2008

Digital watermark technology in security applications

Xu, Xin

http://hdl.handle.net/10026.1/335

http://dx.doi.org/10.24382/3925

University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.
Digital Watermark Technology in Security Applications

by Xin XU
January, 2008

A thesis submitted to the University of Plymouth
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
PAGE NUMBERING AS ORIGINAL
Abstract of

Digital Watermark Technology in Security Applications

by X. XU

With the rising emphasis on security and the number of fraud related crimes around the world, authorities are looking for new technologies to tighten security of identity. Among many modern electronic technologies, digital watermarking has unique advantages to enhance the document authenticity. At the current status of the development, digital watermarking technologies are not as matured as other competing technologies to support identity authentication systems. This work presents improvements in performance of two classes of digital watermarking techniques and investigates the issue of watermark synchronisation.

Optimal performance can be obtained if the spreading sequences are designed to be orthogonal to the cover vector. In this thesis, two classes of orthogonalisation methods that generate binary sequences quasi-orthogonal to the cover vector are presented. One method, namely "Sorting and Cancelling" generates sequences that have a high level of orthogonality to the cover vector. The Hadamard Matrix based orthogonalisation method, namely "Hadamard Matrix Search" is able to realise overlapped embedding, thus the watermarking capacity and image fidelity can be improved compared to using short watermark sequences. The results are compared with traditional pseudo-randomly generated binary sequences. The advantages of both classes of orthogonalisation methods are significant.

Another watermarking method that is introduced in the thesis is based on writing-on-dirty-paper theory. The method is presented with biorthogonal codes that have the best robustness. The advantage and trade-offs of using biorthogonal codes with this watermark coding methods are analysed comprehensively. The comparisons between orthogonal and non-orthogonal codes that are used in this watermarking method are also made. It is found that fidelity and robustness are contradictory and it is not possible to optimise them simultaneously.

Comparisons are also made between all proposed methods. The comparisons are focused on three major performance criteria, fidelity, capacity and robustness. From two different viewpoints, conclusions are not the same. For fidelity-centric viewpoint, the dirty-paper coding methods using biorthogonal codes has very strong advantage to preserve image fidelity and the advantage of capacity performance is also significant. However, from the power ratio point of view, the orthogonalisation methods demonstrate significant...
advantage on capacity and robustness. The conclusions are contradictory but together, they summarise the performance generated by different design considerations.

The synchronisation of watermark is firstly provided by high contrast frames around the watermarked image. The edge detection filters are used to detect the high contrast borders of the captured image. By scanning the pixels from the border to the centre, the locations of detected edges are stored. The optimal linear regression algorithm is used to estimate the watermarked image frames. Estimation of the regression function provides rotation angle as the slope of the rotated frames. The scaling is corrected by re-sampling the upright image to the original size. A theoretically studied method that is able to synchronise captured image to sub-pixel level accuracy is also presented. By using invariant transforms and the "symmetric phase only matched filter" the captured image can be corrected accurately to original geometric size. The method uses repeating watermarks to form an array in the spatial domain of the watermarked image and the the array that the locations of its elements can reveal information of rotation, translation and scaling with two filtering processes.
Acknowledgement

This thesis that can be available is not only a product of hard working, also a combination of inspirations and supports that have been offered to me during my research. I would express my sincere gratitude to everyone who has helped and supported me.

First of all, it has been a prestigious privilege to have Martin Tomlinson as my director of study. Thanks for all his support, guidance and encouragement during the time of my research. With the help of Martin, the research is not only a academic activity, rather an adventure to the uncharted territory. With his support the journey has never been fearful. With his encouragement, the future is always bright.

Thanks to my other supervisors, Marcel Ambroze and Mohammed Zaki Ahmed, for their experienced advices. In every discussion of the research, their opinions always provide new visions to lead me going further efficiently. They are not only my tutors, but also close friends.

My colleagues, with whom I have spent wonderful time in our office, SMB209, Evangelos Papgiannis (Greece), Purav Shah (India), Jing Cai (China), C.J. Tjhai (Indonesia), Ismail Isnin (Malaysia), Li Yang (China), Andrew Rogers (England), Keiko Takeuchi (Japan) and Andria Tieghi (Italy), I thank you all for making my experience with you memorably. We had great conversations over wide range of topics. What they have shared with me is not only academic knowledge, also we exchange the culture from all over the world. We also shared a lot of fun time.

I also wish to thank the people who have supported my research from outside the university. Garth Zambory and Tony Rodriguez from the DigiMarc Inc., I thank you both for the constructive conversation during my research. I also thank Andersen Cheng for his interest of my research and his suggestions for study directions.

At last, I must thank my parents. Without their support and love the six years of study would have been impossible. Without their encouragement, I would have never come this far.
ALL MISSING PAGES ARE BLANK IN ORIGINAL
Author’s declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

A programme of advanced study was undertaken, which included intensive reading of literature relevant to the research project and attendance of international conference on signal processing, communication systems. The author also participated university organised skill development programs.

Conference publications:


Patent participated:

- “Impression of information on noise or an independent signal”, UK patent GB2427800, application filed June, 2005.

Word count of main body of thesis: 34500 (approximated)

Signed: ...........................................

Date: 23 JAN. 2008

xi
## Contents

1 Introduction ........................................... 1
   1.1 Objectives ...................................... 2
   1.2 Scope ........................................... 3
   1.3 Original contribution ............................. 4
   1.4 Organisation .................................... 4
   1.5 Notations ....................................... 6

2 Background ............................................ 9
   2.1 Digital Watermarks and Human visual system .......... 9
      2.1.1 History and applications ........................ 10
      2.1.2 Human visual system and perceptual modelling ... 11
      2.1.3 Performance criterion .......................... 13
      2.1.4 Classifications ................................ 15
      2.1.5 System design ................................ 17
   2.2 Communication theories and techniques ............... 18
   2.3 Signal processing techniques ........................ 21
      2.3.1 Discrete cosine transform ....................... 22
      2.3.2 2D filter design ................................ 23
   2.4 Watermarking algorithms ............................ 24
      2.4.1 Early development ............................. 24
      2.4.2 Informed embedding ............................ 25
   2.5 Watermark synchronisation ........................... 29

3 Orthogonal Spread Spectrum ............................ 33
   3.1 Conventional Spread Spectrum ....................... 33
   3.2 Orthogonal Spread Spectrum ........................ 35
   3.3 Equal energy approach ............................. 37
      3.3.1 Sorting and Cancelling ......................... 37
      3.3.2 Orthogonalisation remainders ............... 40
      3.3.3 Sequence length .............................. 43
      3.3.4 Group search ................................ 44
List of Tables

1.1 Common abbreviations used in this thesis and their definitions. 6
1.2 Common notations used in this thesis and their definitions. 7

3.1 Example of SC proceeding. 39
3.2 Error free detection in spatial domain using 3×3 highpass filter (3HP) and 45×45 bandpass filter (45BP), respectively. 54
3.3 Error free detection in DCT domain, using all DCT coefficients. 55
3.4 Performance of orthogonalisation schemes, affected by 3×3 highpass and 45×45 bandpass masking filter. 56
3.5 Orthogonalisation remainder energy when zero padding for SC is used at different thresholds. 56
3.6 Comparison on fidelity (DWR and αmin), in DCT domain for L = 1024, L = 1008 and L = 512. 57
3.7 Comparison on SNR and orthogonalisation remainder, in DCT domain for L = 1024, L = 1008 and L = 512. 58
3.8 Average performance of fidelity (DWR). 59

4.1 Comparison of term Ω and Ω in Equation 4.22. 80
4.2 Comparisons of the error free DWR and Watson Distance. 84

5.1 Target image object keywords, in a PDF file. 106
5.2 Three major embedding parameters, function and values. 107
# List of Figures

2.1 Performance trade-off triangle .............................................. 13
2.2 Watermarking system presented in form of a communication system ............................................. 18

3.1 Binary sequences and an arbitrary vector in $\mathbb{R}^2$. ................... 36
3.2 Descending sorted 1024 DCT AC coefficients. .......................... 38
3.3 System diagram of SC method. .............................................. 39
3.4 Cover vector is modelled into a constant function. ...................... 40
3.5 Cover vector is modelled into a linear function. ......................... 41
3.6 Cover vector is modelled as a rational function. ....................... 42
3.7 Simulation results of orthogonalisation remainder for rational modelled cover vectors. ......................... 43
3.8 The comparison between CDMA and orthogonal matrix watermarking. ......................................................... 47
3.9 Double rate sampling. ......................................................... 49
3.10 2D band pass filter with $45 \times 45$ coefficients. The normalised cutoff frequencies are 0.1 and 0.2 .................................................. 53
3.11 Experiment results with $L = 128$ .................................. 61
3.12 Experiment results with $L = 1024$. ................................ 62
3.13 Performance of single sequence embedding and eight-sequence overlap embedding. ......................................................... 64

4.1 Dirty paper channel model, modified for watermarking .................. 67
4.2 Complete watermarking channel model, based on Dirty Paper channel. ......................................................... 68
4.3 Examples for encoding watermarks with proposed codes ................ 75
4.4 Encoding situations for dirty paper coding. ............................ 79
4.5 Incomplete example of two classes of binning schemes. ............... 81
4.6 Probability of error based on various DWR. .......................... 83
4.7 The Watson distance produced at various DWR by different classes codebooks. ......................................................... 84
6.6 New design of locations of synchronisation markers (shadowed areas). ×’s in the figure, show the location of correlation peaks. 124
6.7 The marker coverage percentage histogram on 55 test images. 127
6.8 SNR curves. 128
6.9 The histogram of α which obtains the best SNR. 129
6.10 Average signal energy is compared with average noise energy. 129
6.11 SNR curves with difference marker power (variance). 130
6.12 The power comparison between signal with different variance and the noise, which is constant for all different signal power. 130

B.1 Ideal filter response in frequency domain. 151
B.2 Frequency response of a high-pass filter with 125 coefficients applied with Rectangular window, Hamming and Blackman-Harris. 153
B.3 Two Hamming 1-D filter multiplied to be a 2-D window function. 154
B.4 Ideal 2-D bandpass filter responses in frequency domain. 155
B.5 2D band-pass filter frequency response, with normalised cut-off frequencies of 0.1 and 0.2, the number of coefficients is 51. 156
Chapter 1

Introduction

In the year 2005, the United States President George W. Bush signed "REAL ID Act 2005" (U.S. Public Law 109-13 Division B, 2005). The machine-readable identity cards and Driving Licenses became a public law. Despite strong criticisms shadowed the Act, the Congress voted 100 to 0 approved the Bill\(^1\).

Digital watermark, by definition, is imperceptible but machine readable signal sequences covered by human perception oriented media. For still images, like photos, the signal consists of nearly imperceptible alterations of the cover image pixels. When the watermarked image is transferred to other formats, the watermark signal remains inseparable from the cover image. If properly designed digital watermarks are used, attempts at removal only produces evidence of malicious activities by destroying the integrity of the cover image. As a sophisticated technology to protect copyright and intellectual property of video/audio media, digital watermarking techniques have been adopted for only a few years. The first use of the term digital watermark is around the early 1990s. Since then, numerous research efforts have improved this technology for innovative applications.

The first time digital watermark technologies were employed in applications of authority document protection was driving license verification in the United States (Digimarc Inc., 2004). The watermarks strongly link classic printing and electronic identification. The modern digital watermarking algorithms are based upon sophisticated secure communication technologies, thus the resulting security enhancement watermarks have some unique prop-

properties, including stealth, robustness, high data capacity and long life span. Supported by cryptographic theory, counterfeiting could take much longer than the life time of a particular set of watermarks (Cox et al., 2001). Another advantage of the technology is that there is no dependency on expensive materials and special equipment. What necessary is good quality printing and capturing facilities. If some security documents are compromised, emergency document replacement is able to restore the security level to minimise the damage, quickly and cheaply. With these advantages, the possibility of successful forgery and fraud is significantly reduced. These unique properties give digital watermark considerable advantages over other newly proposed electronic identification schemes. A comprehensive comparison of digital watermarking and other electronic identification schemes is included in Appendix A.

1.1 Objectives

This research work is aimed at developing new core watermarking techniques that possess improved performance and are suitable to support a security scheme that can be deployed in the applications of passports, access control cards and banking documents. The potential improvements are concentrated on robustness, data capacity and fidelity.

Capacity is defined as the maximum number of information bits that the watermark is capable of carrying and being correctly detected. The maximum capacity is, firstly, dependent upon the bits to pixel ratio. Secondly, the level of attack tolerance is also required to be improved. Fidelity measures the distortion made by the watermark. The improvement on fidelity is aimed at approaching imperceptibility and it is assisted by well established visual perceptual models. Robustness represents the recovery reliability in the presence of additional noise. Robustness is measured by the probability of correct detection at a certain noise level.

The performance is evaluated by theoretical analysis confirmed by experimental results, in order to discover the limitations and to identify trade-offs. Suitable application models are also discussed.

A complete watermarking system for physical media must be assisted by a good synchronisation procedure, which allows the watermark to be detectable after the degradation resulting from image capture. Therefore developing state-of-the-art watermark synchronisation methods is also an objective.
1.2 Scope

In this work, the watermarking system is generally modelled as a communication system. The watermark is modelled as a transmission signal, which contains a number of information bits. The cover image is modelled as a noise source. Since the encoder has complete knowledge of this noise source, it can be also modelled as side information to the encoder. The embedder and detector are modelled as transmitter and receiver, respectively.

The embedding operation consists of two steps, perceptual masking and watermark signal coding. Perceptual masking is applied through a frequency selective filter or through visual sensitivity models. The design work is mainly focused on the second step of embedding, that is watermark signal encoding algorithms. To optimise the performance, sophisticated communication channel coding theorems are adopted. Since the embedding operation is only required once for the issuance of security documents, the embedding process is allowed to be complex, in order to obtain optimal performance. On the contrary the detection algorithm is aimed at a lower level of complexity, because the detection process is the main operation of the system. Maximum likelihood detection and correlation demodulation techniques are used to deliver optimal performance.

In addition, the design of the watermark synchronisation scheme must include a combination of robust signal processing techniques which are able to synchronise rotated, scaled and translated (RST) watermarks to sub-pixel precision. This operation must be simple and fast, as it is required with every detection operation.

Due to the limitation on time, the research work has concentrated on techniques which improve the performance of digital watermarking, including data capacity, fidelity, robustness and synchronisation. Some assumptions have been made about the watermarking channel in order to simplify the model. The combined distortion of printing and capturing plus considerable signal processing in detection of watermarks is assumed to be additive white Gaussian noise (AWGN). This assumption is supported by the central limit theorem.

Apart from the technical aspect of identification technologies, there are many important practical issues, including legalisation, cooperation and distribution. These topics exceed the subject of this thesis.
1.3 Original contribution

The research work presented in this thesis has made the following contributions, which are to the knowledge of the author, original:

- Proof that optimal performance is achieved by orthogonal spread spectrum watermark coding.
- Three different orthogonalisation procedures which produce binary sequences quasi-orthogonal to the cover vector.
- A biorthogonal dirty paper coding scheme with good corresponding binning strategy.
- Thorough in-depth error probability analysis showing that the detection performance is optimal for the biorthogonal matrix.
- A dirty paper coding method using Goppa codes, that trades robustness against capacity.
- A feature of the Goppa dirty paper codes is that it allows multi-level watermark encoding.
- A synchronisation scheme using image registration techniques.
- A multipurpose watermark embedding strategy for watermark synchronisation.
- Comprehensive simulation results for all of the proposed methods and comparisons of the different performances.

1.4 Organisation

In Chapter 2, the background of research project is reviewed, and includes a brief history and common applications of digital watermarking technologies, some related communication and signal processing techniques. Current state-of-the-art watermarking techniques together with reference to key contributing papers is given in Chapter 2.

In Chapter 3, one of the original contributions is investigated, that is orthogonal spread spectrum watermarking which is shown to have theoretically optimal fidelity, because of the minimum requirement for watermark power. The orthogonalisation algorithms and their performance are described in
detail. Three different algorithms are proposed. These are termed sorting-cancelling, Hadamard matrix search and group search. Also described in Chapter 3 is the degree of orthogonalisation achieved, and its dependence on the sequence length. At the end of Chapter 3, the simulation results and discussions are presented.

In Chapter 4, the proposed dirty paper coding algorithm is discussed in terms of probability of error. The proposed method adopts a biorthogonal matrix and the rows of the matrix are used as codewords which are independent of the cover image. In this case, this detection is blind detection. Analysis shows that for the dirty paper coding scheme performance is optimal achieved by biorthogonal codes. The optimal detection performance is restricted by the limited availability of biorthogonal sequences. The dirty paper performance is a function of the binning strategy and differences between the two classes of binning strategies are analytically compared. The chapter is concluded by simulation results and discussions.

In Chapter 5, the two types of watermarking algorithms, orthogonal spread spectrum and writing on dirty paper watermark coding are compared in terms of fidelity, capacity and robustness. The comparisons are made from two different viewpoints. One is based on the visual quality, indexed by the Watson distance. The second comparison is based on the power ratio between cover vector and watermark signal. The two comparisons clearly indicate the differences among new schemes. An application example is described which uses the Hadamard biorthogonal matrix dirty paper coding algorithm to embed secret messages in an image that is nested in a PDF file and the details of this are given in Section 5.5.

In Chapter 6, the issue of synchronisation of the watermarked image is addressed. Visible borders and edge detection techniques are used to detect the rotation angle and the scaling is corrected by sampling obtained pixels to a known geometric size. A novel watermark synchronisation scheme is also proposed. This scheme includes a multipurpose watermark array. Detection of these array elements utilises invariant transforms that are used for image registration applications. A symmetric phase only matched filter (SPOMF) ensures the quality of detection. The effectiveness of the SPOMF in the detection of synchronisation markers is evaluated by computer simulations and the results are shown in Section 6.2.4.

Finally, the thesis conclusions are given in Chapter 7 with a discussion of the original contributions made to this research and their relationship to practical applications of watermarking. Some ideas for future research directions are also presented.

At the end of the thesis, several Appendices are attached. A report on global intentions of enhanced identity security and comparisons of modern
technologies adopted by enhanced identity security schemes is presented at Appendix A. A detailed design process for the 2D filter described in Chapter 2 is listed in Appendix B. A collections of the papers published during this research work is attached at Appendix B.

1.5 Notations

Generally, throughout the thesis, the notations and expressions are intended to be meaningful both literally and visually, as well as being consistent. However, some exceptions apply to the content in Chapter 2. In order to honour original authors, and to avoid conflict to this thesis, some of the notations are adopted from the cited works.

Vectors are denoted by the beginning letter of the names of the vectors and presented in bold italic font. A matrix is presented by bold and italic font of first letter of the matrix name in capital. Individual columns or rows of matrices are presented in vector format with subscript of either c or r. The Table 1.1 shows the major abbreviations used in this thesis and Table 1.2 shows the major notations.

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Descriptions</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWGN</td>
<td>Additive white Gaussian noise</td>
<td>-</td>
</tr>
<tr>
<td>BER</td>
<td>Bit error rate</td>
<td>-</td>
</tr>
<tr>
<td>DCT</td>
<td>Discrete Cosine transform</td>
<td>Eq. 2.21, p.22</td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier transform</td>
<td>Eq. 2.18, p.21</td>
</tr>
<tr>
<td>DWR</td>
<td>Document to watermark ratio</td>
<td>Eq. 2.4, p.15</td>
</tr>
<tr>
<td>GS</td>
<td>Group search</td>
<td>Sec. 3.3.4, p.44</td>
</tr>
<tr>
<td>HMS</td>
<td>Hadamard matrix search</td>
<td>Sec. 3.4, p.46</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean squared error</td>
<td>Eq. 2.1, p.14</td>
</tr>
<tr>
<td>OSS</td>
<td>Orthogonal spread spectrum</td>
<td>Chapter 3, p.33</td>
</tr>
<tr>
<td>PN</td>
<td>Pseudo-random noise</td>
<td>-</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak signal to noise ratio</td>
<td>Eq. 2.2, p.14</td>
</tr>
<tr>
<td>RST</td>
<td>Rotation, scaling and translation</td>
<td>Chapter 6, p.111</td>
</tr>
<tr>
<td>SC</td>
<td>Sorting and Cancelling</td>
<td>Sec. 3.3.1, p.37</td>
</tr>
<tr>
<td>SPOMF</td>
<td>Symmetric phase only matched filter</td>
<td>Chapter 6, p.111</td>
</tr>
<tr>
<td>WNR</td>
<td>Watermark to Noise Ratio</td>
<td>Eq. 2.5, p.15</td>
</tr>
</tbody>
</table>

Table 1.1: Common abbreviations used in this thesis and their definitions.
<table>
<thead>
<tr>
<th>Notations</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>Cover vector</td>
</tr>
<tr>
<td>$\mathcal{C}$</td>
<td>Complex number space</td>
</tr>
<tr>
<td>$d_{\text{min}}$</td>
<td>Minimum Hamming distance</td>
</tr>
<tr>
<td>$d_E$</td>
<td>Euclidean distance function</td>
</tr>
<tr>
<td>$d_H$</td>
<td>Hamming distance function</td>
</tr>
<tr>
<td>$d_W$</td>
<td>Watson distance (Eq. 2.3, p.14)</td>
</tr>
<tr>
<td>$\mathcal{F}$</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>$F(u, v)$</td>
<td>DCT coefficients</td>
</tr>
<tr>
<td>$H$</td>
<td>Hadamard matrix</td>
</tr>
<tr>
<td>$h_c$</td>
<td>$c$-th column vector of Hadamard matrix</td>
</tr>
<tr>
<td>$h_r$</td>
<td>$r$-th row vector of Hadamard matrix</td>
</tr>
<tr>
<td>$k$</td>
<td>key vector</td>
</tr>
<tr>
<td>$L$</td>
<td>Length of $c$ and $w$</td>
</tr>
<tr>
<td>$N(0, \sigma^2)$</td>
<td>AWGN noise with 0 mean and $\sigma^2$ variance</td>
</tr>
<tr>
<td>$\mathbb{R}$</td>
<td>Real number space</td>
</tr>
<tr>
<td>$\mathcal{U}$</td>
<td>Codebook</td>
</tr>
<tr>
<td>$U^s$</td>
<td>$s$-th bin of $\mathcal{U}$</td>
</tr>
<tr>
<td>$u_i^s$</td>
<td>$i$-th codeword of $U^s$</td>
</tr>
<tr>
<td>$w$</td>
<td>Watermark vector</td>
</tr>
<tr>
<td>$x$</td>
<td>Watermarked vector</td>
</tr>
</tbody>
</table>

Table 1.2: Common notations used in this thesis and their definitions.
Chapter 2

Background

Following the recent emphasis on security, a wide range of social activities require enhanced security, such as national border control, access control to the restricted areas and civil identity. Governments around the world and some international organisations are rapidly proposing advanced technology enhanced identity/travel documentation schemes, to reduce the number of crimes related to the identity fraud activities. Significant changes are happening and many modern technologies are adopted to enhance the identity security (reviewed in Appendix A).

Among other technologies, the digital watermarks have significant advantages. The unique properties of stealth, low cost and long life span will attract strong attention for security and identity applications. However, the digital watermarking technologies are still underdeveloped, especially in the application of the identity documentation. The major application and development milestones are reviewed in this chapter, along with the supporting subjects, such as models of human visual system, signal processing techniques and related communication techniques. More importantly, how this research is derived from the background is also introduced.

2.1 Digital Watermarks and Human visual system

The concept of watermarking can be dated centuries ago in the ancient time. After the revolutionary development of digital communication, such as broadcasting networks and the Internet, this concept has been brought to a new life from its antiquity. In this section, the history and applications of digital watermarking technologies and phenomena of human visual system under which the watermarks retain the stealth are reviewed.
2.1.1 History and applications

Watermarking techniques have a long history. The physical forms of watermark are very common to see in the bank notes and official paper stationary. Generally, watermarks can be divided into two major categories, visible and imperceptible. The visible watermarks normally appear as lightly shaded text or graphics in the background. If one holds a typical large amount note, for example a 10 pound note issued by the Bank of England, against a light source, a portrait of the Queen becomes visible at an appeared blank area on one side of the note. (For more examples, see Cox et al., 2001, Chapter 1).

On the contrary, the digital watermarks are not commonly detectable by human eyes. Normally a digital watermark refers to a sequence of signal to be recognised by computer programs. It was originally designed to protect the copyright of digital multimedia contents, such as images, video and audio. The signal is commonly an identity sequence, which is unique between the owners. Watermarks can also be secret messages, which are coded through digital communication methods and cryptography algorithms. When the code is retrieved from the cover image, corresponding decoding processes can reveal the intelligible message.

In some applications, the digital watermarks are used to identify the buyer of a particular multimedia content, so that the work's original creator, or the seller, can trace the illegal copy back to the buyer, who distributed the merchandise without proper authorisation (Tsolis et al., 2004). Digital watermarks are also used in broadcast monitoring (Salmon, 2001), content authentication (Chang et al., 2005; Eggers et al., 2001), copy control (Petitcolas, 2003) and ownership identity (Craver et al., 1998). Apart from those classic applications, which are further reviewed in (Cox et al., 2001; Zeng, 1998; Eggers et al., 2001), some new applications with the latest mobile and Internet development have been proposed. Digimarc Inc. announced ImageBridge™ (Alattar, 2000) for mobile and Internet commerce. Mobile networks and phone manufactures in Japan have developed a catalogue shopping service using camera mobile phones, as recent BBC report ¹. Some of the broadcast applications of watermarking are described in Mason (2004).

The properties of digital watermarks include:

- Imperceptibility. The watermark should not alter the cover content with a noticeable difference by the viewers. “Fidelity” is a measurable expression and it describes this property quantitatively. The imperceptibility of a digital watermark is measured by its Fidelity Index (FI).

ceptibility is an ideal situation (Cox and Miller, 1997). Commonly the watermark power is very low, comparing with the image power, thus it allows imperceptibility. Human visual (auditory) system models can provide optimal watermark power distribution mask over the cover content.

- Robustness. Watermarks like other signals, are subject to intentional or non-intentional distortions after the embedding. Commonly described non-intentional distortions include compression algorithms, for example the JPEG compression to images and MPEG for video. Intentional distortions are normally described as attacks to watermark signal. The attacks are mainly subject to partially or completely degrade the detection quality (Detailed analysis of attack and counterattack techniques can be found in Hartung et al. (1999)).

- Capacity. The amount of information a watermark signal carries, often measured in the number of binary bit. The maximum theoretical amount is measured by bit per pixel, that may be called “bit-rate”.

2.1.2 Human visual system and perceptual modelling

Modern digital watermarking technologies rely on the properties of human perceptual systems to hide the existence of the watermark. In the development of image watermarking techniques, the human visual system (HVS) and perception model must be referred (Cox and Miller, 1997). To measure the perception quality and the fidelity of watermarking algorithms, visual perception test can be used, but it is very expensive and unlikely to be repeated. Therefore the perceptual characteristics are modelled to simplify the assessment. Several human eye perception phenomena have been studied and noted by the watermarking algorithm designers. In Cox et al. (2001, Chapter 7) and Zeng and Lei (1999), the human visual and auditory perception properties are introduced and discussed. The visual perceptual phenomena are generalised as follows:

- Human eyes are more sensitive to luminance difference at middle range spatial frequencies and less sensitive to very low and high range. This phenomenon is normally described in Contrast Sensitivity Function (CSF) ((see Cox et al., 2001, Figure 7.5, page 211)).
  - It is called spatial frequency sensitivity or contrast sensitivity since the spatial frequency is described in the change of spatial luminance.
This sensitivity is a very important property for designing watermarking algorithms, because watermark signal changes the pixel intensity of the cover image, and the "just noticeable difference" (JND) is defined in contrast sensitivity.

