can be expressed as $\Sigma_1 = \sum_{n=0}^{N-1} b(i) p_n^{\,2} = b(i) \sum_{n=0}^{N-1} p_n^{\,2}$. If N is sufficiently large and p_n is ergodic, it is given that

$$E[p_n^{\,2}] = \frac{1}{N} \sum_{n=0}^{N-1} p_n^{\,2}. \tag{2.7}$$

Then we have $\sum_{n=0}^{N-1} p_n^{\,2} = NE[p_n^{\,2}] = N(\mu_p^{\,2} + \sigma_p^{\,2})$, and hence

$$\Sigma_1 = b(i) \sum_{n=0}^{N-1} p_n^{\,2} = Nb(i)(\mu_p^{\,2} + \sigma_p^{\,2}). \tag{2.8}$$

Substituting (2.6) and (2.8) into (2.5), the probability of detection error is obtained as

$$\begin{aligned} P_e &= P(|\Sigma_2| \geq \Sigma_1) \\ &= erfc\left(\frac{\Sigma_1}{\sqrt{2}\sigma_{\Sigma_2}}\right) \\ &= erfc\left(\frac{Nb(i)(\mu_p^{\,2} + \sigma_p^{\,2})}{\sqrt{2}\frac{1}{\alpha}\sqrt{N(\mu_p^{\,2} + \sigma_p^{\,2})\sigma_d^{\,2}}}\right) \\ &= erfc\left(\frac{\alpha b(i)}{\sqrt{2}}\sqrt{\frac{N(\sigma_p^{\,2} + \mu_p^{\,2})}{\sigma_d^{\,2}}}\right). \end{aligned} \tag{2.9}$$

Equation (2.9) shows that the error probability gets smaller when the power of the PN sequence is large or when the power of the distortion is small. In other words, the larger the ratio between the power of the watermark signal and the power of the distortion, the smaller the value of the error probability is. If we define this ratio as SNR, we have

$$SNR = \frac{\sigma_P^{\,2} + \mu_P^{\,2}}{\sigma_d^{\,2}}. \tag{2.10}$$

Then the probability of detection error can be written as

$$\begin{aligned} P_e &= erfc\left(\frac{\alpha b(i)}{\sqrt{2}}\sqrt{\frac{N(\sigma_p^{\,2} + \mu_p^{\,2})}{\sigma_d^{\,2}}}\right) \\ &= erfc\left(\frac{\alpha b(i)}{\sqrt{2}}\sqrt{N \cdot SNR}\right). \end{aligned} \tag{2.11}$$

It is desirable that the SNR is as large as possible. However, for the watermark application there is a practical upper limit on the power of a PN sequence. It cannot be too large; otherwise the watermark will be perceptually visible in the image.

According to (2.11), the length of the PN sequence, N, also has an impact on the error probability. The smaller the error probability is, the longer the sequence required. However, N is also limited by the size of the original image. For instance, given a 256×256 image, if a PN sequence of $N=1024$ is used, the maximum number of information bits that can be embedded into the image is $L=\frac{256\times 256}{1024}=64$.

It should also be noted that (2.11) is derived under the condition of perfect synchronization. In other words, (2.11) is the ideal performance when there is no synchronization error. However, in many practical situations such as the well-known attack "StirMark" [38], in which pixels in the derivative image are placed at subtly distorted positions relative to those in the original watermarked image, synchronization is not quite possible. In these cases, the undistorted reference image is needed to identify the distortions that have been made. Or all possible shifts of the PN sequence have to be evaluated to identify the correct shift for PN sequence synchronization. No matter which method is employed, the synchronization procedure is cumbersome and complex, especially for PN sequences with a very large cycle.

2.3 Conventional Spread Spectrum Oblivious Watermarking

In the oblivious watermarking the original image is not available in the watermark detection process. The correlation is so applied between the watermarked pixel $\mathbf{w}'$ and the PN sequence $\mathbf{p}$. The response of the correlator is given by

$$\hat{b}(i)=sign\left(\frac{1}{N}(\frac{1}{\alpha}\mathbf{w}'\cdot\mathbf{p})\right) \qquad (2.12)$$

If no manipulation has been applied to the watermarked image, we have the extracted watermarked pixel equal to the original one. The information bit is retrieved as

$$\begin{aligned}\hat{b}(i) &= sign\left(\frac{1}{N}(\frac{1}{\alpha}\mathbf{w}'\cdot\mathbf{p})\right)\\ &= sign\left(\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{\alpha}(\alpha x_n(i)+v_n)p_n\right)\\ &= sign\left(\frac{1}{N}\sum_{n=0}^{N-1}(b(i)p_n+\frac{1}{\alpha}v_n)p_n\right)\\ &= sign\left(\frac{1}{N}(\sum_{n=0}^{N-1}b(i)p_n^{\,2}+\sum_{n=0}^{N-1}\frac{1}{\alpha}v_n p_n)\right)\\ &= sign\left(\frac{1}{N}(\Sigma_1+\Sigma_2)\right).\end{aligned} \tag{2.13}$$

Apparently, (2.13) is very similar to (2.4). The two summations, $\Sigma_1 = \sum_{n=0}^{N-1} b(i)p_n^{\,2}$ and $\Sigma_2 = \sum_{n=0}^{N-1}\frac{1}{\alpha}v_n p_n$, contribute to the correlation results. The only difference is that the distortion d_n in (2.4) is replaced by the original image pixel v_n. In other words, the image is basically considered as channel noise and cannot be deducted in the oblivious watermarking.

Assume that the image pixels being watermarked have the mean μ_v and variance σ_v^2. Following (2.11), the probability of error for the SS oblivious watermarking when no distortion is applied can be written as

$$\begin{aligned}P_e &= erfc(\frac{\Sigma_1}{\sqrt{2}\sigma_{\Sigma_2}})\\ &= erfc\left(\frac{\alpha b(i)}{\sqrt{2}}\sqrt{N\cdot SNR}\right).\end{aligned} \tag{2.14}$$

But here, the SNR is different from the one in (2.10) because the noise here is the image pixel v_n instead of the distortion d_n. Thus, the SNR that is the power ratio of the PN sequence and the noise can be presented as

$$SNR = \frac{\mu_p^{\,2}+\sigma_p^{\,2}}{\mu_v^{\,2}+\sigma_v^{\,2}}. \tag{2.15}$$

Equation (2.14) shows that the error probability of detection in the oblivious watermarking depends on the power ratio between the PN sequence and the image pixels if no extra distortions are introduced. Since the amplitude of a watermark should be kept

low enough to be imperceptibly embedded in images, the power of image pixels will be very significant compared to the power of watermark signals. Thus, the probability of detection error may be very large even when no distortion is applied to the watermarked image. This explains why watermark detection is normally more robust with the original image available.

When the watermarked image is subjected to distortions, the performance of the SS oblivious watermarking becomes even worse. The extracted watermarked data will be

$$w'_n = w_n + d_n = v_n + \alpha x_n(i) + d_n. \tag{2.16}$$

And the correlation is applied as

$$\begin{aligned}
\hat{b}(i) &= sign\left(\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{\alpha}w'_n p_n\right) \\
&= sign\left(\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{\alpha}(v_n + \alpha x_n(i) + d_n)p_n\right) \\
&= sign\left(\frac{1}{N}\sum_{n=0}^{N-1}(b(i)p_n + \frac{1}{\alpha}v_n + \frac{1}{\alpha}d_n)p_n\right) \\
&= sign\left(\frac{1}{N}(\sum_{n=0}^{N-1}b(i)p_n^2 + \frac{1}{\alpha}\sum_{n=0}^{N-1}(v_n + d_n)p_n)\right).
\end{aligned} \tag{2.17}$$

Following (2.4) and (2.11), (2.17) can also be written as

$$P_e = erfc\left(\frac{\alpha b(i)}{\sqrt{2}}\sqrt{N \cdot SNR}\right). \tag{2.18}$$

However, as shown in (2.17) the noise in the detector here is the combination of the image pixel and the distortion. Thus, the power of noises here will be even stronger compared to the one in (2.13) where only the image pixel is considered as noise. Accordingly, the value of SNR will be even smaller. And so there will be more detection errors.

2.4 Performance Test

In order to evaluate the performance of the conventional correlation-based SS watermarking scheme, a binary information sequence is generated. The most widely used PN sequence, *m*-sequence, is employed in this study to spread out the information bits. The image "Lena" is used as the host data as shown in Figure 2.3, and the watermark signal is added to it.

The scalar α is selected equal to 1 for this host image. And the watermarked image is shown in Figure 2.4. It is seen that the watermarked image has a little bit darker background than the original one due to the embedding of watermarks. However, if another image is used, saying the image "Slope" as shown in Figure 2.5, the watermarked image as shown in Figure 2.6 seems almost the same as the original one with the use of the same watermark signal and the same α. Hence, for different host images, all the watermarking procedure is the same except that the value of α should be evaluated to guarantee the embedding of watermarks does not introduce any perceptible distortion into the host image. For the same host image, it is obvious that the larger the α is, the more distortion is introduced to the image.

Figure 2.3 Original image of "Lena"

The watermarked image will be subjected to a series of image processing and attacks, including image resizing, cropping, rotating, median filtering and JPEG compression. These manipulations are preliminary, but show resilience to certain types of common processing. And they will be used through the whole thesis as the tests of robustness performances for different watermarking schemes.

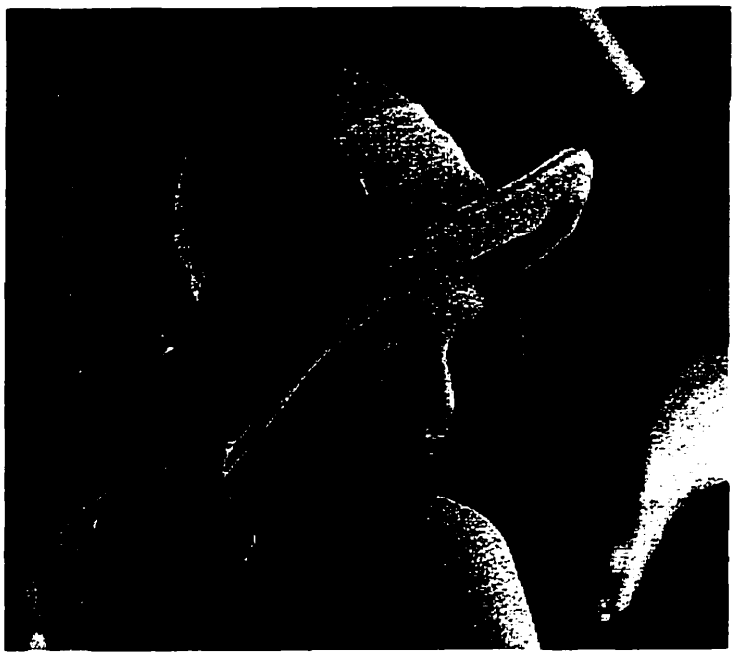

Figure 2.4 Watermarked image of "Lena" with $\alpha = 1$

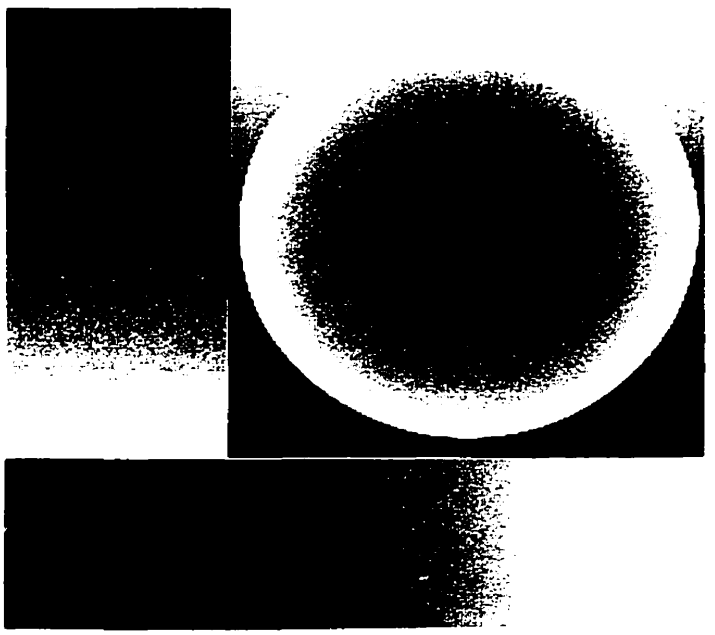

Figure 2.5 Original image of "Slope"

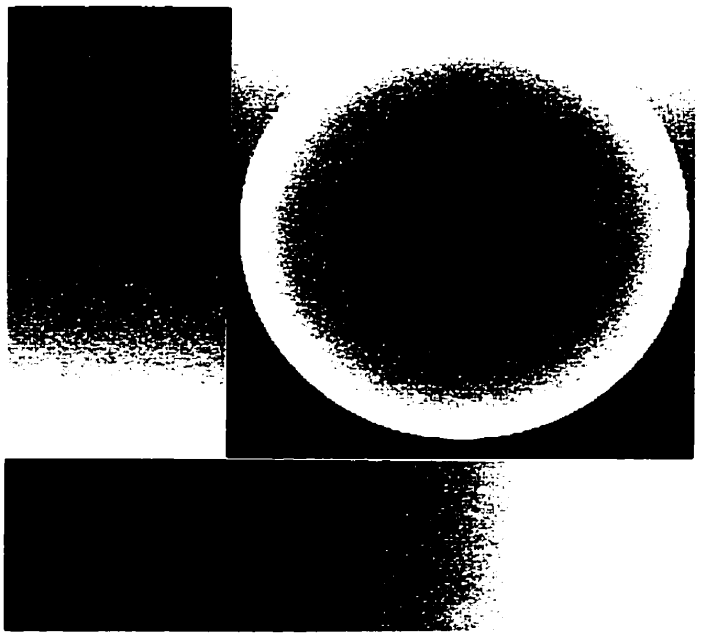

Figure 2.6 Watermarked image of "Slope" with $\alpha = 1$

Since robustness of a watermark refers to how much copyright information still can be retrieved after the original watermarked image is distorted by some manipulation, we use bit error rate (BER) to evaluate robustness of the watermarking scheme for binary information sequences. BER is calculated by comparing the original binary information sequence with the retrieved one, and dividing the number of the error bit by the total amount of the information bits. Each application has its own requirement for an acceptable value of BER. There has not been a standard for BER values in all watermarking applications. According to our experience, the acceptable value is set below 0.1 because under this value we can use error control codes to correct the wrong retrieved bit [35]. The results from simulations are compared to the analytical expressions of (2.11) and (2.18).

A. Image resizing

Image resizing changes the size of an image by using a specified interpolation method. The simplest one is the nearest neighbourhood interpolation [39]. When the size of the original image is enlarged, the extra pixels having the same values as those of the nearest pixels are interpolated. When the size is shrunk, the original pixels are sampled to

create a new one. By doing so, some image pixels are removed, and part of the watermark signal that is added to these removed pixels is removed as well. To analyse this effect we resize the image to 3/4 of its original size, as shown in Figure 2.7. That is, for every 4 columns or rows, there is one column or row of pixels being removed from the original image. This process results in a considerable loss in fine details, when we compare Figure 2.7 with Figure 2.4.

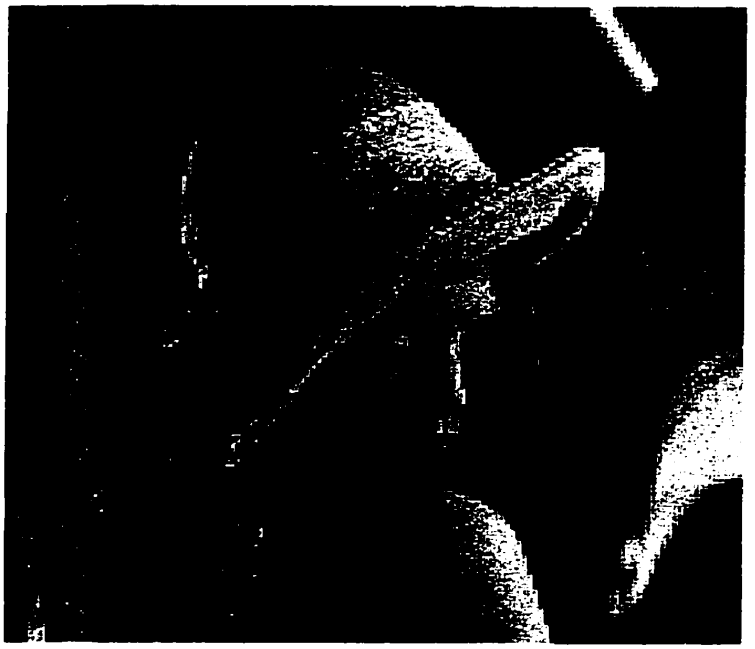

Figure 2.7 The image being resized to 3/4 of its original size

In this study, seven resizing parameters are set as 0.375, 0.5, 0.75, 0.875, 2, 4, and 8 respectively, showing how many times the watermarked image is resized to its original size. Corresponding to the large value of the resizing parameter the SNR is also high. When we have the length of the PN sequence as a constant, saying 1023, the empirical and theoretical BER curves of using the SS nonoblivious and oblivious watermarking techniques are depicted in Figure 2.8.

It is seen from Figure 2.8 that the curve of the empirical BER values is very close to the theoretical one which is obtained by the analytical expressions of (2.11) and (2.18). When the resizing parameter is large, the BER has a small value. In other words, when

the SNR is high there are less detection errors. Meanwhile, the performance of the SS nonoblivious watermarking is much better than the SS oblivious watermarking. This is because, as we have mentioned before, the channel noise, which is introduced by the original image pixels, can be deducted in the nonoblivious watermarking. Hence, the SNR in the nonoblivious watermarking is much lower than the one in the oblivious watermarking.

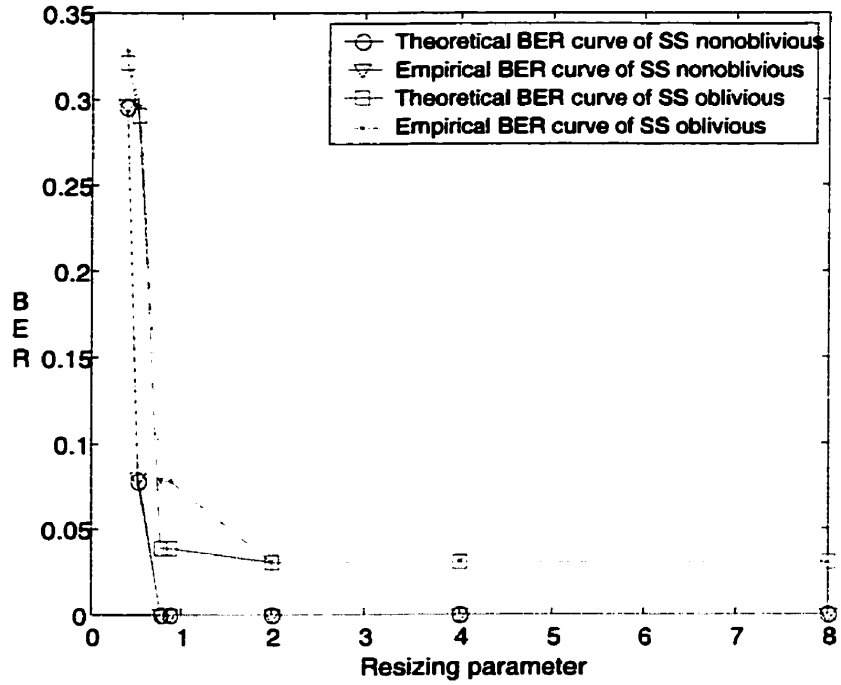

Figure 2.8 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the resizing attack for different resizing parameters

The second test is based on the assumption that the SNR is a constant. For instance, we set the resizing parameter as 0.75, and the BER values can be obtained given different sequence lengths N. Considering the case of N = 63, 127, 255, 512 and 1023, the theoretical and empirical BER curves for both SS nonoblivious and oblivious watermarking are depicted together in Figure 2.9. It is seen that as the increase of the sequence length the detection error becomes less, which is consistent with the expression of the error probability in (2.11) and (2.18). Especially for the nonoblivious watermarking, the empirical BER curve is very close to the theoretical one.

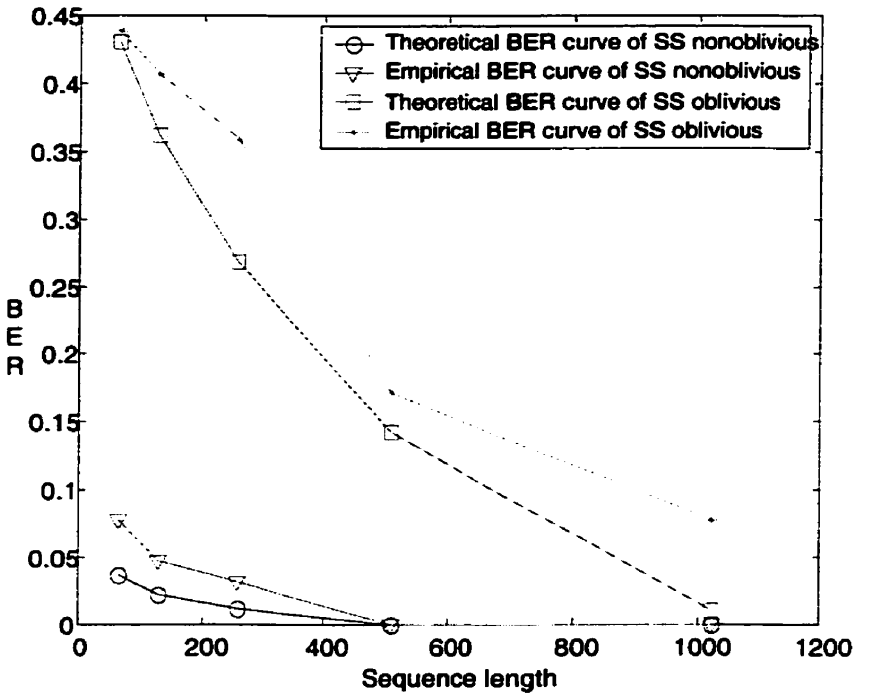

Figure 2.9 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the resizing attack for different sequence lengths

B. Image cropping

The second kind of attacks due to image processing that a watermarked image usually encounters is image cropping. Figure 2.10 shows a cropped version of the watermarked image in which only the left corner of the image survives and the rest part is cut off. The watermark signals added to these removed image pixels are also lost. The larger part of the image is being cropped, the more watermark information will be lost. This explains why a good strategy for image watermarking is to spread out information signals everywhere of the image so that the damage caused by cropping will not be too severe.

The seven cropping parameters are set as 0.1, 0.2 0.3, 0.4, 0.5, 0.7, and 0.9. Thus, after cropping only 0.1, 0.2 0.3, 0.4, 0.5, 0.7, and 0.9 times of the original watermarked image is left untouched. The larger the cropping parameter is, the less the image is distorted. Both the theoretical and empirical BER curves versus cropping parameters are plotted in Figure 2.11. It is shown that the SS nonoblivious watermarking is pretty robust

to image cropping. There is no error in the detection. For the oblivious watermarking, only when the cropping parameter is equal to 0.7, the BER value starts to be less than 0.1.

Figure 2.10 The image being cropped to 1/2 of its original dimensions

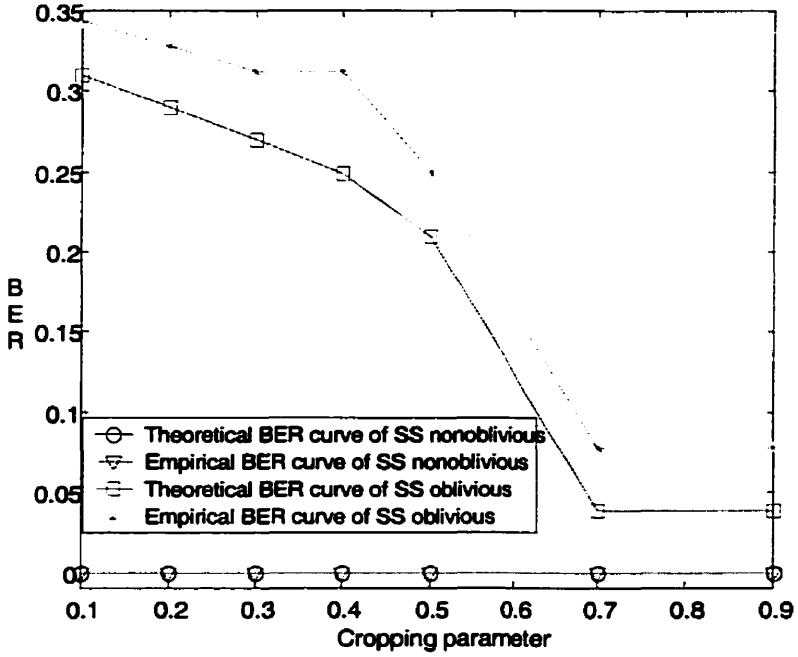

Figure 2.11 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the cropping attack for different cropping parameters

The BER values versus different sequence lengths are shown in Figure 2.12. It is shown that as the decrease of the sequence length the performance of the SS oblivious watermarking degrades greatly. However, for the nonoblivous watermarking the BER values are still all zeros no matter how long the sequence length is used.

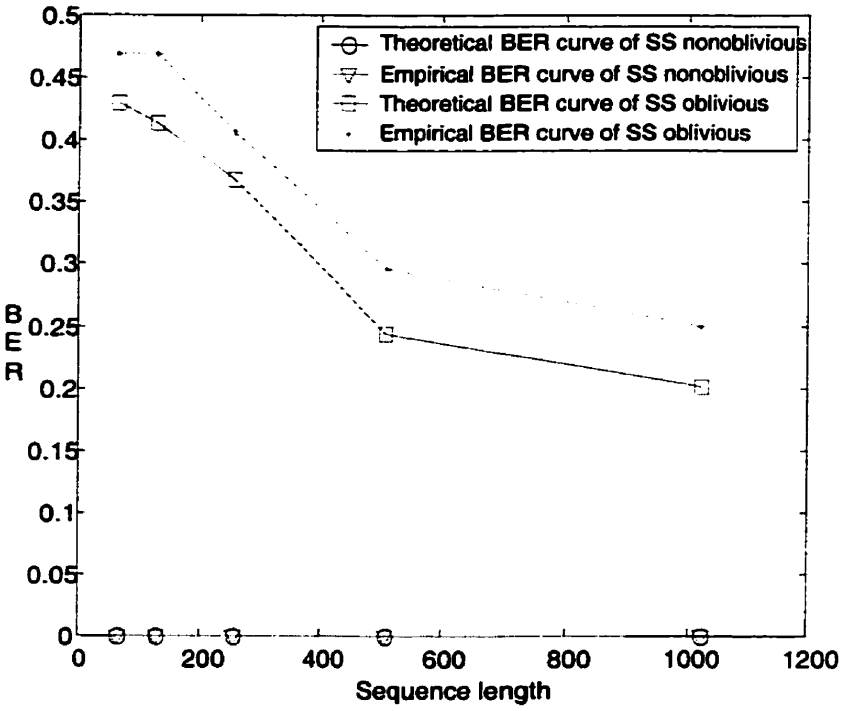

Figure 2.12 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the cropping attack for different sequence lengths

C. Image rotating

Rotation of pixels in the image will change the pixel positions. Figure 2.13 shows the watermarked image being rotated by 10 degrees. After rotation, synchronization error of the spreading sequences is introduced to the watermarked image. For the SS watermarking scheme, this results in a great degradation of the performance. In order to minimize this kind of errors, all possible shifts of the PN sequence should be evaluated in the correlator to find out the correct one.

The seven rotating parameters are selected as -45, -35, -30, -25, -15, -10 and -5 respectively, which is the number of degrees by which the image is rotated. The negative sign implies the image is rotated in a counter-clockwise direction. The large absolute value of the rotating parameter corresponds to the small SNR. The BER curves versus

different rotating parameters are plotted in Figure 2.14. For the oblivious watermarking it is seen that the performance of the SS watermarking can be pretty good when the synchronization error is minimized by trying all possible shifts of the spreading sequence. Even though the image is rotated by 45 degrees, the BER value can be achieved as small as 0.156. For the nonoblivious watermarking even though the detection errors still exit for the large absolute value of the rotating parameter, the largest BER value is only 0.0625, which is very small.

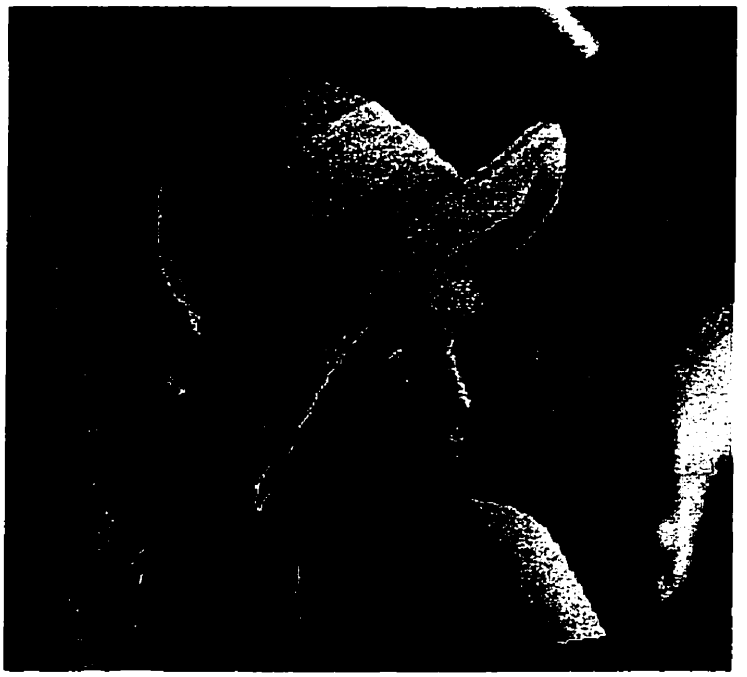

Figure 2.13 The image being rotated by 10 degrees

The BER curves versus different sequence lengths are shown in Figure 2.15. In this test, the rotating parameter is set as a constant equal to 10. It is seen that as the decrease of the sequence length, the BER value increases greatly. For the nonoblivious watermarking, there is a sharp increase of the BER when the length is reduced to 255. Again, it shows that the sequence length plays an important role in the detection process of the SS watermarking.