The property can be explained as, when a small change (JND) $\Delta I$ is applied partially to a background of intensity $I$, the ratio of $\Delta I/I$ is called Weber fraction, and it is nearly constant at 0.02 for a wide range of intensity $I$, but not for very high and very low intensities. This ratio is also determined by the surrounding of the pattern\(^2\). Another example of contrast sensitivity can be shown in the Mach band (shown in Banks, 1990, Figure 11.4, page 155) and (Pratt, 2001, Figure 2.3-2 at page 31).

- Human eyes also have different sensitivity in different spectral frequencies, perceived as colour.

  - It is less sensitive to blue light than red and green, because of the type and number of cones in retina (Pratt, 2001, Figure 2.2-4 Page 29).
  
  - The perception of colour is measured in terms of brightness, hue and saturation (Banks, 1990, page 152).
  
  - The brightness is the measurement of "absolute intensity" of the light source and it is proportional to the energy radiated by the source.
  
  - The hue is used to distinguish the colour.
  
  - The saturation, measures the quantity of white light added to a pure spectral colour.

- Human eyes have a non-linear response to different intensity of illumination, and this response can be modelled with a logarithm relationship, cube root relationship and more complicated models (Cox et al., 2001). This is called brightness sensitivity. Human eyes are more sensitive to the difference in the low brightness than in the brighter area.

- Contrast masking describes the perceptual phenomenon that one signal can hide in the existence of other, such as, a texture pattern is easier to see when it is isolated than it is presented in a highly textured image.

\(^2\)the detailed description can be found at (Pratt, 2001, pp. 30-31)
In this research work, the brightness values of colour or grayscale images will be used to embed watermarks, since brightness presents the power intensity of the image and the contrast sensitivity models are well established. The visual perception model proposed by Watson (1993) and also described in Cox et al. (2001), the *Watson distance* combines the contrast sensitivity and brightness sensitivity together to measure the fidelity of the designed watermarking algorithms and it is also possible to implement the model to reduce visual impact of the watermarks (Miller et al., 2004). However certain sacrifices on data error rate or payload bit-rate is inevitable. Alternatively additional error correction coding schemes must be involved to preserve the communication quality.

### 2.1.3 Performance criterion

The three major properties of watermarking systems, *bit rate*, *robustness* and *fidelity* are recognised as the performance measurement factors. (Decker, 2001) recognised the properties as the "trade-off triangle". The three performance categories cannot be optimised all together, as illustrated in Figure 2.1. A good balance can be achieved with particular applications.

![Performance trade-off triangle](image)

Figure 2.1: Performance trade-off triangle

To measure the fidelity, the following measurement may be used:

- **MSE**, Mean Squared Error, defines the absolute difference between two signals. If two images $I_{x,y}$ and $K_{x,y}$ both have the size of $M \times N$, it is
defined as:

$$\text{MSE} = \frac{1}{MN} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} [I_{x,y} - K_{x,y}]^2$$  \hspace{1cm} (2.1)$$

- **PSNR**, Peak Signal to Noise Ratio, measures the maximum possible power of signal and corrupting noise power to affect the fidelity of the signal's presentation. It is commonly used to measure the quality of image reconstruction by image compression schemes. The common range is between 30dB and 40dB. The definition of PSNR is

$$\text{PSNR} = 10 \log_{10} \left( \frac{\text{MAX}_I^2}{\text{MSE}} \right)$$  \hspace{1cm} (2.2)$$

where the \text{MAX}_I^2 is the maximum power. Normally \text{MAX}_I takes the value of 255. PSNR is an alternative presentation of MSE.

- **Watson Distance**, \(d_w\), is DCT (Discrete Cosine Transform) based perception model proposed by (Watson, 1993). Watson distance takes both luminance and frequency into consideration, and it is defined as:

$$d_w(C_w, C_0) = \left( \sum_{i,j,k} \left( \frac{|C_w(i,j,k) - C_0(i,j,k)|^p}{c(i,j,k)} \right)^\frac{1}{p} \right)$$  \hspace{1cm} (2.3)$$

where the indexes \((i, j, k)\) present the \(i\)-th column and \(j\)-th row of the \(k\)-th DCT block. \(C_w(\cdot)\) and \(C_0(\cdot)\) denote the DCT coefficients of the watermarked and the original image, respectively. The \(c(\cdot)\) is the contrast mask for every DCT coefficient, and it is derived from

\[c[i, j, k] = \max\{t_L[i, j, k], |C_0[i, j, k]|^{\omega[i,j]} t_L[i, j, k]^{1-\omega[i,j]}\}\]

\(\omega[i, j]\) is suggested as 0.7 for all DCT coefficients. \(t_L\) is the luminance mask and it derives from

$$t_L[i, j, k] = t[i, j] \left( \frac{C_0[0,0,k]}{C_{0,0}} \right)^{a_T}$$

where \(t[i, j, k]\) is the sensitivity table for each \(8 \times 8\) DCT block (Cox et al., 2001). \(a_T\) is a constant, with a suggested value of 0.649 (Watson, 1993). \(C_0[0,0,k]\) is the DC coefficient of \(k\)-th block, and \(C_{0,0}\) is the average of DC coefficients in the image.

Robustness is described in terms of Document to Watermark power Ratio (DWR), Watermark to Noise power Ratio (WNR) and Bit Error Rate (BER). Where,


• DWR is described in decibel as

\[
DWR = 10 \log_{10} \left( \frac{\|c\|^2}{\|w\|^2} \right)
\] (2.4)

This ratio describes the power of watermark in terms of cover signal power. This ratio describes the watermark power independent of cover image.

• WNR is described in decibel as

\[
WNR = 10 \log_{10} \left( \frac{\sigma_w^2}{\sigma_z^2} \right)
\] (2.5)

Where, \(\sigma_w^2\) is the power of watermark, and \(\sigma_z^2\) is the power of additive channel noise. This ratio is a direct description of the channel in which the watermark message is transmitted. Therefore this ratio is commonly used to derive the robustness of watermarking algorithms. Especially in the informed embedding scenarios, where the signal (watermark) power will only be distorted by the unknown noise power, such as malicious attacks, the WNR can be seen as the channel SNR.

• BER, is commonly used in communication systems. This performance measures the number of error bit in a unit time, but more commonly present as a possibility of error occurrence in a very large binary data set. The common range in communication systems is between \(10^{-6}\) and \(10^{-8}\) (Proakis, 2001).

Capacity is expressed by bit rate (BR). The factor BR, in this work, is measured as the ratio of one bit of watermark information and the number of pixels used to carry it. Capacity can also be seen as the capacity of communication channel, in the terms BER if the watermark power constraint is applied.

Since this research work models the watermarking system as a communication system. The common communication benchmark systems are used, for example the reliability of the embedding and detection schemes is measured in terms of BER when the watermark signal is subject to any distortion. Important measurement systems that are popular for the watermarking research are also adopted, such as the fidelity of watermarks is measured as MSE and more perception sensitive measurement Watson distance.

### 2.1.4 Classifications

Watermarking algorithms can be classified according to their robustness into fragile, semi-fragile and robust watermarks. Fragile watermarks will become
invalid when a small change is applied to the watermarked content. Semi-fragile watermarks are designed to survive a predefined distortions, but they will not be detectable once the distortion strength exceed the predefined limits. Both kinds of watermarks are normally used in content authentication applications to protect the integrity of the cover content Cox et al. (2001, Chapter 10). Robust watermarks are designed to survive a number of unknown attacks, intentional and non-intentional. The difference between semi-fragile and robust watermarking scheme is that the robust watermarks are expected to survive a wider range of attacks which are not expected. For example, the robust watermarks are expected to survive arbitrary geometric distortions, large additive noise and signal processing attacks (e.g. filtering, brutal force attacks, etc.). However, any robust watermark has a “breaking point”, like semi-fragile watermarks. It is assumed that the robust watermarks are detectable, when the watermarked content is intelligible. It is also believed that the robust watermarks cannot be reproduced by unauthorised parties. The anti-forgery ability exceeds the scope of this thesis, and it is only addressed qualitatively. In this research work, the watermark is considered to be robust, since the life span of a security document may last ten years or more, and many elements may affect the detection quality.

Watermarking systems can also be classified into “public” and “private”, depending on how decoding keys and decoders are distributed, whether publicly available or secretly shared between senders and receivers. In copyright protection applications, it is likely that the public watermarking systems are used (Hartung and Girod, 1997). According to the cryptographic theories, the public watermarking systems are much more difficult to design than private systems (Natarajan, 1997). But for applications that use watermarks to transmit secret messages the decoding keys and algorithms are more likely to be private. The distribution of the watermark keys affect the robustness of the watermark systems. Public watermarking systems are more vulnerable than the private systems.

According to the detection algorithms, the watermarking system may also be classified as blind detection and non-blind detection. Non blind detections depend on the original cover media to decide the watermark signal. Normally a subtractive process is applied between the obtained target medium and the original one. Some early development of watermarking algorithms are non-blind detections. Blind detections are independent of original cover media. It is clear that blind detection is more difficult to design, since some estimation processes have to approximate the watermark signal. A middle state can also be defined, which is semi-blind detection. The semi-blind detection does not require the presence of original cover medium, but a matching decoding key is required. It is believed that the semi-blind detection is easier to design than
the blind detection. From the secrecy point of view, semi-blind detections are more secure than blind detections, because the decision can be made according to a much firmer match. Normally the distribution of a large number of cover images or the decoding keys is not feasible, especially for authenticating passports. Therefore in this work, both semi-blind and blind detections are analysed and discussed for the defined applications.

2.1.5 System design

Digital watermarking techniques have two close relatives, steganography and data hiding. Steganography has a Greek origin. "stegano" means covered, and "graphia" means writing. Data hiding is a similar concept of hiding the existence of the message in the cover content. As defined by Cox et al. (2001), an information hiding system normally involves a database or a codebook that is restricted to unauthorised users. Cox et al. (2001) provide more examples of the three systems.

Considering the nature of this research work, all three concepts may contribute to a stable and reliable watermark based identity security system. The hidden messages in ID photos, in this research work are generally called "watermarks", regardless under which category the embedding techniques may fall. Because the hidden message carries more purposes than one kind of system could describe. For example, deciding the existence of the hidden message is watermarking, and decoding the content of the message is a steganographic approach, even more the retrieval method may have fallen under the category of data hiding.

The design of watermarking techniques are normally classified into two aspects, communication theoretic and information theoretic (Sequeira and Kundur, 2001). Cox et al. (2001) described the watermarking systems into geometric model and communication model. This research work is more concentrated on the communication theoretic design. A watermarking system should consist an embedder (transmitter), a cover image (channel 1), some distortions/attacks (channel 2) and a detector (receiver). The diagram of the system can be illustrated in Figure 2.2. However, the information theoretic approaches and the geometric modelling approach are also regarded as analytic tools to provide diverse views.

Previous research treated the cover image as noise, and others saw it as the side information at the embedder. Because the role of the cover image is unique in the watermarking communication system, it is considered differently, in order to provide optimised performance.

Many communication theories and technologies are used in the digital watermarking system design. They are briefly reviewed in the next section.
The detailed descriptions are available at classic texts, such as Proakis (2001); Sklar (2001).

2.2 Communication theories and techniques

Spread spectrum is a classic communication channel coding technique. It is proposed for secure and reliable military communications. Principle of the spread spectrum coding is to spread the signal energy to a large bandwidth, so that the existence of signal is not noticeable, thus the secrecy is ensured (Proakis, 2001, Chapter 13). The spreading also prevents hostile jamming over the wireless connections. Because jamming requires high power intensity, applications over a wide bandwidth is considered unfeasible. The significant development of the spread spectrum communications is the CDMA systems used in mobile communications and satellite navigation.

If the information data rate $R$ is presented in bits/second and the transmission bandwidth $W$ is denoted in hertz, the value of $W/R$ is the "spreading factor" or "processing gain". This value normally ranges between one hundred and one million, 20dB and 60dB in decibel (Viterbi, 1995). Low detectability of transmission signal by eavesdropping receivers and higher capacity than conventional communication methods through better allocated resources for multiple access also have attractions to the watermarking system designers.

In the wireless communications, Phase Shift Key (PSK) and Frequency Shift Key (FSK) modulations may be used. When PSK is used in conjunction with a pseudo random noise-like (PN) sequence, the resulting signal is called
Direct Spread. When FSK is used, the resulting signal is called Frequency Hopping (Proakis, 2001).

The input signal of a receiver in a time interval $0 \leq t \leq T$ is normally presented as:

$$r(t) = s_m(t) + n(t)$$

(2.6)

where $n(t)$ denotes a noise signal added to the signal during one interval. The signal receiver can be divided into two parts (Proakis, 2001). One is a demodulator that maps the received signal into N-dimensional vector, and another is a detector that makes the decision of which of the $M$ possible waveforms has been sent in $s_m(t)$. Two types of demodulator, optimised for the additive white Gaussian noise (AWGN) channel, are the signal correlator and the matched filter.

The received signal can be seen as a combination of linearly weighted orthonormal functions $f_k(t)$, which span the signal space but does not span the noise space. The correlation demodulator decomposes the received signal $r(t)$ into an N-dimensional vector as:

$$\int_{0}^{T} r(t)f_k(t)dt = \int_{0}^{T} [s_m(t) + n(t)]f_k(t)dt \quad (2.7)$$

$$r_k = s_{mk} + n_k, \quad k = 1, 2, \ldots, N \quad (2.8)$$

Now the signal is presented as a vector $s_m = \{s_{m1}, s_{m2}, \ldots, s_{mN}\}$. By doing so, the noise factor is broken into

$$n(t) = \sum_{k=1}^{N} n_k(t) + n'(t) \quad (2.9)$$

$n_k$ consists of uncorrelated variables, so it is also statistically independent. $n'$ is Gaussian and it is not correlated with $r_k$ so they are also statistically independent. The output of correlation demodulator is sufficient statistics to make decision on which of the $m^{th}$ waveform has been sent.

If the $f_k(t)$ are considered as linear filters, their impulse responses of $N$ filters are:

$$h_k(t) = \begin{cases} f_k(T-t), & 0 \leq t \leq T \\ 0, & \text{elsewhere} \end{cases} \quad (2.10)$$

Then the correlator demodulator becomes a matched filter demodulator. The frequency response of the matched filter is the complex conjugate of the Fourier Transform of the transmitted signal times $e^{-j2\pi ft}$, and their magnitude responses are identical.

Both matched filter and signal correlator demodulators have optimised performance and following properties (Proakis, 2001):
If the signal is corrupted by AWGN, the matched filter demodulator can maximise the SNR.

The maximum output SNR depends on the energy of \( s(t) \), but does not depend on the characters of \( s(t) \).

After the received signal being transferred into a \( N \)-dimensional vector by either correlator or matched filter demodulator, optimal decisions can be made according to the posterior probabilities that are calculated as:

\[
P(s_m|r) = \frac{p(r|s_m)P(s_m)}{p(r)}
\]  

(2.11)

where \( p(r|s_m) \) is the conditional probability density function (PDF) of the received vector given \( s_m \), and \( P(s_m) \) is the \textit{a priori} probability of the \( m \)th waveform being sent. In memory-less detector, the probability of each of the \( M \) signals being sent is treated equally, so \( P(s_m) = 1/M \). The criterion that selects the signal that corresponds to the maximum posterior possibility in set \( P(s_m|r) \) is called maximum a posteriori probability (MAP) criterion. In Equation 2.11, \( p(r|s_m) \) is called \textit{likelihood function}. The decision criterion based on maximum of the the likelihood function is called \textit{maximum likelihood (ML) criterion}. In the case of all \( M \) waveforms being equiprobable, the results of MAP and ML detector are identical (Proakis, 2001). If the signal is sent over the AWGN channel, \( N(\sigma^2, 0) \), the likelihood function is

\[
p(r|s_m) = \frac{1}{(\pi\sigma^2)^{N/2}} \exp \left[-\frac{1}{\sigma^2} \sum_{k=1}^{N} \frac{(r_k - s_{mk})^2}{\sigma^2}\right] \quad m = 1, 2, \ldots, M
\]  

(2.12)

By taking natural logarithm of \( p(r|s_m) \) the Equation 2.12 becomes

\[
\ln p(r|s_m) = -\frac{1}{2} N \ln(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{k=1}^{N} (r_k - s_{mk})^2
\]  

(2.13)

The signal \( s_m \) which minimises the Euclidean distance

\[
d_E(r, s_m) = \sum_{k=1}^{N} (r_k - s_{mk})^2 \quad m = 1, 2, \ldots, M
\]  

(2.14)

maximises the likelihood function.
2.3 Signal processing techniques

Many image processing techniques are derived from their signal processing counterparts. For example, the Fourier transform always used for image spectrum analysis, and the Discrete Cosine Transform (DCT) is a major part of JPEG/MPEG image/video compression algorithms. In this section some common signal processing techniques used in image processing are reviewed.

Fourier transform is commonly used to analyse the frequency spectrum of signal. In image processing, the two dimensional Fourier transform are used to derive the spatial frequency spectrum of images. Two dimensional Fourier transform is defined as:

\[ I(\omega_x, \omega_y) = \mathcal{F}[i(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i(x, y)e^{-j(\omega_x x + \omega_y y)} \, dx \, dy \]  

(2.15)

and its inverse transform is

\[ i(x, y) = \frac{1}{4\pi^2} \mathcal{F}^{-1}[I(\omega_x, \omega_y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(\omega_x, \omega_y)e^{j(\omega_x x + \omega_y y)} \, d\omega_x \, d\omega_y \]  

(2.16)

For digital images, the processing is applied in the digital format, so that the input of the Fourier transform is discrete data. The discrete Fourier transform (DFT) is defined as:

\[ I_{u,v} = \mathcal{F}[i_{x,y}] = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} i_{x,y}e^{-j2\pi(xu/M+yu/N)} \]  

(2.17)

and its inverse transform is

\[ i_{x,y} = \mathcal{F}^{-1}[I_{u,v}] = \frac{1}{NM} \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} I_{u,v}e^{j2\pi(xu/M+yu/N)} \]  

(2.18)

The frequency spectrum of an image is important for the applications of filtering, visual quality enhancement, edge detections, feature extractions and
other image characterising processes. As discussed earlier, HVS has different 
sensitivities on different spatial frequency component, so that processing in 
frequency domain has unique advantages to apply the HVS characteristics. 
The most beneficial advantage of Fourier transform to the image processing 
applications is its relationship to convolution and correlation operations. It 
is well known that the convolution in the spatial domain is the multiplication 
in the frequency domain such that

\[
g_{n,m} = i_{x,y} * h_{x,y} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} i_{x,y} h_{m-x,n-y} = \mathcal{F}^{-1}[I_{u,v}H_{u,v}] \tag{2.19}
\]

Since the Fourier transform can be implemented through fast computational 
algorithms, the convolution related operations can take this advantage. Es-
pecially, the image processing tasks normally involve large amount of data.

2.3.1 Discrete cosine transform

Discrete Cosine transform (DCT) is very popular for image processing appli-
cations. For example, it is the fundamental function of the image compression 
algorithm JPEG. The popularity of DCT is due to the performance advan-
tages over other linear transforms, such as Fourier Transform and Walsh-
Hadamard transforms. Some of these advantages are also beneficial to wa-
termarking algorithms.

For one dimensional signal the DCT is defined as:

\[
I_k = F[i_n] = \sum_{n=0}^{N-1} i_n \cos \left( \frac{\pi(2n + 1)k}{2N} \right) \tag{2.20}
\]

This definition is extensible to two-dimensional signal.

\[
I_{k,l} = F[i_{x,y}] = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} i_{x,y} \cos \left( \frac{\pi(2n + 1)k}{2N} \right) \cos \left( \frac{\pi(2m + 1)l}{2M} \right) \tag{2.21}
\]

The advantages of the DCT include:

- Decorrelation. The transform can successfully remove the correlation 
  between neighbouring pixels. The image decorrelation ability is almost 
  optimal (Clark, 1985).
• Energy compaction. This transform can compact image energy into as few coefficients as possible to allow the quantiser to discard insignificant coefficients without affecting the reconstruction quality.

• Separability and symmetry. The transform can be separated to row transforms and column transforms, and the transform matrix can be precalculated to simplify the process.

• Orthogonality. The transform matrix is orthogonal so that $A^{-1} = A^T$.

DCT has significant advantages in applications of image processing, comparing with Karhunen-Loève (KL) transform (KLT) and Discrete Fourier transform (DFT). KLT's basis functions are formed by statics of image data. It is adaptive, thus it has optimal energy compaction. However, the kernel is generally not separable such that the fast algorithm is not easily possible. The overall computational complexity of KLT is significantly higher than DCT and DFT. The DFT has linear kernel, and it is separable and symmetry, like DCT. However the DFT has complex form, meaning both magnitude and phase of the image will be coded. Studies show that the DCT has better energy compaction than DFT (Clark, 1985).

2.3.2 2D filter design

The perceptual quality is decided by the HVS phenomena, especially the frequency sensitivity. The watermarking algorithms must be able to select perceptually non-significant frequency component to apply the watermarks. The sophisticated signal processing algorithms, such as frequency selective filters are capable to make the selection. Different from ordinary communication signals, an image is a 2-dimensional data set, so is the frequency spectrum. Therefore the filter should also be designed in 2-dimension. There are some different ways to design a 2D filter, in the Appendix B, one of the design processes is described in details.3

Generally, the impulse response of a 2D frequency selective bandpass filter is described as,

$$h(n_x, n_y) = \text{sinc}(\omega_{2x}n_x) \cdot \text{sinc}(\omega_{2y}n_y) - \text{sinc}(\omega_{1x}n_x) \cdot \text{sinc}(\omega_{1y}n_y)$$ \hspace{1cm} (2.22)

Where

$$\text{sinc}(x) = \begin{cases} 1, & x = 0 \\ \frac{\sin(x)}{x}, & x \neq 0 \end{cases}$$ \hspace{1cm} (2.23)

---

3Since the volume of description, it is excepted from the main context.
This bandpass filter has passbands at:

\[ \omega_x = \begin{cases} 
1, & |\omega_{c1x}| < |\omega_x| < |\omega_{c2x}| \\
0, & \text{elsewhere}
\end{cases} \quad (2.24) \]

and

\[ \omega_y = \begin{cases} 
1, & |\omega_{c1y}| < |\omega_y| < |\omega_{c2y}| \\
0, & \text{elsewhere}
\end{cases} \quad (2.25) \]

It is worth noting that the system designed here is a linear time-invariant (LTI) system with a finite number of impulse responses (FIR). FIR filters have guaranteed linear phase output, which is very important for image perceptual intelligibility. Additionally, it's structural and computational simplicities make it a satisfying choice. Non-linear systems are also available to image processing, but their complexity and computational expense do not have significant advantage over the linear systems.

Another point worth noting is that the system shown here is based on 2D rectangular shaped ideal spectral response, since the images are normally presented as a matrix data set. Other design methods, for example the 2D circle shaped ideal response template, are also available (Rabiner and Gold, 1975, pp. 445-446).

2.4 Watermarking algorithms

In this section, the milestone works are generally reviewed to establish the technical background of this work. Watermarks had been designed as random small alters to the pixel values of images. Only until the process is designed as the communication system with side information (Cox et al., 1999), the performances are remarkably improved.

2.4.1 Early development

Cox et al. (1997) proposed a secure spread spectrum coding scheme for digital watermarking multimedia content. The modulation method could have the existence of the transmitted signal hidden from malicious receivers. By embedding the watermark signal in much wider bandwidth and very low power in each frequency bin, the distortion led by the additional watermark energy is significantly reduced. Only the intended receiver have the knowledge of how the signal is spread. A correlation or matched filter demodulator can be used to demodulate the signal. Then a detection scheme is applied. An easy and fast decision scheme is hard decision according to preset threshold. For
optimal performance, a soft decision may be used. In Cox et al. (1997), they use the correlation coefficient function and hard decision.

It is worth noting that the embedding functions proposed in Cox et al. (1997)

\[ x_n = c_n + \alpha w_n \]  \hspace{1cm} (2.26)
\[ x_n = c_n [1 + \alpha w_n] \]  \hspace{1cm} (2.27)
\[ x_n = c_n (e^{\alpha w_n}) \]  \hspace{1cm} (2.28)

The Equation 2.26 is the most straightforward implementation, but it is also the least robust. On the contrary the Equation 2.27 takes the possible large magnitude difference into consideration, that the embedding strength is proportional to the magnitude. The Equation 2.28 changes the large difference in logarithm scale, \( \ln[x_n] = \ln[c_n] + \alpha w_n \). Cox et al. (1997) described the watermark sequence \( w_n \) as AWGN with \( \mu = 0 \) and \( \sigma^2 = 1 \), \( N(0,1) \).

Smith and Comiskey (1996) proposed a similar scheme. However, in their work, the spreading sequences are like the class used in wireless communication modulation, a bipolar binary sequence with +1 and -1 values. They also introduced tiled version of direct spread spectrum modulation, which is claimed to be more robust and computationally less complex. Smith and Comiskey (1996) also proposed a "dual-rail" modulation method.

Bender et al. (1996) described a method that alters a set of pixels with +\( \Delta \) and another set with same size is changed with -\( \Delta \). It is same as the spread spectrum watermarking using \( \pm 1 \)s with strength \( \Delta \).

Some authors proposed the new design of the watermarking sequences. For example Mayer et al. (2002) proposed a blind watermarking detection method using an orthogonal sequence set. The decision is made in favour of the sequence that has the maximum correlation to the estimated watermark. Mayer et al. (2002) only assumed the sequences are uncorrelated to image and the prediction error. However, the proof of the noncorrelation is not presented.

2.4.2 Informed embedding

Costa (1983) proved that the channel capacity is not affected if one of the additive noise sources is completely known to the transmitter. The encoder can choose codes for the signal vectors in the direction of the known noise vectors rather than fight to cancel it. Cox et al. (1999) have derived this communication model to watermarking applications, namely informed embedding. Later, Miller et al. (2000) discussed four embedding strategies presented in the pseudo vector space.
In recent years the concept of informed embedding has been exploited by a number of researchers who realised this concept through various ways. In general, three classes of informed embedding techniques include:


- Straightforward dirty-paper coding schemes proposed by Miller et al. (2004) and Abrardo and Barni (2005).

- Dirty-paper code based on quantisation/restoration schemes proposed by Chen and Wornell (2001), Eggers et al. (2003) and Pérez-González et al. (2005).

In the following content, reviews of significant contributions of these works are given in the order as listed above.

Malvar and Florêncio (2003) proposed an improved spread spectrum watermarking method. A unique embedding function was proposed. An improvement of robustness was introduced by weighting the watermark signal with a function of cover signal vector, namely \( s = x + \mu (xb)u \). Where \( \mu \) is a weighting function of the projection \( x \) of cover signal vector \( x \) on the watermark vector \( u \), where \( x = (x - u)/||u|| \), and \( b \) is information bit. A simple implementation is letting \( \mu = (ab - \lambda x) \), namely linear approximation. This model can analogue the traditional spread spectrum embedding, when \( \alpha = 1 \) and \( \lambda = 0 \). Through observing the statistics of received signal

\[
r = \frac{(y · u)}{||u||} = \alpha b + (1 - \lambda)x + n \tag{2.29}
\]

the more \( \lambda \) is set close to 1 the more interference of the cover signal vector is removed from \( r \). It is also pointed out this model does not limit the embedding distortion level. Therefore, two non-linear processes were used to limit the maximum distortion. Although this algorithm was claimed to have superior reliability performance, this result is achieved through reduction on payload. Pérez-González et al. (2005) pointed out this weakness of the ISS algorithm that the spreading factor is 3000 times smaller than the scheme in (Pérez-González et al., 2005) at the same BER performance.

Mayer and Silva (2004) proposed a spread spectrum based informed watermarking method. The watermark vector \( w \) of length \( M \) is presented as:

\[
w_j = \alpha b_j s_j \tag{2.30}
\]

where \( s_j \) is one of a set of orthogonal sequences having values of \{-1, +1\}. \( b_j \) is the \( j \)th of \( M \) bits of the message. The detection decision is made upon the
sign of the inner product of the watermarked vector and the corresponding reference vector.

To improve the embedding efficiency and minimise the interference of the cover vector, a reference watermark vector \( \mathbf{w}_j^* \) is constructed. A set of local embedding strength \( \alpha_j \) is used to weight individual the watermark vector.

\[
\mathbf{w}_j^* = \alpha_j \mathbf{b}_j \mathbf{s}_j
\]  

(2.31)

The weighting function is obtained by solving the following linear matrix equation for the desired robustness \( \beta \):

\[
\begin{bmatrix}
A_{11} & A_{12} & \cdots & A_{1N} \\
A_{21} & \ddots & \cdots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
A_{N1} & \cdots & A_{NN}
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_N
\end{bmatrix}
=
\begin{bmatrix}
\beta b_1 - R_1 \\
\beta b_2 - R_2 \\
\vdots \\
\beta b_N - R_N
\end{bmatrix}
\]  

(2.32)

Where \( A_{ij} = b_j (s_i \cdot m_j \ast s_j) \) and \( R_j \) is the projection of \( j^{th} \) reference vector \( \mathbf{s}_j \) onto the \( j^{th} \) cover vector. Through solving the Equation 2.32, the local weighting factor \( \alpha_j \) can be obtained. The weighting factors, \( \alpha_j \), have two purposes:

- Making the cover signal vector has zero projection to the relevant watermark \( \mathbf{s}_j \).
- Providing enough robustness energy.

This scheme has a major flaw, according to the definition in Equation 2.32, the matrix \( A \) must span the \( \mathbb{R}^N \) space to obtain solutions of \( \alpha_j \) (Lay, 2002, Theorem 4, pp.42-44). But there is no guarantee that the vectors in \( A \) are base vectors.

Miller et al. (2004) proposed a trellis based dirty-paper coding system. Their trellis codes each watermark bit with 64 possible codewords, which are randomly drawn from independent, identically distributed Gaussian distributions. One codeword which is closest to the cover vector is chosen and is added to the cover vector. Since all codewords are randomly generated, the robustness analysis is difficult, so that an iterative process has to be used to adjust the watermarked vector into the decoding region. This informed embedding method produces very high perceptual distortion (\( d_W=161 \)) before they applied perceptual shaping algorithm.

Abrardo and Barni (2005) proposed informed embedding methods with spherical codewords. This method adopts Costa's (Costa, 1983) channel model and coding/decoding algorithms. Two dirty-paper coding methods
were introduced and compared, namely, orthogonal dirty-paper code and quasi-orthogonal dirty-paper code (Gold sequence). The code book for orthogonal codewords is a non-binary $n$-square unitary matrix. The codeword that has the maximum correlation with the cover vector is chosen from the predefined code collection which is associated with the desired message.

The analysis was focused on minimising the embedding distortion at a fixed robustness level, presented by the probability of error. It is estimated by the maximum pairwise error probability estimation (Proakis, 2001).

$$P_e(m, q) \approx \frac{1}{2} \exp \left( - \frac{c_m^T(u_m - u_q)}{\sqrt{2}\sigma_n\sqrt{||u_m - u_q||}} \right)^2$$  \hspace{1cm} (2.33)

Where $\sigma_n^2$ is the estimated attack power which the watermark is designed to be robust. The embedding distortion constraint $S$ is derived from the maximum pairwise error probability (the maximum value of Equation 2.33 is denoted by $P_e^*$, and the robustness requirement, $\sigma_n^2$).

$$S = 2\sqrt{\sigma^2\log\left(\frac{1}{2P_e^*}\right)}$$  \hspace{1cm} (2.34)

The algorithm compared with quantisation based code is robust to gain attack. The system also shows a significant improvement when concatenated with Turbo code as outer coding to the dirty-paper code.

The results presented in that paper are obtained by Monte Carlo simulations rather than real images. It is well known that the images are not simply statistics. Every image is different in sense of visual information rather than statistical information. The analysis provided in their paper did not show any advantages of orthogonal codes. Even more, since the real images were absent, it's impossible to discover the truth of visual distortion. However these questions will be answered in this thesis (Chapter 4).

Chen and Wornell (2001) introduced a different class of high performance watermarking algorithm, namely “Quantisation Index Modulation”. This method alters the cover signal vectors to a nearest quantisation reconstruction point, which is one of an ensemble of quantisation functions. The number of quantisation functions in one ensemble determines the embedding information rate. The quantisation error determines the embedding distortion. The minimum distance between different sets of quantisation reconstruction points, $d_{\text{min}}$, determines the robustness of embedding. This method has better robustness performance to additive noise than spread spectrum methods, by reducing or eliminating the interference of cover signal. However, this
class of informed watermarking algorithms (Chen and Wornell, 2001; Eggers et al., 2003), applying lattice codes for dirty paper coding, as pointed out by Malvar and Florêncio (2003) and Abrardo and Barni (2005), is not robust to scaling attacks.

Since this work has started with spread spectrum embedding methods and it is expanded to orthogonal dirty paper coding, this research work has little connection with the quantisation based watermarking algorithms.

2.5 Watermark synchronisation

Some schemes have been proposed to extract geometrical invariant vectors from the cover image and these vectors are modified according the watermark vectors. After inverse extraction process, the watermark information is embedded to the cover image. By applying the same extraction process prior to detection, the watermarked vectors can be retrieved or the presence of watermarks can be verified. For example, Lin et al. (2001) have illustrated this method in great details. To avoid repeating, only several points are remarked here. Lin et al. (2001) mentioned, that the inverse log-polar mapping is inherently unstable. An iterative interpolation must be used “three or four” times to satisfy the robustness requirement. Along with the inaccuracy of this estimation, the computational cost is also expensive. It is worth noting that the extracted signal is invariant to both translation and scaling, but the rotation parameter has to be estimated to the nearest degree, through an exhaustive search for 90 degrees. When an ID document is scanned, the most likely geometrical distortion is the rotation. If an algorithms cannot discover the rotation in sub-pixel level precision, and repeating the operation so many times to complete the exhaustive search, it is not reliable and fast enough for ID document authentications. Finally, since the scheme is established in the geometric invariant Fourier-Mellin domain, to avoid the severe implementation difficulty of inversion transform, the objective is set to approximate the stegoimage and the watermark, thus only a single bit output is possible. The scheme can only accomplish one purpose defined this thesis, whether the watermark presents.

Schemes like the one described above, assume that the watermarked image is the only object captured for detection. That captured black background is also commonly assumed. The black background is equivalent to zero-padding. If a white or textured background is obtained, the characteristics of the spectral power density will be changed. But for the application of watermarks in ID documents, the watermarked images rather need to be located and cropped from such background. A scheme which can identify
the location, size and RST parameters of the image is required for this application.

Honsinger and Daly (1998) claimed a scheme to re-synchronise the geometrical distortions that have applied to the image. The claimed scheme has two invisible markers embedded in two corners of an image. After calculating the autocorrelation, two peaks can be generated. Because the cover image power is significantly higher than the power of markers and commonly image energy concentrates at the low frequency band, a simple high-pass filter is applied to reduce the effect of image energy. The relative locations of markers are used to calculate the rotation angle and scaling ratio of the image. Thus, appropriate transforms are able to reverse the distortion. The reverse processes are commonly called rectification. Some experimental results indicated that the proposed filter is not sufficient to distinguish the autocorrelation peaks from background noise. This scheme also requires to share some knowledge to the detectors, such as the distance and the relative angle.

Kutter (1999) proposed embedding same watermark to shifted multiple locations in the cover image. The autocorrelation of the estimated watermark can reveal the geometrical distortions applied to the stegoimage. Kutter (1999) defined a $7 \times 7$ prediction function to estimate the embedded watermark. The function is a simple high-pass filter, and it is used to estimate the weak, but normally high frequency watermark signal from much higher image energy that normally concentrates in low frequency bands. The fundamental principle of the both early schemes (Kutter, 1999; Honsinger and Daly, 1998) is very close. Since the geometric transforms apply to all identical watermarks, thus the watermarks will remain identical after the transform. The autocorrelation function can review the peaks and the locations of the peaks reveal the RST parameters. Both schemes used similar linear high-pass filters to isolate the watermark signal from the image energy. One drawback of these schemes is that after the detection of autocorrelation peaks, the geometric distortions are not directly revealed. Linear regression approaches must be applied to the locations to calculate the number of rotation and scaling. Wrongly matched peak locations will result in inaccurate calculations.

Alattar and Meyer (2003) proposed a similar autocorrelation scheme. Differently the image is converted to the log-polar mapped Fourier domain. The image energy is reduced by a “non-linear high-pass filter”. Then a Phase-Only-Matched filter, which is equivalent to the phase-only autocorrelation, is used to detect the correlation peaks. Because this scheme is commercialised\textsuperscript{4}, in Alattar and Meyer (2003), the detail of the filtering process

\textsuperscript{4}Both authors are employees of Digimarc Corp. Evidently, this scheme is used in a
are not clearly described and there is no comprehensive measurement of the effectiveness presented.

Liu et al. (2005) proposed a new filtering scheme to recover the RS information of the distorted image. Like (Alattar and Meyer, 2003), the phase-only matched filter is used. It is because of a well known fact that the phase information maintains more intelligibility information than the amplitude information. Unlike Alattar and Meyer (2003), this scheme uses a small part of the original work as a re-synchronisation template instead of an additional watermarks. Therefore, their work is evidently closer to the image matching techniques. Obviously, using a part of the image as template is too expensive for ID document application. Rather, the mass distribution of millions templates is not feasible.

The schemes described above are not suitable to the application defined in this work. The re-synchronisation must be performed prior to watermark information detection, and the synchronisation markers must be additional watermarks. Liu et al. (2005) provided the closest solution, but the implementation must be carried out with watermarks that have much low power than an image template. Since it is only the spectral magnitude that is invariant to the geometric distortions, to extract synchronisation watermark signal from watermarked image is more difficult than using phase function. By combining the advantages of above schemes, a new re-synchronisation scheme using watermarks as re-synchronisation markers is proposed in Chapter 6.

commercial product of Digimarc Inc., ImageBridge™ product.
Chapter 3

Orthogonal Spread Spectrum

3.1 Conventional Spread Spectrum

The conventional spread spectrum model of watermarking uses a correlation based demodulator and hard decision detector. The embedding and detection of the $i$-th bit can be described as

**Embedding:**

$$x_i = c_i + \alpha \cdot s_i \cdot b_i$$  \hspace{1cm} (3.1)

where, $i = 0, 1, \ldots, M - 1$

**Detection:**

$$\hat{b}_i = \text{sign} (x_i \cdot s_i)$$  \hspace{1cm} (3.2)

where $c_i$ and $x_i$ are the $i$-th original and watermarked cover vectors respectively. $\{b_i\}$ is the watermark message with $M$ binary bits, and $s_i$ is the spread spectrum sequence for the $i$-th bit and it consists of pseudo-random generated bipolar binary values $\{+1, -1\}$. $c_i$, $x_i$ and $s_i$ all have the same length, denoted by $L$. $\alpha$ denotes the embedding strength. It can be a scalar that is uniformly applied to throughout the entire of cover image or a vector consisting scalar factors for individual cover vectors or pixels. For simplicity, $\alpha$ is treated as a fixed scalar for any given cover vector in the following analysis.

Equation 3.2 is the linear correlation receiver, providing an estimation of data $\hat{b}$. By substituting Equation 3.1 into Equation 3.2, one obtains the
energy distribution of this watermark communication channel.

\[ \phi = [c + a \cdot b_i \cdot s] \cdot s \]  
\[ = I_i + S_i, \quad i = 0, 1, \ldots, M - 1 \]  

(3.3)  
(3.4)

where:

\[ I_i = \sum_{j=Lx_i}^{Lx(i+1)-1} [c_j \times s_j], \quad i = 0, 1, \ldots, M - 1 \]  
\[ S_i = \sum_{j=Lx_i}^{Lx(i+1)-1} [a \times b_i \times s_j \times s_j] \]  
\[ = L \alpha^2 b_i, \quad i = 0, 1, \ldots, M - 1 \]  

(3.5)  
(3.6)

The correlation between the spreading sequence and the cover vector is unknown, and the value of \( I_i \) is random. Thus \( I_i \) can be seen as a source of interference. The energy of \( I_i \), \( E_I \), for detection of one bit is

\[ E_I = \sum_{j=0}^{L-1} |c_j \times s_j|^2 \]  

(3.7)

The energy of signal \( E_S \) is

\[ E_S = \sum_{j=0}^{L-1} |a \times s_j \times s_j|^2 \]  
\[ = L \cdot \alpha^2 \]  

(3.8)  
(3.9)

From Equation 3.3, Equation 3.5 and Equation 3.6, the detection error probability is defined as:

\[ P_e = \Pr\{E_I \geq E_S\} \]
\[ = \Pr\{\langle c \cdot s \rangle \geq L\alpha^2\} \]  

(3.10)

To maximise the detection reliability, the inner product \( \langle c \cdot s \rangle \) must be minimised. In the presence of noise, the minimised term can provide more robustness reliability, since it is clear that

\[ P_e = \Pr\{\langle c \cdot s \rangle + \langle s \cdot n \rangle \geq L\alpha^2\} \]  

(3.11)

The minimised inner product term can be obtained through orthogonalisation operations.
3.2 Orthogonal Spread Spectrum

From Equation 3.7, if the spreading sequence is orthogonal to the interference, the \( E_I \) will have no contribution to error detection, such that the detection is optimal. The orthogonal spread spectrum (OSS) watermarking has the detection performance better than the QIM algorithms proposed by Chen and Wornell (2001). Because of the use of correlation demodulator, the embedding occurs entirely in the projections of \( c \) onto \( w \), so that the projection of embedded signal vector is

\[
x = \tilde{c} + \alpha
\]

Since \( c \) and \( w \) are orthogonal, the term \( \tilde{c} = 0 \) in Equation 3.12. Then the SNR of OSS is

\[
\text{SNR}_{\text{oss}} = \frac{4LD_s}{P(\tilde{n})}
\]

and the SNR of STDM is

\[
\text{SNR}_{\text{stdm}} = \frac{3LD_s}{P(\tilde{n})}
\]

where \( D_s \) is the expected watermark distortion \( D_s = E[1/L\|c - x\|^2] \) (Chen and Wornell, 2001). \( P(\tilde{n}) \) is the energy of noise projection on watermarks. Since the noise terms in both Equation 3.13 and Equation 3.14 are the same, the OSS algorithm exhibits slight advantage over the STDM.

Theoretically, the existence of sequences that are orthogonal to the cover vector in \( \mathbb{R}^L \) is guaranteed, because for any value of \( L \), there are \( L \) sequences mutually orthogonal in \( \mathbb{R}^L \). One of the \( L \) sequences is given as the power normalised cover vector. The other \( L-1 \) sequences can be found through one of the orthogonalisation schemes, such as Gram-Schmidt, Householder transformation and Givens rotations. However, these sophisticated methods do not guarantee uniform amplitude outputs. Since the spreading sequences are added to the image, the sequence amplitude contributes to the image pixel intensity. Using sequences having non-uniformed amplitude will produce irregular alteration on image pixel intensity.

Geometrically, the cover vectors are in the \( \mathbb{R}^L \) space. A binary bipolar sequence is in the same space. The inner product of two vectors (\( c \) and \( s \)) is expressed as

\[
\sigma = \langle c \cdot s \rangle = |c| \cdot |s| \times \cos \theta
\]

where \( \theta \) is the angle between two vectors \( c \) and \( s \). If the angle \( \theta \in \{-\pi/2, \pi/2\} \), the inner product \( \sigma = 0 \). For each cover vector, the minimum correlation can
be obtained by finding the sequence vector that minimises $\cos \theta$. It is that the spread sequence vector has an angle with cover vector as close to $-\pi/2$ and $\pi/2$ as possible.

For some cover vectors the orthogonal binary sequence may not be available. Example is shown in 2-dimension. The total available bipolar binary sequences are $2^2 = 4$:

$$
\begin{bmatrix}
1 & 1 \\
-1 & 1 \\
1 & -1 \\
-1 & -1 \\
\end{bmatrix}
$$

(3.16)

In Figure 3.1, $c$ is an arbitrary vector in $\mathbb{R}^2$ and $s_1$, $s_2$, $s_3$ and $s_4$ are four possible sequences, given in Equation 3.16. The projection of sequences onto the vector $c$ are also shown. In the illustrated case, it is not possible to find an orthogonal sequence in the four vectors shown in the figure. Recall the aim of the orthogonalisation is not constrained to find a zero projection, rather the aim is to minimise the projection. Among all sequences in Equation 3.16, the one that has the minimum projection on the cover vector can be used as the watermark.

![Figure 3.1: Binary sequences and an arbitrary vector in $\mathbb{R}^2$.](image)

The case illustrated in Figure 3.1 is in the lowest dimension, normally $L \gg 2$. Exhaustive search of $2^L$ possible sequences involves unaffordable computations. In the following sections, three methods that generate binary bipolar sequences that approach orthogonal to the cover vectors are introduced.
3.3 Equal energy approach

Considering the use of correlation receiver and bipolar sequences separates the cover vector into two groups. Both groups have opposite signs, \{+1, -1\}. If both groups have equal energy, the output of the correlation detector will be zero, then the sequence that organises the cover vector elements into this status is orthogonal to the cover vector. In this section two methods that are designed under this principle are introduced.

3.3.1 Sorting and Cancelling

One method with simplicity advantage is introduced, and it is named as "Sorting and Cancelling" (SC). Given a \( L \)-length cover vector, the SC method consists of the following steps:

1. The elements of the cover vector are sorted, in descending order, according to magnitudes. An example of the sorted cover vector is shown in Figure 3.2.

2. A running inner product is carried out of the sorted samples so that at each stage, the sample is either added or subtracted from the running total. Between two values the smaller one will be updated a new running correlation. The sign of the operation is recorded as spreading sequence, in permuted order.

3. Repeating Step 2, until the spreading sequence reaches \( L \)-bit long, the operation terminates. The recorded sequence is inversely permuted according to the original locations of corresponding cover vector elements.

The method is explained as following. Let \( i \) denote the index of descending sorted version of \( c \), according to the magnitude, so that \( |c_{i-1}| \geq |c_i| \geq |c_{i+1}| \). Mathematically, the Step 2 is

\[
s_i = \min_{a \in \{1,-1\}} \arg \left( |\sigma_{i-1} + a \cdot c_i| \right)
\]

where \( \sigma_i \) is the running correlation. Since orthogonalisation procedure is designed to minimise the magnitude of the correlation between the sequence and the cover vector, thus the \( \sigma_i \) is described as

\[
\sigma_i = |\sigma_{i-1}| - |c_i|
\]
Equation 3.18 is the core function of SC. Because both $|\sigma_{i-1}| \geq 0$ and $|c_i| \geq 0$ are true, $|\sigma_i| < \max(|\sigma_{i-1}|, |c_i|)$ holds. Since the $\{c_i\}$ is convergent, thus $|c_i| \to 0$ as $i \to \infty$, furthermore Equation 3.18 becomes:

$$\lim_{i \to \infty} \sigma_i = \lim_{i \to \infty} |\sigma_{i-1}| - |c_i| = 0$$ (3.19)

Thus, the SC method is able to produce a sequence that is orthogonal the cover vector, if the number of elements tends to infinity.

From the mathematical definition of the SC method (Equation 3.18), the sequence can also be produced in the following manner. Let $s_i^c = \text{sign}[c_i]$ then $|c_i| = s_i^c c_i$, thus

$$\sigma_i = |\sigma_{i-1}| - s_i^c c_i$$

The following derivation is to find the $s_i$ in terms of operation signs. To produce a convergent sequence $|\sigma_i|$, it is obvious the $\sigma_{i-1}$ and the $|c_i|$ must have opposite signs, so, the $|\sigma_i|$ can be presented as

$$|\sigma_i| = \begin{cases} |\sigma_{i-1} + s_i^c c_i|, & \sigma_{i-1} < 0 \\ |\sigma_{i-1} - s_i^c c_i|, & \sigma_{i-1} \geq 0 \end{cases}$$ (3.20)

Let $s_i^c = \text{sign}[\sigma_{i-1}]$, Equation 3.20 can be expressed as:

$$\sigma_i = \sigma_{i-1} - s_i^c s_i^c c_i$$ (3.21)
In Equation 3.21, the left hand side and the first term on the right hand side are independent from any sign change operation, because only the magnitude information will be carried further. The second term on the right hand side of Equation 3.21 is the interest of operation. Three sign change operations alter the sign of each cover vector element. Thus the elements of the spreading sequence in the permuted order, $s_i$, can be derived from

$$s_i = -s_i^o \times s_i^c$$

$$s_i = -\text{sign}[\sigma_{i-1}] \times \text{sign}[c_i]$$ \hspace{1cm} (3.22)

The above generation scheme can be applied to an example shown in Table 3.1. The sum of the products $s_i$ and $c_i$ at each $i$ is zero. From the Table 3.1, the initial status of $s_i^o$ can be set as the information bit $b_j$, in order to complete the sequence search and information modulation in just one pass-through. The diagram is shown in Figure 3.3

<table>
<thead>
<tr>
<th>$i$</th>
<th>$c_i$</th>
<th>$\sigma_i$</th>
<th>$s_i^o$</th>
<th>$s_i^c$</th>
<th>$s_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>×</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>-7</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>-5</td>
<td>-5</td>
<td>-2</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-2</td>
<td>-2</td>
<td>-1</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3.1: Example of SC proceeding.

Figure 3.3: System diagram of SC method.
3.3.2 Orthogonalisation remainders

As discussed earlier in this chapter, binary sequences that are ideally orthogonal to the cover vector may not exist. Practically, the orthogonality is tolerable to a certain small value. The difference between the produced inner product and zero is named “orthogonalisation remainder” and it is presented as $\epsilon$. This is a quantitative measurement of the degree of orthogonalisation that achieved by methods. The minimum value of $\epsilon$ of a cover vector is denoted by $\epsilon^*$ and $\epsilon^* \sim 0$.

The equal energy orthogonalisation approach is based on the fact that under certain conditions, the cover vector elements can be separated into two groups so that the energy of both groups is almost equal. Orthogonalisation methods apply different strategies to implement the separation. The minimum difference between the energy of two groups is only dependent on the characteristics of cover vectors, such as distribution, limit and length.

As indicated by Equation 3.19, $\{|c_i|\}$ is monotonically convergent to 0. Reorganising the cover vector elements in such a way is optimal, because monotonically convergent sequence minimises the difference between neighbour elements. Three convergence models of the cover vector are used to evaluate the orthogonalisation remainder through SC method.

The first model shown in Figure 3.4 models the convergence of $\{|c_i|\}$ into a constant function $|c_i| = f(i) = a$. In this model, magnitude of all cover vector elements are assumed to be equal. That gives

$$\sigma_i = |\sigma_{i-1} - |a||$$  \hspace{1cm} (3.23)

This model analogs one kind of situation that spatial pixel values are used. In this case, it has 50% chance that the orthogonalisation remainder is
as large as the the last element of the cover vector. That is when the length of cover vector is an odd number. The optimal result can be obtained by letting the length of cover vector to be an even number.

The second model shown in Figure 3.5 models the convergence into a linear convergent function, $|c_i| = f(i) = a - \alpha \cdot i$, where $\alpha$ is the slope of the linear function and $a$ is a constant that $a = c_0$. Thus

$$
\sigma_i = |\sigma_{i-1} - |a - i\alpha|| \quad (3.24)
$$

This model can present situations that coloured DCT spectrum and non-constant pixel values. As indicated in Equation 3.24 the difference between any neighbouring pair of elements is identical. Therefore if the cover vector has even pair of elements the orthogonalisation remainder will be zero. It is proved as following:

$$
\begin{bmatrix}
c_0 & -c_1 & -c_2 & -c_3 & \cdots \\
\alpha & & & & \\
-c_3 & & & & \\
0 & & & & 
\end{bmatrix}
$$

It is worth noting that this model must be evaluated under one condition that $c_0 \geq \alpha \cdot L$. Only when this condition is satisfied, the evaluation holds, because if $c_0 < \alpha \cdot L$, $|c_i| = f(i)$ does not convergent when $i > c_0/\alpha$.

The third model is similar to Figure 3.2, and the convergent function is
modelled as $|c_i| = f(i) = a/i$ (Figure 3.6). In this case

$$\sigma_i = |\sigma_{i-1} - \frac{a}{i}|$$

(3.25)

Since all $f(i)$ is positive, the absolute value of $a/i$ in Equation 3.25 can be removed. Then the running correlation for rational model is derived from $f(1) = a$ onwards

$$\epsilon = |\cdots | a - \frac{a}{2} - \frac{a}{3} - \frac{a}{4} \cdots - \frac{a}{L}|

= a |\cdots | 1 - \frac{1}{2} - \frac{1}{3} - \frac{1}{4} \cdots - \frac{1}{L}|$$

(3.26)

From Equation 3.26, the desired $\epsilon$ can be obtained by solving this equation, and it is only dependent on the maximum value and sequence length. The solution for this model is not so trivial, thus a computer simulation is used to derive the answer. Results are shown in Figure 3.7, presented in logscale.

It must be considered that, numerically, the running inner product $\sigma_i$ produced by SC method possibly reaches zero before the sorted cover vector runs to the end, $\sigma_i = 0$ when $i < L$. If an extra state other than binary is allowed, the zero padding may be used to the bipolar binary sequence to protect the zero orthogonalisation. In more general cases, since the orthogonalisation is applied to real numbers, the zero value is unlikely to be produced. The zero padding protection may be applied to a considerably small value.

Another suboptimal situation is that the cover vector consists very a few high magnitude elements but largely with zero values. This situation
likely analogues the DCT coefficients of an area of pixels having low spatial frequency variations. Facing this situation, the SC method could not supply an optimal orthogonal sequence, rather zero padding the high magnitude elements that cannot be cancelled. However, such areas are visual sensitive, thus this situation can be avoid through high fidelity masking.

### 3.3.3 Sequence length

Let $s^*$ denotes the sequence that produces the minimum correlation with the cover vector $c$. The element-wise production is denoted by $q_i^* = c_i \times s_i^*$, so that

$$\sum_{i=0}^{\infty} q_i = \epsilon^* \approx 0$$  \hspace{1cm} (3.27)

$\epsilon^*$ is dependent on the distribution of the cover vector elements. The orthogonalisation algorithms are used to approach this minimum bound, thus $\epsilon \rightarrow \epsilon^*$.