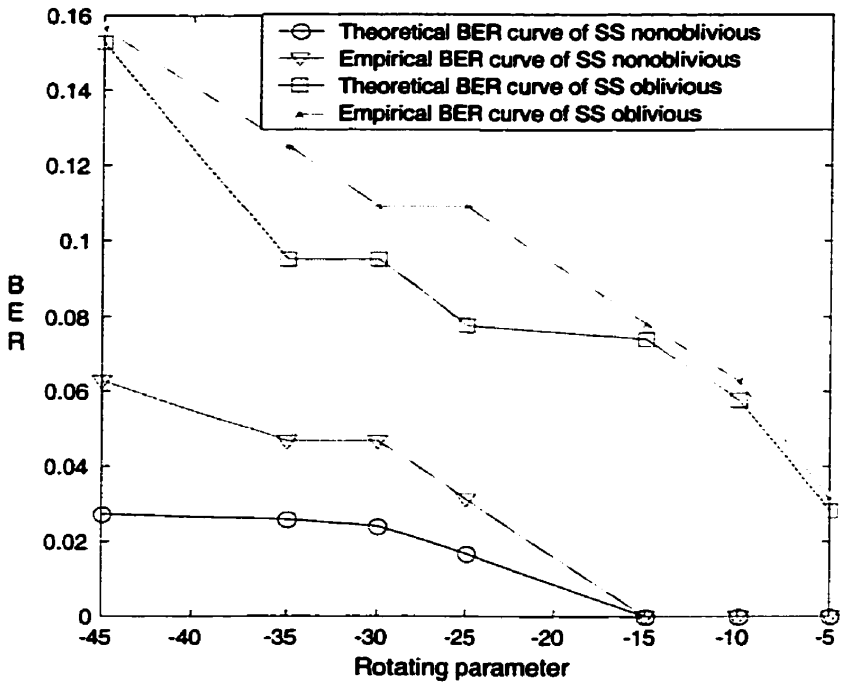

Figure 2.14 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the rotating attack for different rotating parameters

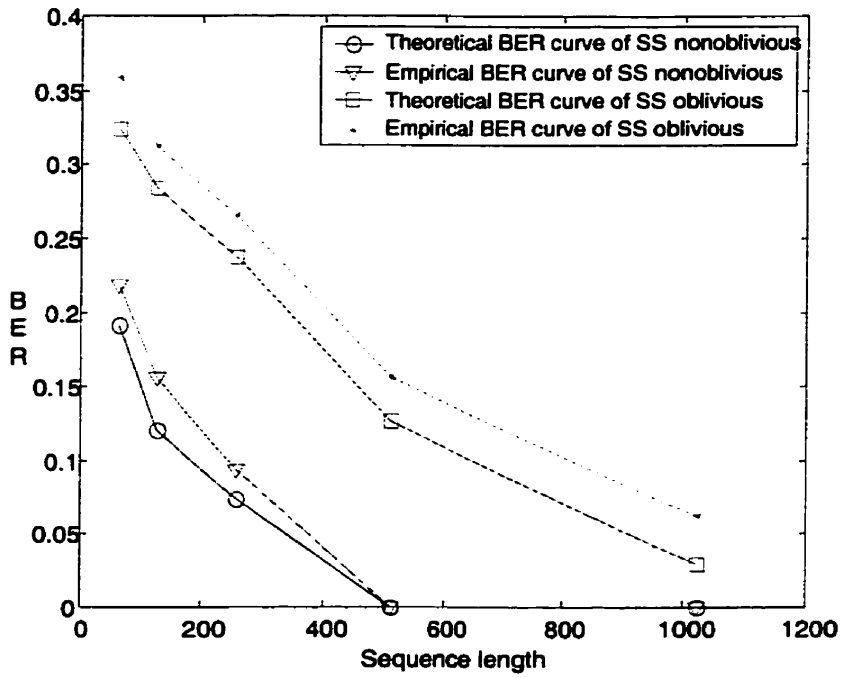

Figure 2.15 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the rotating attack for different sequence lengths

D. Median filtering

A median filter is often used to reduce noises without blurring edges and losing other sharp details. That is, the grey level of each pixel is replaced by the median of the grey levels in a neighbourhood of that pixel. Median filtering is quite effective when the noise pattern consists of strong, spike-like components and the characteristic to be preserved is the edge sharpness. However, when it is applied to the original image, this processing can be considered as introducing nonlinear distortions to the image pixels. Figure 2.16 shows the image after it is passed through a median filter with the window size of 3×3. As shown, the sharp details in the original image are being smoothed.

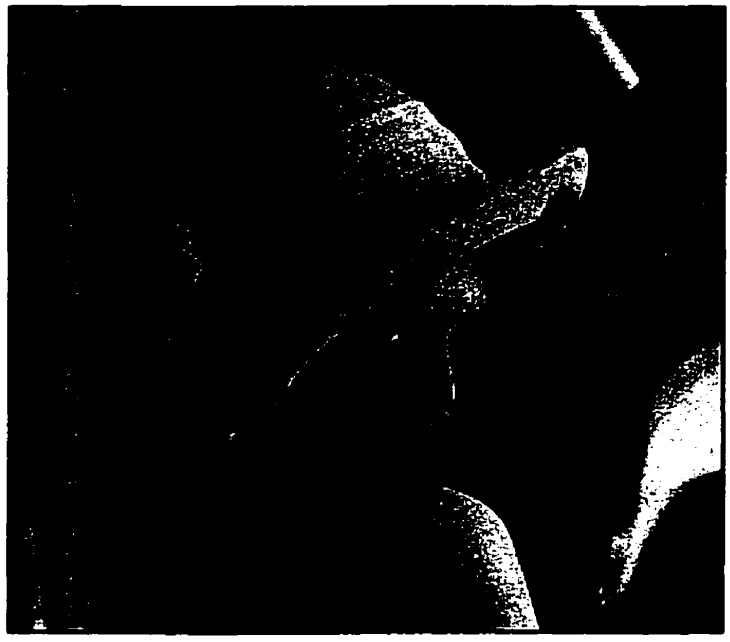

Figure 2.16 The image that has passed through a median filter with a window size of 3×3

The BER curves versus the filtering parameters by using the two watermarking techniques are plotted in Figure 2.17. The filtering parameter is the window size of the median filter. They are set as 11, 9, 7, 6, 5, 3 and 2 respectively. The larger the value of the window size is, the more the distortion is introduced, and so the larger the SNR. It is seen that the performance of the oblivious watermarking is really bad under this attack. Even though the window size is as small as 2, the BER gets a very large value of 0.3. For

the nonoblivious watermarking the performance degrades greatly when the window size goes to 6 or more. Thus, the SS watermarking is not very robust to this kind of nonlinear distortions.

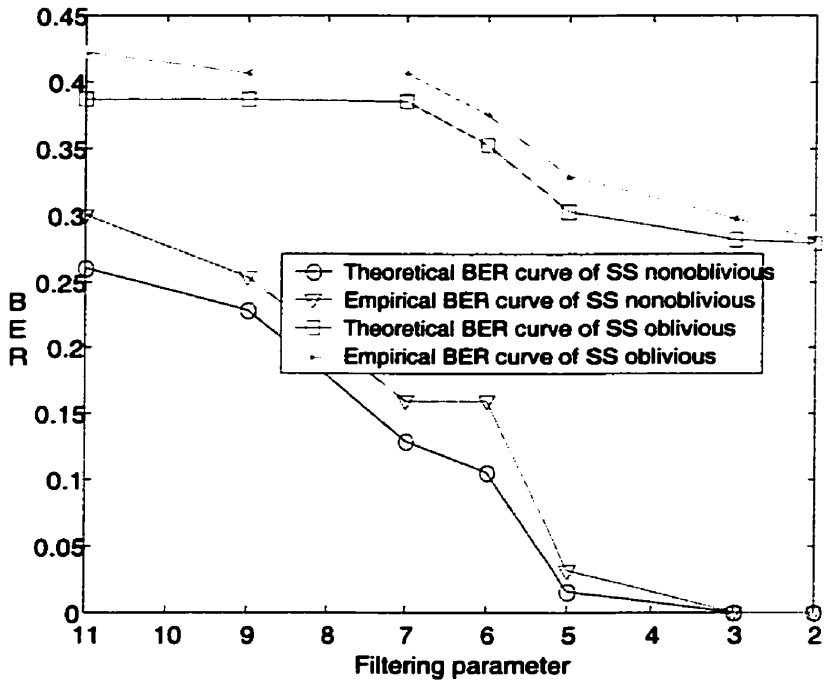

Figure 2.17 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the median filtering attack for different filtering parameters

When the filtering parameter is set as 3, the BER curves versus the sequence lengths are plotted in Figure 2.18. It shows that there is a big difference between the theoretical value and the empirical for the oblivious watermarking. This is because the analytical expression of (2.14) is obtained based on the assumption that the image pixel is Gaussian distributed. However, this may not be the exact model after the sequence length is reduced significantly. For the nonoblivious watermarking this difference is very small because the original image is deducted in the detection. Only the distortion introduced by filtering should be considered. And the power of this kind distortion is relatively small.

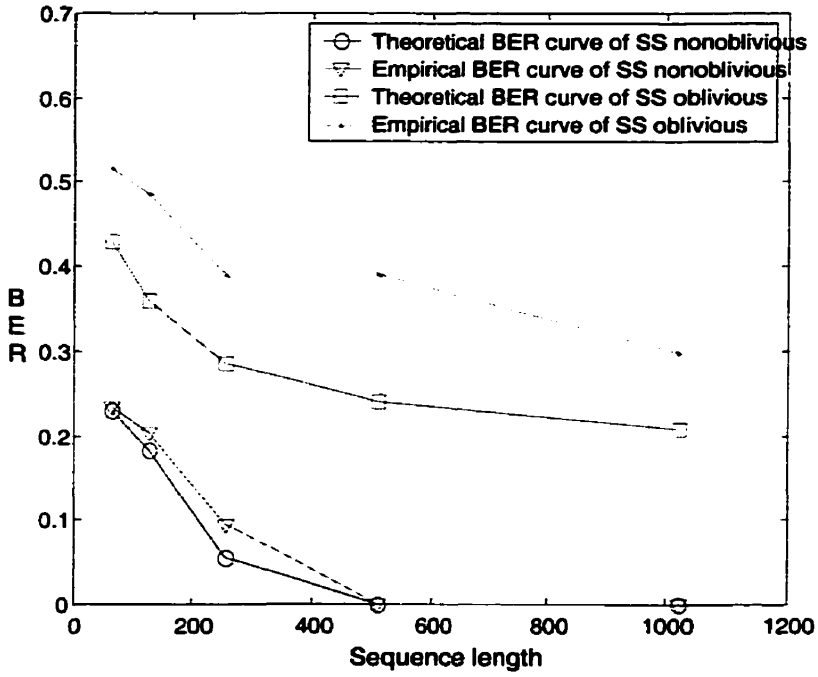

Figure 2.18 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the median filtering attack for different sequence lengths

E. Image compression

Image compression addresses the problem of reducing the amount of the data for transmission and is widely used in to-date digital image technologies. JPEG is probably the most popular compression method for still images. During the process, 8×8 blocks of the input image are formatted by scanning the image left to right and top to bottom. Each block of 64 samples is transformed to a block of 64 coefficients by the forward DCT. The coefficient in the immediate top-left corner is called the direct current (DC) coefficient. Alternative current (AC) coefficients corresponding to increasingly higher frequencies of the sample block progress away from the DC coefficient. Only the nonzero AC coefficients are entropy-coded and data are lost among the high frequency components. When the watermarked image is compressed, the watermark signals added to the high frequencies will be lost. To show the impact of image compression on the image, the compressed watermarked image with the quality factor 50% is shown in Figure 2.19.

The BER curves versus the compression parameters are shown in Figure 2.20. The compression parameter is defined as the quality factor, which is the value representing how many percent the data size of a digital image is compressed to of the

original. In this study, they are set 30, 40, 50, 60, 70, 80, and 90 respectively. The larger the value is, the less the watermarked image is distorted, and so the less the watermark signal is lost. It implies the SNR gets large when the compression parameter is also large. It is shown that for the SS nonoblivious watermarking even though the compression parameter is as small as 30, the BER is still equal to 0. Hence, the SS nonoblivious watermarking is very robust to JPEG compression. However, for the oblivious watermarking, its performance is still much worse than for the nonoblivious watermarking.

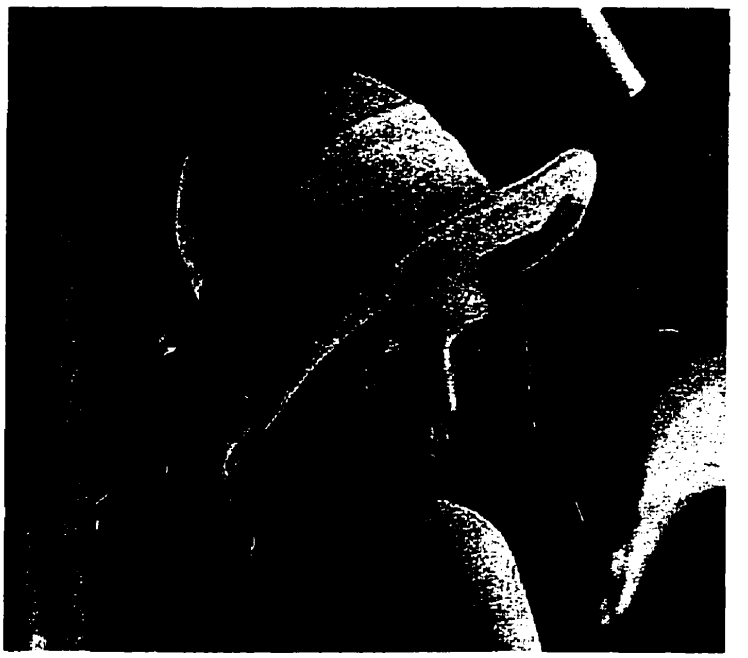

Figure 2.19 The image under JPEG compression with the quality factor 50%

The BER curves versus the sequence lengths are plotted in Figure 2.21 while the compression parameter is set as a constant equal to 50. It shows that the nonoblivios watermarking is very robust to JPEG compression. Only when the sequence length is reduced to 64, the detection error comes out. It is also shown that the performance of the SS watermarking depends greatly on the sequence length. The longer the PN sequence, the better the performance achieved.

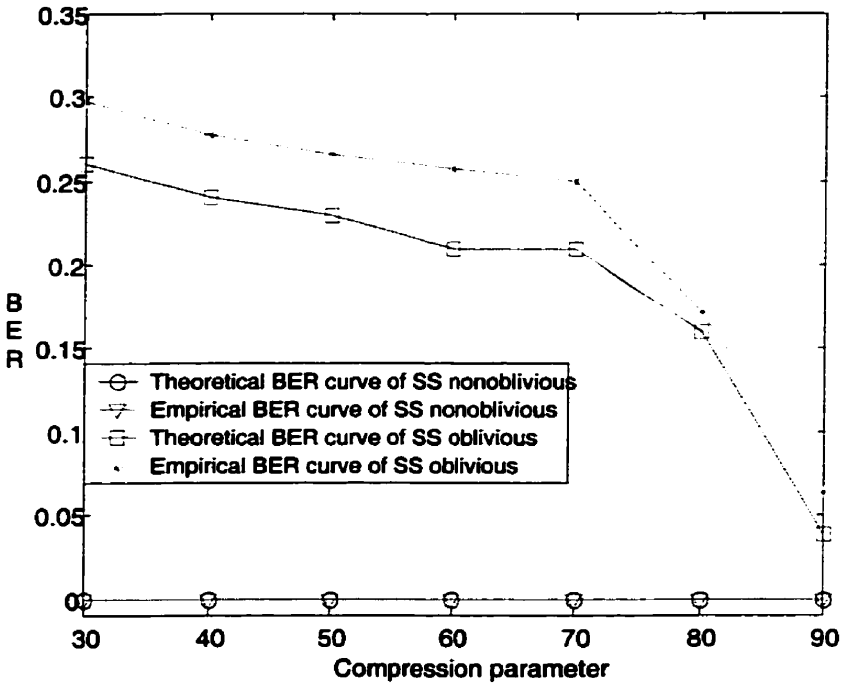

Figure 2.20 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the JPEG compression for different compression parameters

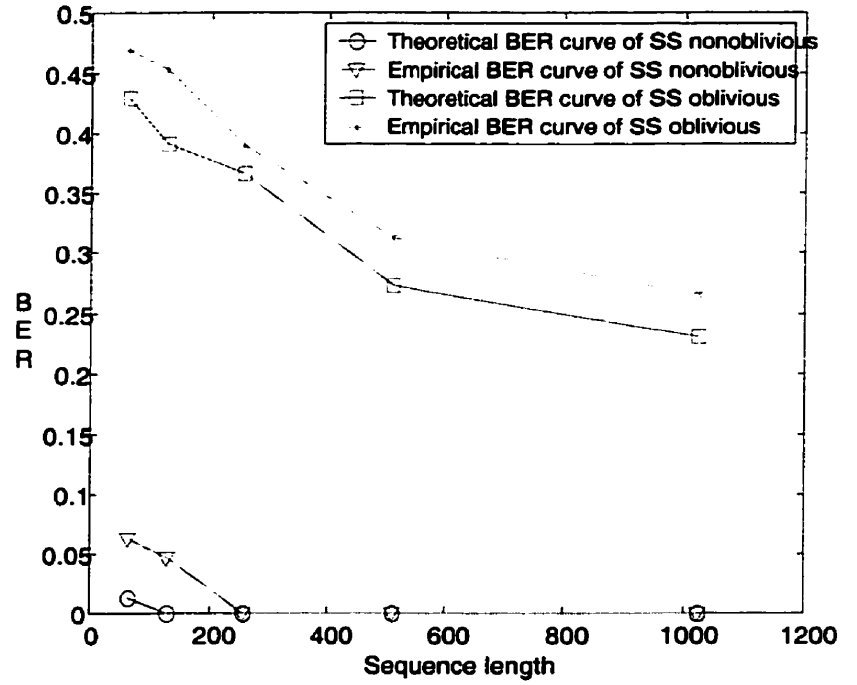

Figure 2.21 The BER curves of using the SS nonoblivious and oblivious watermarking techniques under the JPEG compression for different sequence lengths

To summarize the results of performance tests, we can say that the conventional correlation-based SS nonoblivious watermarking is very robust to cropping, and JPEG compression. However, the length of the PN sequence should be set very long, saying 255. As for the SS oblivious watermarking, the performance of it is so bad that it cannot be used in real practice if robustness is the most important concern for that specific application.

2.5 Summary

The conventional correlation-based spread spectrum watermarking technique is described in this chapter. Both the nonoblivious and oblivious watermarking techniques are discussed along with the performance analysis. It is shown that the performance of the SS watermarking scheme is determined by the length of the spreading sequence and the power ratio between the spreading sequence and the noise, which we define as SNR.

In the nonoblivious watermarking, the watermark signal can be extracted with the original image available. And this makes it possible to achieve the detection and inversion of distortions introduced by image processing or attacks. Therefore, the SS nonoblivious watermarking has a lower SNR compared to the oblivious watermarking. This explains why the SS nonoblivious watermarking is more robust than the oblivious watermarking.

However, in some applications the access to original images is not allowed because it raises a twofold problem. The set-up of a watermarking system becomes more complicated, and the owners of the original images may be compelled to insecurely share their works with people who want to check for the existence of the watermark. Thus, the selection of the nonoblivious watermarking or the oblivious watermarking scheme is really dependent on the specific application.

CHAPTER 3

APPLICATION OF CHAOTIC SPREADING SEQUENCES TO SPREAD SPECTRUM WATERMARKING

3.1 Introduction

In the conventional correlation-based SS watermarking technique, a classical spreading sequence, such as *m*-sequence, is used to spread out the copyright information to provide watermarks. Classical spreading sequences are distinguished by their wideband, flat spectrum, and pseudo-randomness [40]. Recently, the application of chaos to SS communications has attracted attentions because of the noise-like appearance, nonlinear characteristics of the time series generated by chaotic systems, and also because of the simple implementation of chaotic systems [41, 42, 43].

A chaotic system is a deterministic dynamical system whose states change with iterations in a deterministic way [44], i.e., a nonlinear dynamical system model based on its N previous values can be described as:

$$c_n = f(c_{n-1}, c_{n-2}, \cdots, c_{n-N}, \lambda), \tag{3.1}$$

where λ is the bifurcating parameter. A chaotic system generates a set of aperiodic signals with a "noise-like" and broad power spectrum. The system is very sensitive to initial conditions. A slight difference in initial conditions will produce totally different sequences, which possess good correlation properties [45].

These characteristics of a chaotic signal are very helpful in digital watermarking applications. The noise-like and wideband output of a chaotic system can be used as the spreading sequence to spread out copyright information. Because of the good correlation properties of chaotic signals, the inserted information can be retrieved properly when the correlation detector is applied. Furthermore, the correlation of certain chaotic spreading sequences is close to the optimal correlation performance for the application of image watermarking. Thus, it is reasonable to expect that using chaotic spreading sequences is

superior to widely used classical spreading sequences, such as *m*-sequences and Gold sequences in terms of robustness and security.

In this chapter, we review the model of the correlation-based spread spectrum watermarking system and report the derivation of performance indexes for general spreading sequences. A brief overview of some classical spreading sequences along with a discussion of their main characteristics and limitations is given. Following that, some chaotic maps, which can generate well-behaved spreading sequences, are introduced. At the end the performance tests are applied to different spreading sequences. The results are also shown.

3.2 Convention Spread Spectrum Watermarking System with Spreading Sequences

The correlation property of a spreading sequence plays an important role in the detection process when the correlator is applied to detect the copyright information. No matter whether the conventional SS nonoblivious or the SS oblivious watermarking is used it is desirable to use spreading sequences with low cross-correlation and exponentially vanishing autocorrelation so that only when the correct sequence and correct shift is used, a high correlation value can be observed and the information bit $b(i)$ can be retrieved properly.

3.2.1 Analysis of cross-correlation performance

In Chapter 2, the probability of detection error presented in (2.9) and (2.18) is derived under the assumption that the correct spreading sequence is used as well as the correct shift. Actually, when the correlation is applied between the watermark and a wrong PN sequence, it is still possible to obtain the correct result. That is, when the correlation sum $\frac{1}{N}(\mathbf{x}(i)\cdot\mathbf{p})$ is applied between the watermark signal and a wrong sequence, a large correlation value can be obtained and the sign of it is the same as the information bit $b(i)$. We define the probability of this kind detection error as ρ_c.

Assuming that the PN sequence used to generate the watermark is expressed as $\mathbf{p}^x = (p_0^x, p_1^x, \cdots, p_{N-1}^x)$, the correlation applied between the watermark signal and a wrong PN sequence $\mathbf{p}^z = (p_0^z, p_1^z, \cdots, p_{N-1}^z)$ can be written as

$$\gamma^{xz} = \frac{1}{N}(\mathbf{x}(i) \cdot \mathbf{p}^z) = \frac{1}{N}\sum_{n=0}^{N-1} b(i) p_n^x p_n^z. \tag{3.2}$$

Defining the partial cross-correlation function between the two spreading sequences $\mathbf{p}^x$ and $\mathbf{p}^z$ as

$$\Gamma_{N,\tau}(\mathbf{p}^x, \mathbf{p}^z) = \begin{cases} \sum_{n=0}^{N-\tau-1} p_n^x \cdot p_{n+\tau}^z & for \quad \tau = 0,1,\cdots,N-1 \\ \Gamma_{N,-\tau}(\mathbf{p}^x, \mathbf{p}^z) & for \quad \tau = -1,-2,\cdots -N+1 \\ 0 & if \quad |\tau| \geq N, \end{cases} \tag{3.3}$$

we may write the correlation sum as $\gamma^{xx} = \frac{1}{N} b(i)\Gamma_{N,0}(\mathbf{p}^x, \mathbf{p}^x)$ when the correct spreading sequence is used; and $\gamma^{xz} = \frac{1}{N} b(i)\Gamma_{N,\tau}(\mathbf{p}^x, \mathbf{p}^z)$ when a wrong spreading sequence is used. Employing Gaussian approximation which considers the values of the spreading sequences as a random process [46, 47] and following the results obtained in [48], the variance of the correlation between the wrong PN sequence and the watermark signal can be written as

$$\begin{aligned} \sigma_{xz}^2 &= E[(\gamma^{xz})^2] \\ &= E\left[\frac{1}{N^2} b^2(i)\Gamma_{N,\tau}^2(\mathbf{p}^x, \mathbf{p}^z)\right] \\ &= \frac{b^2(i)}{6N^3}\sum_{\tau=1-N}^{N-1}(2E[\Gamma_{N,\tau}^2(\mathbf{p}^x, \mathbf{p}^z)] + E[\Gamma_{N,\tau}(\mathbf{p}^x, \mathbf{p}^z)\Gamma_{N,\tau+1}(\mathbf{p}^x, \mathbf{p}^z)]). \end{aligned} \tag{3.4}$$

Thus, we can have ρ_c expressed as

$$\rho_c = \frac{1}{2} erfc\left(\sqrt{\frac{(\gamma^{xx})^2}{2\sigma_{xz}^2}}\right) = \frac{1}{2} erfc\left(\sqrt{\frac{b^2(i)\Gamma_{N,0}^2(\mathbf{p}^x, \mathbf{p}^x)}{2N^2\sigma_{xz}^2}}\right). \tag{3.5}$$

Substituting (3.4) to (3.5), it may be eventually written as $\rho_c = \frac{1}{2} erfc\sqrt{\frac{1}{R_c}}$, where the quantity

$$R_c = \frac{2N^2\sigma_{zz}^{\ 2}}{b^2(i)\Gamma_{N,0}^{\ 2}(\mathbf{p}^x,\mathbf{p}^x)}$$
$$= \frac{2N^2\dfrac{b^2(i)}{6N^3}\displaystyle\sum_{\tau=1-N}^{N-1} 2E[\Gamma_{N,\tau}^{\ 2}(\mathbf{p}^x,\mathbf{p}^z)] + E[\Gamma_{N,\tau}(\mathbf{p}^x,\mathbf{p}^z)\Gamma_{N,\tau+1}(\mathbf{p}^x,\mathbf{p}^z)]}{b^2(i)\Gamma_{N,0}^{\ 2}(\mathbf{p}^x,\mathbf{p}^x)}$$
$$= \frac{1}{3N\Gamma_{N,0}^{\ 2}(\mathbf{p}^x,\mathbf{p}^x)}\sum_{\tau=1-N}^{N-1} 2E[\Gamma_{N,\tau}^{\ 2}(\mathbf{p}^x,\mathbf{p}^z)] + E[\Gamma_{N,\tau}(\mathbf{p}^x,\mathbf{p}^z)\Gamma_{N,\tau+1}(\mathbf{p}^x,\mathbf{p}^z)] \qquad (3.6)$$

can be used as a reasonable choice for the cross-correlation performance index of the spreading sequences in watermarking applications. The smaller the R_c is, the less the error probability ρ_c.

For the ideal random sequence it has been known that

$$E[\Gamma_{N,\tau}(\mathbf{p}^x,\mathbf{p}^z)\Gamma_{N,\tau+1}(\mathbf{p}^x,\mathbf{p}^z)] = 0,$$
$$E[\Gamma_{N,\tau}^{\ 2}(\mathbf{p}^x,\mathbf{p}^z)] = (N-|\tau|)\times const,$$
$$\Gamma_{N,0}^{\ 2}(\mathbf{p}^x,\mathbf{p}^x) = N^2\times const. \qquad (3.7)$$

A simple substitution of (3.7) in (3.6) leads to the cross-correlation performance index of the ideal random sequence having $R_c = 2/(3N)$.

3.2.2 Analysis of auto-correlation performance

Similarly, the use of the correct PN sequence but with a wrong shift also has the possibility to provide the correct result when the correlator is applied. That is, when the correlation sum is calculated between the watermark signal and the correct PN sequence but with a wrong shift τ presented as $\frac{1}{N}\sum_{n=0}^{N-1} b(i)p_n^x p_{n+\tau}^x$, a high correlation value is observed and the sign of it is the same as the information bit *b(i)*. The probability of this kind detection error is defined as ρ_a, which is determined by auto-correlation performance of a spreading sequence.

According to (3.3), the auto-correlation of the spreading sequence $\mathbf{p}^x$ can be expressed as $\Gamma_{N,\tau}(\mathbf{p}^x,\mathbf{p}^x)$. Then the correlation with the correct shift can be written as

$\frac{1}{N}b(i)\Gamma_{N,0}(\mathbf{p}^x,\mathbf{p}^x)$, while the output of the correlation with a wrong shift τ as $\frac{1}{N}b(i)\Gamma_{N,\tau}(\mathbf{p}^x,\mathbf{p}^x)$. Following the same procedure as deriving ρ_c, the error probability ρ_a can be expressed as $\rho_a = \frac{1}{2}erfc\sqrt{\frac{1}{R_a}}$. If all the sequences can be thought of as independent and if they satisfy a mild form of second-order stationary so that $E[p_m^x p_n^x] = E[p_0^x p_{|m-n|}^x]$, only a small number of algebraic manipulations [49] are necessary to arrive at an expression for R_a according to (3.6) as

$$R_a = \frac{2}{3N} + \frac{4}{3N^3}\sum_{n=1}^{N-1}[(N-n)^2 A_n^{\,2} + \frac{(N-n+1)(N-n)}{2}A_{n-1}A_n], \qquad (3.8)$$

where $A_n = E[p_0^x p_n^x]$. Therefore, (3.8) can be used as the auto-correlation performance index of a spreading sequence in watermarking application. Still, the smaller the R_a is, the less the error probability ρ_a.