The orthogonalisation algorithms can be seen as a special case of polynomial approximation algorithms. The $\{c_i\}$ is a subset of polynomials that is used to approximate the $q^*$ and $s^*$ is a set of weighting coefficients. The
orthogonalisation process can be expressed as

\[ \sum_{i=0}^{\infty} q_i = \sum_{i=0}^{\infty} s_i c_i \]  

(3.28)

Each \( c \) in Equation 3.28 is modelled as an impulse function with a magnitude. The magnitude difference between two neighbouring \( c \) is assumed to be arbitrary small. Thus, the combination of infinite number of \( c \) can achieve the optimal result \( q^* \), where \( \sum q^* \sim 0 \). However, practically, the length \( L \) must be finite. Thus \( c \) can only be a subset of the infinite expansion, such as

\[ \sum_{i=0}^{L-1} q_i = \sum_{i=0}^{L-1} s_i c_i = \sum_{i=0}^{L-1} q_i^* + \epsilon = \epsilon^* + \epsilon \]

This is similar to truncate the Fourier coefficients, and the "truncation error" is \( \epsilon \). Since the subset is obtained not from simple truncation, rather it can be simulated as a random sampling process from the complete set of polynomials, therefore the minimum orthogonalisation remainder \( \epsilon^* \) is dependent on the characteristics of the cover vector elements. It is clear that given a sufficiently large \( L \), the truncation error is minimised.

Another benefit for long sequence is derived from the power of information bit which is the autocorrelation over the vector length:

\[ \hat{b} = \alpha L + P(n) \]

where the noise term \( P(n) \) is the combination of orthogonalisation remainder and transmission noise. Optimally, higher vector dimension leads to higher signal energy, and lower noise energy. On the contrary, low signal energy and high noise energy are the consequences of low cover vector dimension. The SC method ties the fidelity and robustness. The signal to noise ratio (SNR) increases when the length of vector increases, so that less watermark energy is required thus preserving the fidelity. Meanwhile, since the SNR is enhanced, the robustness is also improved.

### 3.3.4 Group search

Another binary sequence orthogonalisation method based on equal energy principle is introduced in this section, namely "Group Search" (GS).

First, all coefficients are sorted according to magnitude. Then the ordered coefficients are separated into two groups, by grouping coefficients as even and odd index. Both groups are assigned with either a plus or a minus sign respectively. Through iterative swapping of the coefficients between two
groups, the difference between the sums of each group can be minimised.

The detailed procedures are:

- Separate cover vector elements into two groups. The optimal method is by sorting the elements in magnitude descending order. Put odd ranked elements in one group and even ranked elements into another (\{p_i\} and \{m_i\}).

- Assign one group with a plus sign and another with a minus sign.

- Calculate a look up table (LUT), which contains the difference produced by every pair of coefficients of both groups.

\[
\begin{array}{c|cccc|c}
   & p_0 & p_1 & \cdots & p_{l-1} & \sum p_i \\
\hline
m_0 & a_{0,0} & a_{0,1} & \cdots & a_{0,l-1} & \\
\vdots & a_{1,0} & \ddots & & \\
\vdots & & \ddots & & \\
m_{l-1} & a_{l-1,0} & \cdots & a_{l-1,l-1} & \Delta \\
\end{array}
\]  

where \( p_i = +c_i \) and \( m_i = -c_i \), and \( l = L/2 \).

- The difference (\( \Delta \)) between the sums of both groups is compared with all entries of the LUT.

- The pair producing the LUT entry that is closest to \( \Delta \) are swapped.

- The LUT is updated to respond the swap. It is followed by another swap operation, until the difference is minimised, then the iteration is terminated.

- The spreading sequence is produced by reordering the signs of both groups according to the corresponding coefficients.

Both SC and GS are designed under this same principle, with only operational differences, such as:

- In the SC, there is no constraint on the number of elements in both groups.

- In the GS, an iteration process optimises the result.
If the number of each group in GS is not required to be constant and equal during the search, the method is likely to reach a good result. But it dramatically increases the number of the searches through increasing the complexity of the LUT.

For GS, the elements are grouped by taking adjacent elements from the convergent sequence, hence the difference between two groups is minimised at the formation of the groups. Therefore, the number of swapping operations is minimised.

The operation is optimised through iteration processes. However, to minimise the number of comparison and swap, the length of both groups is fixed. Therefore, when the difference between both groups is very close to one half of an element magnitude, the GS does not change its group (sign). In this situation, the remainder is as big as half of the magnitude of this element.

### 3.4 Hadamard Matrix Search

Another orthogonalisation method is to find one sequence that produce the minimum correlation to the cover vector from a set of vectors which span \( \mathbb{R}^L \) space. Because of the orthogonality of basis vectors, the embedding can be considered as a CDMA communication system. As shown in Figure 3.8a, if there is no other noise source in the channel, the interference signal, which occupies the same band as the signal transmission, is the only interference affecting reception. It is the ensemble of all other communications but the desired one. In order to identify each transmission, a unique binary spreading sequence, chosen from a set of orthogonal or quasi-orthogonal sequences is allocated to each transmission. As shown in Figure 3.8b, because the cover vector is in \( \mathbb{R}^L \), and a set of binary basis vectors, \( \{s_0 \ldots s_{n-1}\} \) spans the same space, an arbitrary cover vector can be reconstructed by the linear combination of the basis vector set. The vector corresponding to the minimum weight has the least significance. The minimum coefficient present the minimum orthogonalisation remainder that the orthogonalisation method achieved. The alteration on the least significant component can minimise the effect to the whole combination and the correlation detector cancels out all other more significant components. Therefore, this vector can be used to carry the watermark information bit.

The Hadamard Matrix (\( H \)) is an orthogonal sequence set, and it has often been used to analyse signal "sequency spectrum" (Beauchamp, 1984). This matrix is used to decompose the cover vector. If one of the Walsh-Hadamard Transform (WHT) coefficients is very close to zero, the corresponding sequence in \( H \) has the quasi-orthogonality.
Figure 3.8: The comparison between CDMA and orthogonal matrix watermarking.

The Hadamard matrix search (HMS) is applied to the cover vectors through following steps:

1. Apply the WHT to a cover vector, using an $L$ ordered $H$.

2. Identify the minimum coefficient and the corresponding sequence.

3. Repeating 1 and 2 to all cover vectors constructed from the cover image.

A unique property of HMS is that the sequences in the set can be overlapped multiple times in the same cover vector. Since the orthogonality
between the sequences in the set, the overlap embedding will not affect the detection quality of embedded sequences in the cover vector. The implementation of the overlapping embedding only requires changing the step 2 to select \( k \) sequences which produce \( k \)-least inner products instead of one.

It is believed that the Maximum Length sequences used in CDMA communication systems can also be used in this method. The HMS method proposed here is only an example. Any other sequence set that spans the vector space of \( \mathbb{R}^L \) can also be used. The binary requirement is possibly relaxed if the application permits.

### 3.4.1 Orthogonalisation remainders

The HMS method is based on the WHT. The inner product of the chosen sequence and the cover vector is the minimum WHT coefficient. Since the Hadamard Matrix is a set of binary basis sequences which span the \( \mathbb{R}^L \) space, it can fully decompose any cover vector in \( \mathbb{R}^L \). A significant advantage of HMS is that WHT can be implemented through fast algorithms (Beauchamp, 1984). Since WHT have similar properties as the frequency transforms (DFT, DCT), and it is well known that images hardly have a wide spectrum, finding small coefficients is very likely. The following examples show the upper and lower bound for magnitude of WHT coefficients.

The first model of the cover vector consists values that have small variance or even constant, the WHT coefficients will contain a few high magnitude low sequency coefficients followed by small or even zero values at high sequency components. Such as:

\[
c = \{a, a, \ldots, a\}
\]

\[
c^T H = t
\]

Since all rows except row 0 of \( H \) have balanced \( \{-1, +1\} \)

\[
t = \{aL, 0, 0, \ldots, 0\}
\]

This kind of cover vectors can be formed by constant pixel values or transform coefficients of pixels that have a wide range of spatial frequencies, equivalently a wide spectrum in DCT (or DFT). This situation can be seen as the "performance upper bound" of the HMS.

On the contrary, if the cover vector is modelled as a sequence that consists of large number of zero values and a few high magnitude non-zero values, such as:

\[
c = \{a, 0, 0, \ldots, 0\}
\]

\[
c^T H = t
\]
where

\[ t = \{a, a, \ldots a\} \]

This example can be seen as the "performance lower bound" of HMS, since all WHT coefficients are same and are with high magnitude. This example is often found as DCT coefficients of an area with constant pixel values, thus \( a = DC \).

The above two examples are two extreme cases for the best and the worst. The WHT coefficients of cover vectors formed by the DCT coefficients commonly have values in between both situations. This also proves that the HMS orthogonalisation remainder is bound between \( \|c\|_1 = \sum c_i \) and 0.

When the overlapped embedding is applied to the same cover vector, it is believed that the orthogonalisation remainders will be larger, because the sub-optimal sequences are selected. The exceptions can be found near the performance upper bond. It is noteworthy that most images have high energy concentration in the low frequency band. White spectrum for images is not common. Thus HMS likely provides smaller orthogonalisation remainder in spatial domain embedding than frequency domain.

### 3.4.2 Sequence length

Imagine the cover vector is a set of discrete samples of a continuous signal (Figure 3.9). The sampling interval \( T = 1/L \) is presented in solid arrows. The dashed arrows present doubled sampling frequency, where \( T = 1/2L \).

![Double rate sampling](image)

Figure 3.9: Double rate sampling.

The WHT on this discrete sample set can be expressed as

\[ \gamma_c^L = \sum_{i=0}^{L-1} \left[ h_i^c \int_0^T c(t - iT) dt \right] \]
If the sampling frequency is doubled, the WHT coefficients

\[ \gamma^c_{2L} = \sum_{i=0}^{2L-1} \left[ h^c_{2i} \int_0^{T/2} c(t - iT/2) dt + h^c_{2i+1} \int_{T/2}^T c(t - iT/2) dt \right] \]

\[ \vdots \int_0^T c(t - iT) dt = \int_0^{T/2} c(t - iT/2) dt + \int_{T/2}^T c(t - iT/2) dt \]

\[ \therefore h^c_i \int_0^T c(t - iT) dt \geq h^c_{2i} \int_0^{T/2} c(t - iT/2) dt + h^c_{2i+1} \int_{T/2}^T c(t - iT/2) dt \]  \hspace{1cm} (3.30)\\

In above equations, \( \gamma^L_c \) and \( \gamma^{2L}_c \) denote the \( c \)-th WHT coefficient of the Hadamard transforms, using \( L \) and \( 2L \) length Hadamard matrices respectively. The \( h^c_i \) denotes the digit located at the \( i \)-th column and the \( c \)-th row of the Hadamard matrix. In Equation 3.30, the equality holds when \( h^c_{2i} = h^c_{2i+1} \), and the inequality holds when \( h^c_{2i} = -h^c_{2i+1} \). The Equation 3.30 proves that a smaller WHT coefficient can be found if the length of the cover vector is doubled.

### 3.5 Detection implementation

An orthogonal watermarking system, despite the orthogonalisation methods, requires to pass the exact watermark decoding sequences to the detector. The classic spread spectrum watermarking, such as the one proposed by Cox et al. (1997), was aimed to identify one watermark among hundreds of others. The watermarking system designed in this chapter aims to transmit a large amount of data through measuring the correlation peaks. The sequences that are designed orthogonal to cover vectors are able to minimise the detection noise caused by the correlation of cover vectors and decoding sequences. Since the watermarks are additional energy to the least correlation in each cover vector, the correlation output at detector may not be the least correlation after embedding. Due to this fact, estimations of the watermark without decoding keys are not reliable.

Therefore, the orthogonal watermarking is suitable to the one-to-one match authentications. By combining decoding sequences into one sequence, it can be used to determine the existence of the watermark, thus authenticating the watermarked document. Furthermore, breaking the combination, individual decoding sequences can be used to detect embedded information bits. All retrieved information bits, after decoding, are obtained for further processes. The SC and GS methods introduced earlier in this chapter are suitable to this kind of applications.
When the application consists a very large number of users, the distribution of sequences can be difficult. For example, in the application of watermarking the passport photos, the number of users can be millions, and the database containing information in this scale must be distributed to every entry port of the country or possibly every entry port in the world. The deployment and update to this database do not only rise logistical problems but also security alerts. More comprehensive way for large scale implementation is standardised codebooks securely combined with detector software, so that the detection process is complete with a single piece of software and it is scalable.

The HMS scheme has the potential to be implemented through codebook detection as well as the one-to-one match detection. Since the watermark is generated by sequences from a standard matrix, the embedder is only required to send the indexes of the decoding keys to the detector. The standard matrix has been referred to the Hadamard matrix earlier, but it can be replaced by any orthogonal sequence set. The index can also be replaced by the initial state of a sequence generator that is shared between issuers and inspectors. The matrix that is strictly shared between embedders and detectors is seen as the codebook. Comparing with decoding SC and GS sequences, the required side information of HMS detection is much less.

3.6 Simulation results

3.6.1 Preliminary results

The preliminary results are generated by a small number of simulations in order to verify the schemes, to determine optimal parameters of the different algorithms and generally model the channel.

10 600 \times 600 digital photos are used as cover images. A sequence of 341 randomly generated bits as the input message to simulate the (341, 205) LDPC error correction code, $M = 341$. Cover vectors are formed by reorganising tiled cover image pixels or tiled DCT coefficients. The length of cover vector is derived from the number of $8 \times 8$ tiles,

\[ E = \left\lfloor \frac{\text{Height}}{8} \right\rfloor \times \left\lfloor \frac{\text{Width}}{8} \right\rfloor \]
\[ = \left\lfloor \frac{600}{8} \right\rfloor \times \left\lfloor \frac{600}{8} \right\rfloor \]
\[ = 5625 \]
Thus the cover vector length for the preliminary results is:

\[ L = \left\lceil \frac{E}{341} \right\rceil \times 64 \]

\[ = 1024 \]

Each image is divided into 8 x 8 non-overlapping tiles. If spatial domain embedding is applied, every luminance pixel value is subtracted by the mean of the tile. \( J = 64 \) pixels are selected to form a tile. \( K = L/J = 1024/64 = 16 \) tiles form a spatial cover vector. If DCT domain embedding is used, the DCT is applied on each tile. \( J = 64 \) DCT AC coefficients, in every tile, are chosen. \( K = 16 \) tiles form a cover vector. Tiles are interleaved sequentially, in groups of \( M \) tiles. The interleave can minimise effect on information detection caused by image visual characteristics. Sequential interleave can minimise additional information required to reorganise tiles into cover vector, before detections.

The visual masking is implemented by two filters. One is a 3 x 3 high pass filter, as shown in Equation 3.31.

\[
\frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}
\]  

(3.31)

Another filter used in the comparison is constructed by the method described in Appendix B, and the magnitude of frequency response is shown in the Figure 3.10. The output magnitude of any visual filter is normalised between 0 and 1, then it is used to multiply \( \alpha \) as a mask of watermark power.

In the DCT domain, the coefficients of 8 x 8 blockwise DCT transformation are zigzag scanned into a linear order. By selecting part of coefficients from each block, frequency components are “filtered”. Therefore, no filter is necessary for embedding in DCT domain.

A pseudo-random binary generator, referred as “PN”, along with the three proposed watermark coding strategies, are evaluated and compared. The proposed schemes include the Sorting and Cancelling, referred as “SC”, the Hadamard Matrix search referred as “HMS” and the Group Search, referred as “GS”.

**Minimum power to error free detections**

This experiment is aimed at comparing the advantage of orthogonalisation schemes with traditional PN sequence embedding. Results are generated by
Figure 3.10: 2D band pass filter with $45 \times 45$ coefficients. The normalised
cutoff frequencies are 0.1 and 0.2
Table 3.2: Error free detection in spatial domain using 3x3 highpass filter (3HP) and 45x45 bandpass filter (45BP), respectively.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Filter</th>
<th>$\alpha_{\text{min}}$</th>
<th>PSNR</th>
<th>DWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN</td>
<td>3HP</td>
<td>5.050</td>
<td>-20.676dB</td>
<td>36.538dB</td>
</tr>
<tr>
<td>SC</td>
<td>3HP</td>
<td>0.200</td>
<td>16.145dB</td>
<td>66.617dB</td>
</tr>
<tr>
<td>HMS</td>
<td>3HP</td>
<td>5.050</td>
<td>-16.519dB</td>
<td>36.538dB</td>
</tr>
<tr>
<td>GS</td>
<td>3HP</td>
<td>0.100</td>
<td>11.690dB</td>
<td>70.518dB</td>
</tr>
<tr>
<td>PN</td>
<td>45BP</td>
<td>5.050</td>
<td>-15.219dB</td>
<td>32.203dB</td>
</tr>
<tr>
<td>SC</td>
<td>45BP</td>
<td>0.150</td>
<td>19.103dB</td>
<td>64.605dB</td>
</tr>
<tr>
<td>HMS</td>
<td>45BP</td>
<td>5.050</td>
<td>-9.729dB</td>
<td>32.203dB</td>
</tr>
<tr>
<td>GS</td>
<td>45BP</td>
<td>0.100</td>
<td>17.147dB</td>
<td>66.182dB</td>
</tr>
<tr>
<td>PN</td>
<td>X</td>
<td>5.050</td>
<td>-20.676dB</td>
<td>36.538dB</td>
</tr>
<tr>
<td>SC</td>
<td>X</td>
<td>0.050</td>
<td>11.602dB</td>
<td>71.715dB</td>
</tr>
<tr>
<td>HMS</td>
<td>X</td>
<td>0.050</td>
<td>13.013dB</td>
<td>71.259dB</td>
</tr>
</tbody>
</table>

embedding watermarks with gradually increased watermark power and detection results are examined for errors. If no error is detected the iteration will be stopped and the watermark power that produced the error-free detection is considered as the minimum. Comparisons are made in terms of PSNR and DWR. The PSNR is given by:

$$\text{PSNR} = \frac{1}{M} \sum_{i=0}^{M-1} \frac{P(S)^2}{(r_i - P(S))^2}. \quad (3.32)$$

where $r_i$ is the receiving correlation of each cover vector and it can be seen as the orthogonalisation remainder for each operation. $P(S)$ denotes the maximum signal power and it is a function of embedding strength, $P(S) = \alpha^2$. Since no additional noise is added, the difference between the received correlation and the signal power is the orthogonalisation remainder.

In Table 3.2, results of spatial domain embedding are shown with two different visual filters. From the values of watermark power, the advantage of orthogonal embedding is very clear. The PN sequence failed to reach error free within the fidelity restriction, $\alpha \leq 5.05$. From the value of PSNR, 16dB and 19dB are the highest value for both visual masked pixel domain embedding, and both fidelity values are produced by SC scheme. HMS produced 13dB which is the highest PSNR for non-filtered embedding. However the HMS does not produce good results in both spatial filtered conditions.

In the case of DCT embedding, the cover vector is formed by all DCT coefficients. Since the magnitude of coefficients varies dramatically, the em-
Table 3.3: Error free detection in DCT domain, using all DCT coefficients.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Power</th>
<th>PSNR</th>
<th>DWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC DCT</td>
<td>0.050</td>
<td>28.030dB</td>
<td>74.266dB</td>
</tr>
<tr>
<td>HNIS DCT</td>
<td>0.050</td>
<td>11.955dB</td>
<td>74.313dB</td>
</tr>
<tr>
<td>GS DCT</td>
<td>0.050</td>
<td>24.441dB</td>
<td>74.302dB</td>
</tr>
</tbody>
</table>

bedding strength $\alpha$ varies dependent on the magnitude of the corresponding cover vector coefficient. If any cover vector coefficient is smaller than 1, $\alpha$ will be weighted as by the coefficient, as Equation 2.27, otherwise unchanged, as Equation 2.26. Results are shown in Table 3.3. Better orthogonalisation performance than spatial domain embedding is seen. Both SC and GS reached over 20dB of PSNR, at 28dB and 24dB respectively. HNIS also produced 12dB of PSNR. It is clear that the embedding in DCT domain has better orthogonalisation performance and less visual distortions by all orthogonalisation schemes.

It is worth noting that since the visual masked embedding requires higher power to reach error free detection, the DWR values shows that fidelity is compromised. One can improve this shortfall by tolerating the compromised fidelity or reducing capacity requirement.

Visual masking effect

In this section, performance is evaluated under visual masking filters. The watermark power is fixed at a unit, $\alpha = 1$. The orthogonalisation schemes are examined under 3 x 3 highpass filter, 45 x 45 bandpass filter.

From Table 3.4, SC demonstrates the ability to work with different masking function. The visual masking only degrades PSNR by 2dB when the 45 x 45 bandpass filter is used, and by 7dB when the 3 x 3 highpass filter is used. On the contrary, HNIS is sensitive to both kinds visual masking filters. The degradation on the PSNR is around 60dB. GS also does not respond well with visual masking operations. The PSNR is degraded more than 50dB by 3 x 3 filter and 37dB by 45 x 45 bandpass filter. From the results, the comparisons of performance with or without the visual masking filter is clear, that for every scheme, the DWR is only increased by less than 10dB with either kind of filter, but the performance degradation is more than the gain on visual quality. Thus, the visual masking can be excluded from future test. Exception applies if visual quality has much more significant priority than data capacity.
Table 3.4: Performance of orthogonalisation schemes, affected by $3 \times 3$ highpass and $45 \times 45$ bandpass masking filter.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Filter</th>
<th>SNR</th>
<th>DWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC</td>
<td>3HP</td>
<td>30.101dB</td>
<td>52.638dB</td>
</tr>
<tr>
<td>SC</td>
<td>45BP</td>
<td>35.558dB</td>
<td>48.127dB</td>
</tr>
<tr>
<td>SC</td>
<td>x</td>
<td>37.624dB</td>
<td>45.694dB</td>
</tr>
<tr>
<td>HMS</td>
<td>3HP</td>
<td>-30.499dB</td>
<td>50.518dB</td>
</tr>
<tr>
<td>HMS</td>
<td>45BP</td>
<td>-23.709dB</td>
<td>46.182dB</td>
</tr>
<tr>
<td>HMS</td>
<td>x</td>
<td>39.035dB</td>
<td>45.238dB</td>
</tr>
<tr>
<td>GS</td>
<td>3HP</td>
<td>-17.246dB</td>
<td>50.541dB</td>
</tr>
<tr>
<td>GS</td>
<td>45BP</td>
<td>-3.756dB</td>
<td>46.254dB</td>
</tr>
<tr>
<td>GS</td>
<td>x</td>
<td>34.506dB</td>
<td>51.592dB</td>
</tr>
</tbody>
</table>

Table 3.5: Orthogonalisation remainder energy when zero padding for SC is used at different thresholds.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non</td>
</tr>
<tr>
<td>Spatial</td>
<td>50.678dB</td>
</tr>
<tr>
<td>DCT</td>
<td>41.808dB</td>
</tr>
</tbody>
</table>

Zero-padding for SC

As discussed earlier, to boost the SC performance, the cancelling operation can be terminated when a preset threshold is reached. Zero-padding is used to fill the sequence to the required length. In Table 3.5, three pre-set thresholds are compared with SC operation that generates full length sequence. The reduction on interference energy is dramatic, from 50dB to 11dB in spatial domain and 40dB to 11dB in DCT domain. However, since the sequence length is cut short, the signal energy is also reduced. The overall SNR performance does not achieve satisfying improvement.

Selection of DCT cover vector

The above results that are generated in DCT domain are produced by the cover vectors that are formed by all DCT coefficients. Since commonly the DC term in every DCT block possesses a high proportion of the block energy, all AC terms may not sufficient to cancel its energy. Thus it is believed that orthogonalisation without the DC term can reduce orthogonalisation
Table 3.6 shows results of the minimum power for error free detection and DWR. The error free detection can be reached by all orthogonalisation schemes at minimum strength where the increasing power iteration starts, despite the different cover vector length. However, The DWR values suggest that using shorter cover vector reduces watermark energy, thus it reduces embedding degradation to the cover image.

Table 3.7 shows results of orthogonalisation performance that are obtained by a unit power of watermark embedding.

Some remarkable results are shown in Table 3.7, such as the SC scheme can minimise the interference as low as -24dB, at $L = 1008$ and -14dB at $L = 512$. Shorter watermark reduces the interference energy by 3dB, for...
Table 3.7: Comparison on SNR and orthogonalisation remainder, in DCT domain for $L = 1024$, $L = 1008$ and $L = 512$.

HMS, but because of less signal energy caused by shorter watermark, the improvement on SNR is not significant. It is worth noting that the GS method has same principle as SC and different implementation, but the GS method is outperformed by the SC method 20dB at $L = 1008$ and 10dB at $L = 512$. Another drawback of the GS scheme is that its operation time and memory consumption. To finish one $600 \times 600$ the GS may take more than 1 minute. This is due to the update of LUT after every swap.

Remarks

After reviewing the preliminary results, some directions for further investigation can be drawn:

- Orthogonalisation procedures have better performance in the DCT domain than in spatial domain. Additionally, frequency domain embedding is more robust to additional attacks (Cox et al., 1997). Thus, future studies should concentrate on DCT domain embeddings.

- Visual masking filters can reduce embedding distortion, but the compromise to the detection quality is too high comparing to the gain of visual quality. Thus, no visual mask will be used in future investigation.

- GS has same principle as SC, but it is more complicated, and GS’s performance is no better than SC. Thus, GS is no longer required for future studies.

- DCT cover vectors lead better performance without the DC coefficients. Thus, DCT embedding should not include DC coefficients.
Shorter DCT cover vectors have good performance, and the shorter length reduces embedding distortion. Thus, low frequency DCT coefficients are used to form cover vectors.

3.6.2 Simulation results

Theoretical analysis of the performance has been shown earlier. But since definitive models of cover vector is hard to establish, the evaluation is preferred to be performed by computer simulation with real images. Based on the findings of the preliminary results, the performance is evaluated.

The objective is to successfully embed a payload of 1024 bits in a 512×512 image, with minimal distortion to the cover work. A typical length of cover vector is chosen as 128. As theoretically having better performance, the cover vector length of 1024 is also tested as a comparison. In order to obtain unbiased results a set of 50 ID photos, approximately 200,000 8×8 blocks, were used.

Embedding fidelity

The first result shows the fidelity of orthogonal sequence embedding, represented by DWR. For comparison purposes a pseudo-random binary generator, referred to PN, along with orthogonalisation methods, SC and HMS.

From a low level, the watermark embedding strength is increased, until the first error-free detection occurs, then the DWR is measured. Table 3.8 shows the result as the vector length \( L = 128 \) and \( L = 1024 \) respectively.

<table>
<thead>
<tr>
<th></th>
<th>PN</th>
<th>SC</th>
<th>HMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L = 128 )</td>
<td>22.6dB</td>
<td>73.7dB</td>
<td>51.3dB</td>
</tr>
<tr>
<td>( L = 1024 )</td>
<td>32.4dB</td>
<td>74.0dB</td>
<td>66.6dB</td>
</tr>
</tbody>
</table>

Table 3.8: Average performance of fidelity (DWR)

Among three algorithms, the SC sequences offers the highest fidelity, with the least DWR 73.3dB for 128-bit sequences and 74dB for 1024-bit sequences. The PN sequence, for comparison, gives the lowest fidelity, shown as the highest DWR. HMS sequences produces quite high fidelity. Although it not the best in the experiment, with 51dB and 67dB DWR the distortion made by the watermark is almost imperceptible. Both orthogonalisation procedures show a distinct advantage over the standard PN sequence method.

It is interesting that the HMS method requires \( L = 1024 \) for best results, whereas the SC method is already producing good results when \( L = 128 \).
The PN sequence cannot offer the similar performance, in the terms of fidelity, although it is the easiest method to implement. SC sequences provide the best performances overall. The HMS procedure shows good performance that increases with $L$. When cover vector length increases from 128 to 1024, the HMS sequences enabled the watermark energy to be reduced by 15dB, from 51dB down to 66dB.