It is known that when $\partial R_a/\partial A_n = 0$ R_a has a unique minimum, i.e.,

$$\frac{\partial R_a}{\partial A_n} = 2(N-n)^2 A_n + \frac{(N-n+1)(N-n)}{2}A_{n-1} + \frac{(N-n)(N-n-1)}{2}A_{n+1} = 0 \qquad (3.9)$$

for $n = 1,\ldots,N-1$ with $A_N = 0$. This set of simultaneous equations can be solved by setting [49]

$$A_n = \frac{\delta_n(N-1)A_1 - \delta_{n-1}N}{N-n}, \qquad (3.10)$$

where δ_n are coefficients such that $\delta_{-1} = -1$, $\delta_0 = 0$. Substituting (3.10) into (3.9), we can have

$$\delta_{n+1} = -4\delta_n - \delta_{n-1}, \quad n = 0,\ldots,N-2. \qquad (3.11)$$

The equation $\partial R_a/\partial A_{N-1} = 2A_{N-1} + A_{N-2} = 0$ can be solved by using (3.11) to obtain $A_1 = \frac{N\delta_{N-1}}{(N-1)\delta_N}$, and thus $A_n = \frac{N(\delta_n\delta_{N-1} - \delta_{n-1}\delta_N)}{(N-n)\delta_N}$. Furthermore, the Fibonacci-like

recursion of δ_n can be used to give $\delta_n = (-1)^{n+1}\frac{r^{-n}-r^n}{r^{-1}-r}$ with $r = 2-\sqrt{3}$, hence we can obtain

$$A_n = (-1)^n \frac{N}{N-n}\frac{r^{n-N}-r^{N-n}}{r^{-N}-r^N}, \tag{3.12}$$

which is the auto-correlation for a minimum R_a. Note that A_n has a alternating sign so that the positive contribution to R_a due to A_n^2 are counterbalanced by the negative contribution of $A_{n-1}A_n$. Substituting (3.12) into (3.8), the minimum autocorrelation performance index becomes

$$R_a = \frac{\sqrt{3}}{3N}\frac{r^{-2N}-r^{2N}}{r^{-2N}+r^{2N}-2}, \tag{3.13}$$

which approaches to $\sqrt{3}/(3N)$ for large N (actually $|R-\sqrt{3}/(3N)|<10^{-12}$ for $N \geq 10$). This value is less than the classical $R_a = 2/(3N)$ obtained for ideal random sequences where $A_n = 0$ for $n > 0$.

From the above discussion for auto-correlation, it is found that a spreading sequence with a auto-correlation performance index close to $R_a = \sqrt{3}/(3N)$ will minimize the error probability ρ_a caused by using the wrong shifts of the spreading sequence.

3.3 *m*-Sequences and Gold Sequences

It is seen that the choice of a proper set of spreading sequences is a central issue to minimize ρ_a and ρ_c in the correlation-based SS watermarking scheme. Among the many proposals, one of the most studied options is the maximum-length sequence.

Maximum length sequences [50] usually referred to as *m*-sequences can be generated by using an *m*-stage shift register. They are periodic with the period $2^m - 1$. Maximum length sequences work well in practice as it can be proved that if $N = 2^m - 1$, we have $R_c = R_a = 2/(3N)$, i.e., *m*-sequences behave like if they were perfectly random. Nevertheless, for any assigned N only a limited number of sequences can be generated.

The spectrum and the autocorrelation of one m-sequence with $N = 2^{10} - 1$ are plotted in Figure 3.1 and 3.2 respectively.

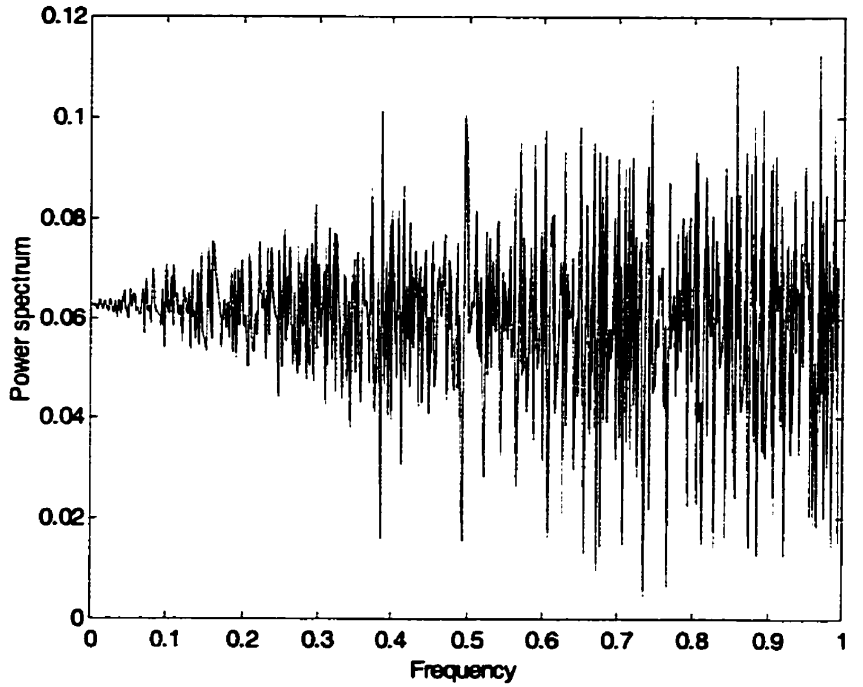

Figure 3.1 The power spectrum of the m-sequence with N =1023

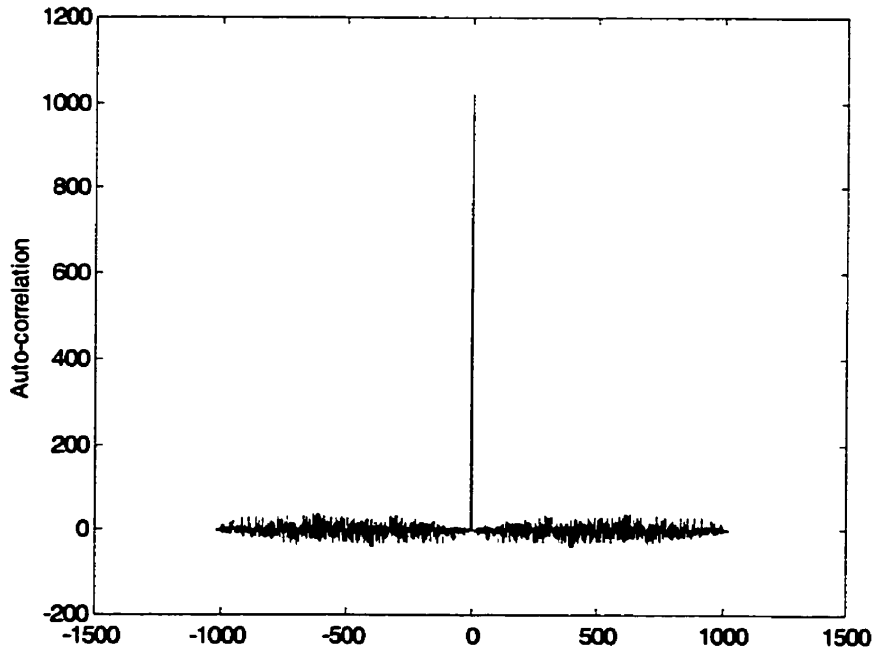

Figure 3.2 The auto-correlation of the m-sequence with N =1023

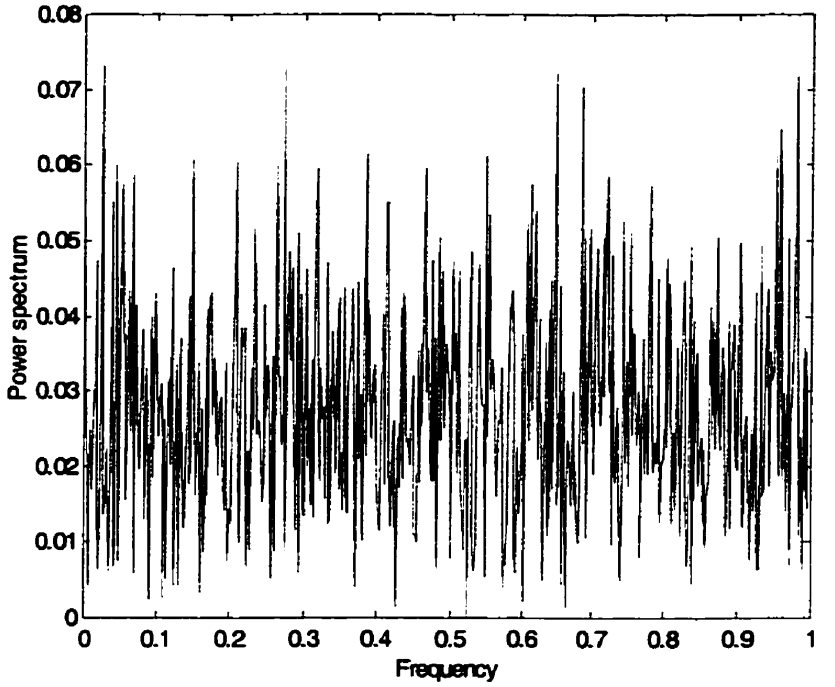

Figure 3.3 The power spectrum of the Gold sequence with $N=1023$

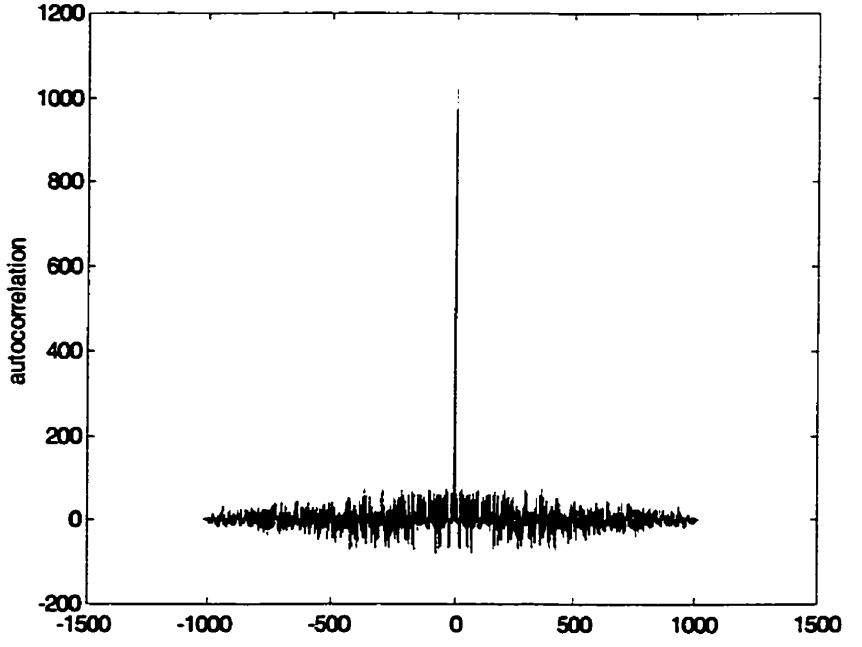

Figure 3.4 The auto-correlation of the Gold sequence with $N=1023$

Gold sequences get around the problem of *m*-sequences as they maintain good correlation properties even if their number is greatly increased. To generate Gold sequences a couple of *m*-sequences must be selected according to specific criteria [52]. The two elements of this preferred pair are then added shifting one of them for all the possible integer shifts. Hence, $N+1$ sequences can be provided. The spectrum and the autocorrelation of one Gold sequence are depicted in Figure 3.3 and 3.4.

It is also noted that although the good auto-correlation and cross-correlation can be guaranteed, *m*-sequences and Gold sequences provide limited security as they can be identified with a number of samples, which is much less than their actual length by means of linear regression models [53]. Moreover the generation of *m*-sequences and Gold sequences is not that easy. It requires having a large number of registers to store the seeds for the generation of them especially when the sequence length is very long.

3.4 Chaotic Spreading Sequences

Chaotic sequences have been widely used in spread spectrum communications. Through the sensitive dependence of chaotic systems on their initial conditions, a large number of uncorrelated, random-like, yet deterministic signals can be generated. Further, the quantization does not destroy the desirable properties of these kinds sequences.

According to the discussion of auto-correlation performance in Section 3.2.2, we can see that the most significant elements of A_n are the larger values, i.e. those with $n \ll N$. In these cases, the expression for A_n in (3.12) reduces to $A_n \approx (-r)^n$. And following the earlier results on exponentially vanishing auto-correlation [54], we may reasonably expect that chaos-based spreading sequences can be designed to tend towards the optimal performance.

To do so, consider a function $f:[0,1] \rightarrow [0,1]$ iterated starting from an initial condition c_0 uniformly distributed in [0,1] to produce the sequence $c_{n+1} = f(c_n)$. This sequence is quantized by the bipolar threshold function $Q:[0,1] \rightarrow \{1,-1\}$ centered at 1/2 and the spreading sequences are taken to be $p_n = Q(c_n)$ for $n = 0,\ldots,N-1$. These sequences are then repeated periodically. In this study, the family of (l, t)-tailed shifts

with l even and $t<l/2$ [49] is employed to illustrate that the chaotic spreading sequence can achieve nearly optimal auto-correlation performance.

(l, t)-tailed shifts are affine in each of the intervals $X_j=[(j-1)/l, j/l]$. The intervals from X_1 to X_{l-t} are mapped onto $X_{t+1}\cup\cdots\cup X_l$ while the last t intervals are mapped onto $X_1\cup\cdots\cup X_t$. Such a construction is clarified by Figure 3.5, in which a (4,1)-tailed shift is reported.

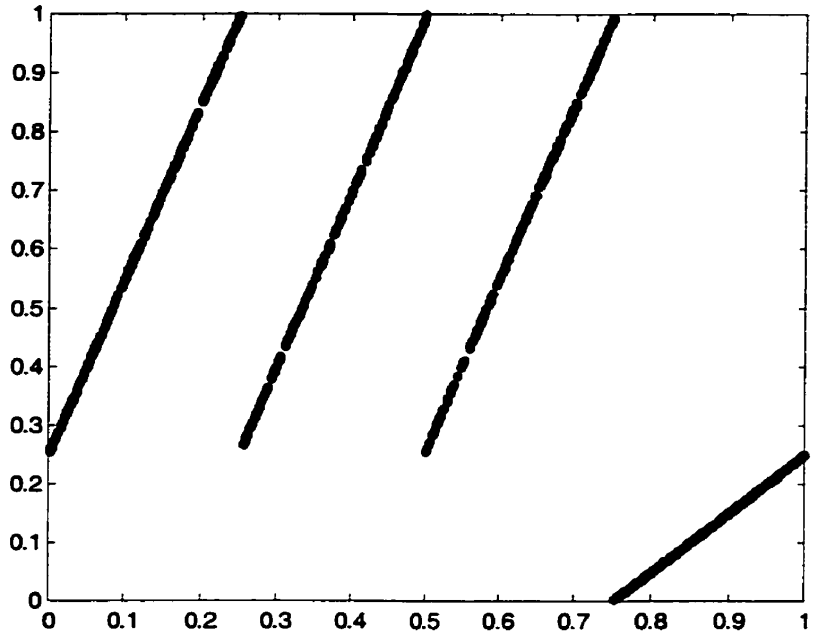

Figure 3.5 The graph of (4,1)-tailed shift map

The auto-correlation function of the tailed shift map can be written as [49] $A_n=\frac{1}{l}\sum_{i=1}^{l}\sum_{j=1}^{l}Q(X_i)Q(X_j)\mathbf{K}_{ij}^n=h^n$, where $Q(X_j)$ indicates the value of Q for all the points in X_j, $\mathbf{K}_{ij}^n=\frac{1}{l}\begin{bmatrix}1-h^{n-1} & 1-h^{n-2}\\ 1-h^n & 1-h^{n-1}\end{bmatrix}$ and $h=-t/(l-t)$. For any given l, this trend approximates the optimal $R_a=\sqrt{3}/(3N)$ choosing t as the integer closest to $lr/(1+r)$. The accuracy of this approximation increases as $n\to\infty$. Since $r=2-\sqrt{3}\approx 0.2679$, for $l=4$ we have $t=0.8452\approx 1$. So the (4,1)-tailed shift has the auto-correlation performance index R_a close to the optimal.

Now we analyze the cross-correlation performance of chaotic spreading sequences. According to the theorem obtained in [55], the cross-correlation performance index R_c of chaotic spreading sequences has a bound $R_c \le \frac{1}{6N^3}[(4B_1+2B_2)N^2 - B_2N - B_2]$, where $B_1 = 1+\frac{2C^2}{1-r_{mix}^{\ 2}}$ and $B_2 = r_{mix}(2C+\frac{2C^2 r_{mix}^{\ 2}}{1-r_{mix}^{\ 2}})$. The bound cannot guarantee optimum performance $R_c \le 2/(3N)$ for all chaotic maps. This dependence on the map is hidden beneath C and r_{mix}. However, for the family of piecewise-affine Markov (PWAM) maps, which the tailed shift map belongs to, it is known that C is equal or close to zero. Then we can obtain $B_1 = 1$ and $B_2 = 0$. So it leads to $R_c = 2/(3N)$ for this kind of maps. Thus, they may behave as well as purely random sequences.

From the above discussion we can see that the well-designed chaotic spreading sequences from certain PWAM maps have a better performance than classical spreading sequences. Besides this, the use of chaotic spreading sequences has some other advantages [42] in the application of watermarking compared to classical sequences. First, they are very easy to generate and store because only the knowledge of the chaotic map as well with initial conditions are required to be obtained for the generation, no matter how long the sequences will be. However, for the classical spreading sequence the longer the sequence length is, the more the registers have to be used. Second, the use of chaotic spreading sequences is more secure. A large number of uncorrelated sequences can be generated by simply changing the initial condition. Therefore, an eavesdropper would have a much larger number of sequences to search. Although the generation of chaotic sequences is simple for the authorized parties, who know the parameters and functions involved, the exact regeneration is very difficult for others that have to estimate them. A slight error in the estimation leads to exponentially increasing errors. This is due to the sensitive dependence of chaotic systems on their initial conditions. Moreover, the generation of the chaotic sequences can easily be made as complicated as desired. For instance, multi-dimensional chaotic maps may be used instead of the one-dimensional ones considered here. Also, several chaotic systems may be cascaded to increase the

number of parameters involved. This will further reduce the chance of detection by unauthorized parties.

3.5 Performance Test

To evaluate the performance of different spreading sequences, the SS oblivious watermarking scheme described in Section 2.3 is employed for embedding and detecting watermarks. The performance of classical spreading sequences is illustrated by using *m*-sequence and Gold sequence. Thus, we have the auto-correlation and cross-correlation performance indexes as $R_a = R_c = 2/(3N)$ for them. To evaluate the performance of chaotic spreading sequences, three kinds of chaotic systems are employed here. All of them are PWAM maps. Since the embedding and detection procedure are the same except that different spreading sequences are used, the difference of the performance can only be explained by the different correlation properties they possess.

Renyi sequence is generated by the Renyi map, which can be expressed by $c_{n+1} = \lambda c_n (\mathrm{mod}\, 1)$, where λ is a parameter that controls the chaotic behavior of the system. In our tests a threshold is selected in such a way that, after thresholding the sequence numbers, a bipolar spreading sequence is produced with approximately equal number of -1s and 1s. The parameter λ controls the frequency characteristics of the chaotic sequence, i.e., the frequency of the transitions $-1 \to 1$ and $1 \to -1$. For $\lambda > 1$ and values chose to 1, we get a spreading sequence with low number of transitions and, thus, lowpass properties, whereas when $\lambda \approx 2$ the transitions are very frequent. The map of $\lambda = 1.2$ is depicted in Figure 3.6.

The other two chaotic sequences are 3-way Bernoulli shift sequence and (4,1)-tailed shift sequence. They are expected to have the performance indexes as $R_c = 2/(3N)$ and $R_c = \sqrt{3}/3N$. The form of 3-way Bernoulli shift map is

$$c_{n+1} = \begin{cases} 3c_n & 0 \le c_n < 1/3 \\ 3(c_n - 1/3) & 1/3 \le c_n < 2/3 \\ 3(c_n - 2/3) & 2/3 \le c_n < 1 \end{cases}, \tag{3.14}$$

and the form of the (4,1)-tailed shift function is

$$c_{n+1} = \begin{cases} 3c_n + 1/4 & 0 \le c_n < 1/4 \\ 3c_n - 1/2 & 1/4 \le c_n < 1/2 \\ 3c_n - 5/4 & 1/2 \le c_n < 3/4 \\ 3c_n - 3/4 & 3/4 \le c_n < 1 \end{cases} \quad (3.15)$$

Figure 3.7 depicts the 3-way Bernoulli map. The (4,1)-tailed shift map has been shown in Figure 3.5.

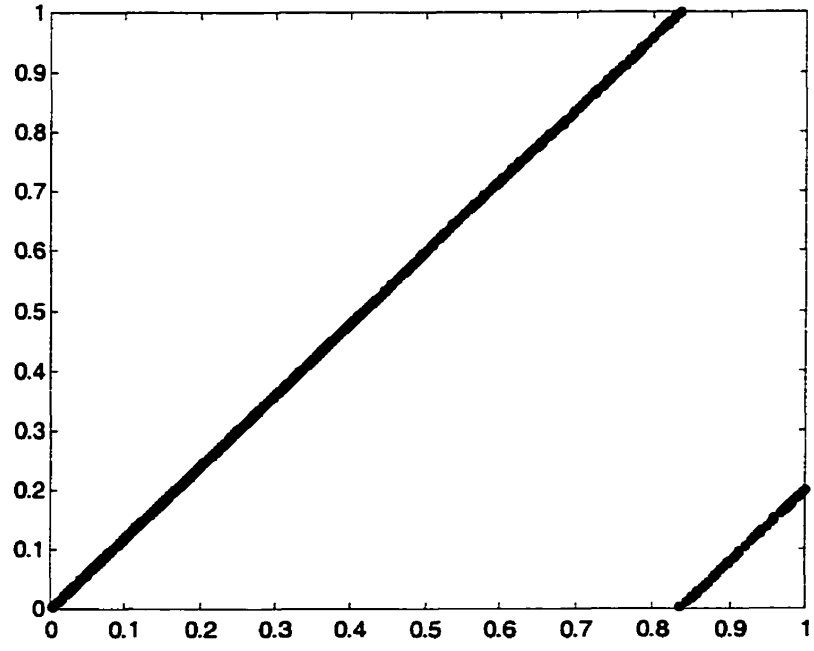

Figure 3.6 The graph of Renyi map with $\lambda = 1.2$

The watermarked image will undergo the same processing as used in Section 2.4 along with the same processing parameters. The processing includes image resizing, cropping, rotating, median filtering and JPEG compression. The BER curves are shown in Figures 3.8 to 3.12. We can see from these figures that there are not many differences in the BER performances of using different spreading sequences. This is because the correlation performance indexes, R_a and R_c of different spreading sequences are very close. Even though there is no distortion, the expected value of auto-correlation performance index for m-sequences and Gold sequences is $2/(3N)$, while for the chaotic spreading sequences generated by the well designed PWAM maps, this value is equal to

difference in the performance indexes is very small. For instance, if we have $N = 1023$, the difference of the two auto-correlation indexes will be 8.7308×10^{-5}. However, for all five kinds of manipulations, the sequence generated by the (4,1)-tailed shift sequence always has the best results.

It is also noted that the BER curves cross at some points. This is because the correlation performance indexes presented in (3.6) and (3.8) are obtained by averaging all possible sequences and all possible shifts. However, only a specific sequence can be used in the tests. Therefore, the performances of these different spreading sequences may not be exactly consistent with the theoretical value calculated by (3.6) and (3.8). Furthermore, as mentioned before, even the theoretical values of the performance indexes are very close. All of these reason result in the crossing of the BER curves.

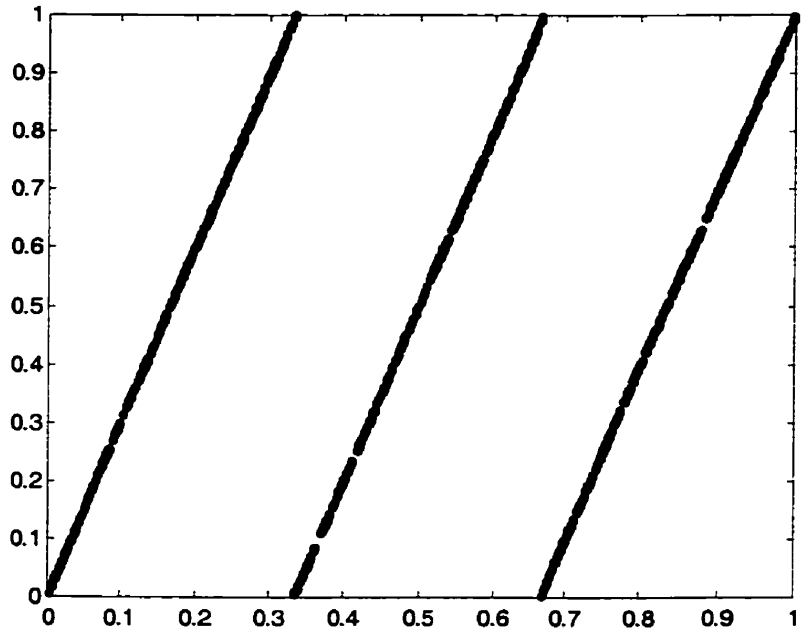

Figure 3.7 The graph of 3-way Bernoulli shift map

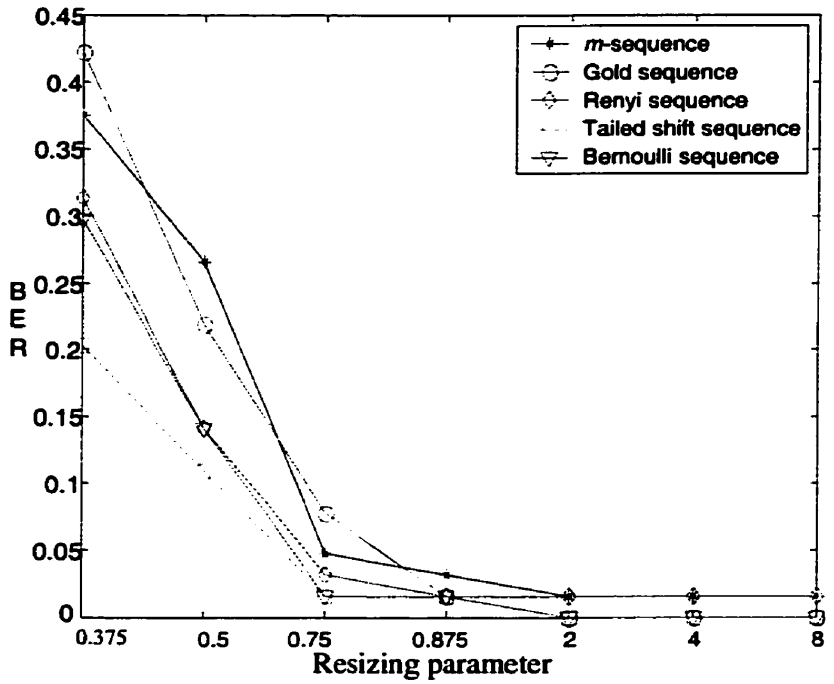

Figure 3.8 The BER results of using different spreading sequences under the resizing attack for different resizing parameters

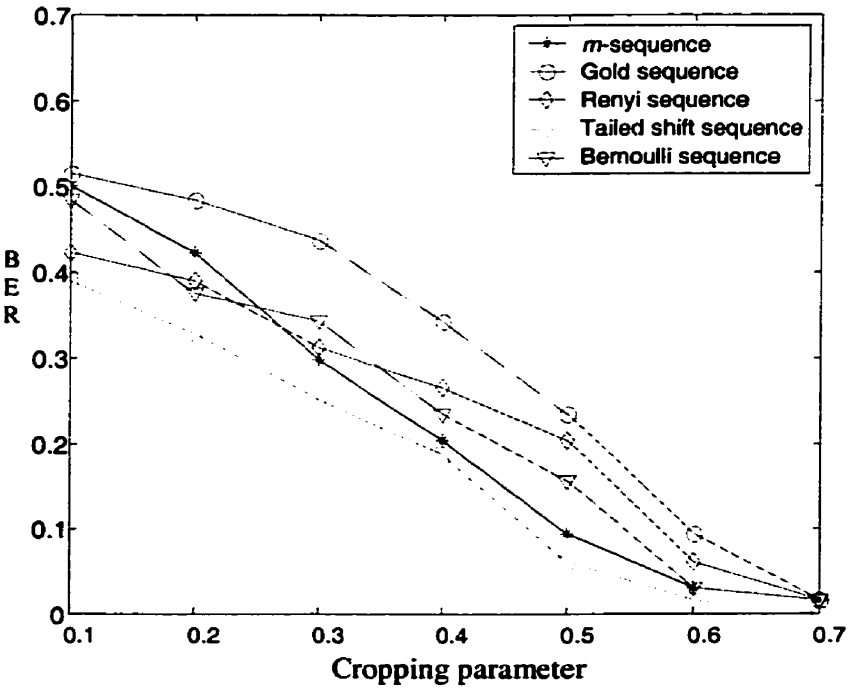

Figure 3.9 The BER results of using different spreading sequences under the cropping attack for different cropping parameters