**Statistics of minimum cross-correlation values**

This section gives results of the interference energy $E_I$, which is produced by the correlation of watermark decoding sequence and cover vector. The results can also be seen as statistics of the degree of orthogonality achieved by proposed orthogonalisation methods.

The results presented in Figure 3.11 are pdf's and cdf's of the orthogonalisation remainders produced by three different methods at $L = 128$. The inner product, $\langle c \cdot s \rangle$ of every cover vector is generalised between $10^{-9}$ and $9 \times 10^{-1}$.

In Figure 3.11a, particular interest focuses on the area where the peaks lie.

- For PN sequences, the resulting cross correlation values are concentrated between $10^{-1}$ and 1. In this region, the detector is likely to be confused between the signal and noise, and detection errors are likely.

- For HMS sequences, most of the minimum cross correlation values shows at one order of magnitude behind the PN curve. By a factor of 10, the detector can distinguish the signal from noise.

- For SC sequences, the majority of cross correlation values concentrated three orders of magnitude lower than PN, between $10^{-4}$ and $10^{-3}$. The margarine leads at least 1000 times less interference compared to the PN sequences.

Figure 3.11b shows the percentage of sequences can reach at desired levels of minimum cross-correlation. For example, if the interference is desired to be as low as $10^{-3}$, approximately 92% of the sequences generated by SC method can achieve this level. Only less than 3% of sequences generated by HMS can achieve the same level.

As discussed before, the different length of $L$ has an impact on the degree of orthogonalisation achieved by a method. For $L = 1024$, the minimum cross-correlation histogram and cdf are shown in Figure 3.12a and Figure 3.12b, respectively.
Comparisons can be made instantly, that the rank of performance generated by the three orthogonalisation methods does not change.

- PN sequences merely increase performance noticeably.
• HMS sequences reduce the minimum cross-correlations by one order of magnitude, comparing with the performance at $L = 128$.

• SC sequences make more improvement, reducing the minimum cross-correlation.
correlation by two orders, comparing to previous results.

The improvement on efficiency is shown as:

- PN sequences, have little improvement, comparing with results at $L=128$.
- HMS method has an improved result. Originally, only 20% sequences have cross-correlation smaller than $10^{-2}$. As the increased length, over 90% of sequences have achieved this level.
- Impressive result is produced by SC method. Almost all sequences produce cross-correlation magnitude below the $10^{-4}$ mark.

Overlapped embedding

The improvement led by the increased watermark sequence length is clearly displayed. However, the loss of capacity is inevitable. To overcome the reduction, overlapped embedding can be adopted, if it does not degrade the established performance. Because SC method has achieved remarkable performance at short length of sequence, so, the overlapped embedding will only be applied to the HMS method. The following results show practical feasibility of 8 sequence overlap embedding, using a modified HMS.

In both histograms Figure 3.13a and the cdfs Figure 3.13b, three lines are laid side by side, comparing the performance of overlapped and non-overlapped embedding. 8 sequence overlap embedding expectedly produces slightly higher cross-correlation values, and they are bounded by the performance of 1024-bit and 128-bit HMS.

The minimum cross-correlations histogram and cdf of overlapped embedding has a performance upper bound (on the left) $L = 1024$ single sequence embedding, but it is further away from the lower bound of $L = 128$ single sequence embedding. The cdf shows that at desired performance level, the number of sequences is lower than the high bound. However, despite the degraded performance, the overlapped embedding is very close to the high performance bound. The fidelity of the eight-sequence embedding is DWR=42.1dB, comparing with the single-sequence embedding’s DWR=51.3dB.

3.7 Summary

In this chapter, spreading sequences are designed to be orthogonal to the cover image as carriers of the watermark information. The sequences of this kind are able to minimise interference energy caused by correlation between
Figure 3.13: Performance of single sequence embedding and eight-sequence overlap embedding.

decoding sequence and cover image. It is shown that the energy of the cover vectors can be significantly reduced through various orthogonalisation
schemes that have been proposed in this chapter. Therefore the minimum requirement on watermark signal power can also be reduced, providing much better fidelity. Reduced interference energy also allows more advantage of signal energy against additional noise.

Three orthogonalisation algorithms have been presented. The schemes show good ability to reduce the native interference from the cover image. The SC procedure demonstrates the best performance, and it produces many impressive results. However, the large quantity of the spreading sequences have to be saved and be redistributed for detections and authentications. The cost efficiency and flexibility are restricted. HMS procedure provides good performance. Restricted by the size of Hadamard matrix, commonly quasi-orthogonal sequences are used with slightly big margins to orthogonality. But it is revealed that overlapped embedding technique by applying a modified HMS algorithm improves detection quality with a little compromise made to the image fidelity.

Generally, embedding in DCT domain allows better orthogonalisation performance and less visual distortions by all orthogonalisation schemes. The visual masked embedding requires higher power to reach error free detection, the DWR values shows that fidelity is compromised.

To make a better orthogonalisation, there is another very important factor need to be considered, the operation time. The GS method has good performance on DCT domain vectors. However the time performance is significantly higher than SC. For example to complete the operation on a 600 x 600 image, GS method takes more than a minute. Additionally, lack of flexibility is a major disadvantage comparing to GS. In GS, the numbers of elements in both groups are fixed to minimise the computational consumption of updating the LUT and exhaustive search.

Conclusively, this chapter provides a new direction of spread spectrum watermarking techniques. A new branch of informed embedding technology is identified. The improvement over conventional spread spectrum watermark embedding techniques is clear, in terms of capacity, fidelity and robustness.
Chapter 4

Biorthogonal Dirty Paper Code

4.1 Problem formulation

In recent years the research in watermarking and information hiding has been extended to the applications of writing on dirty paper model. In this model the signal is coded in favour of the known noise, thus the channel capacity is only dependent on other noise sources. This model in watermarking is referred to watermarking with side information (Cox et al., 1999). Some authors proposed new algorithms (Miller et al., 2004; Chen and Wornell, 2001; Malvar and Floréncio, 2003; Eggers et al., 2003) and new coding methods (Mayer and Silva, 2004; Abrardo and Barni, 2005). This chapter investigates the biorthogonal dirty paper coding method.

As illustrated in Figure 4.1, if one noise source \( c \) is known to the encoder, the channel capacity is solely determined by the ratio of signal power \( X \) to the power of additional noise, \( N \) (Costa, 1983). In this model the information is coded according to \( c \). Costa (1983) choose a codeword \( \mathbf{u} \) from \( \mathcal{U} \) according

![Figure 4.1: Dirty paper channel model, modified for watermarking](image-url)
\[ |(u_t - \alpha c) \cdot c| \leq \delta \]  

(4.1)

where \( \delta \) is an arbitrary small value, and \( \alpha \) can be optimised according to the channel situations (Costa, 1983). Equation 4.1 indicates to search a codeword so that \( u_t - \alpha c \) is nearly orthogonal to the cover signal vector. This also means that the codeword contains a component that is nearly \( \alpha \)-times of the cover signal vector.

The transmitted signal, in watermarking notation is \( x = w + c \) where \( w = u_t - \alpha c \). This transmitted signal leads to a distortion energy constraint at \( P = \frac{1}{L} \sum_{i=0}^{L-1} w_i^2 \). If \( N \) is designed to be the maximum channel noise power against which the watermark is robust, the optimal factor \( \alpha \) is obtained at

\[ \alpha = \frac{P}{P + N} \]  

(4.2)

The image watermarking system can be modelled having two receivers (Cox et al., 1999), but both receivers have different detection interests, so that Figure 4.1 can be changed to Figure 4.2.

![Figure 4.2: Complete watermarking channel model, based on Dirty Paper channel.](image)

In Figure 4.2, the “detector1” is a perceiving detector, most likely a human viewer. This detector is interested in the perception quality, presented as MSE and Watson distance \( (d_W) \). The interest of “detector2” is the message embedded in the cover image. The parameters that detector2 considers are bit error rate or probability of error \( (P_e) \) and communication channel capacity. Unlike electronic communication system, a watermarking system must protect both detectors interests. The encoder has to consider two constraints. As emphasised in Figure 4.2, the perceptual quality is constrained by DWR or \( d_W \), and the message detection quality is constrained by the channel noise power of \( N \). With Costa’s model, message detection quality is
independent of the interference led by the cover image energy but solely by
the distortion added after the message is embedded.

With all background information described, a complete "Writing on Dirty
Paper" watermarking scheme should consist

1. A codebook, \( U \), containing many codewords as candidates of the wa-
termark signal.

2. An algorithm which can effectively choose the right codeword \( u_i \) or a
combination of codewords from codebook \( U \), known as encoding rules.

3. A perception quality bound that restrains watermark energy for making
unacceptable distortions, visual masking.

4. A channel robustness constraint to maximise the watermark energy
ensuring the message communication quality.

5. And a message decoding algorithm to match the encoding algorithm.

In above list, item 3 and 4 are contradictory. The watermark power is de-
sired to be minimised by visual quality constraint, while it is desired to be
maximised by transmission quality constraint. The solution is to find an
appropriate trade-off, so that both conditions can be balanced.

It is necessary that to find a codebook \( U \) or an algorithm to construct a
codebook to produce optimised both message detection and visual quality.
In Costa (1983) the codebook is randomly generated Gaussian values. Al-
though the original paper did not obvious indicate, so, it is believed that the
codebook consists of real numbers. The codebook is searched to select one
codeword \( u_i \) makes \( (u_i, c) \) "jointly typical".

By definition, for an arbitrary ensemble \( \mathbf{x} \) with alphabet \( \mathcal{A}_x \) having \( N \)
symbols, the typical elements of \( \mathcal{A}_x^N \) have probability close to \( 2^{-NH} \) (MacKay,
2003). This typical elements set \( T_{NB} \) can be expressed with an arbitrary small
value \( \beta \) as

\[
T_{NB} \equiv \{ x \in \mathcal{A}_x^N : \left| \frac{1}{N} \log_2 \frac{1}{P(x)} - H \right| \leq \beta \}
\]  

(4.3)

This definition is sometimes called "asymptotic equipartition principle" and
it is equivalent to Shannon's source coding theorem (MacKay, 2003, p.165).

The "jointly typicality" is defined under the following conditions. Firstly,
define codeword \( \mathbf{x} \) from an ensemble \( X^N \). Consider that \( y \) is the correspond-
ing channel output as one random codeword is sent. The joint ensemble is
defined as \( (XY)^N \). This pair of sequences \( x, y \), both having length \( N \), are
defined to be jointly typical (to tolerance $\beta$) if (MacKay, 2003)

$$x \text{ is typical of } P(x) \rightarrow \left| \frac{1}{N} \log \frac{1}{P(x)} - H(X) \right| < \beta \quad (4.4)$$

$$y \text{ is typical of } P(y) \rightarrow \left| \frac{1}{N} \log \frac{1}{P(y)} - H(Y) \right| < \beta \quad (4.5)$$

$$x, y \text{ is typical of } P(x, y) \rightarrow \left| \frac{1}{N} \log \frac{1}{P(x, y)} - H(XY) \right| < \beta \quad (4.6)$$

In Costa’s model, the code book $U$ shared by encoder and decoder contains a portion of side information $c$, which is $u = w + \alpha c$. The amount of information of $c$ is determined by the factor $\alpha$. The value of $\alpha$ that was derived to reach the maximum channel capacity is $P/(P+N)$. Costa’s model used a jointly typical decoder. However, the typical set decoding algorithm is not optimal (MacKay, 2003, p.165).

In this chapter the dirty-paper coding method using biorthogonal codes and maximum likelihood decoder (Tomlinson et al., 2005) is discussed.

### 4.2 Biorthogonal dirty paper code

The biorthogonal dirty paper coding scheme consists a code set and an encoding/decoding algorithm. The major difference between other writing on dirty paper scheme and biorthogonal dirty paper coding scheme is that the code set is not a randomly generated binary set, rather a biorthogonal matrix, with distinguishable orthogonality structure. A trivial advantage of biorthogonal matrix over random generated codewords is that the Euclidean distance between code words is maximised thus detection robustness is maximised. Another advantage of biorthogonal coding is that the interference between codewords is zero, subject to appropriately designed binning scheme. The dirty paper codebook in this chapter is formed by rows of a biorthogonal matrix. The entire codebook is divided into $B$ bins. Each bin presents one symbol of $\log_2 B$ bits.

The encoding algorithm selecting one codeword, in one bin of codewords ($U^m$) that is associated to message $m$, is selected because it has the minimum Euclidean distance between $c$, so that

$$w = \arg \min_{u^m \in U^m} \| u^m_m - c \|$$

$$= \arg \max_{u^m \in U^m} (u^m \cdot c) \quad (4.7)$$

70
After embedding, assuming no additional noise, the information is retrieved by an optimal decoder (Cover and Thomas, 2006, p.201), namely Maximum likelihood decoder (ML):

$$
\hat{b} = \arg \min_{s \in \{0,1,\ldots,B\}} \| u_s^* - c \|
$$

(4.8)

The watermarking system is seen as dual receiver communication channel (Figure 4.2). Between two quality measurement, the message detection quality has priority. The design is focused on binary code, thus the biorthogonal Hadamard matrix (denoted by $bH$) is used as an example. $bH$ is built based on $L \times L$ Sylvestre-Hadamard matrix $H$

$$
bH = \begin{bmatrix} H \\ -H \end{bmatrix}
$$

(4.9)

To extend the number of codes as well as to provide secrecy of the codes, a binary scramble key $k$ is used. The key sequence has the length of $L$, and it consists $\{+1, -1\}$. One key will be used to alter all rows of $bH$ producing one code matrix, $U$ as the codebook. The scramble action is performed in such a way that in every row of $bH$, each element multiplies the corresponding element in the key sequence.

$$
u_r = h_r \cdot k
$$

(4.10)

where $u_r$ denotes the $r$-th row of the codebook, and $\cdot$ denotes the Hadamard product. It is important that the resulting code matrix $U$ keeps mutual orthogonality between all the rows. The following proof ensures that the scramble operation does not affect the Euclidean between and Hamming distance between codewords.

The $bH$ is a sphere in $\mathbb{R}^L$ space having the radius of $\sqrt{L}$. One scramble key rotates this sphere. Since the scramble key is also on the surface of the sphere, the rotation can take the reference of the first row of $bH$, $h_0$ and the scramble key. Since the row 0 of $bH$ is an all-one sequence, after the multiplication of a scramble key, it equals to the scramble key. In other words the $h_0$ is rotated to the location of the scramble key. All other rows of $bH$ are rotated by the same key vector, thus the code set structure remains unchanged.

It is well known that the $d_{\min}$ between any pair of $bH$ rows is half of the length (MacWilliams and Sloane, 1977, p.49). Hence the Euclidean distance is $\sqrt{4L/2} = \sqrt{2L}$. The the mutual distance of $bH$ based codes are

71
independent to the binary scramble key.

\[ d_E(u_i, u_j) = \| (h_i \cdot k) - (h_j \cdot k) \| \quad \forall i \neq j \]

\[ = \| (h_i - h_j) \cdot k \| \]

\[ = \sqrt{\sum_{c=0}^{L-1} (h_i[c] - h_j[c])^2 k[c]^2} \]

\[ = \sqrt{\sum_{c=0}^{L-1} (h_i[c] - h_j[c])^2} \]

\[ = \sqrt{2L} \quad (4.11) \]

where \( d_E(\cdot) \) denotes Euclidean distance function. The Equation 4.11 indicates that the scramble keys do not affect the mutual Euclidean distance between codewords in the same code set.

Hamming distance represents the bitwise disagreement between two sequences. With same index in both sequences, two bits have a distance of 1, if they are not same, otherwise, the distance is counted as 0.

\[ d_H(u_i, u_j) = \sum_{n=0}^{L-1} u_i^j \odot u_j^i \]

where \( d_H(\cdot) \) is the Hamming distance function.

\[ = \sum_{n=0}^{L-1} h_n^i \cdot k_n \odot h_n^j \cdot k_n \]

The sign of \( h_n^i \) and \( h_n^j \) are flipped by the same bit \( k_n \), therefore, \( h_n^i \cdot k_n \odot h_n^j \cdot k_n = h_n^i \odot h_n^j \)

\[ = \sum_{n=0}^{L-1} h_n^i \odot h_n^j \quad (4.12) \]

holds for all \( i, j \in \{0, 1, \ldots 2L - 1\} \) and \( i \neq j \).

Orthogonal sequence set is an efficient code structure. Since there is no mutual projection between codewords. The space is spanned with equally maximum distance. Because all codes are maximally separated, the robustness of the watermark is maximised. It is worth noting that the orthogonality is important to ensure this coding method optimal. If all codewords
are mutually orthogonal, not only the $d_{\text{min}}$ is maximised, also embedding one codeword has minimal affect on others. With the encoding and decoding rules, Equation 4.7 and Equation 4.8, watermark vector $w$ is coded with information bit $m \in \{0, 1 \ldots B\}$ as

$$w = \arg \min_{u^m \in U^m} d_E(c, u^m) \quad (4.13)$$

Let the output $w = u_j^m$. After the embedding $w$ then $x = c + w$, the decoding decision is made based on

$$\hat{b} = \arg \min_{s \in \{0, 1 \ldots B\}} d_E(d, U^s)$$

Referring to Equation 4.13, the watermark $w$ is the codeword that has the maximum cross-correlation with cover vector $c$. Thus

$$= \arg \max_{s \in \{0, 1 \ldots B\}} \{ \langle c \cdot U^s \rangle + \langle w \cdot U^s \rangle \}$$

$$= \arg \max_{s \in \{0, 1 \ldots B\}} \{ \langle c \cdot U^s \rangle + \langle u_j^m \cdot U^s \rangle \} \quad (4.14)$$

In Equation 4.14, the first term in the bracket produces the highest correlation as defined in Equation 4.13, which is the desired result. The second term, as the watermark codeword is either orthogonal to most codewords or produce $\pm L$. The $j$-th codeword in the $m$-th bin will produce the maximum correlation $L$. Another codewords $(-u_j^m)$ produces $-L$. Let $\langle c \cdot u_j^m \rangle = a$, where $a$ is a positive real number, and $L$ is also a positive real number.

$$a + L > a - L$$

holds for all $a > 0$ and $L > 0$. Therefore the message can be successfully decoded. The zero and negative correlations of the $w$ and other codewords will not affect the decision of Equation 4.14.

As the most robust choice of code, biorthogonal Hadamard matrix has the maximum $d_{\text{min}}$ that is linked to the watermarking performance in the following aspects:

- $d_{\text{min}}$ determines the robustness. Higher $d_{\text{min}}$ would prevent higher channel noise to move the watermark vector outside the desired decoding region, consequently, providing better detection reliability.

- To move the cover vector closer to the desired codewords yields embedding distortion that is constrained to $\Delta^2 = P$. If the cover vector is
closer to the rival codewords, \( d_{\text{min}} \) reflects the maximum power required to move cover vector to the desired decoding region.

\[
\begin{align*}
    d_E(u_i, u_j) &= \begin{cases} 
    0, & d_H(u_i, u_j) = 0 \\
    2\sqrt{2d_{\text{min}}}, & d_H(u_i, u_j) = L \\
    2\sqrt{d_{\text{min}}}, & d_H(u_i, u_j) = L/2
    \end{cases}
\end{align*}
\]

It is desired to be smaller, so that a small distortion is able to give reliable detections.

Both conditions are contradictory to each other. This contradiction reflects the nature of watermarking system (Decker, 2001).

### 4.3 Embedding and Detection

In this section the encoding and decoding algorithms are generally discussed, in order to establish the platform for discussions of code design and performance analysis.

As discussed in the last section, the scramble key has no effect on the \( sH \) code structure. Embedding with the code set has merely additional complication. The code word selection, described in Equation 4.14 can be examined at element level as following. The column vector of Euclidean distance is

\[
d_r = \sum_{c=0}^{L-1} c_h \cdot h^r_c \cdot k_c \tag{4.15}
\]

The last two terms in Equation 4.15 form the codebook, \( U \). The term \( g_r \) is independent from the scramble key \( k \). Therefore, all row vectors of \( G \) are independent of \( k \). Thus matrix \( G \) can be calculated before the calculation of distance vector.

\[
G^r_c = c_h \cdot h^r_c \tag{4.16}
\]

so that, the column vector of Euclidean distance is

\[
d = Gk^T \tag{4.17}
\]

Each entry in this vector indicates the Euclidean distance between the cover vector \( c \) and the \( r \)-th row of \( sH \) scrambled by the key, respectively. From Equation 4.16, it is clear that the distance vector is a special Walsh Hadamard Transform (WHT). If all elements of that key are are 1's, the Equation 4.17 becomes a column vector of WHT coefficients.
In Figure 4.3 the outer circle and all marks in the figure presents the projection of the $L$-dimensional codebook sphere over 2-D plane. The rotation caused by scrambling action is shown at the circle's border with the relative locations of the $\times$ and $\circ$, presenting the $\mathbf{H}$ and $\mathbf{U}$ respectively. $r$ is the distance between any point on the surface of the sphere to the origin. The $+$ and $\circ$ marks, inside the circle, present two code bins respectively to encode a binary message.

![Figure 4.3: Examples for encoding watermarks with proposed codes](image)

In the Figure 4.3, two examples are illustrated. Both are intended to encode bit "0", presented at $\circ$ points. Firstly, the watermarked vector $\mathbf{A}$ is mapped from the cover vector $\mathbf{a}$ through the encoding process described as following. Given $\mathbf{a}$, the encoder can obtain two closest points from both bins. The encoding process adding $\Delta$ times $\mathbf{u}_0^q$ to $\mathbf{a}$ obtains $\mathbf{A}$ closer to $\mathbf{u}_0^q$ and further away from $\mathbf{u}_1^q$, so that the maximum embedding constraint $P$ is satisfied. The distance between $\mathbf{A}$ and the desired codeword is $r^2_a$. If the watermarked vector $\mathbf{A}$ is inside the decoding zone, $z(\mathbf{u}_0^q)$, the decoding will be successful. The decoding zone is defined as the minimum Euclidean distance between $\mathbf{u}_0^q$ and neighbour codewords belonging to another bin. Note that for biorthogonal code sets the $z(\mathbf{u}) = d_{\text{min}}/2$, except $d_{\text{E}}(\mathbf{u}, -\mathbf{u})$.

The second example shows mapping cover vector $\mathbf{b}$ to $\mathbf{B}$. This example shows an error detection. The maximum embedding distortion constraint is not enough to alter $\mathbf{B}$ inside $z(\mathbf{u}_0^q)$. Therefore an error is declared by the encoder. Since the error is known to the encoder, so that some appropriate actions can be taken to avoid loss information. For example a special de-
coding mark can be added to notice the decoder that no information can be retrieved from this vector.

It is noteworthy that in Figure 4.3, the decoding zone \( z(\cdot) \) is shown in dashed circle, but it should be an \( L \)-dimensional polygon. Each side of a zone is at the middle of the distance between two codewords. When orthogonal codes are used, all the distances are equal to \( d_{\min}/2 \). However the accurate shape and its projection are very hard to estimate, so the circled approximation was drawn.

### 4.4 Design and performance

In this section, the encoding performance and code design are theoretically evaluated to find the error probability. There are two major factors could affecting the detection reliability.

- Channel noise \( N = \sigma^2 \), that is the combination of all intentional and non-intentional alterations to the watermarked image. Ruled by the Central Limit Theorem, the summary of these mutually independent variables can be surely modelled as the AWGN, \( N(0, \sigma^2) \).

- Embedding distortion constraint, \( \Delta^2 = P \), is the limit on watermark power. This constraint is quantitatively regulated by DWR.

#### 4.4.1 Code designs

Based on the analysis above, it is possible to derive practical code designs. One code book contains one code set and its negative counterpart. Due to the special properties of biorthogonal matrix, two classes of binning methods can be considered for binary messages. They are referred as \textbf{Class1} and \textbf{Class2} and described below.

In \textbf{Class1} codebook, the original code set is binned for one message bit, and its negative counterpart is binned for the other bit. In each bin the \( d_{\min} \) uniformly equals to \( \sqrt{2L} \), and it is independent from the scramble key, as proved in Equation 4.11. This class satisfies the condition that all codewords in one bin are mutually orthogonal, and two bins produce the exactly the same correlation with the cover vector but to opposite directions. The decoding algorithm determines the correct bit according to the best correlation. This codebook construction method maintains one full set of original orthogonal matrix in one bin. The codewords in one bin can fully decompose any cover vector in this space.
The Class2 codebook is constructed by grouping half of the orthogonal set and its negative counterpart to one bin and the other half and the negative counterpart to another bin. In such a way, both most correlated and uncorrelated codewords are in the same bin. One code bin covers opposite directions may eventually get closer to any cover vector than the Class1 codebook. However, half code set does fully span the space. It means that the codewords may have significant larger correlation in the rival code bin. If it is the case, this effect may cause the embedding distortion to severely high. However, since the codewords are equally spaced, this undesired effect should become less likely to happen when the dimension of the coding space growing higher. With the binary message level this class has very good performance (Xu et al., 2006a).

### 4.4.2 Probability of error

The following analysis starts with the noise free environment and it is followed by the error probability if noise presents. Apparently, if the $\sigma^2 < x(u_i^n) - \tau^2$, the error probability is of little difference from noise-free transmission. $\Delta^2$ is determined before the embedding. Let $U^m$ and $U^p$ denote the desired and rival code set respectively. The noise free decoding error probability $P_{e^0}$ can be derived:

\begin{align*}
P_{e^0} &= \Pr\{\min[d_E(U^m, x)] \leq \min[d_E(U^p, x)]\} \\
&= \Pr\{\min_i[d_E(u_i^n, c + \Delta u_j^n)] \leq \min_i[d_E(u_i^p, c + \Delta u_j^p)]\} \\
&= \Pr\{\min_i[(u_i^n)^2 - 2u_i^n(c + \Delta u_j^n) + (c + \Delta u_j^n)^2] \\
&< \min_i[(u_i^p)^2 - 2u_i^p(c + \Delta u_j^p) + (c + \Delta u_j^p)^2]\}
\end{align*}

Eliminate all squared terms since they are independent to $u_i^p$ or $u_i^n$

\begin{align*}
&= \Pr\{\min_i[-2u_i^n(c + \Delta u_j^n)] \leq \min_i[-2u_i^p(c + \Delta u_j^p)]\} \\
&= \Pr\{\max_i((u_i^n \cdot c) + (\Delta u_j^n)) \leq \max_i((u_i^p \cdot c) + (\Delta u_j^p))\}
\end{align*}

Using Equation 4.22, differences between Class1 and Class2 codebooks are discussed. In Class1 codebook, all codewords in $U^m$ are mutually orthogonal, the term $\otimes$ produces $LA$. The term $\oplus \leq 0$ as codewords in $U^m$ and $U^p$ are either mutually orthogonal or negative for all combinations of $i$ and $j$. For Class2 codebook, $\otimes$ produces non-positive values and $\odot$ always produces
zero. Since the Equation 4.22 derives its answer through maximum values, hence the non-positive values are not affecting the results. That lets the difference between Class1 and Class2 codebooks can only be distinguished through $\Theta$ and $\Psi$. Their values defines the performance between Class1 and Class2 codebook construction methods.

Since the value of term $\Theta$ is defined as $L\Delta$ despite the codebook construction method and $\Psi$ is established as a non-positive value that produces the maximum contribution to the $P_e$ when it is 0. Therefore, Equation 4.22 can be further reduced to:

$$P_e \leq \max_{i} \left\langle u_i^n \cdot c \right\rangle + L\Delta \leq \max_{i} \left( \left\langle u_i^p \cdot c \right\rangle \right)$$

$$P_e \leq \max_{i} \left( \left\langle u_i^p \cdot c \right\rangle \right) - \max_{i} \left( \left\langle u_i^n \cdot c \right\rangle \right) \tag{4.23}$$

From Equation 4.23, the decoding would be correct if the correlations of the cover vector with the most correlated desired codeword plus the watermark energy is higher than the correlation of the cover vector with the most correlated rival codeword. Let $u_i^p$ and $u_j^n$ be two codewords that have the maximum correlation with the cover vector in either code bins respectively. Therefore, in the absence of noise the probability of error is

$$P_e^0 = \Pr\{L\Delta < c \cdot (u_k^p - u_j^n)^T\} \tag{4.24}$$

According to Equation 4.24, if the embedding distortion $L\Delta$ is fixed, the only term decides the probability of error is $c \cdot (u_k^p - u_j^n)^T$. The encoder would prefer to find two codewords in either bins have similar correlation with the cover vector, hence the right hand side of Equation 4.23 is minimised. This can be done by using nonorthogonal codes. Hence, distances between codewords are smaller. It makes finding a pair of similar correlations more possible. Additionally, between nonorthogonal codewords, there are more places in common. Hence there is less $c$ elements contribute to right hand side term of Equation 4.24. The number of nonorthogonal codewords is large, hence finding a pair of codes satisfies this condition can be easier. However it also increases the possibility of error decoding because the correct codeword is easier to be flipped to the rival codeword by additional noise, and using nonorthogonal codewords, the term $\Theta$ of Equation 4.22 may also has positive contribution to the $P_e^0$.

The main advantage of the orthogonal codewords over nonorthogonal codewords is the minimised interference between codewords that is introduced by embedding. This interference affects the detection quality. The orthogonal codewords can have an disadvantage. All codewords are mutually common in $L/2$ places, since the equal Hamming distance between
codewords, except the negative self, having $L$ disagreements. The orthogonal codes have maximum number of disagreement. Thus more coefficients in the cover vector will contribute the second term on the left hand side that may lead to errors at encoder. If the cover vector $c$ and $(u^*_k - u^*_m)$ are statistically independent, their product can be very small. Especially with $L \to \infty$, this term will tend to zero. However with a finite $L$, there are three possible conditions the encoding algorithm can encounter, and they are listed in Figure 4.4.

![Figure 4.4: Encoding situations for dirty paper coding.](image)

In figures of Figure 4.4, the dashed circle denotes embedding constraint, $c$ in the middle present a cover vector. Other two vectors presented as "+" are the most correlated codewords from both code bins. As illustrated in Figure 4.4a both code sets have similar correlations. In Figure 4.4b and Figure 4.4c one codeword is more correlated with $c$ than another. These two figures are opposite to each other and they can simulate embeddings of alternative binary message bit to one cover vector, respectively. The minimum distance between orthogonal codes is maximum, thus the situation shown in Figure 4.4a is less likely than Figure 4.4b and Figure 4.4c. If Figure 4.4b happens, the detection is correct. However, if Figure 4.4c happens, an error will occur. There are two possible solutions for the error situation. The first one, allow codewords closer together, hence difference between the watermark and the rival codewords can both have closer distance to the cover vector. Closer distances can only be achieved by nonorthogonal codewords. Balancing between orthogonal codes and nonorthogonal codes is equivalent to choose between known and unknown errors. Obviously, for better transmission quality the error locations are preferred to be known at encoder. The second solution is to increase the size of the decoding range, that is determined by $P$. Therefore, more codewords can included in the decoding
Table 4.1: Comparison of term $\Theta$ and $\Theta$ in Equation 4.22.

<table>
<thead>
<tr>
<th></th>
<th>Class1</th>
<th>Class2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\max \in U^m$</td>
<td>$\max$</td>
<td>$\min$</td>
</tr>
<tr>
<td>$\max \notin U^m$</td>
<td>$\min$</td>
<td>$\max$</td>
</tr>
</tbody>
</table>

range. However, more embedding distortion is the inevitable consequence. That question becomes the balance between the detection qualities of visual and hidden information, mentioned earlier. The decision may only be made for particular applications.

If additional noise presents, the equation Equation 4.24 can be evolved to

$$P_e^N = \Pr\{L\Delta \leq c(u_k^p - u_j^m)^T + n(u_k^p - u_j^m)^T\}$$  \hspace{1cm} (4.25)

According to Equation 4.25, the robustness to the channel noise has a new definition as:

$$P_e = \Pr\{L\Delta \leq (c \cdot u_k^p) - (c \cdot u_j^m) + n(u_k^p - u_j^m)^T\}$$  \hspace{1cm} (4.26)

Referring to Figure 4.4a, $\Theta \sim 0$, hence $L\Delta$ can be fully used to against the noise term $\Theta$. In the condition illustrated in Figure 4.4b, $\Theta < 0$ thus the detection is error free for $L\Delta > \Theta$. Finally, as shown in Figure 4.4c, the term $\Theta > 0$, the robustness is compromised.

Derived from above, the direct deciding factors for the probability of error is the difference between term $\Theta$ and $\Theta$ of Equation 4.22. The values are listed in Table 4.1 where "max." and "min." refer to the maximum and minimum values of WHT coefficients. Since Class2 does not include a full set of basis sequences in one bin, some of the values are not possible to obtain, hence "-" is used to denote indefinite values. To obtain optimal performance, the two terms must be as close as possible. In Figure 4.5, two classes binning strategies are compared, generally. In the figure, two solid vectors in the circle present a pair of codewords from the desired codesets, and both dashed vectors denotes the rival codewords. An arbitrary vector $c$ denotes the cover vector. Class1 has no proper correlation with the rival codeset as illustrated, but in case of a negative message bit being embedded in the same vector, the embedding energy is required to be very high. Class1 is likely to simulate the situation illustrated in Figure 4.4b and Figure 4.4c. On
the contrary, Class2 has better constellation. Both code bins have similar correlations with the cover vector, like Figure 4.4a.

There is no means that the Figure 4.5 is a complete analysis, but some idea of the difference can be revealed. The superior between two classes of coding methods can be better distinguished through encoding performance. In next section, this topic is discussed with simulation results.

4.5 Simulation results and discussions

In this section, two coding classes will be compared to each other, in terms of probability of error under fixed embedding distortions. Also the visual distortion which is measured in Watson distance $d_W$ (Watson, 1993) is also shown. The embedding is made in DCT domain. First 32 DCT AC coefficients in zigzag order are selected to form a cover vector, hence, $L = 32$. Performance is affected by selection of the scramble key. This topic will be examined in the future, due to time constraint. The following results are implemented through biorthogonal Hadamard matrix.

In both Figure 4.6a and Figure 4.6b, the error rate is shown against various embedding constraint, ranging between 0 dB to 29 dB. In Figure 4.6a only the desired codeword is added into the cover vector. Class2 codebook
has a clear advantage over the Class1 codebook. Investigations have revealed that in some occasions the rival code set produces much higher correlation that cannot be overcame by the limited embedding power. Thus all rival codewords that produce higher correlations than the desired codeword are removed according to the correlation produced by the desired codewords. The subtraction can ensure the detection quality. However, if the combined energy is more than the constraint, the correct detection under this constraint is hardly possible. Therefore, no watermark is embedded and an encoder error is flagged. The Figure 4.6b shows the results. The Class1 codebook gains more improvement than Class2 codebook. The Class2 codebook has little improvement by subtracting rival codewords, until high level of DWR is given.

In both Figure 4.6a and Figure 4.6b, the error rate is caused by the embedding constraint globally applied to the entire image. The global constraint leads to failed detection at a number of individual vectors. Due to the fact described above, even the rival codewords are subtracted proportionally from the cover vector, the constraint is still not able to give advantage to desired codewords in all blocks. There is a \((B - 1)/B\) (50% for binary) possibility that the desired code set produces lower correlations with one cover vector. In other words, a half of the message bits can be correctly detected without extra embedding power. Among the other half of cover vectors, the difference between correlations of rival code set may be small enough to control. However, investigation has found out that in some cover vectors the difference is very large. It is these vectors that cause the error rate in Figure 4.6a and Figure 4.6b. Along with reducing embedding strength, the advantage that is provided by the watermark energy to the desired codeword vanishes. The error rate in both figures also indicates the proportion of those out-of-control vectors.

The Figure 4.7 displays the Watson distance versus the DWR. The Watson distance produced by different codebooks has little difference. It is unexpected that the Class1 produces even smaller \(d_W\), when some rival codewords are proportionally subtracted.

The results displayed were based on satisfying the global uniformed embedding constraint. Investigations have found out that the uniformed constraint has less power on those uncontrollable vectors as discussed before. Therefore it degrades the local detection quality. As discussed earlier there are 50% of the message bits do not need any extra energy for correct detection. It is interesting to find out that if the global constraint is relaxed and instead, the detection reliability is maximised, what level of global fidelity distortion would be produced. The Table 4.2 shows minimum DWR and Watson distance when the detection is error free. This result shows the level
of global fidelity distortion when the uniformed constraint is relaxed and the
detection reliability is maximised. The error free results are achieved by sub-
tracting the rival codewords which have higher correlation than the desired

Figure 4.6: Probability of error based on various DWR.
Figure 4.7: The Watson distance produced at various DWR by different classes codebooks.

codewords to the level of correlation produced by the desired codeword, and the desired codeword is added to the cover vector with a unit power.

<table>
<thead>
<tr>
<th>Class</th>
<th>DWR</th>
<th>$d_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>44.6 dB</td>
<td>41.6</td>
</tr>
<tr>
<td>Class 2</td>
<td>69.9 dB</td>
<td>22.0</td>
</tr>
</tbody>
</table>

Table 4.2: Comparisons of the error free DWR and Watson Distance

As it can be seen (Table 4.2) the visual quality and the distortion energy are both significantly smaller than the uniformly controlled environment. It is because that in the previous results the embedding distortion is uniformly regulated all over the image to the level satisfying the constraint, as shown in Figure 4.6a and Figure 4.6b a small percentage of blocks require higher embedding power to bring desired codeset into advantage. However, in the constraint free environment, the embedding power is minimised to a level where is just enough to yield desired result. All blocks that have higher correlation with the rival codewords also received just enough power to bring the desired code set in advantage. The trade-off for the minimal detection error is several severely distorted blocks. However from Table 4.2, it is clear
that the overall performance is better than the global constraint is applied.

The above discussion revealed one fact that the detection quality of both visual and information are contradictory, the codewords are required to have closer correlations between codesets, thus only a low power is necessary to bring the desired codeset in advantage over the rival codeset. This requirement is shown in Equation 4.24. However, the codewords also must have significantly large Hamming distance to be distinguished, especially in the presence of noise.

From the above simulation results, it is clear that the Class2 codebook has better performance than the Class1 codebook. The Class2 codebook has lower error rate in both embedding methods (Figure 4.6a and Figure 4.6b). It is because Class2 binning separates the codewords to have closer correlation to more possible cover vectors between code bins (Equation 4.24). The Class1 and the Class2 codebooks both have only half of a complete biorthogonal matrix, but they have different constellation setups.

![Cross-correlation values between two bins, produced by Class1 codebook.](image)

Figure 4.8: Cross-correlation values between two bins, produced by Class1 codebook.

The Figure 4.8 and Figure 4.9 show the averaged cross-correlation values obtained by the Class1 and Class2 codebooks with one image, respectively. Figure 4.9 shows the results produced by two codebooks derived from Class2, respectively. In Figure 4.9a, the codebook consists the first half of the orthogonal matrix and its negative counterpart in one bin and the second
Figure 4.9: Comparisons of cross-correlation values produced by Class 2 codebook construction methods.

half of the matrix and the negative counterpart are in another. In Figure 4.9b the codebook is composed by odd index of sequences and their negative
counterparts in one bin, and the even number of sequences and their negative counterparts in another. Although the shapes are different between the two instances, the properties are same. Between Figure 4.9a and Figure 4.9b, both bins have similar maximum cross-correlation values. This property provides better performance as discussed above. On the contrary, the Class1 codebook cannot produce such optimal cross-correlation values. One bin is negative to another, because of that it has inferior performance, compared to Class2 codebooks.

4.6 Summary

In this chapter the application of biorthogonal matrix as dirty paper codes to digital watermarking is discussed. From the analysis it is obvious that the biorthogonal matrix provides optimal performance in terms of minimum interference between codewords in the same coding bin and different bins. Error probability is solely decided by the correlation between the cover vector and the code bins. Also the robustness is maximised due to the maximum $d_{\text{min}}$. The noise has to flip half of the watermark vector to produce an error.

Dirty paper code embedding performance is very much dependent on the watermark energy distribution. As results shows, solely adding desired codeword to the cover vector is not enough to produce a good performance. As the energy distributed with some subtractions of rival codewords, both code construction methods produced higher performance.

The biorthogonal dirty paper coding has similarity with $M$-array orthogonal modulation technique. But dirty paper coding is required to protect the image quality. The coding must be in favour of first noise source. Costa (1983) provided dirty paper coding condition to optimum channel capacity.
Chapter 5

Comparisons and applications

5.1 Introduction of comparisons

5.1.1 Objective

In this chapter, the designed watermarking schemes are compared in terms of fidelity, capacity and robustness. The aim of the comparison is to evaluate general performance of the schemes and to identify trade-offs.

In previous chapters, where the schemes are initially described, the performance of each scheme is described and tested, but they were tested at different environment with different experimental purposes. In this chapter, the algorithms are tested in a same environment with the same targets, in order to determine the practical performance advantage and drawbacks between each other.

5.1.2 Methodology

The comparison is carried out by simulation of embedding and detection processes at a number real ID photos. Embedding processes are assisted by randomly generated watermark messages \(\{d_i\}\). Measurement of results adopts well-known tools, such as BER, different variations of SNR; and the Watson distance \((d_W)\).

The simulation is divided into two initial points of view, Watson distance based and DWR based. The Watson distance based simulation constraints watermark power referring to amount of visual distortion made by watermarks. The DWR based simulation constraints watermark power to a fixed ratio (DWR) of the power of cover vector.

Watson distance based simulation allows higher watermark power if the power is in favour of visual vector. However, if the watermark power is
on the opposite direction of visual vector, even a low power distortion can make high visual distance. In other words, Watson based measurement may "absorb" more distortion power than DWR based simulation which generates the same Watson distance. On the contrary, the DWR measures the power magnitude only, regardless the relative direction of visual vector, therefore, in the situation when watermark vectors are not in favour of visual direction, DWR based simulation results must show good results.

5.1.3 Cover vector

The cover vector is constructed by DCT AC coefficients because watermark embedding in frequency domain is more robust to distortions (Cox et al., 1997). The tiles are interleaved sequentially, such watermark embedding is benefited by avoiding particular image features that affect detection quality on some watermark bits. It can also allow the cover vector elements to appear more random. The construction of a DCT domain cover vector is shown in Figure 5.1.

Images are divided into 8 x 8 non-overlapping tiles. The DCT is applied on each tile. The first \( J = 32 \) DCT AC coefficients in JPEG zigzag order of each tile are selected. \( K \) blocks form a cover vector. The value of \( K \) is determined by the number of payload bits \( M \) that can be carried by the watermark. The length of the cover vector is denoted by \( L \). The size of all cover vectors is denoted by \( S \). Thus, in an image has a size of \( N \), \( S \leq N \), \( L \times M = S \), and \( J \times K = L \). If the image has the number of rows or columns not 0 mod 8, the maximum number of multiple 8 columns or rows will be sampled out from the centre of the image.

5.2 Comparisons based on Watson distance

5.2.1 Fidelity on error-free detection

The number of payload bits carried by the watermark is calculated by the number of blocks, \( K \), that can exist in one cover vector. For simulations, the parameter \( J = 32 \), and for compatibility with WHM, \( K \) is selected as the number that has integer power of 2.

In the encoder, starting from a low level, the embedding strength is gradually increased, until the first error-free detection occurs. Then the Watson distance and MSE are calculated, as shown in Figure 5.2a and Figure 5.2b. Most of time, the SC orthogonalisation procedure produces the least distortion to the cover image, because the sequences are completely customised
(a) Selection of DCT AC coefficients in one $8 \times 8$ block, $J = 32$.

Figure 5.1: Formation of cover vector in DCT domain

to the cover vector. At $L = 128$, SC produces only 14 Watson distance and 0.1 MSE. The HMS procedure produces similar results with less than 18 Watson distance and the MSE is very stable at 0.2. These two procedures have similar reaction to the variation of cover vector length. Both may be used when fidelity requirement is stronger than the requirement of maximum data payload.

The flexible insertion algorithm of HMDP adds more distortion to the cover image for better detection performance. However, at the shortest vector length, its perceptual fidelity is the highest among the three. It is worth noting that the HMDP is not affected by the length of cover vector as much as other two methods. This characteristics of HMDP allows its applications in high payload and low fidelity requirement.

### 5.2.2 Capacity

Every cover image is tested to determine its maximum capacity for randomly generated message bits. The maximum payload is obtained when $K = 1$. For example, if an cover image has $200 \times 256$ pixels, $S = 51200$, because of $J = 32$, $M = 1600$ bits.

As the tests proceed, the payload bits and detected error are accumulated, respectively, until the payload running total reaches or exceeds $1,000,000$ bits for each cover image. Detected bits in error are given by
Figure 5.2: Fidelity reading of error free detection by the lowest signal power.

\[ e_i = \begin{cases} 
0 & \text{when } \hat{d}_i = d_i \\
1 & \text{when } \hat{d}_i \neq d_i 
\end{cases} \quad (5.1) \]
For each cover image $p$, a bit-error-rate ($BER_p$) is calculated as:

$$BER_p = \frac{1}{M} \sum_{i=0}^{M-1} e_i$$

BER is the mean of all $BER_p$. The Figure 5.3 show the BER of different procedures against the fixed Watson distance.

![Figure 5.3: BER against fixed Watson distances.](image)

The simulation starts with an exhaustive search on Watson distance while increasing watermark strength, until the distortion reaches a predefined Watson distance, shown at the x-axis value. When this Watson distance is reached, the watermark strength is kept as constant, and then payload bits and detection errors start accumulating.

SC procedure produces a good performance at an early stage. Between $d_W \in \{4, 24\}$, it has the best performance. The lowest BER reaches $10^{-3}$ when $d_W = 24$, and at $d_W = 12$ the largest margin leading the second best performance method is almost 10 times better, reaches $3 \times 10^{-3}$. However its advantage does not increase continuously. The slope of the decreasing BER curve reduces after $d_W = 12$. After $d_W = 26$, SC performance is the inferior method. In the span between 24 and 60, the BER reduces from $6 \times 10^{-4}$ to $2 \times 10^{-4}$.

The HMS produces very good results overall. At the beginning, HMS BER rapidly reduces. This BER steadily reduces almost linearly, until error-
free detections are reached at the very end of the chart. Any increase of embedding strength, using HMS procedure, is the increase of power of weak spectrum components. For every image, this increase can continuously grow, until the power of these weak components are significant in the spectrum. So that, the BER curve is a function of the embedding strength, which is the main deciding factor of Watson distance.

At early stages of the simulation, when \( d_w \in \{4, 8\} \), a high fidelity region, the HMDP hardly produces any improvement in the BER curve. The reason we believe is the watermark group \( G_{d(i)} \) has not obtained energy strong enough to be distinguished. Between \( d_w \in \{4, 8\} \), the BER drops gradually. Detailed investigations on the simulation data reveals that the HMDP can produce very low, even zero BER, on some of the images between the fidelity levels mentioned above. Other images may require slightly stronger embedding energy. After \( d_w = 24 \), the curve starts falling sharply, and reaches below \( 2 \times 10^{-5} \) at \( d_w = 28 \). After this fidelity level, the increasing of embedding energy is unnecessary for improving HMDP detection, because detection error no longer occurs. It is that the flexible insertion algorithm effectively pushes the watermark energy whilst restraint the distortion level.

### 5.2.3 Robustness

The simulation is performed in a similar manner to the previous one, except that the Watson distance, is set at \( d_w = 50 \). The BER is measured against AWGN variance, \( \sigma^2 \in \{0, 1, \ldots, 9\} \), as shown in Figure 5.4

It is worth noting that the HMDP yields very good results from error free to \( 3 \times 10^{-4} \) when \( \sigma^2 \leq 3 \), where all three procedures achieve the same performance. The SC procedure starts at a BER \( 3 \times 10^{-4} \), but the performance is not much degraded by the increasing noise level. It is most robust against AWGN attack, when \( \sigma^2 \leq 4 \). Continuing with increasing noise level, the BER generated by SC procedure reaches at \( 8 \times 10^{-3} \) when \( \sigma^2 = 9 \) and it is the best result among the three at this highest noise level. HMS procedure generates the middle performance between the other two, HMS achieves slightly lower BER than HMDP, \( 2 \times 10^{-2} \), at high noise level.

This test illustrates that the Hadamard matrix based algorithms are not very robust against AWGN attack. On the contrary, the customised sequences generated by SC procedure are very robust against AWGN attack, though a refinement should be introduced to reduce the level of BER. It is also worth noting that the results of the last two experiments were derived using the shortest cover vector length. As shown in the fidelity experiment, this performance will be certainly improved by increasing the cover vector length.
Figure 5.4: BER against AWGN variance. Watson distance is fixed at 50.

5.3 Comparisons based on DWR

This section shows the comparison results based on DWR measurement. The watermark power is decided by a fixed ratio to the cover vector power. Compare to the previous comparisons, the results shown in this section are more biased to consider watermarking as a communication system.

It is worth noting that in this section the presentation of DWR is different from the DWR used in earlier chapters. Before, the DWR is uniformly calculated as

$$DWR = 10 \times \log_{10} \frac{\sum i^2}{\sum w^2}$$  \hspace{1cm} (5.3)

where $\sum i^2$ is the squared sum of all pixels of cover image, and $\sum w^2$ is the squared sum of watermark in pixel domain. In this chapter, the DWR is calculated in the following manner

$$DWR = 10 \times \log_{10} \frac{\|c\|}{\|w\|}$$  \hspace{1cm} (5.4)

where $\|c\|$ and $\|w\|$ are the norm of cover vector and watermark vector in DCT domain, respectively.

The main difference between Equation 5.3 and Equation 5.4 is that the document energy is calculated differently. In Equation 5.3, the document
energy is calculated by including entire image pixels, but in Equation 5.4 the document energy does not include the DC term of every block and 31 high frequency AC terms. It is believed that the energy excluded from the document energy is near or over half of the entire pixel energy. Therefore the DWR calculated in this chapter is much smaller than the results shown earlier (especially in Chapter 3). The new calculation matches the popular literature, such as (Abrardo and Barni, 2005).

In this set of results, 5 watermarking schemes are compared. Two orthogonalisation schemes are SC and HMS. The Hadamard code dirty paper scheme operates at Class1 and Class2 binning strategies, as described in Chapter 4 and they are referred as "bdpc1" and "bdpc2" respectively. A new writing on dirty paper scheme is developed to expand the Hadamard biorthogonal matrix based dirty paper coding schemes. The new scheme is named as Algebraic Geometric Dirty Paper Codes (AGDPC). The AG code is also known as Goppa code, that is often used for cryptography and error detection and correction. The set of AG code used in this simulation includes 128 codewords and $L = 32$, $d_{\min} = 10$. The entire collection is divided randomly into two subsets; each presents 1 and 0 bit respectively. Comparing to the Hadamard biorthogonal code, AG code has more codewords and codewords are closer to each other, that provides a good chance to find codewords from each code set that are similarly closer to the cover vector. A significantly larger codebook also provides opportunity for multi-level watermark embedding by increase the number of bins. Since the AG codewords are not mutually orthogonal, embedding one codewords affect the correlations of other codewords, thus AG code does not easily operate in error-free mode like Hadamard code watermark, therefore only the basic embedding structure is used. In order to remark the different simulation environment, in this section, all schemes are indicated in figures with lowercase.

### 5.3.1 Capacity and fidelity

This simulation combined the capacity and fidelity categories. The aim is to find the visual distortion made by watermark signal at various power level while having the largest capacity. It is clear that the fidelity performance is better at lower capacity requirement. Figure 5.5 shows the bit error rate that is produced by watermark embedding schemes, at a particular DWR.

From the left to right of the horizontal axis, the value of DWR decreases, however it is the direction where the watermark power increases. In the Figure 5.5, there are two algorithms show the error free status. They are the SC scheme at DWR=8dB and HMS scheme at DWR=14dB. These two schemes are both OSS coding schemes. It is worth noting that the HMS is
Figure 5.5: Bit error rate produced by different embedding schemes at various DWR.

running at 8 times longer cover vector length with 8 watermark sequences overlapping. The best performance is supported by the claim that was made in Section 3.4. Three writing on dirty paper schemes also shows signs of waterfall, but the required power is higher than the defined scope of the simulation. Among the three, the AG code based coding scheme has marginal advantage over two biorthogonal schemes. With the increasing watermark power, this advantage becomes more significant. It proves that the number of codewords does affect the embedding quality. Finding AG codewords that are closer to cover vector is easier than using Hadamard code (referring to the Equation 4.24).

It is worth noting that both BDPC schemes do not operate at error-free mode in this simulation, since error-free mode is designed to reach the highest detection quality while minimise the embedding distortions to the cover image; it is free from DWR constraint and embedding distortion is only a function of robustness tolerance.

The Figure 5.6 shows the Watson distance reaction of different schemes. Except HMS, other embedding schemes produce similar results. The typical maximum Watson distance is below 200. HMS produces the highest Watson distance near 240 at DWR=0dB. Referring to Figure 5.5, the best capacity performance produced by HMS compromises the fidelity performance. Al-
Figure 5.6: Watson distance produced by different embedding schemes at various DWR.

though the watermark power is same for all schemes, different coding methods produced uncommon visual impact. It is why the comparison must be made with two types of constraint, visual quality and power.

Because the DWR is much easier to operate than the Watson distance, the Figure 5.6 can assist to define an unacceptable DWR region, for example, if the maximum acceptable Watson distance is 80, the acceptable DWR>10dB can be defined. Using the combination of both Figure 5.6 and Figure 5.5, one can define an operation region for acceptable trade-offs between bit error rate and visual distortion level.

5.3.2 Robustness against AWGN

The robustness test is aimed at reveal the detection quality after the watermarked image has undergone a number of signal processing. The watermark power is fixed at DWR=15dB, where the watermark power does not degrade the image visual quality beyond acceptance. The simulation is designed to adding AWGN directly to the watermarked vectors, then a detection routine follows and the difference between the detected and original sequences is accumulated.

The first presentation, shown in the Figure 5.7, gives the general idea of the power of noise to which the watermark signal is robust. Because both OSS
schemes have better BER performance at the noise-free environment (Figure 5.5), the performance maintains the significant advantage with additional noise. Between the two, SC has better robustness than HMS. Observing at $\sigma < 4$, the BER produced by SC barely increases.

The second presentation, shown in the Figure 5.8, displayed the difference of bit error, between noise-free and noisy environment that has increasing noise power. At each noise power ($\sigma^2$), the additional BER is shown. The large increase indicates that the particular method is not robust to noise. From the Figure 5.8, the AGDPC is the least robust method to the AWGN, because the number of additional wrongly detected bits is the largest for $\sigma > 4$. Two classes of biorthogonal dirty-paper coding methods have similar robustness performance, and between both, the Class1 method is more robust than Class2 method. Two OSS methods have exact performance on robustness, and both curves overlap each in figure. Both OSS method only produced less than 10% of additional wrongly detected bits at the highest simulation noise power.