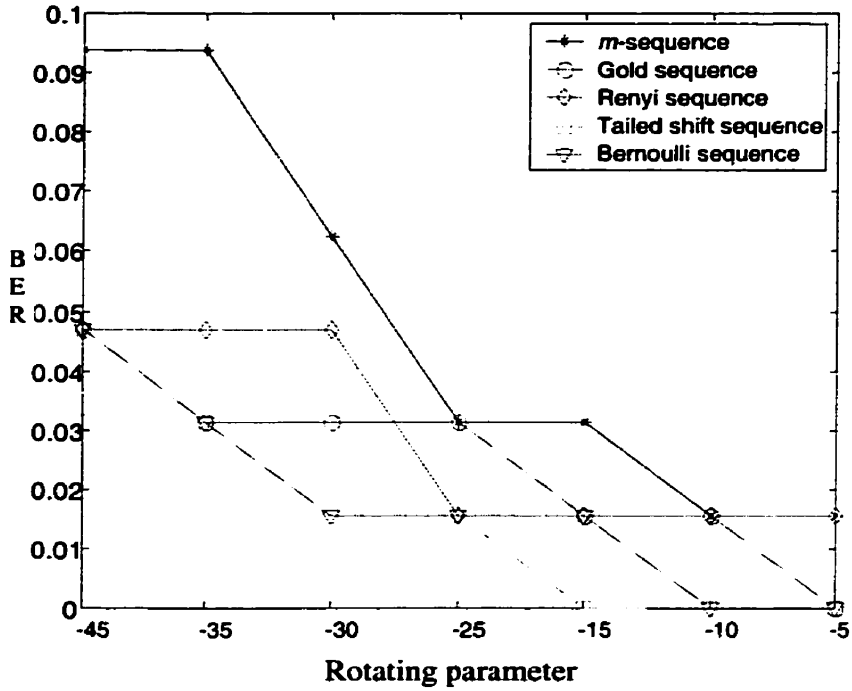

Figure 3.10 The BER results of using different spreading sequences under the rotating attack for different rotating parameters

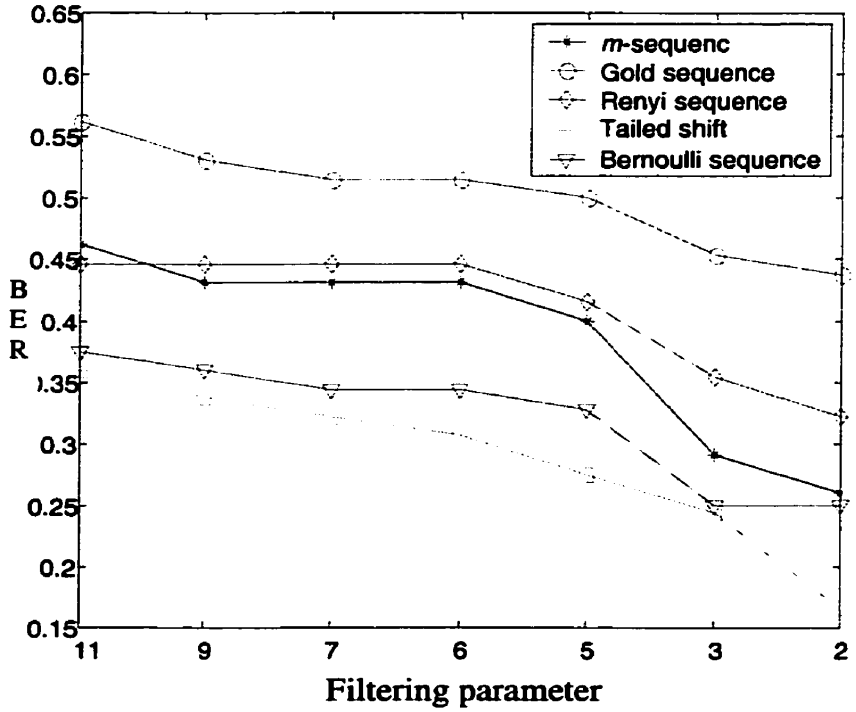

Figure 3.11 The BER results of using different spreading sequences under the median filtering attack for different filtering parameters

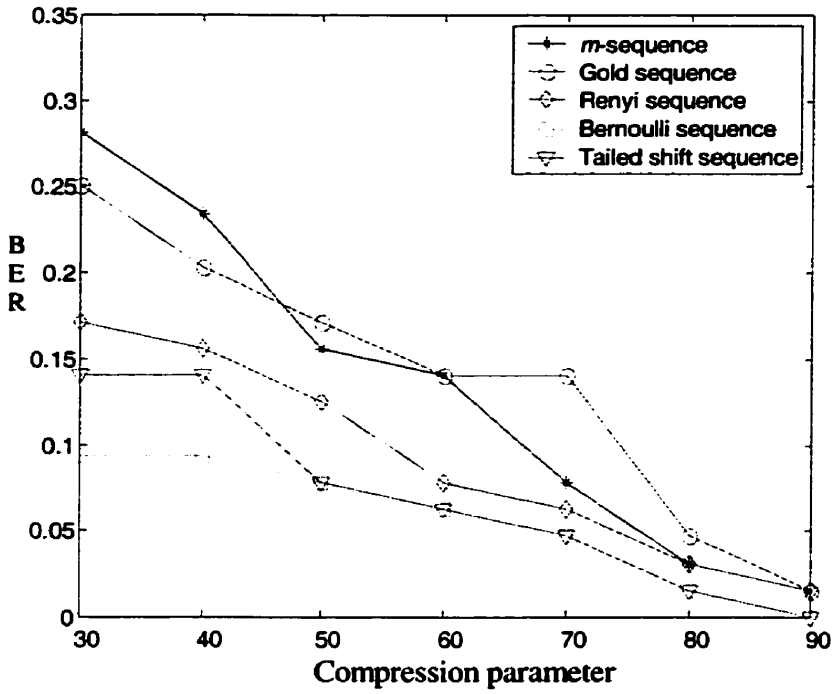

Figure 3.12 The BER results of using different spreading sequences under the JPEG compression for different compression parameters

3.6 Summary

In this chapter, chaotic spreading sequences instead of classical spreading sequences are proposed to be used in the conventional correlation-based spread spectrum watermarking scheme. Chaotic spreading sequences are noise-like, wideband, and possess good correlation properties, which are very important in the watermarking application. According to the analysis of the correlation performance in the watermarking application, it is found that the chaotic spreading sequences generated by PWAM maps have a better auto-correlation performance than the classical ones. Further, their cross-correlation performance is not worse than the performances of the classical sequences. Besides this, the generation and storage of chaotic spreading sequences is much easier compared to classical ones. The security of watermarking is also improved by using chaotic spreading sequences. Therefore the use of chaotic spreading sequences can lead to an improved robustness performance than the classical spreading sequences while they also have many other advantages.

CHAPTER 4

CHAOTIC PARAMETER MODULATION WITH APPLICATION TO DIGITAL IMAGE WATERMARKING

4.1 Introduction

Although the conventional correlation-based SS technique is the most popular method for digital watermarking, it is not a direct conclusion for it. There are some problems that the SS technique might encounter in its watermark application. First, the conventional SS technique usually requires a long spreading sequence for a satisfactory performance. However, the size of the digital media will put a limit on the length of the spreading sequence. For instance, say the size of an image is $N_1 \times N_2$ and the sequence length is N, the maximum number of information bits that can be inserted into the watermark is the $\left[\frac{N_1 \times N_2}{N}\right]$, where $[\cdot]$ is the round off operator to the nearest integer. Apparently there is a tradeoff between the length of the PN sequence and the payload, the number of information bits. Second, the conventional SS system requires a synchronization procedure to align the spreading sequences properly for the correlation detector. The watermark detector has to know both the PN sequence and its possible shift. When the shift is unknown, it will be found by means of a sliding correlator. That is, all possible shifts are experimentally evaluated, and the right shift is claimed when the correlation sum is larger than for all other shifts. Finding correlation is so cumbersome and complex, especially for PN sequences with a very large cycle, and hence synchronization is a complicated procedure for conventional SS systems. Third, conventional SS systems are developed for digital communications and the information sequences are usually binary. In other words, any information that is requested to be inserted in digital media has to be encoded into binary codes before the conventional SS watermarking technique can be applied. Not only is this procedure inconvenient, but it

also reduces the payload of a watermark. It is therefore preferred to use the numerical sequence as the information signal directly. Fourth, although the classical spreading sequences for communications, such as *m*-sequences and Gold sequences, have good auto-correlation and cross-correlation properties, their linearity limits the security of the watermarking. It has been shown that these PN sequences can be identified with a number of samples, which is much less than their actual length by means of linear regression models. Although the use of chaotic spreading sequences can solve this problem, the improvement of the robustness performance is not significant.

In this chapter, we propose using an analog spread spectrum system based on chaotic parameter modulation (CPM) [56, 57] for watermarking. While an analog SS system can process numerical sequences directly and solve the payload problem of the conventional SS watermarking technique, the CPM approach can increase the security of a watermark because of its inherent nonlinearity. In addition, the CPM approach does not require the synchronization process as in the conventional SS system. More precisely, CPM modulates the copyright information by embedding it in some parameter of a nonlinear dynamical system. By keeping the parameter of the nonlinear system in the chaotic regime, the output signal of the nonlinear system is therefore chaotic and hence has a wide bandwidth for SS transmission. In a CPM system, demodulation of information is a parameter estimation process from a noisy chaotic signal. The watermark detection for the CPM approach therefore does not require the synchronization procedure, and its performance is not affected by the correlation of the modulated signals. While the conventional CPM usually modulates the information sequence in the bifurcating parameters of a chaotic system, we propose a new CPM method by embedding the information in the initial condition. Not only does the new approach have a better parameter estimation performance, but it is also shown to perform a more robust watermark than the conventional CPM.

4.2 Chaotic Parameter Modulation Watermarking Based on Bifurcating Parameter

Chaotic parameter modulation (CPM) employs a chaotic dynamical system to modulate a message signal for SS transmission. By keeping the dynamical system in the

chaotic regime, the output wideband signal of the dynamical system may be used as the transmitted signal. When the CPM is applied to watermarking, the copyright information is stored in the bifurcating parameter. The key factor of successfully retrieving the information is to estimate the bifurcating parameter of a chaotic signal perturbed by noise. Figure 4.1 shows the block diagram of the basic structure of the CPM watermarking.

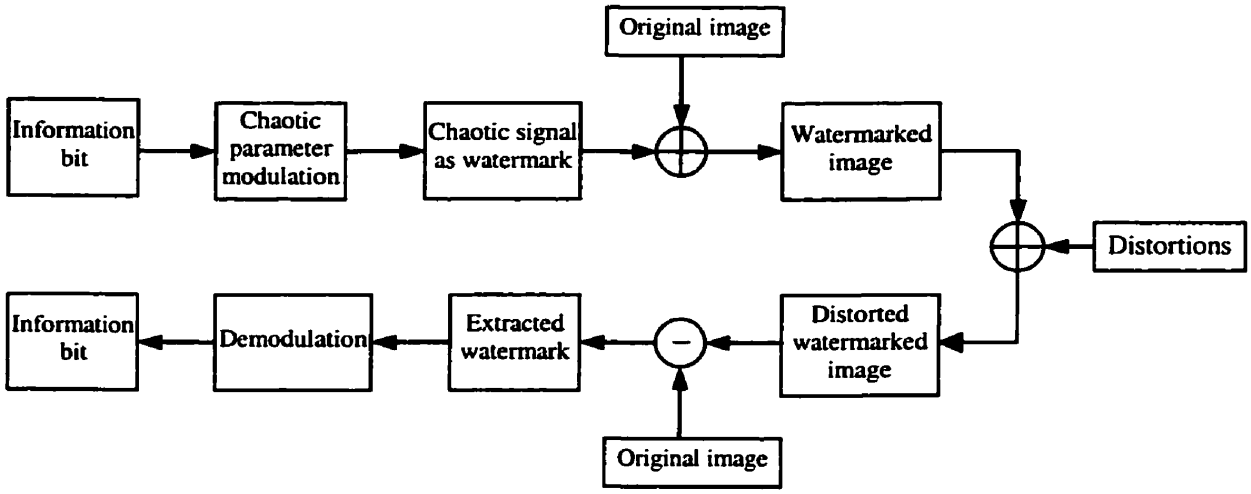

Figure 4.1 A chaotic parameter modulation (CPM) watermarking system

Let $b(i)$ be the information bit to be embedded into a digital image, and $x_n = f^n(x_0, \lambda)$ be the chaotic system to be used for CPM. λ denotes the bifurcating parameter, x_0 is the initial condition, and n represents the number of the iterations. To modulate $b(i)$ into the bifurcating parameter of the chaotic system, we set $\lambda = g(b(i))$ in general. The choice of g is to make sure the mapping of $b(i)$ is inside the chaotic regime of the λ. The logistic map given by $x_n = f(x_{n-1}, \lambda) = \lambda x_{n-1}(1 - x_{n-1})$ is used in this study [58]. The watermark signal is the system output sequence $\mathbf{x}(i) = (x_0, x_1, \cdots, x_n, \cdots, x_{N-1})$, where $x_n = f^n(x_0, g(b(i)))$. Figure 4.2 and 4.3 show the signal waveform and the power spectrum of $\{x_n\}$ for $\lambda = 4$. Apparently, the signal has a noise-like appearance and the power spectrum is also wideband.

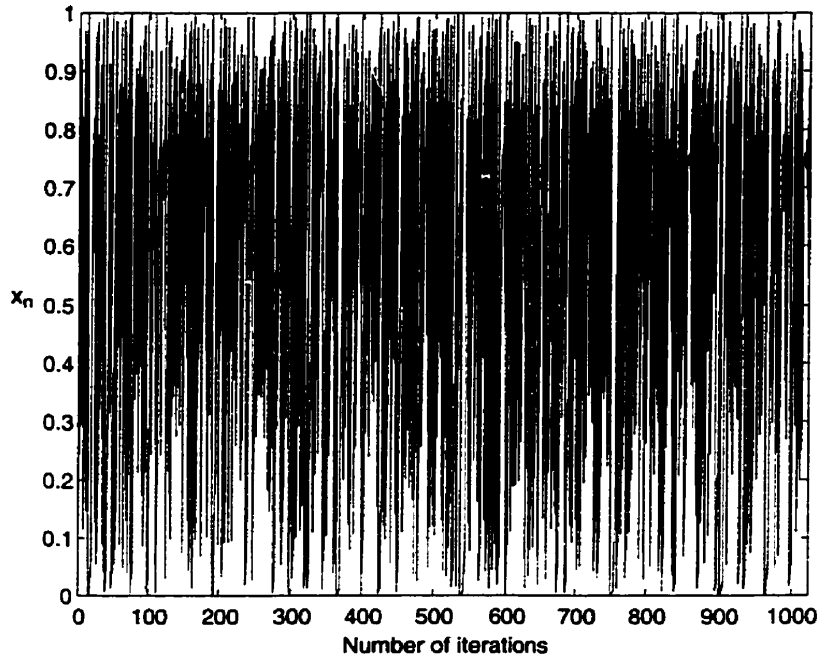

Figure 4.2 The chaotic sequence generated by the logistic map with $\lambda = 4$

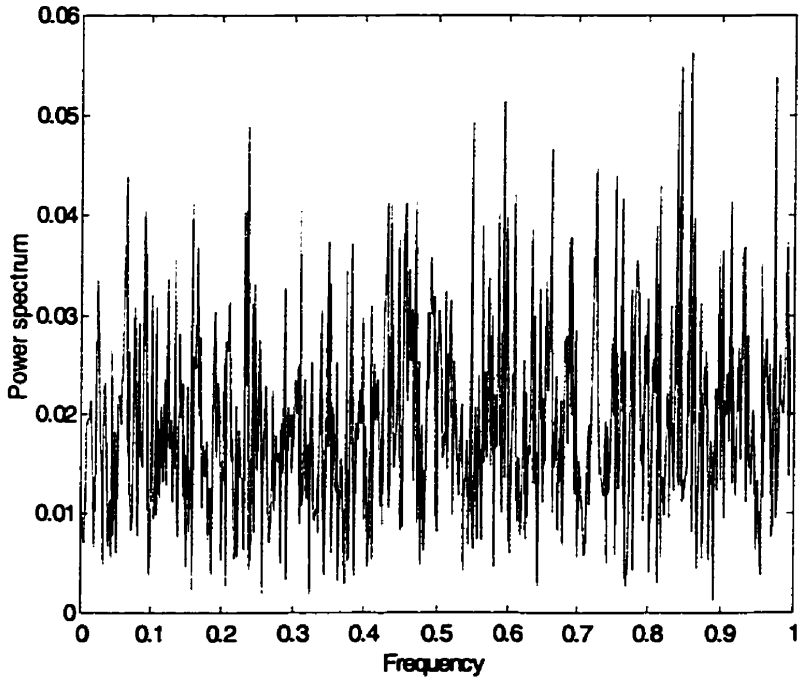

Figure 4.3 The power spectrum of the chaotic sequence generated by the logistic map with $\lambda = 4$

In the nonoblivious watermarking system in which the original image is available at the receiver end, the information retrieval process is given by (2.3). That is,

$$\mathbf{y}(i) = \mathbf{x}(i) + \frac{1}{\alpha}\mathbf{d}(i). \tag{4.1}$$

However, since we are not using a spreading sequence for generating watermarks, $\mathbf{x}(i)$ is not equal to $b(i)\mathbf{p}$ as given in (2.3). Instead, the nth element x_n of $\mathbf{x}(i)$ is generated by the above logistic map. In other words, the retrieved watermark can be expressed as

$$y_n = x_n + e_n, \tag{4.2}$$

where $e_n = \frac{1}{\alpha} d_n$. (4.2) is basically a measurement noise model with error e_n. If $\lambda = g(b(i))$ is given, retrieving the information $b(i)$ in the CPM watermarking system is equivalent to estimating the parameter λ from the following system,

$$\begin{aligned} x_n &= \lambda x_{n-1}(1 - x_{n-1}) \\ y_n &= x_n + e_n. \end{aligned} \tag{4.3}$$

Since λ is a function of the copyright information, and is also a function of time, an adaptive filter is recommended to track the variation of λ rather than just using an off-line estimation method [59, 60]. Using an auto-regression method for λ, (4.3) can be expressed as a state space estimation problem and λ can be tracked using a nonlinear filter such as the extended Kalman filter (EKF). However, it has been shown that the EKF exhibits some aperiodic motion in filtering noisy chaotic systems [61] and hence results in an inferior performance to a linear adaptive filtering algorithm such as the least mean square (LMS) [56].

Applying the LMS algorithm to estimate the λ of (4.3), we have [56]

$$\begin{aligned} \varepsilon_n &= y_{n+1} - \hat{\lambda}_n y_n (1 - y_n) \\ \hat{\lambda}_{n+1} &= \hat{\lambda}_n + \beta y_n (1 - y_n)\varepsilon_n, \end{aligned} \tag{4.4}$$

where $\hat{\lambda}_0$ is set as zero. For the LMS algorithm to converge, the step size parameter β should satisfy the condition $0 < \beta < 2/\beta_{\max}$, where $\beta_{\max} = \mathrm{var}(y_n(1 - y_n))$.

Another linear adaptive filtering implementation for solving (4.3) is based on the recursive least square (RLS) algorithm. The RLS algorithm has a faster convergence rate but the tradeoff is a higher computational complexity as well as a higher sensitivity to numerical instability. In this paper, a numerically stable version of the RLS algorithm is employed [62], and the corresponding RLS for tracking the λ of the logistic map is given as

$$
\begin{aligned}
q_n &= \frac{K_{n-1} y_{n-1}(1-y_{n-1})}{\eta + y_{n-1}(1-y_{n-1})K_{n-1}}, \\
\varepsilon_n &= y_n - \hat{\lambda}_{n-1} y_{n-1}(1-y_{n-1}), \\
\hat{\lambda}_n &= \hat{\lambda}_{n-1} + q_n \varepsilon_n, \\
K_n &= \frac{1}{\eta}(K_{n-1} - q_n y_{n-1}(1-y_{n-1})K_{n-1}),
\end{aligned}
\tag{4.5}
$$

where $\hat{\lambda}_0$ is set as zero, $0 < \eta \leq 1$, and K_0 is a large constant number. The mean-squared error of $\hat{\lambda}_n$ is magnified by the inverse of the smallest eigenvalue $\beta_{\min}$ of the correlation matrix.

However, no matter LMS or RLS is used, they only can work well in a high SNR environment. If the noise is so strong that the SNR is less than 10dB, both of them cannot converge. Therefore, this scheme cannot be applied to the oblivious watermarking because in this case the SNR is as low as around –30dB. This is the main problem of using the CPM watermarking scheme based on bifurcating parameter.

4.3 Chaotic Parameter Modulation Watermarking Based on Initial Condition

In this section, we propose a new CPM approach by embedding the information signal into the initial condition of a chaotic system instead of its bifurcating parameter. When it is applied to image watermarking, the watermark detection problem becomes estimating the initial condition of a chaotic system. Many works have been devoted to this problem, and some of these initial condition estimation techniques have been shown to be close to optimum [63, 64, 65].

The new CPM method modulates the information into the initial condition of a chaotic map, that is $x_0 = g(b(i))$. For the logistic map $x_n = \lambda x_{n-1}(1-x_{n-1})$, $x_0 = g(b(i))$ should be kept inside the regime (0,1), and λ can be any number inside the chaotic regime. Here, we set $\lambda = 4$. The chaotic output $\mathbf{x}(i) = (x_0, x_1, \cdots, x_n, \cdots, x_{N-1})$ generated by $x_n = f^n(g(b_i), 4)$ is then added to the image pixels to provide the watermarked image. The detection problem becomes the estimation of the initial condition $x_0 = g(b(i))$ from the noisy retrieved signal $\{y_n\}$ in (4.3). Two estimation methods are proposed here for solving this watermark retrieval problem. They are the dynamical programming (DP) and the halving method (HM) [59].

4.3.1 Dynamical programming

The estimation of the initial condition is normally obtained as the value of $\hat{x}_0$ that minimizes $J = \sum_{n=0}^{N-1}(y_n - \hat{x}_n)^2$ where $\hat{x}_n = f^n(\hat{x}_0)$. However, a straightforward minimization here requires one to compute $x_n = f^n(\hat{x}_0)$, which will lead to computational errors because of the sensitivity of chaotic systems to the initial condition. Rather, we employ the DP approach, which does not use that forward propagation.

We employ the parameterization given in [66] whereby the chaotic sequence $(x_0, x_1, \cdots, x_{N-1})$ is replaced by $(s_0, s_1, \cdots, s_{N-2}, x_{N-1})$. The sequence $\{s_n\}$ is obtained as $s_n = \begin{cases} 0 & if \quad 0 \le x_n < 0.5 \\ 1 & if \quad 0.5 \le x_n < 1 \end{cases}$ for the logistic map $f(x) = 4x(1-x)$. This itinerary can be used to determine the appropriate preimages of the chaotic signal when propagated backward. The function is $x_{n-1} = f_{s_{n-1}}^{-1}(x_n) = \dfrac{1 + (2s_{n-1} - 1)\sqrt{1 - x_n}}{2}$, which consists of the preimages $x_{n-1} = f_{s_{n-1}}^{-1}(x_n) = \dfrac{1 - \sqrt{1-x}}{2}$ for $s_{n-1} = 0$ and $x_{n-1} = f_{s_{n-1}}^{-1}(x_n) = \dfrac{1 + \sqrt{1-x}}{2}$ for $s_{n-1} = 1$.

Let $J_k = \sum_{n=0}^{k} (y_n - \hat{x}_n)^2$ where $\hat{x}_n = f^{-1}_{\hat{s}_n, \hat{s}_{n+1}, \cdots, \hat{s}_{k-1}}(\hat{x}_k)$ be the inverse function composition for $n \le k-1$. An objective function is defined as $I_k = \min_{(\hat{s}_0, \hat{s}_1, \cdots, \hat{s}_{k-1})} J_k$, and we have

$$
\begin{aligned}
I_k &= \min_{(\hat{s}_0, \hat{s}_1, \cdots, \hat{s}_{k-1})} \sum_{n=0}^{k} (y_n - \hat{x}_n)^2 \\
&= \min_{\hat{s}_{k-1}} \min_{(\hat{s}_0, \hat{s}_1, \cdots, \hat{s}_{k-2})} \sum_{n=0}^{k-2} (y_n - \hat{x}_n)^2 + (y_{k-1} - f^{-1}_{\hat{s}_{k-1}}(\hat{x}_k))^2 + (y_k - \hat{x}_k)^2 \\
&= \min_{(\hat{s}_0, \hat{s}_1, \cdots, \hat{s}_{k-2})} [\min_{\hat{s}_{k-1}} J_{k-1} + (y_k - \hat{x}_k)^2].
\end{aligned}
$$

Since $\hat{x}_k$ does not depend on $(\hat{s}_0, \hat{s}_1, \cdots, \hat{s}_{k-1})$ but only on $(\hat{s}_k, \cdots, \hat{s}_{N-2}, \hat{x}_{N-1})$, we have

$$
I_k = \min_{\hat{s}_{k-1}} I_{k-1} + (y_k - \hat{x}_k)^2 \tag{4.6}
$$

for our DP algorithm. This recursion is computed for $k = 0,1,\cdots,N-1$ with the initialization $I_0 = (y_0 - x_0)^2$. After obtaining I_{N-1}, $(\hat{s}_0, \hat{s}_1, \cdots, \hat{s}_{N-2})$ is searched for its optimal value. Then, we backtrack to obtain the estimation of the $\hat{x}_0$ as $\hat{x}_0 = f^{-1}_{\hat{s}_0, \hat{s}_1, \cdots, \hat{s}_{n-2}}(\hat{x}_{N-1})$. Note that this approach avoids the exponential increase in computational errors since the propagation is along the stable manifold.

4.3.2 Halving method

The second initial condition estimation technique is a sub-optimal but computationally efficient approach called the halving method (HM). It has been proven [59] that the initial condition for the logistic map $f(x) = 4x(1-x)$ can be represented by $x_0 = H(\sum_{n=1}^{\infty} a_n 2^{-n})$, where $H(x) = \sin^2(\frac{\pi}{2}x)$ and $a_n = a_{n-1} \oplus s_{n-1}$. $\oplus$ denotes the exclusive OR operation. s_{n-1} is the itinerary of $H^{-1}(x_{n-1})$: $s_{n-1} = 0$ if $0 \le H^{-1}(x_{n-1}) < 1/2$; $s_{n-1} = 1$ if $1/2 \le H^{-1}(x_n) < 1$. So the expression of a_n can also be written as

$$a_n = \begin{cases} 0 & if \quad 0 \le H^{-1}(x_{n-1}) < 1/2 \quad and \quad a_{n-1} = 0 \\ 1 & if \quad 1/2 \le H^{-1}(x_{n-1}) < 1 \quad and \quad a_{n-1} = 0 \\ 1 & if \quad 0 \le H^{-1}(x_{n-1}) < 1/2 \quad and \quad a_{n-1} = 1 \\ 0 & if \quad 1/2 \le H^{-1}(x_{n-1}) < 1 \quad and \quad a_{n-1} = 1 . \end{cases} \tag{4.7}$$

It is interesting to note that in practice we only have a limited number of the itinerary $(s_0, s_1, \cdots, s_{N-1})$ based on the itineration of the data set $(x_0, x_1, \cdots, x_{N-1})$. Hence, the estimation of x_0 is given by as $\hat{x}_0 = H(\sum_{n=1}^{N} a_n 2^{-n})$, and the estimation of $H^{-1}(\hat{x}_0) = 0.a_1 a_2 \cdots a_N$ will have a maximum error of $1/2^N$. This is why it is called the halving method. When N is large, the estimate produced by the HM is found to be very accurate [63, 64].

In the watermarking application, we obtain the $\{\hat{s}_n\}$ as the itinerary $H^{-1}(y_n)$. With the initialization $\hat{a}_1 = \hat{s}_0$, we have $(\hat{a}_1, \hat{a}_2, \cdots \hat{a}_N)$ by $\hat{a}_{n+1} = \hat{a}_n \oplus \hat{s}_n$. The HM estimate of x_0 is then given by

$$\hat{x}_0 = H(\sum_{n=1}^{N} \hat{a}_n 2^{-n}) = \sin^2(\frac{\pi}{2} \sum_{n=1}^{N} \hat{a}_n 2^{-n}). \tag{4.8}$$

The same problem exists for the CPM watermarking based on initial condition as the one for the CPM based on bifurcating parameter. That is, DP and HM cannot work well in the background of strong noises either. If the noise in the chaotic signal is so strong that the estimated sign of the parameterization sequence $\{\hat{s}_n\}$ is very different from the original $\{s_n\}$, it will lead to a big bias in the estimation of the initial condition. Hence, it is also not applicable to the oblivious watermarking. So in this thesis, the term "CPM watermarking" is only referred to the CPM nonoblivious watermarking.

4.4 Performance Test

To evaluate the performance of the CPM watermarking scheme and to compare it with the conventional SS watermarking technique, two different signals are used as the copyright information. The first one is a binary sequence so that no coding is needed for

the conventional SS technique to apply for. The second one is a sequence of numerical values, which simulates the real application of image watermarking.

The image "Lena" with the size 256×256 is used as the host data to be watermarked. The watermarked image is subjected to the same series of image processing and attacks, including image resizing, cropping, rotating, median filtering and JPEG compression. BER is given to evaluate the performance of a watermarking scheme for binary information sequences. As for numerical information sequences, we use mean square error (MSE) as the performance measure.

4.4.1 Performance test on binary copyright information

For the conventional correlation-based SS watermarking methods as described in 2.2, each information bit is multiplied with a sequence of spreading codes. The m-sequence is employed here as the spreading sequence. The correlation sum is used in the detection process.

For the CPM watermarking scheme based on bifurcating parameter, two parameters are selected to represent the binary symbols 1 and −1. That is $\lambda = g(b(i)) = \begin{cases} \lambda_1, & if \quad b(i) = -1 \\ \lambda_2, & if \quad b(i) = 1 \end{cases}$. The chaotic signal is then generated based on these two bifurcating parameters, i.e., $x_n = \begin{cases} f^n(x_0, \lambda_1), & if \quad \lambda = \lambda_1 \\ f^n(x_0, \lambda_2), & if \quad \lambda = \lambda_2 \end{cases}$, and added to the image pixels. In the detection process, LMS and RLS are applied to estimate λ from the noisy received signal $\{y_n\}$. A threshold λ_t is set as $|\lambda_t - \lambda_1| = |\lambda_t - \lambda_2|$ to decide which symbol is transmitted. Without loss of generality, assuming that $\lambda_1 < \lambda_2$, we have $\hat{\lambda} = \begin{cases} \lambda_1, & if \quad \hat{\lambda} < \lambda_t \\ \lambda_2, & if \quad \hat{\lambda} \geq \lambda_t \end{cases}$.