Finally, the Figure 5.9 shows the BER against the power relation between noise and the watermark. This can be seen as $E_b/N_0$, commonly used in evaluation of communication systems. This figure is the representation of the Figure 5.7, with an reversed and non-linearly scaled horizontal-axis.
Figure 5.8: Additional error bit as the percentage of all simulated bits ($\geq 5 \times 10^6$), caused by increasing noise.

Figure 5.9: Bit error rate produced by different embedding schemes at various WNR, at DWR=15dB.
5.4 Summary

In this chapter the simulations from two different angles present the performance of different watermarking schemes. From visual quality consideration to power ratio consideration, the schemes are analysed and compared.

When the visual quality is the major consideration, the error-free mode Hadamard dirty paper code has the best performance on capacity and fidelity. Although the robustness performance has the fastest growing BER, the HMDP noise tolerance is a design factor at the watermark embedding process. The HMDP can be robust to any large additional noise, as long as the visual quality constraint permits. In this situation, the HMS scheme is the second runner in both capacity and robustness categories. At the last place is the SC. Although the SC has good fidelity rating when cover vector is long, in the high capacity test the SC does no reach error free at the highest fidelity tolerance.

When the major consideration switch to the power ratio, the results suggest otherwise. The SC out performs all other schemes. It reaches error-free detection at the lowest watermark power, and produces the highest robustness against AWGN, even though the SC has to operate at positive region of WNR. However, SC is not very robust to high power AWGN ($\sigma^2 \geq 4$). The BER increases rapidly. The HMS is the second highest capacity algorithm. It also demonstrates robustness in low power noise. Same as the other orthogonal embedding scheme the SC, the BER is rapidly increased by the noise power. However, results of the three writing on dirty paper schemes determine that this class of embedding requires higher power to reach good performance. It is worth noting that the two biorthogonal dirty paper coding schemes do not operate at the optimal mode, since they must have same measurement to be compared with other schemes.

By comparing between two groups of results, it can be seen that visual quality consideration is practical. The results support selections of operable schemes at defined environment. The power ratio consideration is suitable to evaluate the communication channel characteristics. Both groups of results shows different conclusions on SC and Hadamard code schemes. This is because, in fixed visual distortion, power of SC watermark is restricted. Because SC generates sequence has the least correlation, thus the watermark produces the maximum distance to the cover vector. However, at the fixed channel power ratio, the watermark power is fixed to a proportion of the cover vector power. The limited watermark power cannot overcome the orthogonalisation remainder at the tested short vector length. On the contrary the dirty paper coding schemes which produce the most correlated watermarks to the cover vector have the advantage of tolerant more energy without sig-
significant visual distortion. But, at fixed power, they are restricted to bring the desired codeset in advantage. Above all, the schemes have their advantages and drawbacks. Only practical applications can decide what scheme to choose.

5.5 Application example

In this section, a demonstration software is introduced. The software has ability to replace an image of a fixed size with its watermarked replica in a PDF file. Watermark encoding is an implementation of the HMDP operating with flexible margin control. The examples in this section show that watermark carries 2048 bit information which is ASCII coded message and cover vectors are formed by 64 DCT low AC coefficients collected from two 8 x 8 blocks.

One example shown in Figure 5.10 and Figure 5.11 is the PDF file and the decoder Graphic User Interface (GUI). The PDF file shown in Figure 5.10 is the transmitted file\(^1\). The PDF file can be a confirmation of purchase or a product brochure. To avoid unauthorised modification of genuine description, the publisher can add key features of the product as a watermark message, thus it allows customers to check authenticity of the product. In Figure 5.11, shown on the left, the secret message is displayed after a successful decoding operation.

In Figure 5.12, the image used in above example is compared between the original and watermarked.

Other applications of this software include:

- Secret messages can be transferred without attracting any attention from malicious “observers”.

- Photograph index. Location, time and any interesting information of photographs can be stored along with the photographs regardless file format.

- Authenticate companion message. Important content of a message can be highlighted in an attached picture to ensure the integrity and authenticity of the message.

---

\(^1\)The PDF file is created for experiment and demonstration purpose only. The text is a production of fiction. The image is copyrighted by Tiffany & Co.
Tiffany’s Diamond Ring Purchased 14th October 2005

This diamond encrusted ring made from platinum was purchased from Tiffany on 14th October 2005 for $150,000. A summary of the description from Tiffany’s online catalogue is:

Channel-set band ring, full-circle. Round brilliant diamonds, carat total weight .72 (size 4), color grade G, clarity grade VS; platinum. Channel-set band ring, full circle. Baguette diamonds, carat total weight 2.24 (size 4), color grade G, clarity grade VS; platinum. 200mm

Figure 1: Mult-diamond encrusted ring

Figure 5.10: Watermarked image in a PDF file.

5.5.1 The software

Apart from the above applications, the complete software package has an important role in search of optimal operation parameters for the embedding scheme. The demo shown earlier is a standalone decoder. A complete codec GUI is shown in Figure 5.13.

For the embedding process, user can select a PDF file, that must have been created by PDF specification version 1.4 or lower. Since after version 1.5, the PDF file structure become more difficult to manipulate by a simple object replace mechanism. The file is required to contain at least one image in required format. The image must be a colour image with colour depth of 8 bit at 512 × 512 resolution. If the file contains more than one qualified
image, only the first one will be used. The encoder will start to search the object keyword in the PDF file. Since PDF files are plain text structured the search is typical string match operation with the keywords listed in Table
Figure 5.13: Complete GUI of the demo software (The “Retrieving” tab has identical controls as the standalone decoder).

5.1.

The binary data included between stream and endstream keywords are the image data in complete JPEG file format. The key to the replacing operation is that if all these data are extracted from the PDF file and written into a blank file, it will be recognised as a JPEG image. Thus by deleting the data and writing a new JPEG image in the same location, the image is swapped.

It is worth noting that the /Length keyword indicates the size of the binary data in number of bytes. If the number indicated by this keyword does not match the actual data size, the PDF reader software will either read an incomplete JPEG data stream or some irrelevant data will be included into the JPEG decoding process. Thus an error occurs, that could lead the PDF file unreadable.

User may also select an image file in any popular format, that will save the extraction process. The selected image file will be treated as the extracted JPEG data. The entire set of binary data will be passed to Application Program Interfaces (APIs) of a reputable image manipulation software suite.
Table 5.1: Target image object keywords, in a PDF file.

"ImageMagick\textsuperscript{®}. ImageMagick completes the JPEG decoding processes quickly and accurately. Then the image is returned to pixels. At the same time, the demo software copy all the data before the start of image object to a blank file that will become the watermarked PDF file.

At this point, the watermarking module programmed at margin control mode of HMDP can start its operation of coding and embedding. It is followed by returning the embedded pixels to ImageMagick to produce a valid JPEG data stream. The demo software collects the JPEG from a temporary file written by ImageMagick and attaches the new JPEG data stream at the end of the new PDF file. The data that are located after the image object in the original PDF file will be affected by replacing new data. Therefore some changes must be applied.

At the end of a PDF file, there is a special object called cross reference table, led by \texttt{xref} keyword. In the table, all objects featured in the PDF file are indexed by the number bytes between the beginning of the file and the beginning of the object (Adobe System Incorporated, 2004). This table works as the content list of a book that allow random access of any object in the PDF file. Since the new JPEG data very unlikely has identical size as the original one, all objects following the image object will be affected,

\textsuperscript{2}ImageMagick is an open source image manipulation software suite, maintained by ImageMagick Studio LLC, Landenberg, PA, USA and distributed under GPL. The demo uses its C++ API, Magick++, (http://www.imagemagick.org/Magick++)

106
Control | Function | Range | Default | Effect
---|---|---|---|---
Embedding strength | Power of the watermark codeword | 0.1 to 10 | 0.1 | Provide more power to the desired codeword.
Compression quality | Compression level of new JPEG stream | 0 to 100 | 95 | Low value for smaller final file size, high value for less distortion on watermark detection.
Normalised safe detection margin | Separate the desired and rival codewords | 0.1 to 2.0 | 0.1 | High value for more reliable detection, but more embedding distortions added.

Table 5.2: Three major embedding parameters, function and values.

and corresponding xref entries must be change accordingly. The effect can be quantified by the difference between size of the original and watermarked data stream. The change can be made by adding or subtracting the difference to the original indexes.

5.5.2 The interface

Referring to the Figure 5.13, underneath the PDF loading field, is a check box that ask user whether to keep a copy of original file. If this switch is off, the original file will be overwritten. Below the check box, there is a space to load the text file which must be an ASCII coded plain text file. If the message contained in this file is longer than the defined maximum length, the message will be truncated to satisfy the maximum length.

The core operational value for the demo software is nested in the “Embedding Parameters”. In this area, three major embedding parameters are variable through sliding the nobs. The functions of the parameters are listed in the Table 5.2. The software interface allows user to adjust the parameters separately to achieve desired performance and quality.

The codec software also includes a complete decoding routine to monitor the embedding quality and allow user to adjust parameters accordingly. In Figure 5.14, the software reports an initial embedding using default values.

After several different settings, the embedding software can achieve error free detection through different settings. Examples are shown from Figure 5.15 to Figure 5.17.
Figure 5.14: Error message warns the user to adjust controls for error free detection.

Figure 5.15: One example of error free detection, note the parameters.
Figure 5.16: The second example of error free detection, note the parameters.

Figure 5.17: Another example of error free detection, note the parameters.
The software also provides a preview frame to allow user check the embedding visual quality. By clicking on the preview image, a new window will be shown as Figure 5.18 displaying the embedded image in its original size for a closer exam on the quality.

Figure 5.18: Large preview window displays the watermarked image in its original size.
Chapter 6

Synchronisation

Watermarked Images may be subject to rotation, scaling, translation and other geometrical distortions prior to the detection. The watermarked image in this thesis is preferred as the printed image on security documents. The watermarked images on ID cards or passports are commonly subject to RST distortions at capture. Other distortions, such as cropping and shearing are assumed not applicable. If some unrecoverable distortions applied (failed captures), the watermarked image can always be repositioned for better capture quality. This application, particularly, has a unique challenge for synchronisation processes. It is to locate the watermarked image precisely from a captured image, e.g. the information page of a passport or the face of an ID card. The captured image is always larger than the watermarked image. It requires the watermarked image to be sampled as precisely as possible from the captured image. The performance of the earlier proposed schemes in Chapter 3 and Chapter 4 can be degraded by poor synchronisation between the watermarked image and the watermark detection code.

A simple but effective synchronisation scheme is firstly discussed in Section 6.1. This scheme adopts high contrast frames to enclose all pixels of the watermarked image. The borders are served as re-synchronisation markers to aid the detections of RST parameters. Then after the recovery of those parameters, the image is corrected to its original form through geometric transforms. This synchronisation is dependent on regulating or sharing the knowledge of the image size. This requirement is not a restriction to the applications, since most of security documents have regulated aspect ratio and size.

An optimal watermarked image synchronisation scheme is described in Section 6.2. The scheme uses well-defined image invariant transforms that are described in Section 6.2.1 and SPOMF technique. The major difference between this method and other existing image registration methods is that
watermarks are used as templates instead of a portion of the original image. The proposed design of synchronisation watermark pattern is described in Section 6.2.3. This new design has less markers than earlier designs, and it is also capable to locate the watermarked image from the captured image that includes the watermarked image and its background. The size of the watermarked image is not required to be known at the decoder, since the synchronisation watermarks can detect the actual size.

The RST distorted image, in the following context is called captured image, which contains rotated, scaled and translated and all pixels of the watermarked image. The additional watermarks used to synchronise the watermarked image is called synchronisation markers.

6.1 Border and linear regression

Visual features, namely high contrast frames are used. The frames have distinguished pixel value pattern, like the solid bold line shown in Figure 6.1. The thin dashed lines in the figure present the edge of the captured image.

![Figure 6.1: Frames and linear regression synchronisation scanning.](image)

- Through scanning every row and column of the captured image in both directions (the arrows in Figure 6.1, the borders are detected as a set of points.
• Using linear regression function, these points are linked into one continuous line along each side of the watermarked image. One line \((a \rightarrow b)\) along the bottom of the watermarked image is shown in the Figure 6.1 as an example.

• Then the angle between this predicated line and the edge of the captured image presents the rotation angle, the angle \(\alpha\) shown in Figure 6.1.

• After the rotation being corrected, the average distances between the predicated borders and the edges of the captured image are the translation parameters.

• By cutting the background between the lines and edges, the scaled watermark image is obtained.

• The final step is to calculate the scaling factor by dividing the known length of both horizontal and vertical borders to the scaled version of the image. The ratio obtained from the calculation is used to re-sample the pixel values to standard size.

6.1.1 Implementation

The procedure defined in previous section can be implemented for synchronising printed images. An example is shown in Figure 6.2.

The borders used in this example are black lines with 2 pixel thickness. To avoid losing any information from the dithering processes of both printer and scanner, the image is printed with 300dpi and scanned at 600dpi. Therefore, 2 pixel wide borders approximately becomes 4 pixel wide. The frames only partially along the edges, since the detection on the corners may lead to incorrectly identifications. For example, during the horizontal detection the vertical frames on the corner will affect the decision by showing more frame information.

The captured image is passed through a histogram equalisation process to enhance the border contrast. The detection looks for continuous 3 to 5 pixels below the black threshold. Because the printing and scanning may change colour of a pixel, the detection threshold is slightly higher than the value of a black pixel. If the detection for a row or a column gives a positive output of the estimated location of the first pixel below the threshold, the coordinates of this location are recorded as a possible point along the border of target image. However, if the detection could not find enough pixels below the
threshold or more pixels than required, there is no pixel coordinates recorded for this scan.

After the scan of each side of the captured image, the coordinates of all recorded points of each side are used to calculate the rotated image border using least square linear regression function as:

\[
\tan \alpha = \frac{\sum_{i=0}^{N-1}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=0}^{N-1}(x_i - \bar{x})^2}
\]

(6.1)

where \( \bar{x} \) and \( \bar{y} \) are the means of horizontal and vertical coordinates of the recorded points respectively. \( N \) is the total number of recorded points on each side of the captured image. Equation 6.1 produces different values of regression for each side. In general, the regression function results are expressed as

\[
\tan \left( \frac{n\pi}{2} + r \cdot \alpha \right), \quad n \in \{0, 1, 2, 3\} \text{ and } r \in \{+1, -1\}
\]

(6.2)

where variable \( r \) denotes the rotation direction and \( n \) denotes the index of the sides. When \( r = 1 \), the watermarked image appears anticlockwise rotated at an angle of \( \alpha \) and \( r = -1 \) indicates clockwise rotation. It is worth noting that for one captured image, the four results of Equation 6.2 may not be exactly same. The mean of \( \alpha \) for \( n \in \{0, 1, 2, 3\} \) is calculated for better accuracy.

The example shows that the detected angle is 0.46°. The watermarked image has the resolution of 368 x 520. The example shown in Figure 6.2, the trimmed image is approximately twice as the original one. After scale down the trimmed image to its known size, the synchronised image is shown in Figure 6.2d.

### 6.1.2 Improvement

The linear regression used in this scheme is optimal estimator, but the border detection is not optimised. The decision rule for the border detection cannot guarantee accurate detections of the border locations. In digital form, since the pixels have very high contrast and little degradation, the detections are always accurate, but when the image is printed and scanned, the dithering process break the continuous tone of the pixels. Thus some light dots may be added between the black dots of the frames. Better decision quality can be achieved by applying edge detection filters as

\[
\Delta_x = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}
\]

(6.3)
Figure 6.2: An example of synchronising a captured image.

\[
\Delta_y = \begin{bmatrix}
1 & 1 & 1 \\
0 & 0 & 0 \\
-1 & -1 & -1
\end{bmatrix}
\]  
(6.4)
The border scheme is equivalent to create a step edge around the image. The two filters in Equation 6.3 and Equation 6.4 are well known as the Prewitt edge detection filters that measure the 1st derivative of the pixel intensity (Ziou and Tabbone, 1998). These filters detect the step gradient changes at the side where the positive column or row is. Therefore, the horizontal filter shown in Equation 6.3 can only produce the left edge of the borders. Thus the edge produced by this filter at the right hand side border is the inner edge, where the border meets the image. This is not the desired result, since only the outer edges can present the scaling and translation factors of the captured image. Therefore, a minus sign is assigned to the filter, when the scan passes the centre point of the captured image. In Figure 6.3 the outer borders of both horizontal and vertical directions are shown respectively.

![Image of horizontal and vertical edges](image-url)

(a) Horizontal edges  
(b) Vertical edges

Figure 6.3: Output of edge detection filters. (For printing quality, the colour is inverted)

Fast implementation can be achieved through the convolution of two-dimensional filters and selecting the peak values in the region of interest. One of the filters is used as:

$$\nabla = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (6.5)$$

Ziou and Tabbone (1998) indicated that before the edge detection, a smooth
operation can improve the detection quality. In the Figure 6.4 examples are shown the output of the filter (Equation 6.5) with and without a smooth operation. The smooth operation is made by a $5 \times 5$ Gaussian blur IIR filter.

Figure 6.4: Output of the filter (Equation 6.5) with and without a smooth operation. (For printing quality, the colour is inverted)

Comparing between Figure 6.4a and Figure 6.4b the dither effect is reduced through the smooth operation. Thus the edges are much more detectable. However, information loss is inevitable (Ziou and Tabbone, 1998). A good performance balance is required to enhance as well as protect the detection quality.

The frames used in this example is black lines. However more adaptive approach can improve the detection quality even further. Since the edge detection filters produce peaks output at area where sharp contrast presence, the borders can be designed as pixels that have the maximum contrast to the neighbour pixels.

6.1.3 Summary

Frames around the watermarked image are used to mark the location of the watermarked image on the printed media. The frames create a high contrast step function around the image. The edge detection filters show good results to emphasise the contrast shift. After enhancement processes,
the frames are much more detectable. The high contrast steps are made of black frames on assumed white background, during the experiment. More adaptive high contrast methods may improve the performance, for example the frame pixels can be adaptively the highest contrast to the pixels they replace. Hence the edge detection functions are able to utilise the highest contrast to the neighbouring pixels and to produce enhanced results.

Least square linear regression is an optimal process to minimise the detection errors on the frames. The images are scanned from 4 different sides, and the results can further ensure the accuracy of detections and estimations.

The computer simulations do not yield any false detection for all digital images. Due to time constraint, however, comprehensive results on printed image synchronisation are absent from the research.

6.2 Synchronisation using RST invariant descriptors

6.2.1 Geometric invariant transforms

Through investigations, the invariant representations of images are the key to re-synchronise the distorted image accurately. It is similar to pattern matching techniques. For example, comparing Liu et al. (2005) and Wolberg and Zokai (2000) it is clear that they have significant similarity. Wolberg and Zokai (2000) applied the log-polar transform to image registration application that match two image with an arbitrary combination of rotation, translation, and scaling. Liu et al. (2005) applied the log-polar mapping to match the target image and discover the parameters of geometric transformations. If a two-dimensional image is translated, rotated and uniformly scaled, the new image can be described (Liu et al., 2005) as:

\[ i' = i \left[ \sigma (x \cos \alpha + y \sin \alpha) - x_0, \sigma (-x \sin \alpha + y \cos \alpha) - y_0 \right] \] (6.6)

As suggested by Wolberg and Zokai (2000), large scale factors "would alter the frequency content beyond recognition". The maximum scale factor suggested by Wolberg and Zokai is ten-fold. It is believed that this restriction does not effect the performance of this application. The rotation and translation in security document applications are normally caused by careless capture process. Scaling is commonly resulted by the different settings of printers and scanners. Since the security document image is printed on a physical material, the scale is assumed to be uniform along both axises,
and have no effect on aspect ratio. Rotation is originated at the geometrical centre of the image.

The translation invariance of the spectral magnitude was introduced by Bracewell (1986). The theorem is reviewed here in 1D signal, for the benefit of clarity, and it is trivially expendable to multiple dimensions as demonstrated by Ó Ruanaidh and Pun (1997). If the Fourier transform of a signal $f(x)$ is defined as $F(\omega)$, and the shifted version of $f(x)$ is defined as $f_1(x) = f(x - \tau)$ and its Fourier transform is defined as $F_1(\omega)$. The relation between $F(\omega)$ and $F_1(\omega)$ can be found as

$$ F_1(\omega) = \int_{-\infty}^{\infty} f_1(x) e^{-j\omega x} dx $$

$$ = \int_{x=0}^{N-1} f(x - \tau) e^{-j\omega x} dx $$

$$ = e^{-j\omega \tau} F(\omega) $$

then

$$ |F_1(\omega)| = |F(\omega)| $$  \hspace{1cm} (6.7)

Schalkoff (1989) indicates that the rotation in the spatial domain rotates the spectrum by the same amount and the $\alpha$ times scaling in the spatial domain scales the spectral magnitude by $\alpha^{-1}$. Through mapping the spectral magnitude from Cartesian coordinates $|F(u, v)|$ to polar coordinates, $(r, \theta)$, the rotation is presented as the linear shift along polar axis. The mapping is carried out as following:

$$ r = (u^2 + v^2)^{\frac{1}{2}} $$  \hspace{1cm} (6.8)

$$ \theta = \tan^{-1}(v/u) $$  \hspace{1cm} (6.9)

The scale invariant property of Mellin transform has been introduced several decades ago by Casasent and Psaltis (1976), and Cristóbal and Cohen (1997) introduced the scale representation of an image, called “the scale transform”. The Mellin transform is defined as:

$$ M(p) = \int_{0}^{-\infty} f(x)x^{p-1}dx $$  \hspace{1cm} (6.10)

Where $p \in \mathbb{C}$. If the scaled version of $f(x)$ is defined as $f_2(x) = f(\alpha x)$ and the Mellin transform of two signals are presented as $M(p)$ and $M_2(p)$ respectively, then the scale invariant property of Mellin transform is described as follow:
\[ M(p) = \int_0^\infty f(x)x^{p-1}dx \]

\[ M_2(p) = \int_0^\infty f_2(x)x^{p-1}dx \]

\[ = \int_0^\infty f(\alpha x)x^{p-1}dx \]

Let \( u = (\alpha x) \) the \( du = \alpha dx \)

\[ = \int_0^\infty f(u)u^{p-1}\alpha^{-p+1}\frac{1}{\alpha}du \]

\[ = \alpha^{-p+1-1} \int_0^\infty f(u)u^{p-1}du \]

\[ = \alpha^{-p}M(p) \]

then

\[ |M_2(p)| = |M(p)| \] (6.11)

The value of \( p \) has been defined in different ways, such as Cristóbal and Cohen (1997) defines as \( p = -jc + \frac{1}{2} \). Another example, Schalkoff (1989) and Casasent and Psaltis (1976) define the Mellin transform in polar-logarithmic plane, \( p = -j\omega - 1 \). The relations between Mellin transform and Fourier transform benefit the practical implementation as argued by De Sena and Rocchesso (2004), so the definition of \( p = -j\omega - 1 \) is preferred in this following content.

As discussed above, the translation, rotation and scale invariant space is the Mellin transform of polar sampled spectral magnitude of the image. The Mellin transform can be realised through Fast Fourier transform (FFT) (Brandt and Lin, 1996). If a signal \( f_L(x) = f(\ln x) \) is defined, the relations between both are described as

\[ M_L(\omega) = \int_0^\infty f_L(x)x^{-j\omega-1}dx \]

\[ = \int_1^\infty f(\ln x)x^{-j\omega-1}dx \]
Let $u = \ln(x)$ then $x = e^u$ and $du = \frac{1}{x}dx$

\[
\begin{align*}
\int_0^\infty f(u)e^{-ju\omega-u}e^{-u}du \\
= \int_0^\infty f(u)e^{-ju\omega}du \\
= \int_0^\infty f(x)e^{-j\omega x}dx
\end{align*}
\]  

This relation between Fourier and Mellin transforms, described in Equation 6.12, proves the Mellin transform can be implemented by FFT through sampling the uniformly distributed samples with a logarithmic function (Cristóbal and Cohen, 1997). De Sena and Rocchesso (2004) suggested a cubic spline interpolation for the sampling process. De Sena and Rocchesso stated this method has the “linear complexity and resolution of a tridiagonal matrix”.

Above all, an effective re-synchronisation scheme should consist the following steps and illustrated in Figure 6.5:

1. Taking Fourier transform of target image, yielding $I(u,v)$.
2. Taking the spectral magnitude, yielding $|I(u,v)|$.
3. Sampling the $|I(u,v)|$ to polar coordinates, yielding $i_p(r,\theta)$.
4. Logarithmic sampling $i_p(r,\theta)$ along the radius axis, yielding $i_p(\rho,\theta)$, the Fourier-Mellin Invariant descriptor (FMI).
5. Taking 2D Fourier transform of the FMI, yielding $I_m(\mu,\nu)$.
6. Extract phase information from FMI, yielding $\exp{-j\phi_I(\mu,\nu)}$.
7. Apply above procedure to synchronisation markers and yielding $\exp{-j\phi_G(\mu,\nu)}$.
8. Multiplying $\exp{-j\phi_I(\mu,\nu)}$ and $\exp{j\phi_G(\mu,\nu)}$ yielding $Q(\mu,\nu)$. This process is known as Symmetric Phase Only Matched Filter (SPOMF) (Chen et al., 1994).
9. Applying inverse Fourier transform to $Q(\mu,\nu)$ and yields $q(\rho,\theta)$.
10. Locate the peaks at $q(\lambda,\alpha)$ to recover the scaling and rotation parameters.
11. Inverse the rotation and scaling affects using the scaling factor $\sigma = \exp(\lambda)$ and rotation factor $\alpha.$
12. Apply cross-correlation of corrected target image and synchronisation marker using SPOMF to identify the translation parameters and size of the image.