For the CPM watermarking scheme based on initial condition, the information bit is stored in the initial condition by $x_0 = g(b(i)) = \begin{cases} x_{10}, & if \quad b(i) = -1 \\ x_{20}, & if \quad b(i) = 1 \end{cases}$ where x_{10}, x_{20} are two different initial values. Without loss of generality, we assume $x_{10} < x_{20}$ here. In

the demodulation process, the DP or HM approach is applied to provide an estimate of the initial condition x_0. A threshold x_{t0} is then chosen to determine whether $\hat{x}_0$ is equal to x_{10} or x_{20} and afterwards find out what *b(i)* is. Here we set the threshold x_{t0} as $|x_{t0} - x_{10}| = |x_{t0} - x_{20}|$.

We now perform tests on the proposed CPM watermarking schemes. The BER curves under those attacks and processing are depicted in Figures 4.4 to 4.8. In all experiments, different sequence lengths including N = 63, 127, 255, 511 and 1023 are considered.

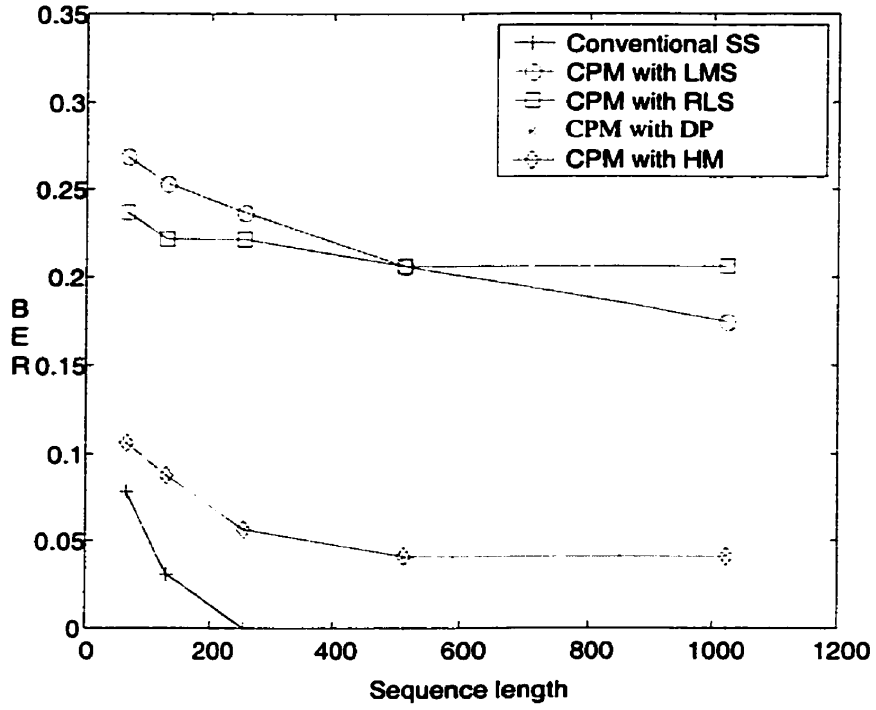

Figure 4.4 The BER curves of different watermarking schemes for different sequence lengths under the resizing attack

In the test of image resizing, the watermarked image is resized to 0.75 times of its original size. It is shown in Figure 4.4 that the conventional SS scheme has the most robust performance under this image processing action. The performance of the CPM based on initial condition is not as good as the SS scheme, but it is close to that of the SS technique. The CPM based on bifurcating parameter has the worst performance for both LMS and RLS methods. It is interesting to note that when the sequence length is short,

say $N=64$, the performance of the conventional SS watermarking has a significant degradation while the CPM watermarking is not that much sensitive to this effect.

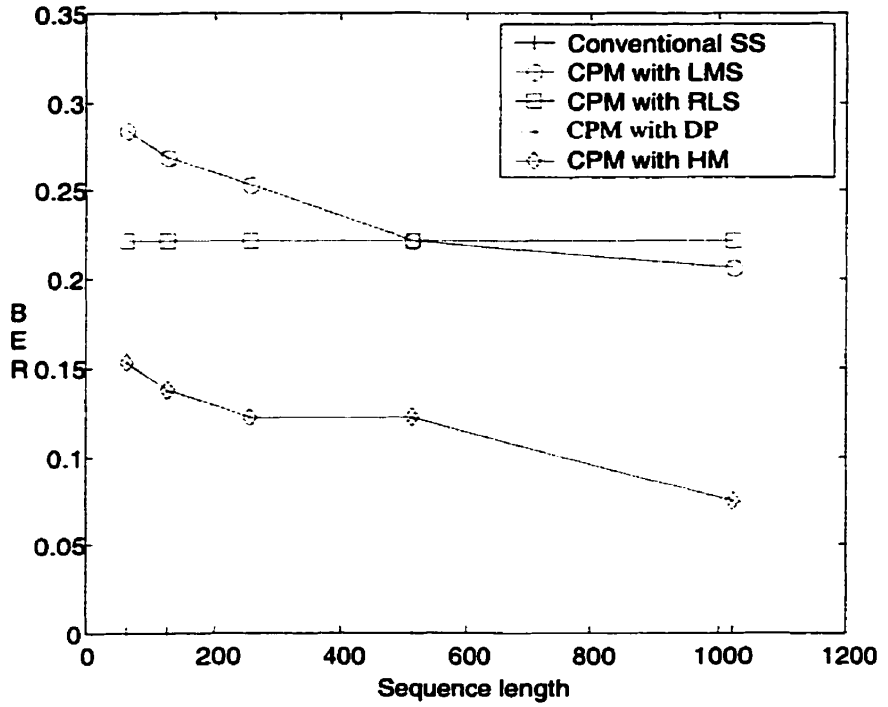

Figure 4.5 The BER curves of different watermarking schemes for different sequence lengths under the cropping attack

In the test of image cropping, the cropping parameter is set as 0.5. That is, the half of the original watermarked image is cut off. In order to extract the watermark from the cropped image, the missing portions of the image were replaced with portions from the original unwatermarked image. As shown in Figure 4.5, the SS watermarking scheme has the best performance. All the BER values of it are equal to zero. The CPM scheme based on initial condition is the second. It is also noted that the SS scheme is more robust to cropping than to resizing. This is because for the cropped part, the extracted watermarks signals can be presented as $y_n = 0$, while for the remaining part the watermark signals are basically undistorted, i.e., $y_n = x_n$. When the correlation sum $\hat{b}(i) = sign(\frac{1}{N}\sum_{n=0}^{N-1} y_n p_n)$ is applied, the correct result still can be obtained. However, for the CPM scheme, it requires the sequence $\{y_n\}$ to do the itinerary and then the

estimation. The more elements of $\{y_n\}$ are lost, the larger the bias is introduced to the estimation.

In the test of image rotating, the watermarked image is rotated by 10 degrees. After rotation, synchronization error of the spreading code is introduced to the watermarked image. For the SS watermarking scheme, this results in a great degradation of the performance. It is seen from Figure 4.6 that the BER curve of the SS technique is not only worse than those for resizing and cropping, but it is also worse than the BER curve of the CPM based on initial condition. It is expected that for larger rotations the SS technique will have higher degradations. However, for the CPM schemes, the performance does not change much. The reason is that the CPM approach does not require any synchronization in the demodulation process. The poor performance of the CPM based on bifurcating parameter is due to the fact that the LMS and RLS cannot work properly in the background of strong measurement noise and is not the result of synchronization errors.

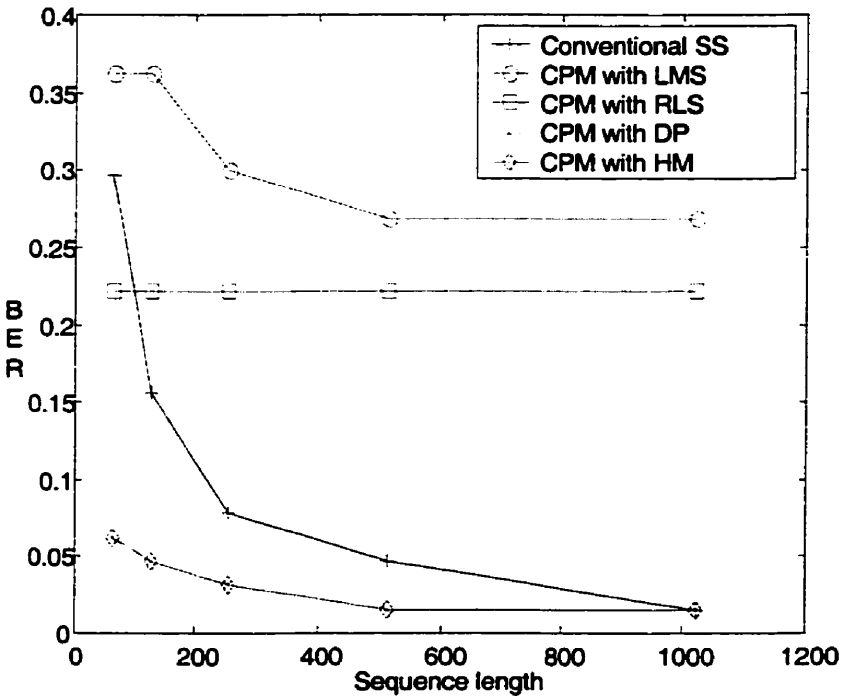

Figure 4.6 The BER curves of different watermarking schemes for different sequence lengths under the rotating attack

In the test of median filtering, a median filter with a window size of 3×3 is applied to the watermarked image. As shown in Figure 4.7, all watermarking schemes are less robust to this nonlinear processing compared to resizing and cropping. For the SS technique, there is a significant increase in BER for $N < 500$. In fact, when $N \leq 128$, the performance of the SS technique becomes even worse than the CPM based on initial condition. For the CPM based on initial condition, the curve is relatively flat which again indicates that the performance is not that sensitive to the sequence length.

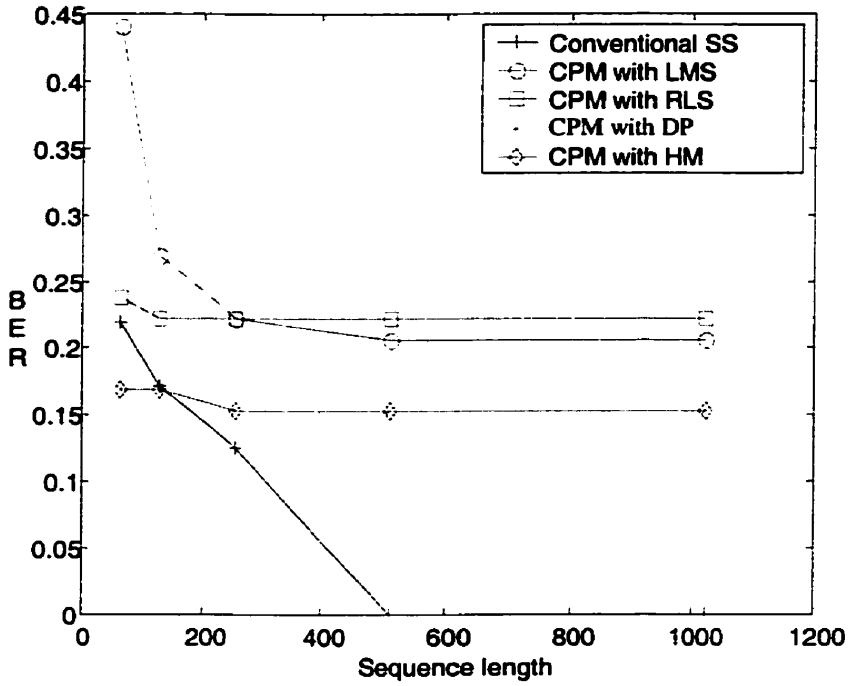

Figure 4.7 The BER curves of different watermarking schemes for different sequence lengths under the median filtering attack

In the test of image compression, the JPEG compression is applied to the watermarked image with the quality factor 50%. In Figure 4.8 the BER curves are plotted for different watermarking schemes. Again, the SS watermarking technique has the most robust performance when the sequence length is long. When the sequence length is short, CPM based on initial condition has a slightly better performance than the SS technique. In fact, when $N = 64$, the SS watermarking technique appears to be even worse than the CPM method based on RLS under this compression attack.

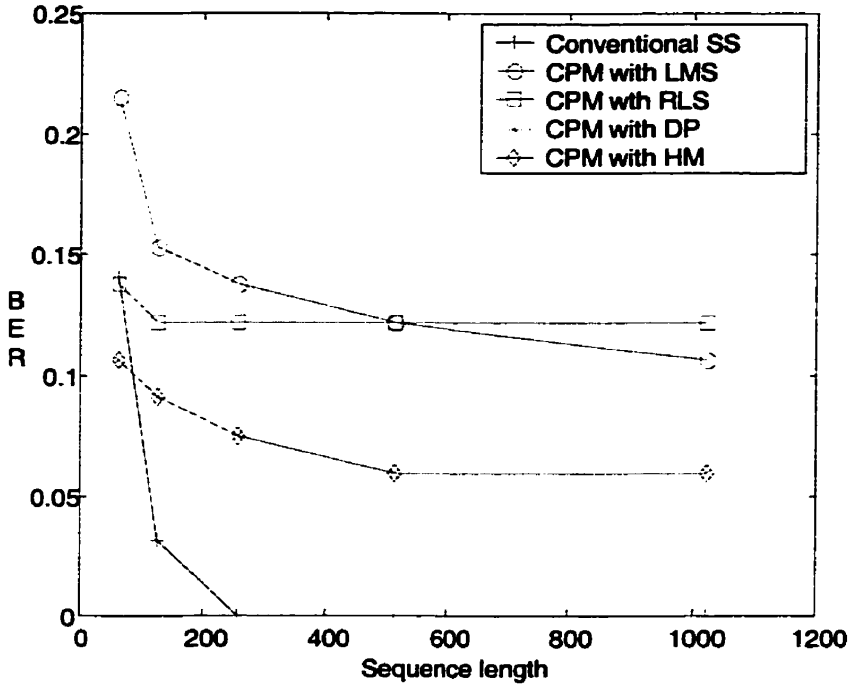

Figure 4.8 The BER curves of different watermarking schemes for different sequence lengths under the JPEG compression

There are some common observations that should be mentioned here. First, there is no apparent difference between the DP and HM estimations in all the tests. This is because the value of the sequence length is set relatively large for the HM method to get accurate estimates. Although the minimum length is only 64, the maximum error of using the HM will be $1/2^{64} = 5.421 \times 10^{-20}$, which is too small to be displayed in the figures.

Second, it is noted that when the sequence length is short, say $N < 511$, the RLS always has a better performance than the LMS. This is consistent with the theory that the RLS algorithm can always reach the least square optimum solution. And because the RLS has a faster converge rate than LMS, the short sequence length may not be enough for the LMS to track the bifurcating parameter and get the correct estimation results. When the sequence length is very long, the performance of the LMS is slightly better than that of the RLS. However, the BER values of them are so close that the difference is only within 0.025 and can be ignored.

4.4.2 Performance test on numerical copyright information

In the real watermarking application, copyright information is usually not in a binary format. The conventional SS technique requires a coding scheme to insert the information into the image. However, it is desirable to have a watermarking technique that can modulate the copyright information directly. For instance, when a real image is to be embedded into the host image, the pixel values which are in numerical values do not need to be converted into binary sequences. Not only does this simplify the watermarking procedure, but it also increases the payload of a watermark.

For the SS watermarking technique, each element of the numerical sequence is still spread out by a spreading sequence denoted as $b(i)\mathbf{p}$. However, the detection process is a little bit different now. Since the information sequence is not binary any more, the sign cannot be used as the threshold to decide which information bit is transmitted. Instead, the information bit is retrieved as

$$\begin{aligned}\hat{b}(i) &= \frac{1}{N}\left(\sum_{n=0}^{N-1} y_n p_n\right) \\ &= \frac{1}{N}\sum_{n=0}^{N-1}\left(b(i)p_n + \frac{1}{\alpha}d_n\right)p_n \\ &= \frac{1}{N}\sum_{n=0}^{N}\left(b(i){p_n}^2 + \frac{1}{\alpha}d_n p_n\right).\end{aligned} \tag{4.9}$$

The spreading sequence used in this study is still an m-sequence, which is binary. Hence, we have ${p_n}^2 = 1$. If there is no distortion, we have $d_n = 0$, and thus $\hat{b}(i) = \frac{1}{N}\sum_{n=0}^{N-1} b(i){p_n}^2 = \frac{1}{N}\sum_{n=0}^{N-1} b(i) = b(i)$. The information bit can be correctly recovered. However, when d_n is nonzero, $\frac{1}{\alpha}d_n p_n$ will introduce errors in the detection process.

For the CPM watermarking based on bifurcating parameter, the modulation is relatively straightforward because the bifurcating parameters used to store the information are real numbers. We only need to control the bifurcating parameters so that the whole numerical sequence well be inside the chaotic regime. That is, $\lambda_i = g(b(i))$ and λ_i is in the chaotic regime for all i. The chaotic signal $\{x_n\}$ generated by $x_n = f^n(x_0, \lambda_i)$

is then used as the watermark. In the detection process, LMS and RLS are applied to estimate λ_i from the extracted signals $\{y_n\}$. Given the estimate $\hat{\lambda}_i$, we can recover the information value *b(i)* as $\hat{b}(i) = g^{-1}(\hat{\lambda}_i)$.

For the CPM watermarking based on initial condition, *b(i)* is stored in the initial condition by $x_{i0} = g(b(i))$. The chaotic watermark signal is generated by the iteration $x_n = f^n(x_{i0}, \lambda)$ with $\lambda = 4$ for the logistic map. The DP and HM approach are employed in the estimation of the initial condition x_{i0}. The estimated *b(i)* can be afterwards obtained by $\hat{b}(i) = g^{-1}(\hat{x}_{i0})$.

To evaluate the performance of the proposed CPM watermarking scheme and to compare it with the SS technique, a 64×64 image of 256 grey levels named "Camera" as shown in Figure 4.9 is used as the copyright information. More precisely, the image "Camera" will be modulated to provide the watermark signal and inserted into the host image "Lena". If we transform this image sample into the binary code and assume that each pixel value is encoded into 8 bits, the total number of the information bits will be $L = 64 \times 64 \times 8 = 32768$. Since the size of "Lena" is 256×256, the maximum length of the spreading sequence is $\frac{256 \times 256}{32768} = 2$. It is too short for both the SS technique and CPM to work properly. Therefore, the copyright information has to use the numerical values and is modulated directly as described above. In this case, the maximum length of the spreading sequence can be extended to $\frac{256 \times 256}{64 \times 64} = 16$. Since 16 points is still too short for the LMS and RLS to converge, the CPM based on bifurcating parameter will not be considered any further.

The two dimensional image "Camera" is first transformed to a one-dimensional signal by scanning from left to right and from top to bottom. Since the pixel value varies between 0 and 255, the map $x_{i0} = g(b(i)) = b(i)/256$ is employed here to make sure that all initial values are within the interval (0, 1). Because the information sequence is of numerical value, the mean square error (MSE), defined by

$$MSE = \frac{1}{L}\sum_{i=0}^{L-1}(b(i) - \hat{b}(i))^2 \qquad (4.10)$$

is used as a measure. The results are summarized in Table 4.1 and the images retrieved by CPM-DP for these various attacks are displayed in Figure 4.10.

Figure 4.9 The real image "Camera" as the copyright information

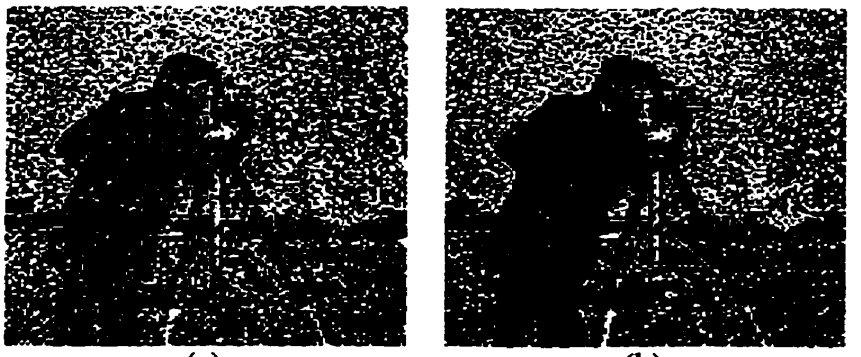

Figure 4.10 The retrieved copyright information by the CPM-DP method after (a) resizing, (b) cropping, (c) rotating, (d) median filtering, (e) JPEG compression

	Conventional SS technique	Chaotic parameter modulation based on initial condition	
		DP	HM
Resizing to 3/4 of the original	0.1586	0.0853	0.0853
Cropping to 1/2 of its original	0.7156	0.2881	0.2881
Rotating to 5 degrees	0.6082	0.0134	0.0134
Median filter 3×3	0.2922	0.1830	0.1830
JPEG compression to 60%	0. 3278	0.1724	0.1724

Table 4.1 The comparison of MSE by using the CPM based on initial condition and conventional SS watermarking techniques for numerical copyright information

Apparently, the CPM watermarking technique outperforms the SS technique. The improvement can be quite significant especially for resizing and rotation. This is because the CPM scheme is originally designed for analog spread spectrum communications and is very suitable for modulating and demodulating numerical information. In addition, both DP and HM work effectively for short data sequence and hence produce a good watermarking performance. For the SS technique, since the sign cannot be used as the threshold to detect the information, it is more sensitive to distortions.

Under the same attacks, the BER performances versus the sequence length are plotted in Figures 11 to 15. It is shown that even though the sequence length is very long, the performance of the conventional SS watermarking for numerical sequences is still worse than of the CPM watermarking based on initial condition. Thus, the approach of the CPM based on initial condition is really superior to the conventional SS technique in watermarking numerical information.

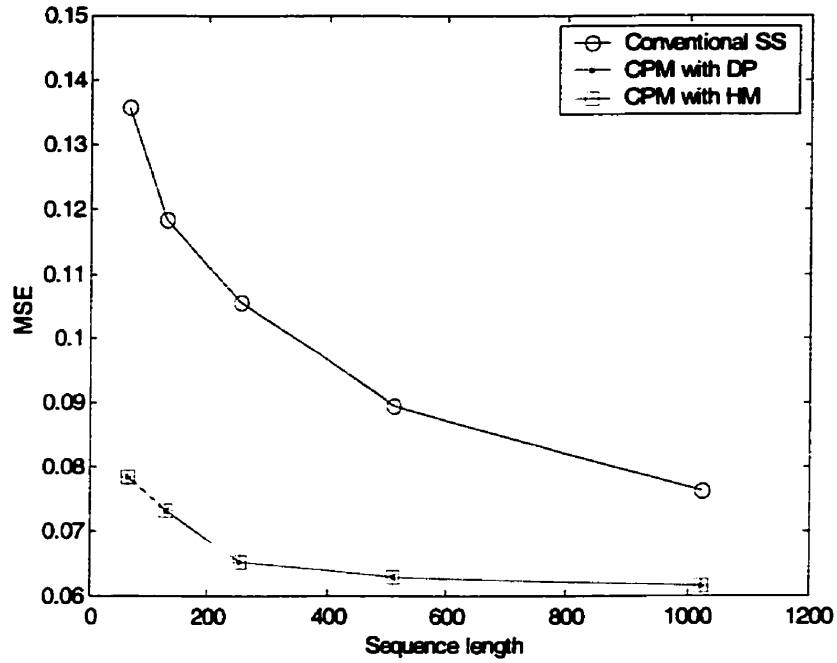

Figure 4.11 The MSE performances of different watermarking schemes for different sequence lengths under the resizing attack

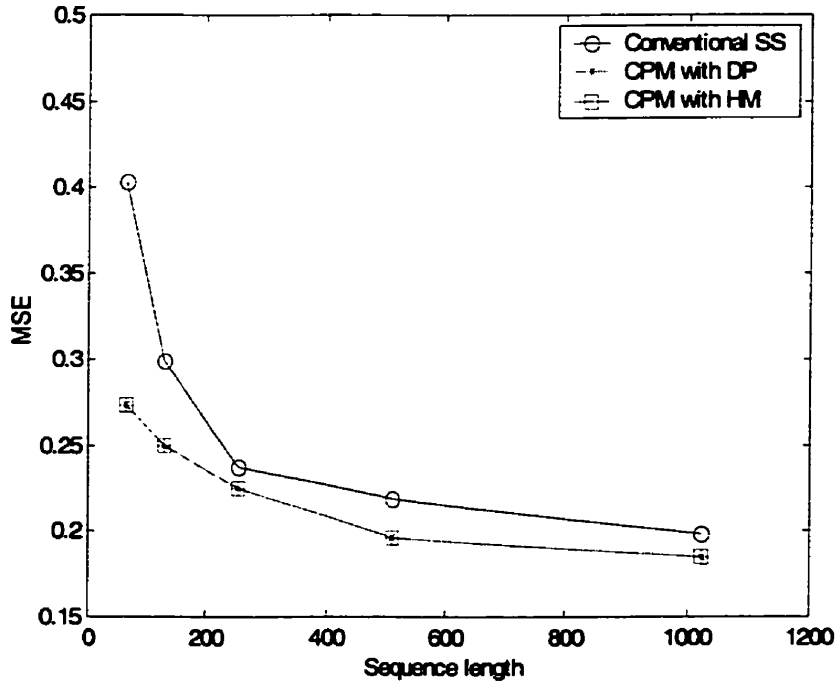

Figure 4.12 The MSE performances of different watermarking schemes for different sequence lengths under the cropping attack

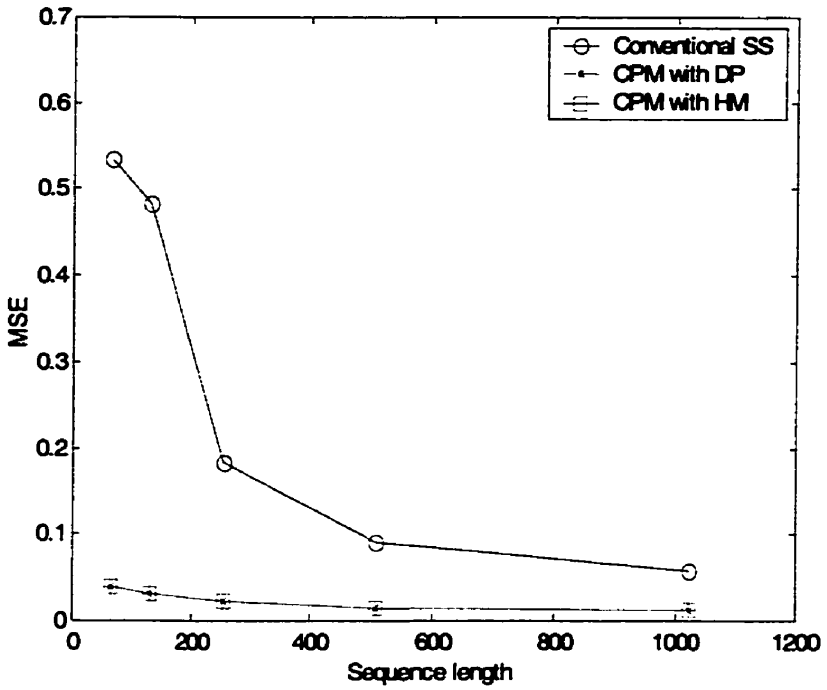

Figure 4.13 The MSE performances of different watermarking schemes for different sequence lengths under the rotating attack

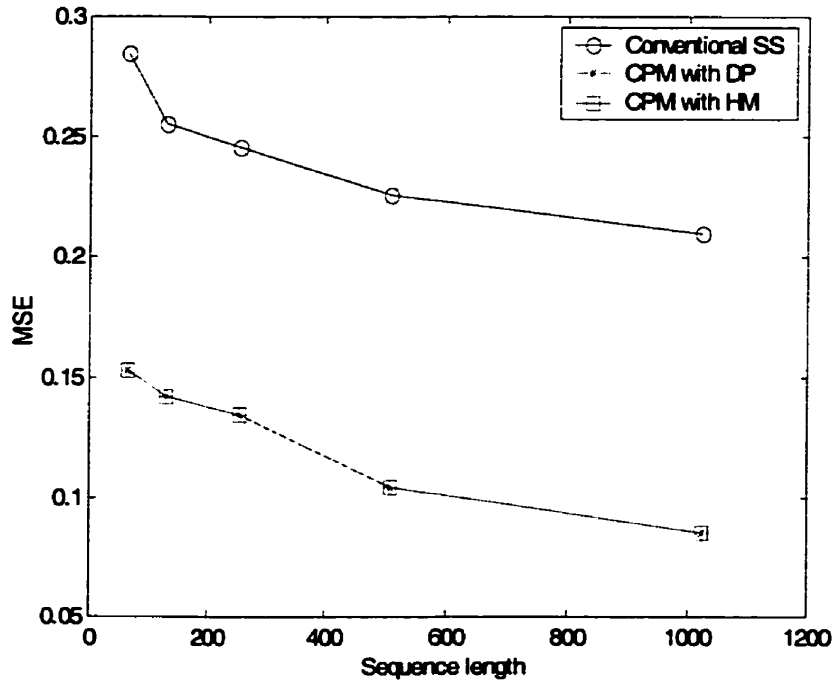

Figure 4.14 The MSE performances of different watermarking schemes for different sequence lengths under the median filtering attack

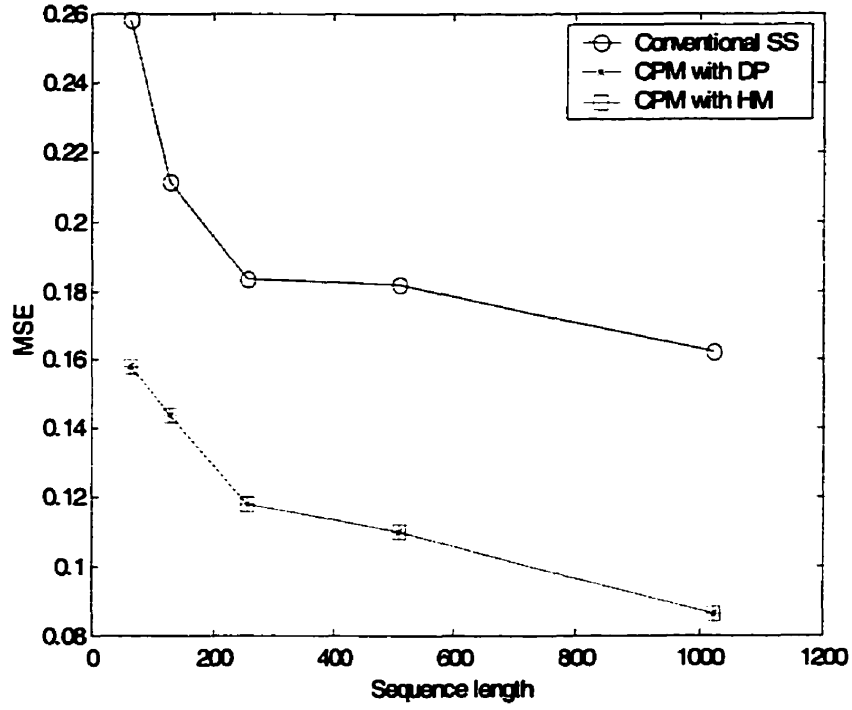

Figure 4.15 The MSE performances of different watermarking schemes for different sequence lengths under the JPEG compression

4.5 Summary

We propose a novel watermarking scheme for digital images by using the chaotic parameter modulation (CPM). The information signal is modulated into the parameter of a chaotic dynamical system. The generated chaotic signal is then used as a watermark signal for insertion into a host image. Retrieval of the copyright information is formulated as a problem of parameter estimation from a noisy chaotic signal. In addition to embedding the copyright information into the bifurcating parameter of a chaotic system, we propose a new CPM scheme based on initial condition, that is, the copyright information is stored into the initial condition instead of the bifurcating parameter. Because of the efficient dynamic programming and halving method for initial condition estimation, the CPM based on initial condition is shown to work more effectively in a noisy environment compared to the CPM scheme based on bifurcating parameter.

Two kinds of information sequences, binary and numerical, are considered in our performance tests. For a binary symbol, the SS nonoblivious watermarking technique has

a better performance than the proposed CPM scheme when the length of the spreading sequence is long enough. But when the length decreases, the performance of the SS nonoblivious technique will greatly degrade and become worse than the CPM scheme based on initial condition. The CPM watermarking scheme also has an advantage that, it does not require the spreading code synchronization. This can greatly simplify the implementation of the watermarking scheme and reduce the potential synchronization errors. It also improves the robustness against the rotation attack. For numerical copyright information, it is a natural choice to use the CPM scheme since it is originally designed for analog spread spectrum. There is no need to have a coding process and it is shown that the performance of the CPM scheme based on initial condition is superior to the conventional SS nonoblivious technique.

However, there are still some problems in the application of the CPM approach to image watermarking. For all LMS, RLS, DP, and HM approach they are not so effective in the background of strong noise compared to the conventional SS technique. This results in that the CPM scheme cannot be applied to the oblivious watermarking, in which the original image cannot be deducted in the detection process so that the SNR is as large as around –30dB. Even though the original image is available, the performance of the CPM watermarking based on initial condition is still not as good as the performance of the conventional SS nonoblivious watermarking except for the processing of rotating. Therefore, the improvement of the estimation performance is very necessary.

CHAPTER 5

CHAOTIC PARAMETER MODULATION WATERMARKING BASED ON MEAN VALUE DETECTION

5.1 Introduction

Even though the chaotic parameter modulation (CPM) method has been proposed to overcome the problems of the conventional SS watermarking technique, the estimation approaches described in Chapter 4 are less robust compared to the correlation detector used in watermarking of binary information sequences. The original image is required in the detection process of the CPM watermarking scheme. Otherwise, adaptive algorithms such as LMS and RLS cannot converge because of the strong noise. Although the DP and HM approaches have better performances, they are still inferior in the oblivious watermarking. We propose a novel watermarking scheme called chaotic parameter modulation based on mean value detection (CPM-MVD), which is more robust to the attacks and image processing in this chapter.

It is shown that for a certain class of chaotic systems, the mean value of a chaotic signal is a monotonic function of bifurcating parameters. Thus, the estimation can be accomplished by obtaining the mean value of the noisy chaotic signal and then employing the monotonic mean value function to find out which parameter is exactly used. Not only does the proposed MVD approach not require the synchronization procedure compared to the conventional SS watermarking technique, but also it can resist much stronger noise compared to those adaptive filtering approaches. It is shown that the proposed new approach can overcome all the problems encountered in the conventional SS and CPM watermarking technique and have superior performances in terms of robustness and payload.

5.2 Watermark Generation Using Chaotic Parameter Modulation

Chaotic parameter modulation (CPM) based on bifurcating parameter is employed here to generate watermarks. Applying the CPM, the information signals are modulated into the bifurcating parameters by setting $\lambda = g(b(i))$ in general. The design of the map, again, is to make sure all the obtained λ will be in a certain regime so that the output signal generated by the map is chaotic and satisfies the requirements of being noise-like and wideband.

The output of the chaotic system, $\mathbf{x} = (x_0, x_1, \cdots, x_n, \cdots, x_{N-1})$, is then used as the watermark signal and added to the original image pixels to provide the watermarked edition denoted as $w_n = v_n + \alpha x_n$. It is seen that the procedure of watermark generation and embedding is exactly the same as described in 4.2.

5.3 Watermark Detection Using Mean Value Detection

5.3.1 Description of mean value detection

Assume $x_n = f(x_{n-1}, \lambda)$ is a chaotic map defined in some closed interval between $[\lambda_{min}, \lambda_{max}]$. Let $\{x_n\}$ be a chaotic signal generated from a certain value of λ. The system from which the bifurcating parameter is estimated can be given by (4.3). That is

$$x_n = \lambda x_{n-1}(1 - x_{n-1})$$
$$y_n = x_n + e_n.$$

According to the Birkhoff ergodic theorem [67], the limit $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n$ exists, and this limit is independent of the initial condition x_0, but only determined by the bifurcating parameter λ. We call this function the mean value function $M(\lambda)$ of a chaotic map f. For many chaotic maps, we have an interesting observation that their mean value functions are in fact monotonic. That is, for each $\lambda \in [\lambda_{min}, \lambda_{max}]$ the chaotic map has a unique mean value $\mu_x = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n$ for the chaotic signal generated by λ.

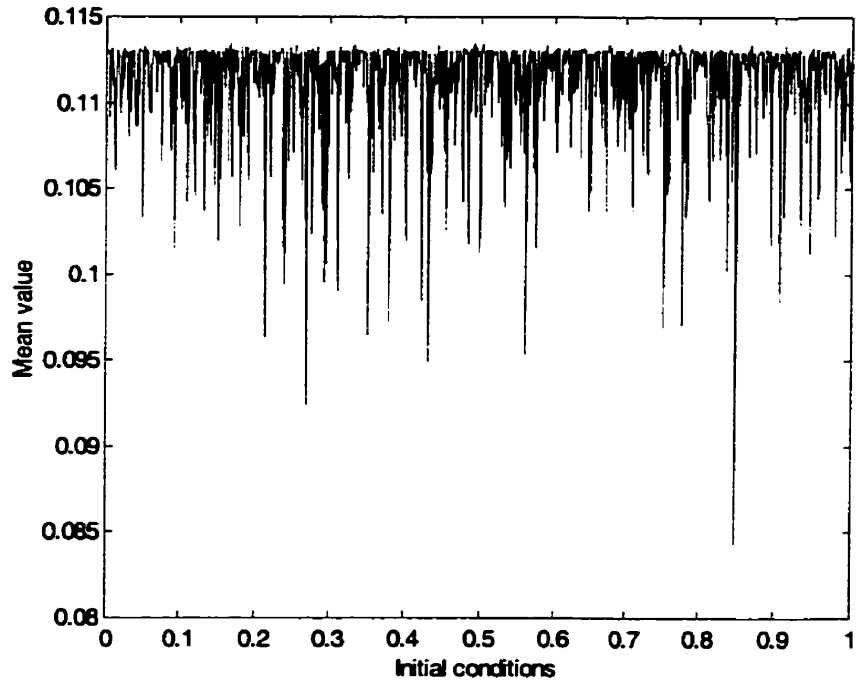

Figure 5.1 The mean value curve of the chaotic signals generated by the Tent map using different initial conditions

To illustrate the above properties of the mean value function, The Tent map

$$f(x,\lambda) = \lambda - 1 - \lambda |x|, \quad x \in [-1,1] \quad \lambda \in (1,2] \tag{5.1}$$

and the Chebyshev map

$$f(x,\lambda) = \cos(\lambda \arccos(x)), \quad x \in [-1,1] \quad \lambda \in (1,2]. \tag{5.2}$$

are used in this study. Figures 5.1 and 5.2 illustrate the relationship between the initial condition and the mean value by using those two chaotic maps. In this experiment, the bifurcating parameter λ is set as a constant, saying 1.3. 1000 initial conditions are selected randomly in the regime of initial conditions [-1,1]. For each initial condition a sequence of chaotic signals is generated, and the mean value is calculated by averaging them. The variance of these 1000 mean values is equal to 903672×10^{-6} for the Tent map, and 1.1078×10^{-5} for the Chebyshev map. The very small value of the variance shows that the mean value of the chaotic signal is very stable for initial conditions. In other words, the mean value is independent of the initial condition once the parameter λ is set.

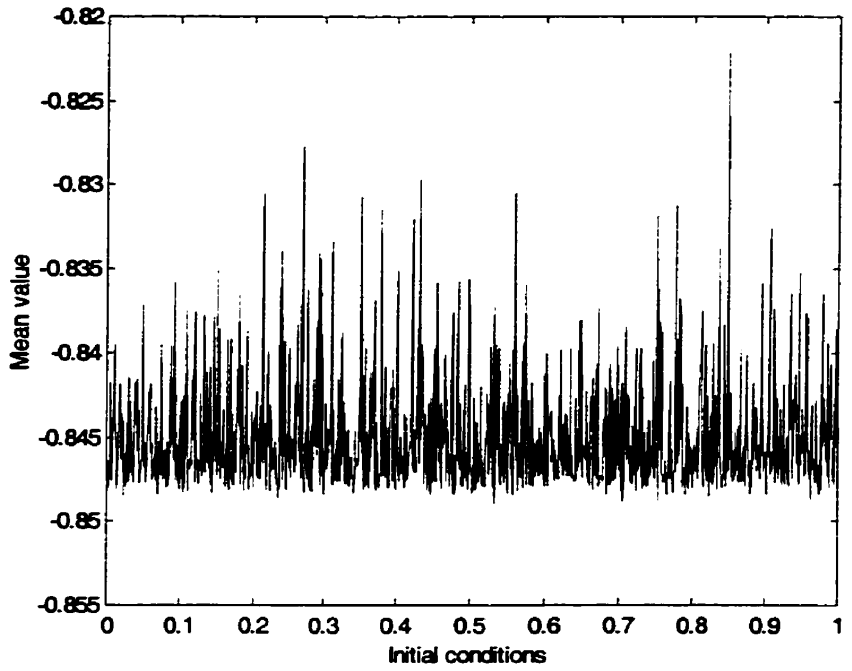

Figure 5.2 The mean value curve of the chaotic signals generated by the Chebyshev map using different initial conditions

The relationship between the mean value and the sequence length is also tested. During the test, the bifurcating parameter and the initial condition are set as the constants. The sequence length varies from 100 to 2000 and the step size is set as 10. The obtained mean values versus the sequence length are plotted in Figures 5.3 and 5.4 for the Tent map and the Chebyshev map respectively. It is seen that an increasing sequence length the mean value approaches a constant. For the Tent map when sequence length $N \geq 300$ the mean value is almost unchanged. And for the Chebyshev map this happens when N is larger than 200. Even though when the sequence length is short, saying $N = 50$, the difference of the mean values deriving from the long sequence length and the short is very small. For the Chebyshev map this difference is around 0.02; for the Tent map, it is only about 0.0025. Therefore, we can say that the longer the sequence is, the more stable the mean value is. In other words when the sequence length is very large, the mean value is close to a constant.

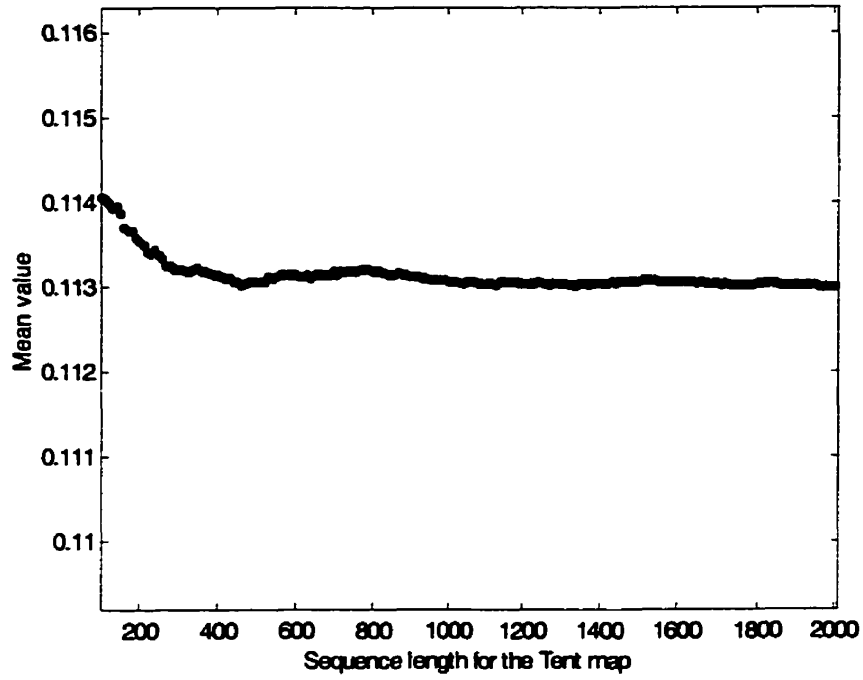

Figure 5.3 The mean value curve of the chaotic signals generated by the Tent map using different sequence lengths

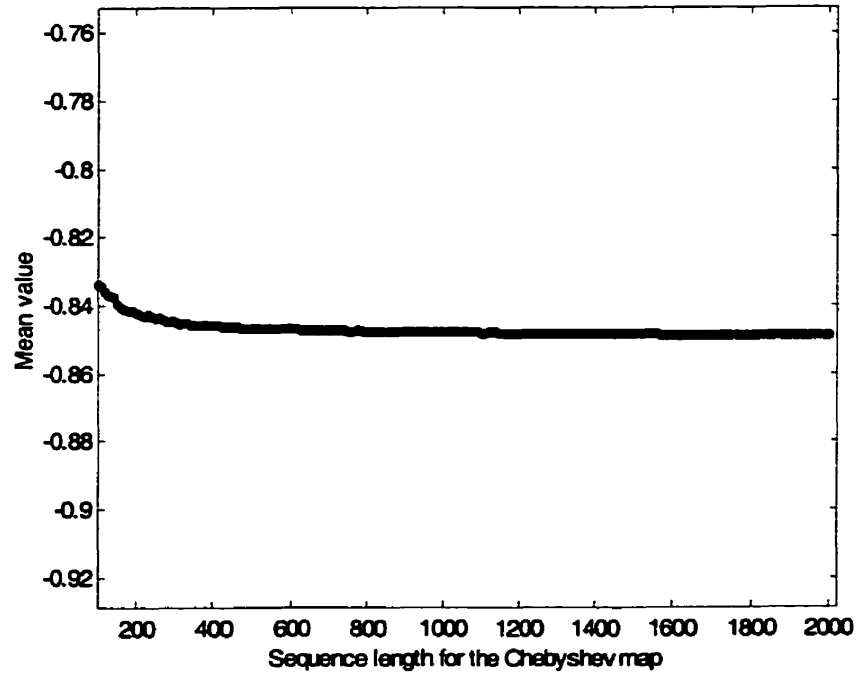

Figure 5.4 The mean value curve of the chaotic signals generated by the Chebyshev map using different sequence lengths

Now the relationship between the mean value and the bifurcating parameter is investigated. In Figures 5.5 and 5.6 the mean value functions of the Tent map and Chebyshev map versus the bifurcating parameter are plotted. The parameters are selected from [1.1, 1.9] for the Tent map, and [1.3, 2] for the Chebyshev map. They are both within their chaotic regime. It is shown that the mean value function $M(\lambda)$ of the Tent map is a monotone function when the λ is inside [1.1, 1.5]; and for the Chebyshev the mean value function $M(\lambda)$ is monotonic in the whole regime [1.3, 2] of the investigation.

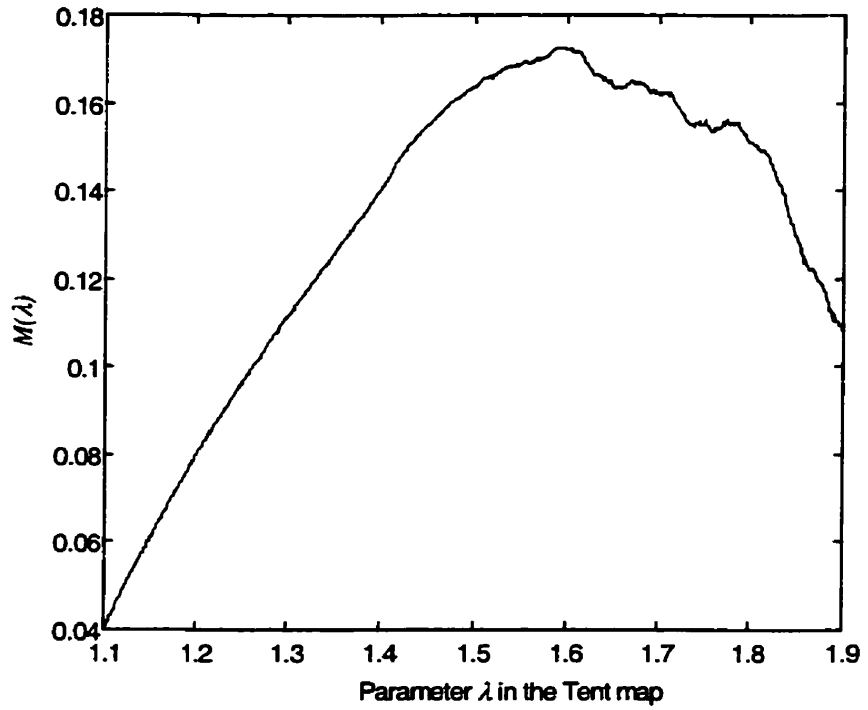

Figure 5.5 The mean value curve of the chaotic signals generated by the Tent map using different bifurcating parameters

Because of this monotone relationship between the mean value and the bifurcating parameter, it is possible to estimate the parameter λ from a chaotic signal by the following procedure. First, we compute the estimated mean value $\hat{\mu}_x$ from the received chaotic signal y_n. Second, we estimate the bifurcating parameter $\hat{\lambda}$ by taking the inverse of the mean value function, i.e., $\hat{\lambda} = M^{-1}(\hat{\mu}_x)$. We call this the mean value detection (MVD) approach.

To avoid deriving the inverse mean value function M^{-1} which may be difficult to obtain analytically, we can get $\hat{\lambda}$ by solving the following optimization problem. Suppose that the mean value function $M(\lambda)$ is continuous and monotone on the interval $[\lambda_{min}, \lambda_{max}]$. If $\hat{\mu}_x$ is given, then λ can be determined by finding the minimum of $J = |M(\lambda) - \hat{\mu}|_x$ for $\lambda \in [\lambda_{\min}, \lambda_{\max}]$.

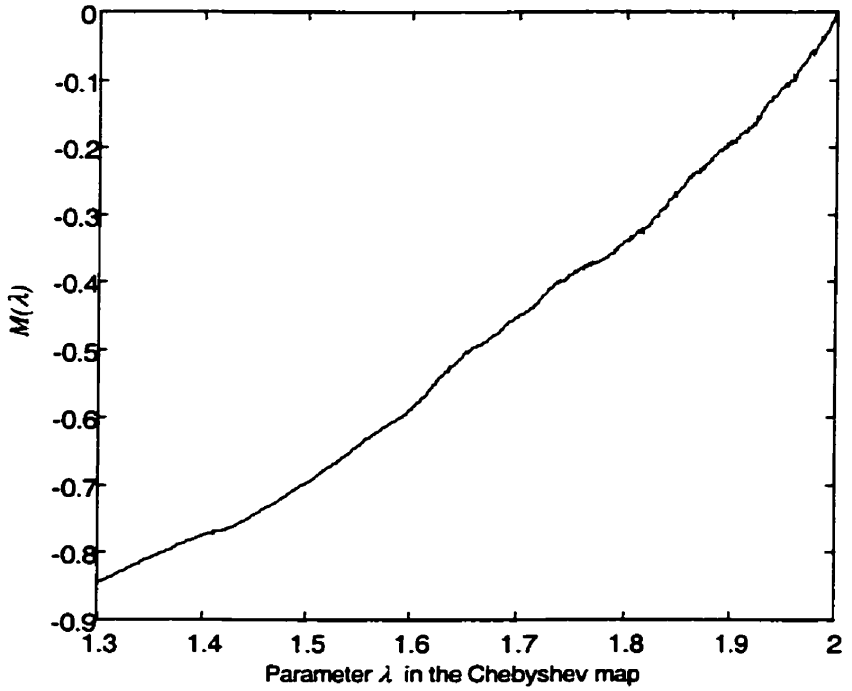

Figure 5.6 The mean value curve of the chaotic signals generated by the Chebyshev map using different bifurcating parameters

5.3.2 Mean value detection applied to image watermarking

Retrieval of the hidden information *b(i)* can be easily accomplished by the MVD approach. Since the location of the watermark signal is known, the watermarked signal w_n' can be extracted first. Due to some unavoidable distortions w_n' may not be equal to the original watermarked one w_n. We can express w_n' as $w_n' = w_n + d_n$. Considering the watermarked pixel w_n as an addition of the watermark and the original image pixel, i.e., $w_n = v_n + \alpha x_n$, we have $w_n' = v_n + \alpha x_n + d_n$. Applying the MVD to the extracted sequence $\{w_n'\}$, we have

$$\mu_{w'} = \frac{1}{N}\sum_{n=0}^{N-1} w_n' = \frac{1}{N}\sum_{n=0}^{N-1}(v_n + \alpha x_n + d_n) = \mu_v + \alpha\mu_x + \mu_d, \qquad (5.3)$$

where $\mu_{w'}$, μ_v, μ_x and μ_d denote the mean values which are the ensemble average of the extracted watermarked pixels w_n', the original pixel v_n, the watermark signal x_n and the distortion d_n respectively. Assuming that the mean of the original pixels is known, we can obtain an estimate of the mean value of the watermark signals as

$$\hat{\mu}_x = \frac{\mu_w' - \mu_v}{\alpha} = \frac{\mu_v + \alpha\mu_x + \mu_d - \mu_v}{\alpha} = \mu_x + \frac{1}{\alpha}\mu_d. \qquad (5.4)$$

If there is no distortion introduced to the watermarked image, we have $d_n = 0$, and hence $\mu_d = 0$. Therefore, we have $\hat{\mu}_x = \mu_x$ and the mean estimation is unbiased. The bifurcating parameter can then be retrieved by $\hat{\lambda} = M^{-1}(\hat{\mu}_x) = M^{-1}(\mu_x)$.

The advantage of the MVD approach is that no matter how strong the distortions are, the correct result can always be obtained if the mean of the distortions is zero. For instance, assume $d_n \neq 0$ in general. However, it is observed that the distortion process has a mean of zero, i.e., $\mu_d = 0$. In this case the estimated mean value of the received signal is $\hat{\mu}_x = \mu_x + \frac{1}{\alpha}\mu_d = \mu_x$, and there will be no bias introduced to the estimated λ. This is quite different from the adaptive filtering approach used in the previous Chapter, in which the noise power plays an important role in the retrieved performance. It is therefore expected that the MVD approach can work efficiently in low SNR environments.

Another advantage of the MVD approach is that it does not require any synchronization. This is a well-accepted property of the CPM method. For the conventional SS technique the watermark detector should know both the correct PN sequence and its possible shift. When the shift is unknown, all possible shifts are experimentally evaluated. The procedure of achieving synchronization is very cumbersome and complex, especially for PN sequences with a very large cycle. However, for the MVD scheme in the application of watermarking, the sequence shifting does not have to be known. The retrieval process only requires summation of all the extracted signals and dividing it by the sequence length N to obtain the estimated mean

value. Compared to the conventional SS watermarking technique and other estimation methods used in the conventional CPM scheme, the MVD approach is very simple and fast to implement. This is very important for watermarking application that requires real-time processing [68].

It should also note that with the MVD approach, the original image is not necessary to be known so that each value of the pixel that the watermark signal is embedded can be extracted. Only the mean value of the original image has to be known in the detection. Therefore, the words "nonoblivious" and "oblivious" do not really apply to this watermarking scheme.

5.4 Application of Chaotic Parameter Modulation Watermarking Based on Mean Value Detection to Binary Information Sequences

No matter the methods of LMS and RLS or DP and HM are used to estimate parameters in the CPM watermarking scheme, the performance of watermarking binary information is not so desirable compared to the conventional SS technique, especially when the sequence length is very long. In order to have an improvement, the MVD is employed to demodulate the bifurcating parameter from the extracted and corrupted chaotic signals.

5.4.1 Discussion of CPM-MVD with the application to binary information sequences

The binary information sequence is composed of only {-1, 1}. Two parameters are selected corresponding to the bits 1 and –1. That is

$$\lambda = g(b(i)) = \begin{cases} \lambda_1, & if \quad b(i) = -1 \\ \lambda_2, & if \quad b(i) = 1. \end{cases} \tag{5.5}$$

The chaotic signal is then generated based on these two bifurcating parameters, i.e., $x_n = \begin{cases} f^n(x_0, \lambda_1), & if \quad \lambda = \lambda_1 \\ f^n(x_0, \lambda_2), & if \quad \lambda = \lambda_2 \end{cases}$. The generated chaotic signal is then added to the image as the watermark.

The detection process therefore does not require estimation of a wide range of parameters but only two values λ_1 and λ_2. To apply the MVD, an obvious necessary

condition is that λ_1 and λ_2 should not have the same mean value, that is, $M(\lambda_1) \neq M(\lambda_2)$. Without loss of generality, we can choose λ_1 and λ_2 such that $M(\lambda_1) < M(\lambda_2)$. To retrieve the information bits *b(i)* we have to determine λ from $\hat{\mu}_x$ given in (5.4). Based on the fact that $M(\lambda)$ is monotonic, λ can be estimated by $\hat{\lambda} \approx M^{-1}(\hat{\mu}_x)$. But since λ only takes two values for binary information sequences, the detection process can in fact be further simplified to a binary decision process. That is,

$$\hat{\lambda} = \begin{cases} \lambda_1, & \text{if} \quad \hat{\mu}_x \text{ is closer to } M(\lambda_1) \\ \lambda_2, & \text{if} \quad \hat{\mu}_x \text{ is closer to } M(\lambda_2). \end{cases} \tag{5.6}$$

More precisely, a threshold of the mean value μ_t is set for making a decision. That is,

$$\hat{\lambda} = \begin{cases} \lambda_1, & \text{if} \quad \hat{\mu}_x < \mu_t \\ \lambda_2, & \text{if} \quad \hat{\mu}_x \geq \mu_t. \end{cases} \tag{5.7}$$

In this study, we choose the threshold μ_t to be the midpoint between $M(\lambda_1)$ and $M(\lambda_2)$. Thus, we have $\mu_t = \dfrac{M(\lambda_1) + M(\lambda_2)}{2}$. Defining the difference between the two mean values as the mean distance of the two correspondent bifurcating parameters, we have the mean distance of the two bifurcating parameters selected to store the information bits 1 and -1 as

$$\Delta M = \left| M(\lambda_1) - M(\lambda_2) \right| \tag{5.8}$$

Since the threshold μ_t is chosen as the midpoint of $M(\lambda_1)$ and $M(\lambda_2)$, the difference between μ_t and $M(\lambda_1)$ or $M(\lambda_2)$ is equal to $\Delta M/2$.

With the thresholding process, the estimate of the watermark mean value is compared to the threshold. The difference between them can be written as

$$\begin{aligned} \Delta\mu &= \hat{\mu}_x - \mu_t \\ &= \mu_x + \frac{1}{\alpha}\mu_d - \mu_t \\ &= M(\lambda) - \mu_t + \frac{1}{\alpha}\mu_d. \end{aligned} \tag{5.9}$$

When the transmitted symbol is "-1", we have $\lambda = \lambda_1$ as given in (5.5). Hence, (5.8) can be written as

$$\Delta\mu = M(\lambda_1) - \mu_t + \frac{1}{\alpha}\mu_d = -\frac{\Delta M}{2} + \frac{1}{\alpha}\mu_d. \tag{5.10}$$

Accordingly, (5.8) is written as

$$\Delta\mu = M(\lambda_2) - \mu_t + \frac{1}{\alpha}\mu_d = \frac{\Delta M}{2} + \frac{1}{\alpha}\mu_d \tag{5.11}$$

for the transmitted symbol "1". It is seen that the sign of $\Delta\mu$ can be used to decide which symbol is transmitted. If there is no distortion introduced or the mean of distortions is zero, we have $\mu_d = 0$, and hence the sign of $\Delta\mu$ is exactly the same as the information bit.

However, in the case of $\mu_d \neq 0$ the value of μ_d may be large enough to change the sign of $M(\lambda) - \mu_t$ so that a wrong detection is provided. A bit error occurs if $-\frac{\Delta M}{2} + \frac{1}{\alpha}\mu_d > 0$ or $\frac{\Delta M}{2} + \frac{1}{\alpha}\mu_d < 0$. That is, $\mu_d < -\frac{\alpha\Delta M}{2}$ or $\mu_d > \frac{\alpha\Delta M}{2}$. Combining them, we can have the probability of detection error as

$$P_e = P(|\mu_d| > \frac{\alpha\Delta M}{2}). \tag{5.12}$$

It should be noted in the application of the MVD approach to image watermarking, μ_d is actually calculated by $\mu_d = \frac{1}{N}\sum_{n=0}^{N-1} d_n$, where N is the sequence length. Thus, μ_d is the estimate of the distortion mean. Assume that the distortion d_n is a random variable which has a normal distribution with the zero mean and variance σ_d^2. The variance of μ_d can be given as

$$\begin{aligned}
\sigma_{\mu_d}^2 &= E\left[\left(\frac{1}{N}\sum_{n=0}^{N-1} d_n\right)^2\right] \\
&= E\left[\frac{1}{N^2}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} d_m d_n\right] \\
&= \frac{1}{N^2}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} E(d_m d_n) \\
&= \frac{1}{N^2} NE[d_n^2] \\
&= \frac{1}{N}\sigma_d^2.
\end{aligned} \tag{5.13}$$

Provided (5.13), (5.12) can be further written as

$$
\begin{aligned}
P_e &= P(|\mu_d| > \frac{\alpha\Delta M}{2}) \\
&= 2P(\mu_d > \frac{\alpha\Delta M}{2}) \\
&= 2\frac{1}{\sqrt{2\pi}\sigma_{\mu_d}} \sum_{\mu_d=\frac{\alpha\Delta M}{2}}^{\infty} \exp(-\frac{\mu_d^2}{2\sigma_{\mu_d}^2}) \\
&= 2\frac{1}{\sqrt{2\pi}\sigma_{\mu_d}} \sqrt{\pi/2}\sigma_{\mu_d} erfc(\frac{\alpha\Delta M}{2\sqrt{2}\sigma_{\mu_d}}) \\
&= erfc(\sqrt{\frac{N}{2}}\frac{\alpha\Delta M}{2\sigma_d}).
\end{aligned}
\tag{5.14}
$$

(5.14) shows that using a large value for α, N, and ΔM will produce a small probability of detection error. It is noted that the tradeoff for a long sequence length is the reduced amount of information that can be inserted into the host data. For the conventional SS watermarking technique, there is no good method to solve this problem. However, for the MVD approach, this problem can be alleviated by using a large mean distance ΔM. That is, we can increase the mean distance and keep the same sequence length unchanged to maintain a desired error probability. By doing that the amount of the inserted information does not have to be sacrificed.

5.4.2 Performance Test

In order to test the performance of the proposed CPM-MVD watermarking scheme, we generate a binary sequence for transmission. The watermarked image is then subjected to a series of image processing and attacks, including image resizing, cropping, rotating, median filtering and JPEG compression. The same processing parameters used in Section 2.4 are applied here again. Since the performance test in Section 4.4.1 have shown that in most cases the conventional SS nonoblivious watermarking has superior performances for resisting image manipulations to the CPM watermarking schemes in watermarking binary information, the comparison is only given between the proposed CPM-MVD approach and the conventional SS nonoblivious watermarking scheme.

The sequence length for both watermarking techniques is set as $N = 1023$. For the conventional SS approach, the m-sequence is employed. For the CPM-MVD technique,

both the Tent map and the Chebyshev map are considered. Two bifurcating parameters $\lambda_1 = 1.3$ and $\lambda_2 = 1.5$ are selected in the modulation. The mean distance of using these two parameters is $\Delta M = |M(\lambda_1) - M(\lambda_2)| = |M(1.3) - M(1.5)|$. For the Tent map this value is equal to 0.10148 while for the Chebyshev map we have $\Delta M = 0.2765$. Thus, the Chebyshev map has a larger mean distance than the Tent map.

The BER curves of using the CPM-MVD and the conventional SS nonoblivious watermarking schemes under various attacks are depicted in Figures 5.7 to 5.11. For the attacks of image resizing, cropping, rotating and JPEG compression, a larger value of the processing parameter leads to the less distorted image. And so the distortion has a smaller value of the variance σ_d^2. However, for the attack of median filtering, the smaller value of the filtering parameter corresponds to the smaller variance of the distortion. Thus, it can be seen that as the increase of the resizing, cropping, rotating and compression parameter, or the decrease of the filtering parameter, in other words, as σ_d^2 gets smaller, we have a reduced BER value. Apparently, it is consistent with the result obtained in (5.14).

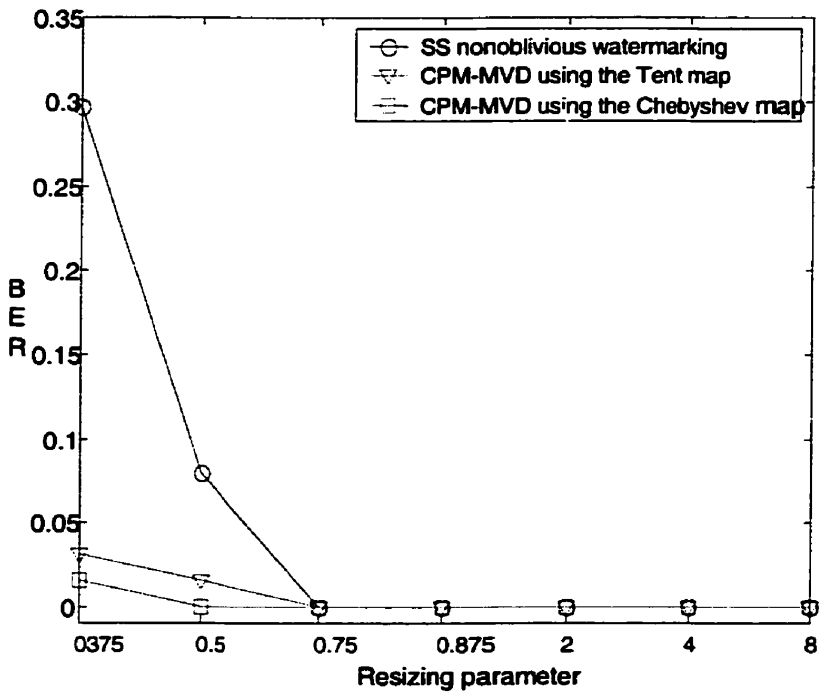

Figure 5.7 The BER curves of using the CPM-MVD and SS watermarking techniques under the resizing attack for different resizing parameters

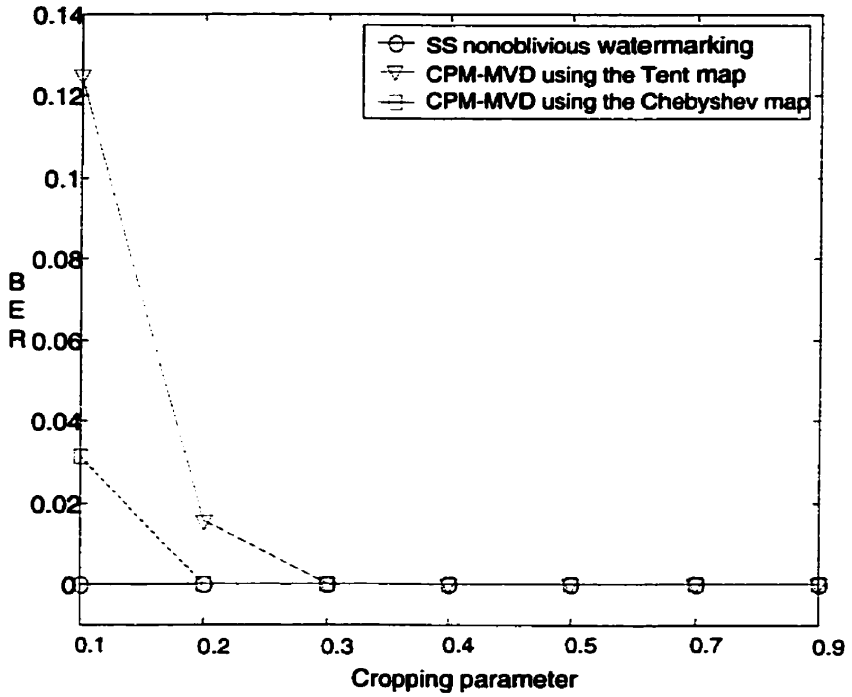

Figure 5.8 The BER curves of using the CPM-MVD and SS watermarking techniques under the cropping attack for different cropping parameters

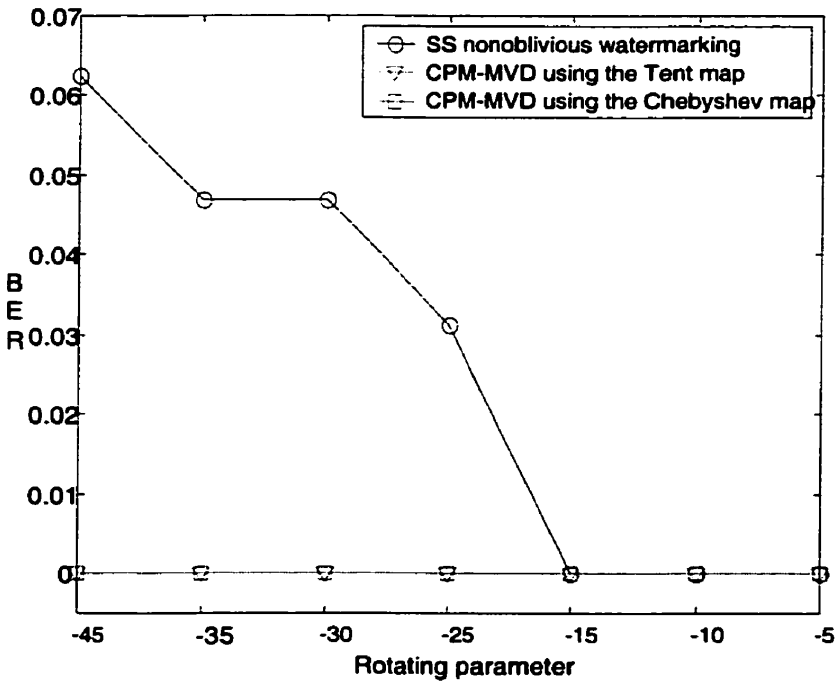

Figure 5.9 The BER curves of using the CPM-MVD and SS watermarking techniques under the rotating attack for different rotating parameters

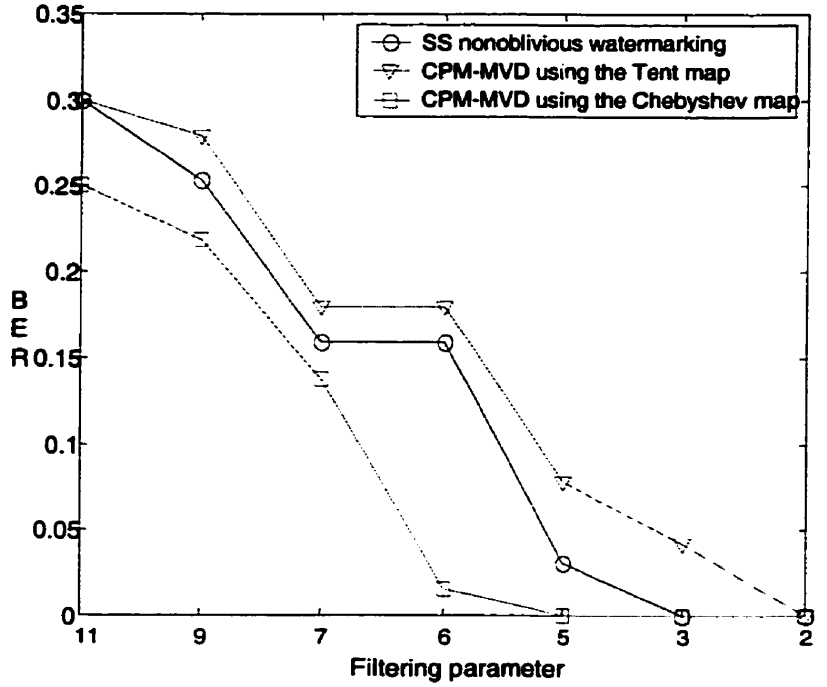

Figure 5.10 The BER curves of using the CPM-MVD and SS watermarking techniques under the median filtering attack for different filtering parameters

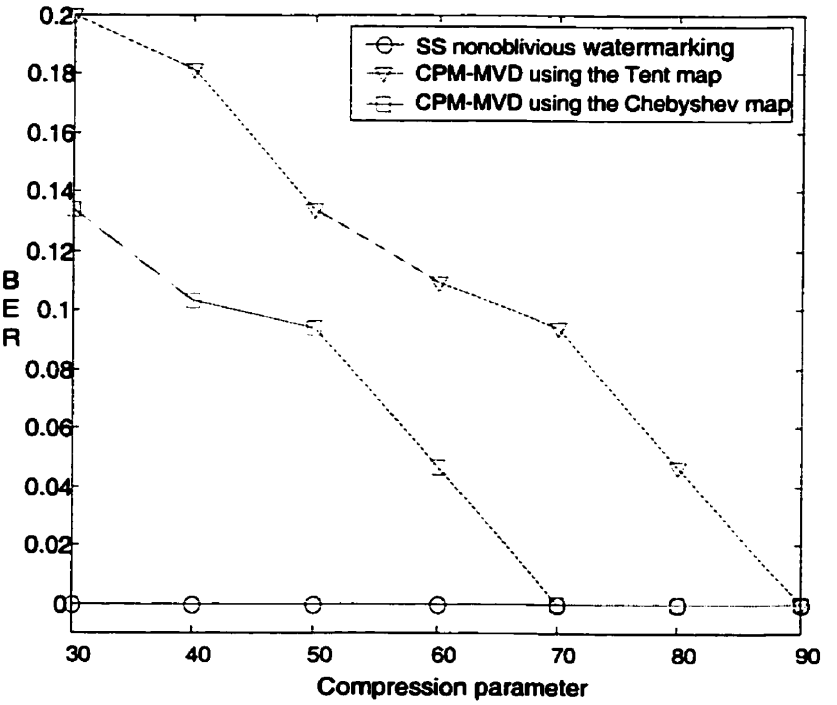

Figure 5.11 The BER curves of using the CPM-MVD and SS watermarking techniques under the JPEG compression for different compression parameters

It is shown that for both the attacks of image resizing and rotating, the BER values of the CPM-MVD approach based on the Tent map and the Chebyshev map are all lower than of the conventional SS nonoblivious watermarking. This implies that the CPM-MVD approach is more robust to these processing. For the attack of median filtering, the performance of the CPM-MVD using the Chebyshev map is better than the SS nonoblivious watermarking while the performance of using the Tent map is the worst.

It is also seen that all the performances of the CPM-MVD based on the Chebyshev map are better than those of the CPM-MVD based on the Tent map. This is because, although the same couple of bifurcating parameters are chosen, the use of the Chebyshev map provides a larger mean distance than the use of the Tent map. In the test, the mean distance for the Tent map is $\Delta M = 0.10148$ while the mean distance for the Chebyshev map has a value as large as $\Delta M = 0.27650$. According to (5.14), the larger the mean distance is, the less the probability of detection. Thus, the results of the performance test are consistent with the expression (5.14).

However, for the attacks of image cropping and compression the SS nonoblivious watermarking still has the best performances, showing the conventional SS nonoblivious watermarking is really robust to these two kinds of processing.

5.4.3 Improved CPM-MVD with the application to binary information sequences

According to (5.14), the probability of detection error can be minimized when the mean distance is as large as possible. An improved CPM-MVD scheme is so proposed to increase the mean distance and enhance the performance.

Instead of using two bifurcating parameters of a chaotic system to provide the mean distance $\Delta M = |M(\lambda_1) - M(\lambda_2)|$, we propose use only one parameter in the following way to represent the symbol "-1" and "1",

$$x_n = \begin{cases} -f^n(x_0, \lambda), & if \quad b(i) = -1 \\ f^n(x_0, \lambda), & if \quad b(i) = 1. \end{cases} \tag{5.15}$$

Thus, the mean value of the chaotic signal is $M(\lambda)$ for the symbol "1", and $-M(\lambda)$ for the symbol "-1". The mean distance is then equal to $2M(\lambda)$. For the Tent map, if two parameters have to be selected to do the modulation, the maximum mean distance can be achieved is $|M(1.1) - M(1.5)| = 0.1196$ since the interval [1.1, 1.5] is the regime for the

mean value function $M(\lambda)$ being monotonic. However, for the improved CPM-MVD the maximum mean distance can be obtained as $2\times M(1.5)=0.3290$. It is seen that the improved CPM-MVD has a much larger range for the mean distance than the conventional CPM-MVD. When the Chebyshev map is used for the improved CPM-MVD, the range of the mean distance is also greatly enlarged.

According to the modulation process, the detection of information bits employs the sign of the estimated mean to decide which symbol is transmitted. That is,

$$b(i)=\begin{cases}-1 & if \quad \hat{\mu}_x<0\\ 1 & if \quad \hat{\mu}_x\geq 0,\end{cases} \tag{5.16}$$

where $\hat{\mu}_x$ is obtained by (5.4).

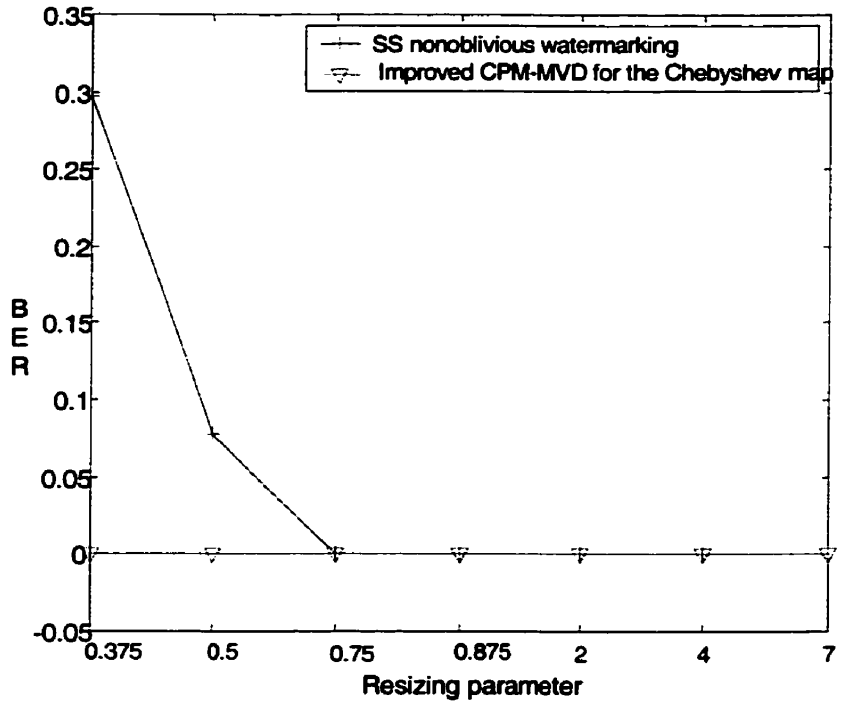

Figure 5.12 The comparison of the BER performance between the SS nonoblivious and improved CPM-MVD watermarking schemes under the resizing attack for different resizing parameters

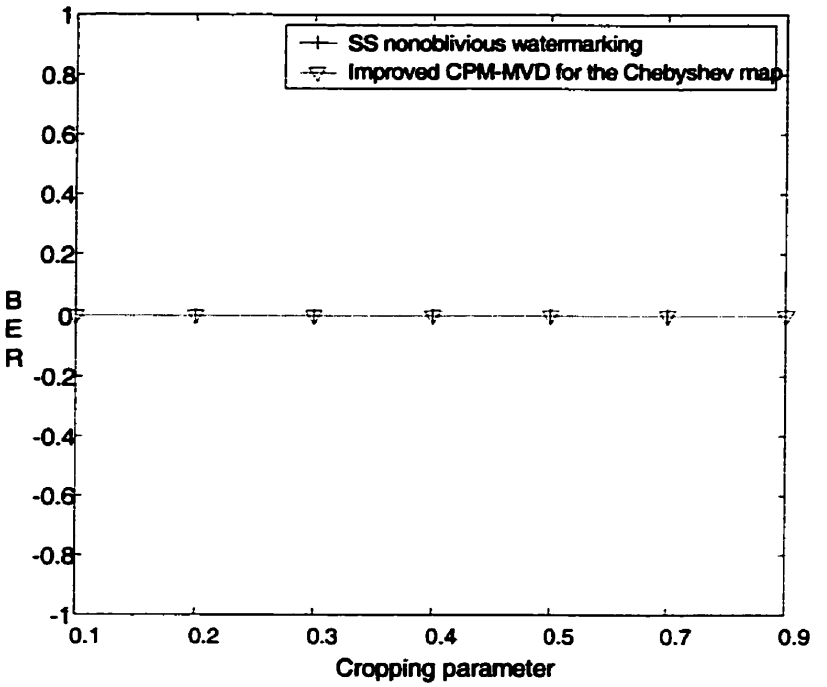

Figure 5.13 The comparison of the BER performance between the SS nonoblivious and improved CPM-MVD watermarking schemes under the cropping attack for different cropping parameters

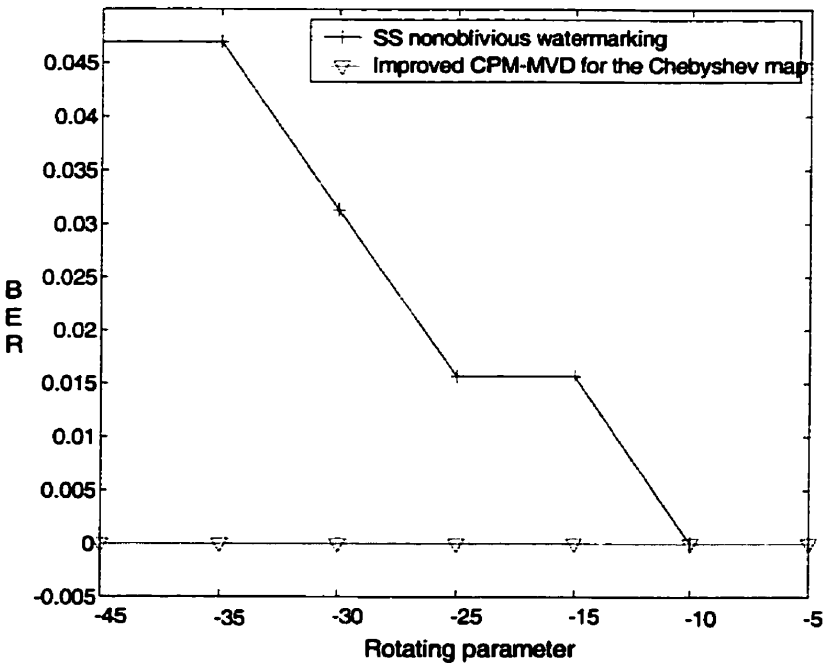

Figure 5.14 The comparison of the BER performance between the SS nonoblivious and improved CPM-MVD watermarking schemes under the rotating attack for different rotating parameters

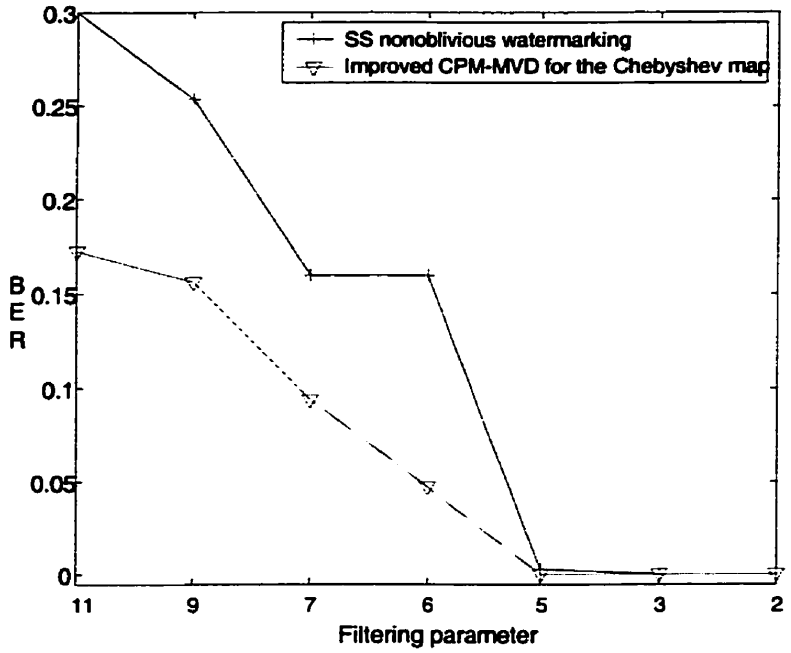

Figure 5.15 The comparison of the BER performance between the SS nonoblivious and improved CPM-MVD watermarking schemes under the median filtering attack for different filtering parameters

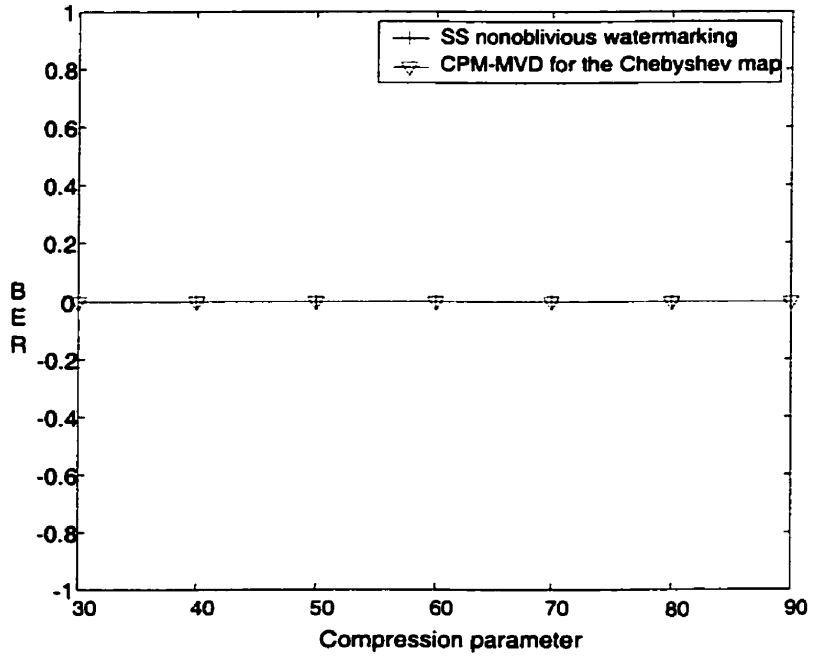

Figure 5.16 The comparison of the BER performance between the SS nonoblivious and improved CPM-MVD watermarking schemes under the JPEG compression for different compression parameters

The BER performances of the improved CPM-MVD watermarking scheme are illustrated in Figures 5.12 to 5.16. For the aim of comparison, the BER curves of the conventional SS nonoblivious watermarking technique are also plotted. According to the results in the previous chapters, the conventional SS nonoblivious watermarking is the most robust scheme for binary information sequences. However, It is shown here that for the attacks of image resizing, rotating, median filtering, the improved CPM-MVD has smaller BER values than the SS nonoblivious watermarking.

For image cropping and JPEG compression the improved CPM-MVD scheme also has zero detection errors as the SS nonoblivious watermarking technique. Thus, to have a better understanding of the two watermarking techniques under the cropping and compression attacks, a further comparison is shown in Table 5.1. In the previous comparison, the smallest value of the cropping parameter is equal to 0.1, and presently it is further reduced to 0.05. In other words only 0.05 times of its original image is left after cropping. Table 5.1 shows that the improved CPM-MVD is more robust than the conventional SS technique. For the compression attack, the compression parameter was set as 30 in the previous study. Now it is reduced to 10, and the digital image is compressed to 10 percent of its original size. This comparison shows no effect in the improved CPM-MVD watermarking scheme but a general error for the conventional SS watermarking technique.

	BER of SS nonoblivious watermarking	BER of improved CPM-MVD watermarking
Cropping to 0.05 times of its original	0.0781	0.0156
JPEG compressed to 10% of its original	0.0469	0

Table 5.1 A further comparison of BER between the SS nonoblivious and improved CPM-MVD watermarking schemes

5.5 Application of Chaotic Parameter Modulation Watermarking Based on Mean Value Detection to Numerical Information Sequences

For real-life watermarking applications, copyright information is usually not in a binary format and hence it is of interest to investigate the proposed watermarking technique in modulating numerical copyright information. Let $(b(1),b(2),\cdots,b(i),\cdots,b(L-1))$ be the sequence of numerical information. For the CPM-MVD watermarking, each element of the information sequence is modulated into the bifurcating parameter of a chaotic system. Since the bifurcating parameter should be controlled inside a certain regime, $\lambda_i = g(b(i))$ is to be designed to guarantee that λ_i is always within the chaotic regime. Given the parameter the corresponding chaotic signals can be generated as $x_n = f^n(x_0,\lambda_i)$, $n\in(0,1,\cdots N-1)$. It is seen that since the CPM is originally designed for analog spread spectrum communications, the modulation of numerical information is very straightforward. The MVD is applied in the detection process and the estimated mean value of the corrupted watermark signal is obtained as (5.4). Provided $\hat{\mu}_x$ we estimate the bifurcating parameter as $\hat{\lambda}_i = M^{-1}(\hat{\mu}_x)$. Since it is very difficult to obtain the analytical expression for the mean value function $M(\lambda)$,we can obtain $\hat{\lambda}_i$ by searching $\hat{\lambda}_i$ in the regime to minimize $J = |M(\lambda)-\hat{\mu}_x|$. After that the information $b(i)$ can be demodulated as $\hat{b}(i) = g^{-1}(\hat{\lambda}_i)$.

To evaluate the robustness performance of these three schemes, the image, "Camera" shown in Figure 4.9 is used again as the copyright information which is to be inserted into the host image "Lena". Since the size of the image "Camera" is 64×64 and the size of "Lena" is 256×256, the sequence length is set as $\frac{256\times256}{64\times64}=16$. For the aim of embedding, the two dimensional image "Camera" is first transformed to a one-dimensional signal by scanning from left to right and from top to bottom. Since the pixel value varies between 0 and 255, we have an information sequence denoted as $\{b(i)\}$, and $b(i)\in[0,255]$, $i=0,1,\cdots,64\times64-1$. The Chebyshev map is used in this study.

Accordingly, $\lambda_i = g(b(i)) = \left\{1.3 + \frac{b(i)}{255}(0.7)\right\}$ is designed to make sure all the bifurcating parameters λ_i are inside the interval [1.3, 2]. For these parameters, the mean value function $M(\lambda)$ of the Chebyshev map is being monotonic. The watermark signal is then generated as

$$x_n = f(x_{n-1}, \lambda_i) = \cos\left[\left(1.3 + \frac{b(i)}{255}(0.7)\right)\arccos(x_{n-1})\right] \qquad (5.17)$$

for all $n = 0,1,\cdots 15$. The mean square error (MSE) defined in (4.10) is used as a measure to evaluate the robustness performance. The lower value of MSE indicates the more robustness of the watermarking scheme.

We also consider the conventional SS and CPM watermarking schemes for comparison. The MSE curves of different watermarking techniques under the five attacks are shown in Figures 5.17 to 5.21. It is observed that the CPM-MVD always has the best robustness performance. The CPM with DP and CPM with HM are more robust than the SS nonoblivious watermarking. However, their performance is not as good as the CPM-MVD's. This is because the SS technique is developed for digital communications. Thus, when the information sequence is of numerical value, the performance of it degrades greatly even though the original image is available in the detection process. But for CPM watermarking, the modulation and demodulation of numerical information sequence is direct. However, DP and HM used in then conventional CPM watermarking cannot work efficiently in the low SNR environments. This explains why the conventional CPM watermarking is only of the nonoblivious scheme. The MVD approach is much more robust to strong noises compared to the DP and HM. Even though the original image is unknown, it still can work properly provided the mean of the original image.

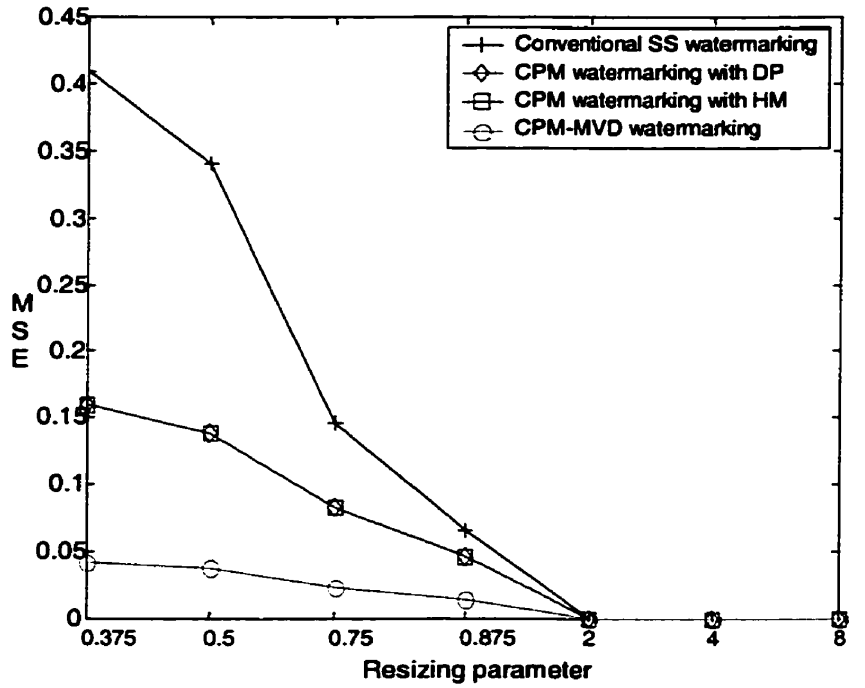

Figure 5.17 The comparison of the MSE performance by using different watermarking schemes under the resizing attack for different resizing parameters

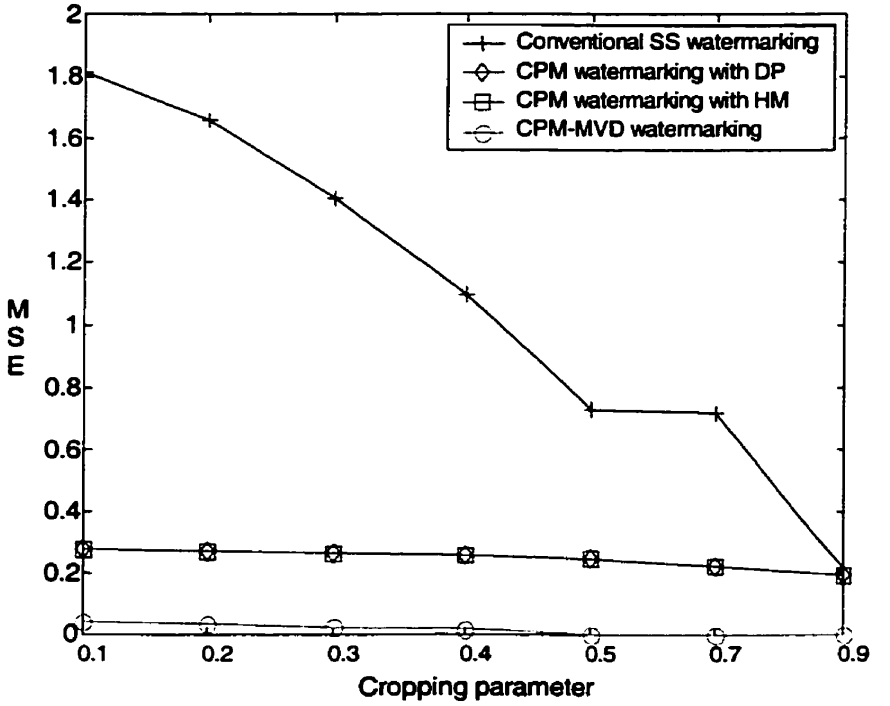

Figure 5.18 The comparison of the MSE performance by using different watermarking schemes under the cropping attack for different cropping parameters

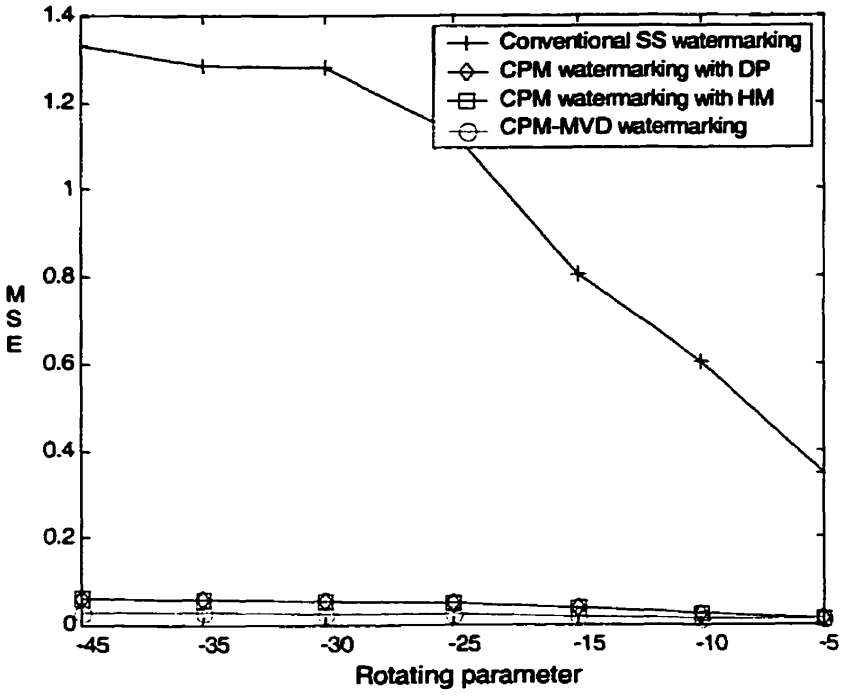

Figure 5.19 The comparison of the MSE performance by using different watermarking schemes under the rotating attack for different rotating parameters

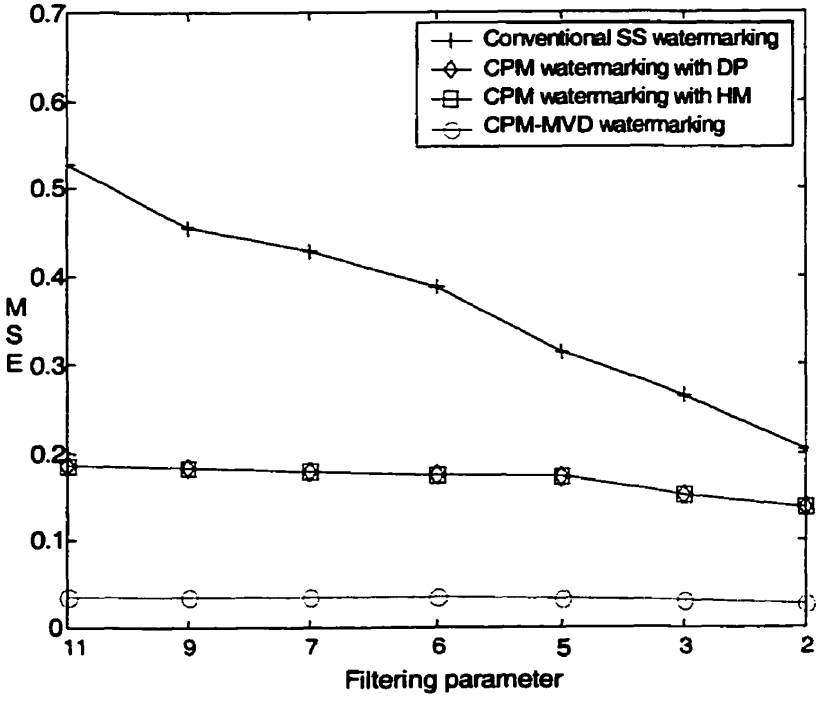

Figure 5.20 The comparison of the MSE performance by using different watermarking schemes under the median filtering attack for different filtering parameters

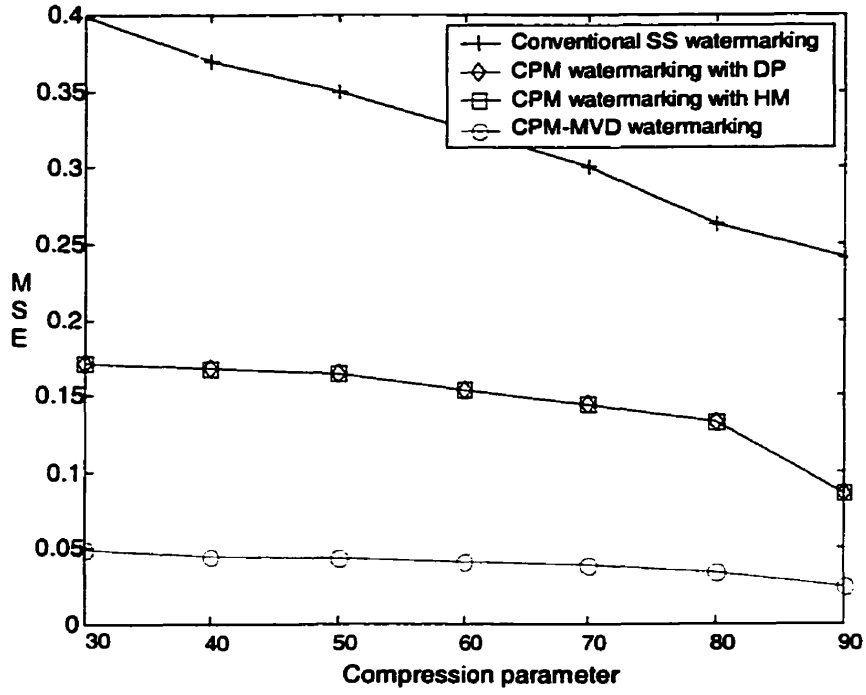

Figure 5.21 The comparison of the MSE performance by using different watermarking schemes under the JPEG compression for different compression parameters

5.6 Summary

We propose a novel watermarking scheme for digital images by using the CPM-MVD scheme. The information signal is modulated into the bifurcating parameter of a chaotic dynamical system. The generated chaotic signal is then used as a watermark signal for embedding into a host image. Retrieval of the copyright information is formulated as a problem of parameter detection from a noisy chaotic signal. Once we get the estimated mean value from the extracted and possibly corrupted watermark signal, the parameter can be easily obtained by making use of the inverse of the mean value function.

Two kinds of information sequences, binary and numerical, are considered in our analysis. For a binary symbol, the mean distance plays an important role in the detection process. The larger the mean distance is, the less the detection error. An improved CPM-MVD scheme is then proposed to widen the mean distance. Meanwhile, the modulation procedure is simplified as well. Only one controlled bifurcating parameter is required in

the modulation. It is shown that the improved CPM-MVD scheme can work efficiently in a very noisy environment and has a superior robustness performance than other watermarking schemes.

The CPM-MVD watermarking scheme also has an advantage that it does not require the spreading code synchronization. This can greatly enhance the speed of the whole watermarking procedure and also reduce the potential synchronization errors as well.

For numerical copyright information, it is a natural choice to use the CPM-MVD scheme as it is originally designed for analog spread spectrum applications. Hence, there is no need to have a coding process. Furthermore, it can resist much stronger noise than the conventional CPM-DP and CPM-HM schemes. It is shown that the performance of the CPM-MVD scheme is much better than the CPM with DP or HM and the SS technique under various attacks and processing.

CHAPTER 6

CONCLUSIONS

Digital watermarking is an enabling technology to prove ownership on copyrighted material, detect originators of illegally making copies, and to monitor the usage of the copyrighted multimedia data. It is required to be robust, imperceptible, secure, and have the payload as much as possible. Even though the spread spectrum (SS) watermarking is the most popular technique nowadays, it is not a direct conclusion for watermarking. There are some problems that the SS technique might encounter in the watermarking applications. First, although the classical spreading sequences, such as *m*-sequences and Gold sequences, have good auto- and cross-correlation properties, their linearity limits their security. Second, the SS technique has a tradeoff existing between robustness and payload. Third, since it is developed for digital communications, its performance degrades greatly when the inserted information is of numerical value. Fourth, synchronization of the spreading sequence is required in the detection process. Finding the correct shift of the correct spreading sequence is very cumbersome and complex especially for those with a very large cycle.

To overcome these problems we propose using chaotic spreading sequences instead of the classical spreading sequences. The chaotic sequences are distinguished by their wideband, noise-like, nonlinear characteristics, simple generation and storage. It is also proven that in watermarking applications they can achieve better auto-correlation and similar cross-correlation performances compared to the classical ones. However, the improvement of the performance is not very significant.

Next, we propose applying chaotic parameter modulation (CPM) to watermarking. The watermark is generated by the CPM method. The copyright information is embedded in the parameter of a chaotic system to generate a sequence of chaotic signals as watermarks. Retrieval of copyright information is then a problem of parameter estimation from the corrupted watermark signal. For the CPM based on

bifurcating parameter, adaptive filters are employed to estimate the parameters from the corrupted watermarked image. For the CPM based on initial condition, two efficient estimation techniques, namely, the dynamical programming (DP) and halving method (HM) are provided for information retrieval. It is shown that when the sequence length is short, the performance of the CPM based on initial condition is better than that of the SS nonoblivious watermarking. It is also noted that when synchronization is not achieved properly, for instance, after the attack of image rotating, the performance of the CPM is better than that of the conventional SS technique. The advantages of using CPM approach are that it does not require any synchronization in the detection process, and its performance does not quite depend on the sequence length. In addition, since the CPM approach is designed for analog SS, it can modulate numerical information directly without converting it to binary symbols. It is shown that the CPM watermarking technique usually has a much large payload than the conventional SS method. However, the CPM watermarking approach is relatively less robust to some distortions, compared to the conventional SS technique. Even when the CPM based on initial condition approach is used, these problems can only be solved at high SNR. Therefore, without the original image, the CPM approach might not work effectively.

To make the CPM technique applicable to both oblivious and nonoblivious watermarking, a novel mean value detection (MVD) approach is proposed. The watermark is generated by the CPM based on bifurcating parameter. Since for some certain chaotic systems there is a monotone relationship between the bifurcating parameter and the mean value of the chaotic signal, the retrieval of the inserted information can be accomplished by calculating the mean value of the possibly corrupted watermark signal and then using the monotone relationship to detect the corresponding bifurcating parameter. It is shown that the proposed CPM-MVD watermarking scheme does not only have all the advantages of the CPM approach mentioned above, but it can also work well in the presence of strong noise. Even though the original image is not accessible, the performance of it is still superior to the conventional SS watermarking techniques for all kinds of attacks. Therefore, the CPM-MVD approach is the best among all the watermarking techniques proposed in the thesis in terms of robustness, payload and security.

REFERENCES

[1] P. Samuelson, "Legally speaking: Digital media and the law," *Commun. ACM*, vol. 34, pp. 23-28, Oct. 1991.

[2] J. Taylor, *DVD Demistified: The Guidebook for DVD-Video and DVD-ROM*. NY: McGraw-Hill, 1997.

[3] S. Rupley, "What's holding up DVD," *PC Mag.*, vol. 15, pp. 34, Nov. 1996.

[4] J.L. Renaud, "PC industry could delay DVD," *Advanced Television Markets*, issue 47, May 1996.

[5] U.S. Copyright Office Summary, "The Digital Millennium Copyright Act of 1998," Dec. 1998. Available: http://lcweb.loc.gov/copyright/legislation/dmca.pdf.

[6] Commission of the European Communities, "Amended proposal for a European parliament and council directive on the harmonization of certain aspects of copyright and related rights in the information society". Available: europa.eu.int/comm/internal_market/en/intprop intprop/copy2en.pdf.

[7] I.J. Cox and M.L. Miller, "A review of watermarking and the importance of perceptual modeling," *Proc. SPIE Electronic Imaging '97, Storage and Retrieval for Image and Video Databases V*, San Jose, CA, Feb. 1997.

[8] G.C. Langelaar, "Conditional access to television service," *Wireless Communication, the Interactive Multimedia CD-ROM*, 3rd ed., Amsterdam, The Netherlands: Baltzer Science, 1999.

[9] G.C. Langelaar, R.L. Lagendijk, and J. Biemond, "Real-time labeling of MPEG-2 compressed video," *J. Visual Commun. Image Representation*, vol. 9, pp. 256-270, Dec. 1998.

[10] R.J. Anderson and F.A.P. Petitcolas, "On the limits of steganography," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 474-481, May 1998.

[11] T. Kalker, G. Depovere, J. Haitsma, and M. Maes, "A video watermarking system for broadcast monitoring," *Proc. SPIE Electronic Imaging '99: Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 103-112, San Jose, CA, Jan. 1999.

[12] R.B. Wolfgang and E.J. Delp, "Fragile watermarking using the VW2D watermark," *Proc. SPIE Electronic Imaging '99: Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 204-213, San Jose, CA, Jan. 1999.

[13] M. Kutter and F. Petitcolas, "A fair benchmark for image watermarking systems," *Proc. SPIE Electronic Imaging '99: Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 226-239, San Jose, CA, Jan. 1999.

[14] J.L. Massey, "Contemporary cryptology: An introduction," *Contemporary Cryptology: The Science of Information Integrity*, pp. 3-39, G. J. Simmons, Ed. NY: IEEE Press, 1992.

[15] K. Tanaka, Y. Nakamura, and K. Matsui, "Embedding secret information into a dithered multilevel image," *Proc. IEEE Military Commun. Conf.*, pp. 216-220, Sept. 1990.

[16] G. Caronni, "Ermitteln unauthorisierter Verteiler von maschinenlesbaren Daten," ETH, Zurich, Switzerland, Tech. Rep., Aug. 1993.

[17] A. Tirkel, G. Rankin, R. van Schyndel, W. Ho, N. Mee, and C. Osborne, "Electronic water mark," *Proc. Digital Image Computing-Techniques and Applications (DICTA '97)*, pp. 666-672, Dec. 1993.

[18] H.D. Luke, *Korrelationssignale* (in German), Berlin, Germany: Springer, 1992.

[19] R.G. van Schyndel, A.Z. Tirkel, and C.F. Osborne, "A digital watermark," *Proc. Int. Conf. on Image Processing (ICIP '94)*, vol. 2, pp. 86-89, Austin, TX, Nov. 1994.

[20] A. Tirkel, R. van Schyndel, and C. Osborne, "A two-dimensional watermark," *Proc. Digital Image Computing-Techniques and Applications (DICTA '97)*, Dec. 1993.

[21] D. Benham, N. Memon, B.L. Yeo, and M. Yeung, "Fast watermarking of DCT-based compressed images," *Proc. Int. Conf. Image Science, Systems, and Technology (CISST '97)*, pp. 243-253, Las Vegas, NV, June 1997.

[22] M. Swanson, B. Zhu, and A. Tewfik, "Multiresolution video watermarking using perceptual models and scene segmentation," *Proc. IEEE Int. Conf. on Image Processing (ICIP '97)*, vol. 2, pp. 558-561, Santa Barbara, CA, Oct. 1997.

[23] R.B. Wolfgang, C.I. Podilchuk, and E.J. Delp, "Perceptual watermarks for digital images and video," *Proc. IEEE*, vol. 87, pp. 1108-1126, July 1999.

[24] V. Darmstaedter, J.F. Delaigle, D. Nicholson, and B. Macq, "A block based watermarking technique for MPEG-2 signals: Optimization and validation on real digital TV distribution links," *Lecture Notes in Computer Science*, vol. 1425, pp. 190-206, 1998.

[25] M. Kutter, F. Jordan and F. Bossen, "Digital watermarking of color images using amplitude modulation," *J. Electron. Imaging*, vol. 7, pp. 326-332, Apr. 1998.

[26] C. Langelaar, J.C.A. ven der Lubbe and R.L. Lagendijk, "Robust labeling methods for copy protection of images," *Proc. Electronic Imaging*, vol. 3022, pp. 298-309, San Jose, CA, Feb. 1997.

[27] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A DCT-domain system for robust image watermarking," *Signal Processing*, vol. 66, pp. 357-372, May 1998.

[28] J.I.K. Oruanaidh, W.J. Dowling, and F.M. Boland, "Phase watermarking of digital images," *Proc. Int. Conf. on Image Processing (ICIP '96)*, vol. 3, pp. 239-242, Sept. 1996.

[29] I.J. Cox, J. Killian, T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for Multimedia," *IEEE Trans. on Image Processing*, vol. 6, pp. 1673-1687, Dec. 1997.

[30] M. Barni, F. Bartolini, V. Cappellini, A. Piva, and F. Rigacci, "A M.A.P. identification criterion for DCT-based watermarking," *Proc. Europ. Signal Processing Conf. (EUSIPCO '98)*, Rhodes, Greece, Sept. 1998.

[31] F. Hartung and B. Girod, "Digital watermarking of raw and compressed video," *Proc. SPIE Digital Compression Technologies and Systems for Video Commun.*, vol. 2952, pp. 205-213, Oct. 1996.

[32] E. Koch and J. Zhao, "Copyright labeling of digitized image data," *IEEE Commun. Mag.*, vol. 36, pp. 94-101, Mar. 1998.

[33] D. Kundur and D. Hatzinakos, "A robust digital image watermarking method using wavelet-based fusion," *Proc. IEEE Int. Conf. Image Processing (ICIP '97)*, vol. 1, pp. 544-547, Santa Barbara, CA, Oct. 1997.

[34] H. Wang and C.C.J. Kuo, "An integrated progressive image coding and watermark system," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP '98)*, vol. 6, pp. 3721-3723, Seattle, WA, May 1998.

[35] J.K. Oruanaidh, T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Processing*, vol. 66, pp. 303-317, Mar. 1998.

[36] J.R. Hernández, F. Pérez-González, J.M. Rodriguez, and G. Nieto, "Performance analysis of a 2-D-multipulse amplitude modulation scheme for data hiding and watermarking of still images," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 510-524, May 1998.

[37] P.H.W. Wong, O.C. Au, J.W.C. Wong, "Image watermarking using spread spectrum technique in log-2-spatio domain," *IEEE Int. Symposium on Circuits and Systems*, pp. 224-227, Geneva, Switzerland, May 2000.

[38] F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn, "Attacks on copyright marking systems," *Information Hiding, Second International Workshop (IH '98)*, D. Aucsmith, Ed., pp. 219-239, Springer, Berlin, 1999.

[39] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, 1992.

[40] G. Proakis, *Digital Communications*, 3rd edition, NY: McGraw-Hill, 1995.

[41] H. Saigui, Z. Yong, H. Jandong, and B. Liu, "A synchronous CDMA system using discrete coupled-chaotic sequence," *Proc. IEEE South-easecon, Bringing Together Education, Science and Technology*, vol. 42, pp. 1524-1527, 1994.

[42] G. Heidari-Bateni, C.C. McGillem, "A chaotic direct-sequence spread spectrum communication system," *IEEE Trans. Commun.*, vol. 42, pp. 1524-1527, 1994.

[43] U. Fildmann, M. Hasler and W. Schwarz, "Communication by chaotic signals: The inverse system approach," *Int. J. Circuit Theory and Applications*, vol. 24, pp. 551-579, May 1996.

[44] D. Ruelle, *Chaotic Evolution and Strange Attractors*. Cambridge, England: Cambridge Univ. Press 1989.

[45] G. Heidari-Bateni, C.D. McGillem, "Chaotic sequences for spread spectrum: An alternative to PN-sequences," *Proc. of 1992 IEEE Int. Conf. on Selected Topics in Wireless Communications*, pp. 437-440, Vancouver, B.C., Canada, June 1992.

[46] M.B. Pursley, "Performance evaluation for phase-coded spread-spectrum multiple-access communication—Part I: System analysis," *IEEE Trans. Commun.*, vol. 25, pp. 795-799, Aug. 1977.

[47] J.S. Lehnert and M.B. Pursley, "Error probabilities for binary direct-sequence spread-spectrum communication with random signature sequences," *IEEE Trans. Commun.*, vol. COM-35, pp. 87-98, 1987.

[48] G. Mazzini, G. Setti, and R. Rovatti, "Chaotic complex spreading sequences for asynchronous DS-CDMA—Part I: System modeling and results," *IEEE Trans. on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 44, pp. 937-947, Oct. 1997.

[49] G. Mazzini, R. Rovatti and G. Setti, "Interference minimization by autocorrelation shaping in asynchronous DS-CDMA systems: Chaos-based spreading is nearly optimal," *Electron. Lett.*, vol. 35, No. 13, pp. 1054-1055, June 1999.

[50] R.E. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1984.

[51] S.W. Golomb, *Shift Register Sequences*, San Francisco, CA: Holden-Day, 1967.

[52] R. Gold, "Optimal binary sequences for spread spectrum multiplexing," *IEEE Trans. Inform. Theory*, vol. 13, pp. 619-621, 1967.

[53] M.K. Simon, J.K. Omura, R.A. Scholtz, and B.K. Levitt, *Spread Spectrum Communications Handbook*. NY: McGraw-Hill, 1994.

[54] R. Rovatti and G. Mazzini, "Interference in DS-CDMA systems with exponentially vanishing autocorrelation: Chaos-based spreading is optimal," *Electron. Lett.*, vol. 34, No. 20, pp. 1911-1913, Oct. 1998.

[55] R. Rovatti, G. Setti, and G. Mazzini, "Chaotic complex spreading sequences for asynchronous DS-CDMA—Part II: Some theoretical performance bounds," *IEEE Trans. CASI-45*, pp. 496-506, 1998.

[56] H. Leung and J. Lam, "Design of demodulator for the chaotic modulation communication system," *IEEE Trans. on Circuits and Systems—I: fundamental Theory and Applications*, vol. 44, pp. 262-267, Mar. 1997.

[57] Z. Zhu and H. Leung, "Adaptive identification of nonlinear systems with application to chaotic communications," *IEEE Trans. on Circuits and Systems—I*, vol. 47, pp. 1072-1080, Jul. 2000.

[58] R.L. Devaney, *An Introduction to Chaotic Dynamical Systems*, MA: Addison-Wesley, 1989.

[59] S. Kay and V. Nagesha, "Methods for chaotic signal estimation," *IEEE Trans. Signal Processing*, vol. 43, pp. 2013-2016, Aug. 1995.

[60] E.J. Kostelich and T. Schrieber, "Noise reduction in chaotic time series data: a survey of common methods," *Phys, Rev. E*, vol. 48, pp. 1752-1783, Mar. 1993.

[61] H. Leung, Z. Zhu. and Z. Ding, "An aperiodic phenomenon in the extended Kalman filter in Filtering noisy chaotic signals," *IEEE Trans. on Signal Processing*, vol. 48, pp. 1807-1811, Jun. 2000.

[62] S. Haykin, *Adaptive Filter Theory, 2nd edition*, Englewood Cliffs, NJ: Prentice Hall, 1991.

[63] S. Kay, "Asymptotic maximum likelihood estimator performance for chaotic signals in noise," *IEEE Trans. on Signal Processing*, vol. 43, pp. 1009-1012, Apr. 1995.

[64] S. Wang, P.C. Yip and H. Leung, "Estimating initial conditions of noisy chaotic signals generated by piecewise linear Markov maps using itineraries," *IEEE Trans. on Signal Processing*, vol. 47, pp. 3269-3302, Dec. 1999.

[65] C. Pantaleó, D. Luengo, and I. Santamaria, "Optimal estimation of chaotic signals generated by Piecewise-linear maps," *IEEE Signal Processing Lett.*, vol. 7, pp. 235-237, Aug. 2000.

[66] H. Papadopoulos and G. Wornell, "Optimal detection of a class of chaotic signals," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. III.117-III.120, Apr. 1993.

[67] E. Ott, T. Saucer and J.A. Yorke, *Coping with Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems*, NY/Chichester: John Wiley & Sons, 1994.

[68] F. Hartung and M. Kutter, "Multimedia watermarking technique," *Proc. IEEE*, vol. 87, pp. 1079-1107, July 1999.