13. Re-sample the watermarked image from the restored target image.

![Diagram](image)

Figure 6.5: Fourier-Mellin Invariant Descriptor diagram

### 6.2.2 Fourier Mellin invariant synchronisation

According to the invariant properties of the transforms, the rotated, scaled and translated image is expressed as Equation 6.6. Since the embedded watermarks are transformed with the cover image, the same expression applies to the watermark patterns. Let the \( p(x, y) \) denotes the synchronisation watermark embedded in the captured image, \( i(x, y) \).

\[
i(x, y) = p[\sigma(x \cdot \cos \alpha + y \cdot \sin \alpha) - x_0, \sigma(-x \cdot \sin \alpha + y \cdot \cos \alpha) - y_0]
\] (6.13)

After the first DFT (Step 1), the synchronisation patterns is

\[
I(u, v) = e^{-j\phi_p(u, v)} \sigma^{-2} |P[\sigma^{-1}(u \cos \alpha + v \sin \alpha), \sigma^{-1}(-u \sin \alpha + v \cos \alpha)]|
\] (6.14)

where \( \phi_p(u, v) \) is the spectral phase of \( P(u, v) \) (Chen et al., 1994). As proved in Equation 6.7 the translation invariant information is retrieved by taking
the spectral magnitude. The rotation is decoupled through re-sampling the magnitude from Cartesian coordinate to polar coordinate using Equation 6.8 and Equation 6.9. The rotation applied to the original pattern is expressed as

\[ i_p(r, \theta) = \sigma^{-2} p_p(r/\sigma, \theta - \alpha) \]  

(6.15)

Through logarithmic sampling of the radial axis, the scaling is decoupled to linear translation along logarithm radial axis.

\[ i_{lp}(\rho, \theta) = \sigma^{-2} p_{lp}(\rho - \lambda, \theta - \alpha) \]  

(6.16)

where \( \rho = \ln(r) \) and \( \lambda = \ln(\sigma) \). Taking DFT of \( i_{lp}(\rho, \theta) \), one obtains the FMI. It is noteworthy that \( |I_m(\mu, \nu)| = |P_m(\mu, \nu)| \). The following step adopts the SPOMF that can be expressed as

\[ Q(\mu, \nu) = \frac{I_m(\mu, \nu)P_m(\mu, \nu)}{|I_m(\mu, \nu)||P_m(\mu, \nu)|} \]

\[ = \exp\{-j(\lambda \mu + \alpha \nu) - j \phi(\rho \mu + \theta \nu)\} \cdot \exp\{j \phi_p(\rho \mu + \theta \nu)\} \]

\[ = \exp\{-j(\lambda \mu + \alpha \nu)\} \]

The output of an inverse DFT transform

\[ q(\rho, \theta) = \sum \sum \exp\{-j(\lambda \mu + \alpha \nu)\} \exp\{j(\mu \rho + \nu \theta)\} \]

\[ = \sum \sum \exp\{-j(\lambda \mu + \alpha \nu) + j(\mu \rho + \nu \theta)\} \]

\[ = \sum \sum \exp\{j(\rho - \lambda)\mu + j(\theta - \alpha)\nu\} \]

\[ = \sum \sum \exp\{-j(\lambda - \rho)\mu - j(\alpha - \theta)\nu\} \]  

(6.17)

Let \( u = \lambda - \rho \) and \( v = \alpha - \theta \)

\[ q(\lambda - u, \alpha - v) = \sum \sum \exp\{-j(um + vn)\} \]

\[ = \delta(u, v) \]

Therefore,

\[ q(\rho, \theta) = \delta(\lambda - \rho, \alpha - \theta) \]  

(6.18)

### 6.2.3 Synchronisation watermarks

Kutter (1999) and Alattar and Meyer (2003) both used 9 identical markers, but as Liu et al. (2005) indicate that detections made in Fourier-Mellin
transform domain only require one. The research developed a scheme only require the pattern repeatedly embedding 4 times to identify the size of the watermarked image and its location in the captured image. The size and the shape of the pattern are arbitrary. Refer to Figure 6.6, each one is located at the top-left and bottom-right corners. Another two are located at the top and left border of the watermarked image right next to the top-left corner marker. Totally four peaks are detectable at the Step 12. As the \( \times \) locations, three are located close to the top-left corner the horizontal and vertical distances, \( a \) and \( b \), between the three peaks decide the distance between the last peak and bottom and right borders.

![Figure 6.6: New design of locations of synchronisation markers (shadowed areas). \( \times \)'s in the figure, show the location of correlation peaks.](image)

The marker signal has been designed as sinusoidal signal (Day et al., 2004). But in this work the marker signal is Gaussian distributed random numbers \( (N(0, \sigma^2)) \) to avoid artifacts. Another reason to use AWGN is that the additional noise is modelled as AWGN, so the additional noise is not correlated with the synchronisation markers. The markers are embedded in the spatial domain. They could also be embedded in the frequency domain (DFT or DCT) for better robustness (Cox et al., 1997), but the Step 8 and Step 12 require both magnitude and phase information of markers respectively. Other solutions, are also possible, for example, embedding one set of markers in the frequency magnitude for FMI-SPOMF and other markers embedded in the spatial domain for Step 12. However due to time restriction the synchronisation watermarks are designed in spatial domain. Although this solution may not be optimal, the better design is left for future research.
All four markers are the pattern repeating at different locations in the spatial domain (Figure 6.6). They have identical frequency components, both magnitude and phase. After first FFT, Step 1, the magnitude information of four markers are overlapped in frequency domain to enhance the strength.

### 6.2.4 Symmetric phase only matched filter

The concept of Symmetric Phase Only Matched Filter (SPOMF) was first proposed for the optical communication described in Chen (1993, Chapter 3) and the performance is compared with other classic filters by Chen et al. (1994). Liu et al. (2005) made a similar comparison and conclusions that the SPOMF has a very sharp discrimination of correlation peak. In this work, the synchronisation marker comprises a set of watermarks embedded in the spatial domain.

The phase information required in Step 8 can be obtained through Fourier transform.

\[
I(u, v) = F[i(x, y)] = A(u, v)e^{-j\phi(u, v)}
\]

where \( F \) is the Fourier transform, \( i \) and \( I \) present the image in spatial domain and Fourier domain. \( A(\cdot) \) is the amplitude function and \( e^{-j\phi(\cdot)} \) is the phase function of \( I \). The energy of image pixels is transformed to amplitude of the Fourier coefficients. Normalising the Fourier coefficients with their amplitude eliminates those energy and yields an optimal operation spectrum for correlation based detector. Hence the resulted function has a unity spectrum, the process is also known as a whitening.

\[
\exp\{-j\phi(u, v)\} = \frac{I(u, v)}{A_I(u, v)}
\]

It is also clear that every image has different spectrum, and any one cannot be characterised individually. It is, therefore, interesting to find out how much the normalisation improves the correlation detection. An exponential factor \( \alpha \) is added to normalise amplitude in gradient levels. As shown below, when \( \alpha = 1 \) the amplitude energy is normalised, or whitened. When \( \alpha = 0 \), both phase and amplitude functions are present in the correlation function.

\[
\gamma_{\phi}(u, v) = \sum_{v=0}^{N} \sum_{u=0}^{M} A_1^{1-\alpha}(u, v)e^{-j\phi(u, v)} A_1^{1-\alpha}(u, v)e^{j\phi_M(u, v)}
\]
where \( \exp(-j\phi_I(\cdot)) \) and \( \exp(-j\phi_M(\cdot)) \) are the phase functions of the target image and the marker respectively. When \( \exp(-j\phi_I(\cdot)) = \exp(-j\phi_M(\cdot)) \) the above correlation function will become the autocorrelation function. The correlation \( \gamma_\delta \) composes several \( \delta \)-functions, yielding peaks at the locations at which the patterns match. To measure the performance of the correlation based scheme, the peaks are considered as signal hence they are where the detection interest lays, and the non-peak energy is considered to be noise. The logarithmic ratio of averaged signal energy and noise energy (SNR) benchmarks the performance. When it has no distortion applied, the results are the performance bound. A simulation is designed to evaluate the detection quality affected by spectrum normalisation. The results are shown in the next section.

### 6.2.5 SPOMF test

55 ID photos are impressed with synchronisation markers. Every image uses the same marker, which is pseudo randomly generated Gaussian White noise, 0 mean and variance is 1, using the algorithm described in (Press et al., 2002), and the same marker is embedded at multiple locations in the same image. The number and size of the markers are uniformly regulated. In the performance evaluation, the size of markers is \( 64 \times 64 \). The markers are allocated as described in Section 6.2.3. 4 markers are embedded. At this stage, the size and locations of the markers are assumed not affecting the SNR performance. According to the reality applications, the marker size and location can be designed to be more effective. To clarify the evaluation situation, the histogram of the percentage of the image that is covered by markers is presented in Figure 6.7.

The Figure 6.8a shows the SNR curves of 55 test images. And Figure 6.8b shows the mean SNR curve. The normalising factor \( \alpha \) ranges between 0 and 2.

As it shows, all curves have similar behaviours. The SNR reduces when \( \alpha \in [0, 0.18] \). It is followed by an increase between \( \alpha \in [0.18, 1.1] \). Then it falls down again. Surprisingly the maximum SNR is not obtained when \( \alpha = 1 \) instead, the average of maximum SNR is at \( \alpha = 1.1 \). The histogram of \( \alpha \) that has produced the maximum SNR is shown in Figure 6.9.

Following results may explain the unexpected behaviour. Figure 6.10 shows the average energy of signal and noise against \( \alpha \). Whilst the noise energy is normalised, the signal energy reduces with noise energy. The reason behind is that the pure amplitude of the cover image is not able to obtain at the detector, because of the blind detection. The amplitude used at this normalisation process contains amplitude functions of both image and marker.
Therefore, the marker is “over normalised” when \((A_I + A_M)\alpha = A_M\). Shown in the Figure 6.10, signal is reducing faster before \(\alpha\) reaches 0.2, and reducing the rate of decreasing, after \(\alpha = 0.2\). This change is likely caused by the energy ratio between signal and noise. After \(\alpha = 0.6\) the signal reduction almost stops, and reaches the biggest gap to the noise curve in the region between 1 and 1.2.

The interest of further study is focused on how the non-linear normalisation would affect the SNR. The SNR curve has shown an interesting non-linear change. Earlier, the affect is related to the power ratio between the signal and noise. To find more results to support the hypnosis, the markers are assigned more energy in the spatial domain, in the term of power that is presented as the AWGN variance.

In Figure 6.11, 5 SNR curves are plotted. They are the average of all 55 test images. The curve of “var=1” has been shown in Figure 6.8b. Other 4 SNR curves are of similar shape. In order to focus on the region of interest the figure is only plotted for \(\alpha \in \{0, 1\}\). According to the figure, the curves show an effect of “shift” from the right to the left with increasing marker energy. Also, the curves are getting more “compacted” with more dramatic changes. After \(\sigma^2 = 4\), the minimum SNR of each curve is believed shifting outside the scope of the figure to the left, but it is also out of scope of the study figure. From the Figure 6.11, it is clear to notice that the power difference
between markers and image is linked to the location where the minimum SNR is produced. Similar to early results, the signal and noise power are compared for helping understand the reason of the effect. The Figure 6.12
Figure 6.9: The histogram of $\alpha$ which obtains the best SNR

Figure 6.10: Average signal energy is compared with average noise energy

shows 5 signal power curves and the noise power curve which is independent of the change of $\alpha$, against the factor $\alpha$.

From the Figure 6.12 and Figure 6.11, the shapes of SNR curves and Sig-
Figure 6.11: SNR curves with difference marker power (variance).

Figure 6.12: The power comparison between signal with different variance and the noise, which is constant for all different signal power.

The signal power curves are also very similar, like the one presented in Figure 6.11. The signal energy curve is shifted to the left and becomes more compact as
the marker energy increases. The similar behaviours can only be resulted by a function. Due to the time constraint, the research on this topic is inconclusive, and further research is required to explain the non-linear behaviour.

6.3 Summary

In this chapter, the question of synchronising watermarked image from a bigger captured image is addressed. The first solution is rather straightforward, using high contrast borders to mark the location and regular size to control the scaling. Edge detection techniques are used to enhance the borders appearance and make easy for detections. Detected locations of the borders are generalised by an optimal linear regression algorithm to produce an estimated rotation angle for all four borders of the image. The produced angles may not be completely same, but an average of all four angles is used to minimise the error. The only part of the scheme which is not so perfect is the border detection method. Although the edges are emphasised, the detection is made by comparing pixel values sequentially with fixed thresholds. The detection apparently lacks flexibility and accurate detection of possible coordinates of the border pixels is not guaranteed.

Theoretically a more accurate scheme that is derived from invariant transforms is also described. The Fourier-Mellin transform decouples rotation and scaling to phase shifts along both axes of the FMI coordinates. Symmetric phase only matched filter (SPOMF) is proved having sharper discrimination of the correlation peaks. Hence the detection of rotation and scaling is proved to accurate to sub-pixel level (Liu et al., 2005). Later in this chapter the benefit of SPOMF is revealed at spatial domain watermark synchronisation that is important to identify the location of watermarked image from its background. Very sharp peaks are produced with undoubted detectability.

The only question left to answer is the use of watermarks. The FMI and SPOMF are often used as image registration application. There is no question about the large energy difference between two patterns to be matched. The synchronisation marks are required to have enough power to alter the magnitude information of the first DFT such that the presence of markers is detectable from the phase information of the Mellin transform. Unfortunately due the limited time, this question can only be answered by future research.
Chapter 7

Conclusions

7.1 Watermarking methods

In this research work, spread spectrum watermark techniques have been examined and it has been shown that the randomness of the sequences often leads to unstable performance, especially, when the length of the sequence is short. Zero correlation between the spreading sequences and cover vectors is only found if the length is long enough to provide sufficient degrees of randomness. This leads to the more satisfying approach that the spreading sequences be designed orthogonal to the cover vector. Hence the output of a correlation receiver can produce much better discrimination.

It is feasible to find non binary sequences orthogonal to the cover vector. However, if the sequences are constrained to binary, it is difficult to determine sequences orthogonal to the cover vector. This research work has presented methods to produce binary sequences quasi-orthogonal to a cover vector consisting of real numbers. The length of both cover vectors and spreading sequences directly affects the capacity of a watermarking system. Hence, it is desired that spreading sequences are as short as possible. As discussed in Chapter 3, it is difficult to produce such sequences, if the length is short. Complete orthogonal binary sequences may not exist for cover vectors in low dimension. Therefore, the research focused on determining a method to derive sequences that have the minimum correlation with cover vectors. It is also necessary to find the magnitude of the minimum correlation, since the orthogonalisation remainder restricts the system performance. Additionally, the conditions to produce required degree of orthogonalisation is also necessary in designing consideration.

In Chapter 3, sequences of this kind produce the optimal performance. Sequences generated by the orthogonalisation methods can significantly reduce
the interference caused by the cover vector. Therefore the minimum requirement on watermark signal power can also be reduced, providing much better fidelity. Reduced interference energy also produces larger margin against additional noise.

Three orthogonalisation algorithms have been presented. Each scheme shows good ability to reduce the native interference from the cover image. It is worth remarking that among the three methods, the simplest implementation, the SC method has been proved to produce the desired results with a number of conditions and it demonstrates the best performance. The method reorganises cover vector energy into two groups having equal energy to be assigned opposite signs, hence the cancellation reduces the energy of interference to a very low level. Comparing SC with its cousin GS, the SC provides an optimal routine to reorganise cover vector elements into two equal energy groups. It is relatively flexible and is very simple and fast. Preliminary and formal results have shown its superior performance of reducing the interference energy. However, the large quantity of the spreading sequences have to be saved and redistributed for use in detection and authentication. Cost efficiency concerns restrict SC to small scale but high performance requirement applications.

Derived from the idea behind CDMA, the HMS procedure provides good performance. Restricted by the size of the Hadamard matrix, quasi-orthogonal sequences are used with slightly higher margins compared to ideal orthogonality. However it was shown that the overlapped embedding technique of applying a modified HMS algorithm improves detection quality with a little compromise made to the image fidelity.

Embedding in the DCT domain has a clear advantage of robustness. It was also shown that it allows better orthogonalisation performance and less visual distortion despite orthogonalisation schemes. It was also found that visual masked embedding requires higher power to achieve error free detection, and the DWR values show that fidelity is compromised. The orthogonal sequences require very low power to provide reliable embedding, hence the visual masking is found not necessary, if the information detection quality has higher priority compared to visual degradation.

Conclusively, this thesis provides a new direction of spread spectrum watermarking techniques. A new branch of informed embedding technology is identified. Improvements over conventional spread spectrum watermark embedding techniques is clear, in terms of capacity, fidelity and robustness.

A completely opposite direction of research on watermark generation scheme has also been explored in this work. Instead of searching for the least correlated sequence, the most correlated sequence from a predefined codebook is used as a watermark. The biorthogonal matrix codebook pro-
vides optimal performance in terms of minimum interference both between codewords in the same coding bin and different bins. The robustness is also maximised due to the maximum $d_{\text{min}}$, for a given code length. The noise has to flip at least one quarter of the watermark vector in order to produce an error.

Analysis provided in this thesis has identified one performance deciding factor that a pair of codewords from either of the code bins are desired to be similarly correlated with the cover vector. This requirement found the only, but significant drawback of the biorthogonal matrix dirty paper coding that is the limited number of codewords and consequently codewords are far apart from each other, making such selection difficult.

A set of Goppa codes (32, 7, 10) was used as a non-orthogonal dirty paper code to compare the performance with the orthogonal codes. A significant advantage was shown as bit error rate that is known to the encoder. The results prove the analysis that non-orthogonal codewords are closer to each other, hence to find similarly correlated codewords is consequently easier. Also as expected, the robustness tests reveal that with additional noise, the errors produced by non-orthogonal codes increase much faster than orthogonal codes. Simulations proceeded from two different viewpoints, and results have been presented of the performance of different watermarking schemes. From visual quality consideration to power ratio consideration, the schemes are analysed and compared. Both viewpoints led to significantly different conclusions. The orthogonal dirty paper code method can implement error free detection at a very low visual distortion level, despite the actual power had added to the image. This indicates for some vectors, that watermark power is allowed to be higher than usual. In other words, the higher power of watermark that is produced by HMMDP causes less visual distortion than other sequences produced by HMS and SC, because the watermark is produced as the best correlated sequence to the cover vector. Orthogonalisation methods, however, produce sequences that have least correlation with the cover vector, and consequently produce higher visual distortions. The conclusion changes if watermark power is solely determined by a fixed ratio to the cover vector power. In this case, the orthogonalised codes demonstrate a significant advantage compared to dirty paper coding methods. Orthogonalisation sequences are optimised for the correlation decoder that decodes message by measuring the direction of correlation peaks, and it is not affected by any other factors. However, the dirty paper methods are restricted as it requires more energy for the watermark to bring the desired code set into advantage.

The dirty paper method described in this work is a framework that has the benefit of simplicity and clarity to implementation. The performance is
solely dependent on the use of code. The focus has been drawn to orthogonal codes. It is clear that orthogonal codes are the most robust choice. The use of the code also includes the binning scheme. Two classes of binning schemes have been discussed and the differences shown. There are other choices of codes and binning methods but robustness and fidelity are contradictory design criteria. Higher \( d_{\text{min}} \) brings better robustness, but requires higher power to achieve error free detection and consequently worse fidelity. On the contrary, smaller \( d_{\text{min}} \) reduces required watermark power, but provides less robust detection against additional noise.

7.2 Synchronisation

The synchronisation problem is solved by using high contrast frames that enclose all pixels of the watermarked image. The frames can be determined by border detection algorithms that are often used in image processing applications. Several border detection filters have been tested. The detected border location coordinates are represented by a series of pairs of coordinates. Locations are processed by an optimal linear regression algorithm. The result is an estimated rotation angle. The high contrast frames are not only used to correct image rotation, they are also used to indicate location of the image in the background. Scaling is decoupled by resampling the scaled image to the known size. In digital format, this scheme is robust and it does not fail. However, for the printed images, the associated dither effect produces many small dark dots on white paper. The black dots confuse the border detection algorithms, unless a smooth filter is applied. Some results have shown that the smooth filter can improve the detection quality, but, the filter also reduces the contrast at the exact location of the border.

The theoretical study of the FMI descriptor and SPOMF has shown that is has a very good potential as a accurate synchronisation scheme. Synchronisation watermarks are placed repeatedly in strategic locations of a watermarked image. Fourier Mellin transform will overlap the spectrum of multiple watermarks. The overlapping in spectrum magnitude ensures the small magnitude change is detectable when correlation detector is applied. The simulation results in marker location detection shows that the peak is very detectable, and requires low marker power.
7.3 Future works

The orthogonalisation scheme described in this thesis is one class of informed watermark coding scheme, but simulations that have been carried are not informed embedding. The latter needs to consider the direction of orthogonalisation remainder. For example if the orthogonalisation remainder is in favour of the information bit, to achieve required robustness at this cover vector requires less power. With this technique, orthogonalisation schemes can also be designed to work in error-free mode at a predefined additional noise level.

The dirty paper coding is a branch of channel coding. Research results developed from error correction coding may be applied to produce better performance watermarking codes. The application of Goppa code is a very good example. The advantage on BER is very clear. It is believed that some other codes that have more codewords than orthogonal codes and larger $d_{\text{min}}$ than the Goppa codes will deliver better performance.

The use of scramble keys that are applied to the orthogonal matrix is one method to produce channel code. From the application point of view, one scramble key improves the security features of the codebook. From performance point of view, one scramble key rotates the code sphere to a different angle. If keys can be found that can produce code spheres that have large angle apart then this is beneficial to providing better embedding performance. Methods that are able to produce this class of keys can be studied further in the future.

The research work is only under the assumption that the watermark decoder is required the minimum complexity and very fast decoding time. Thus the work has been focused on efficient encoding schemes. The decoding method is only designed as the simplest Maximum Likelihood decoder with hard decision. Commonly, soft decision can improve decoding performance noticeably. Further research can do studies on soft decision decoding of watermarking channel.

As the introduction of dirty paper coding scheme, another aspect of the watermarking performance surfaces, dirty paper coding. In the thesis, two classes codes are examined, orthogonal codes and non-orthogonal codes (Goppa codes). Orthogonal codes are very far apart in terms of Euclidean distance. Non-Orthogonal codes have interference between codewords. For orthogonal codes, there is no easy solution to change the far-apart property. But for non-orthogonal codes, the interference may be cancelled through some decoding algorithms. If the interference can be treated as Inter-Symbol Interference (ISI), there are some algorithms can be adopted to avoid any incorrect decoding. Shah et al. (2007) has proposed an iteration scheme for
ISI noise cancelling for magnetic recording channel. Adding watermark to image can also be simulated as adding image noise to watermark signal. The magnetic recording channel has been classified as colour noise (Shah et al., 2007). The watermarking channel can also be classified as coloured noise. Thus the scheme proposed by Shah et al. (2007) is expected to improve the watermark signal detection performance. However, further research is needed to prove the interference between non-orthogonal codes surely can be treated as ISI, furthermore, the watermark channel is required to be more formally classified.

The FMI-SPOMF synchronisation scheme has been theoretically studied, and it has been proved effective with image to image synchronisation (Chen et al., 1994). However, when it is applied to watermark to watermark synchronisation, the energy of correlation peaks has not been tested in FMI. Future study may test the scheme further and evaluate the effectiveness if additional noise is present.
Appendix A

Modern technologies improve identity

In year 2006, International Civil Aviation Organisation (ICAO) published its 6th edition of Machine Readable Travel Document (MRTD, Doc. 9303). This revision remarks the specifications of both conventional passports with a Machine Readable Zone (MRZ) (ICAO, 2006a) and biometrics capable electronic passports (ICAO, 2006b). This document may be well known as the specification of the “Electronic Passport”. At October 2005, the U.S. lawmakers approved “Electronic Passport” ¹. Officially declared that the support of the new international specification of MRTD. According to the ICAO report, there are 41 countries have expressed their intentions to upgrade to the “ePassport” in two years time ².

Tougher border control is only one side of the story. Within the border, the number of crimes by abusing the weakening identification system is growing, namely “ID Theft”. “ID theft” defined by the American law is “a fraud committed using the identifying information of another person” (Federal Trade Commission, U.S., 2004, page 9 §603 (q) (3)). On both sides of the Atlantic, governments launched national campaigns against the latest development ³ providing advice to the public of threats posted by the criminals. The latest report commissioned by the U.S. Federal Trade Commission (FTC), indicates continuously increasing trend of the “ID Theft” crimes since 2003. The number of complaints was 685,000 at 2005, and it

²MRTD Report Vol.1 No.1 2006 Page 34
costs the consumer $680 million. "ID Theft" is still the top Consumer Sentinel complaint category, at 37% (Federal Trade Commission, U.S., 2006). In the UK, an official estimation puts a £1.7 billion price to identity theft, according to Home Office figures.

The criminals use fraudulent identity document do not only engage financial crimes. Investigations, in "The 9/11 Commission Report", indicates that the attacks have evidential links to the flaws in the identification system in the U.S.. Several terrorists used forgery documents to enter U.S. and even claim legal ID documents, like the Driver's License, as Kean et al. (2004) have found out, so that they could skip the security facility before they boarded the flights, that they hijacked.

Same at the year 2005, the "Real ID Act 2005" (the Bill H.R. 418) (U.S. Senate, 2005a) also became a public law. One year later, the British Parliament passed “Identity Cards Act 2006” expressed a strong willing toward tougher identity authentification. At the end of 2006, Chinese government has finished its five year long project, building the largest national identity database, containing 1.3 billion entries.

Facing the threats both in and out the national borders, the Governments have to improve the identity system by applying new technologies to clarify whether the person is indeed whom is being claimed. However, the new amendments in policies were not welcomed by everyone. The limitations and characteristics of a certain technology often brings cheers and criticisms. The following sections explain some of the major technologies under serious debates by the lawmakers and oppositions.

An identity document, regardless its formation, is a presentation of the bearer in other domain, commonly electronically. It is not only required to be capable of automated reading operations, it is also required to reliably identify the bearer.

Contact and contactless smart cards have started being or will be used as ID documents. Whether the cards are secure enough and whether the information in the card or the card itself may be tempered? Biometrics are commonly recognised as the unquestionable identity of individuals. However, whether the present technologies are unquestionable to retrieve the uniqueness of every one safely and quickly? The debates around these two hottest identity technologies are summarised in the following sections.

Not too far from the pitch, another player is warming up to cheer the crowd. Digital watermarks can be placed in almost any media. With the

---

4 published at Feb. 2006, cited from “Identity Fraud” website in the U.K.
lowest cost to deploy, compared to the other two. It is highly temper-prove, due to the information leaves the digital domain after deployment and fixed in the analogue form. Any temper is evidently to be detected. The last section explains why the digital watermark is a prime choice for identification applications, and estimates the possible reasons why the undergoing programmes did not choose this technologies.

A.1 Smart Cards

A.1.1 Background

A smart card normally refers to one or more Integrated Circuit Chips (ICCs), which may contain a micro-control unit and some memory (such as EEPROM, ROM and RAM). The IC chip(s) may be carried in a number of physical protections, for example, the most common form is a plastic card, similar to majority credit cards. A smart card is activated by a reader, to download/upload data. Some of the smart cards use secure file system and cryptographic processor to protect their content. Other kinds of smart cards may solely used as in storage purposes.

The inventor of the first smart card is not commonly agreed, but it is safe to say that the smart card was invented in 1970s. In 1976, French Roland Moreno filed a patent for a memory card. In 1978, Michel Ugon filed a patent includes a microprocessor in a small portable device for data transfer. This marks the complete concept of today's smart card.

The first mass use of Moreno's invention was payment for French public payphone, starting at 1983. Later in early 1990s, the smart cards were widely used as credit and debit cards by European banks. In 1993, Master Card, VISA and Europay agreed to develop a specification of smart card used in payment cards, the EMV. The boom of smart cards came at the later 1990s, when the GSM mobile SIM cards were greatly demanded.

Smart cards using contactless interface became very popular when the world has turned to the new century. Lead by the applications in the mass transportation systems around globe, this fast and easy way of payment has become a part of millions people's daily life. Operating in a near field of the reader, the ICC applies the energy received from the reader via the card's foil or antenna. The card may be a simple model for data storage, or it may

---


7Michel Ugon (Saint-Ouen, France), "Portable data carrier including a microprocessor", United States Patent 4,211,919, July 8, 1980
also have many complicated facilities, such as simple but sophisticated file system and crypto-circuits, to protect the information stored in it memory. An International Organisation for Standardisation (ISO) publication \(^8\) (dated 2004 and partially amended at 2006) establishes the globe trend to fast, easy and secure electronic card systems.

### A.1.2 Pros and cons

The smart cards, regardless the connection interface, have a number of advantages, such as

- **High data transfer rate** in several hundreds kilobits per second.

- **Multifunctional**, on a same card multiple applications such as banking, access control and identity can co-exist, even multiple interfaces, contact or contactless.

- **Large storage space**, the memory size is ranged between 64 bytes to 64K bytes \(^9\).

- **Interoperability**, ensured by well established international standards.

- **Security**, wired logic access control, crypto-processor supports 3DES, AES, ECC and RSA.

- **Secure processing biometrics** by on-card comparing live biometrics and templates (Contactless smart chip technology: the business benefit, 2005).

Most of these advantages lay the foundation on which the secured electronic identity systems can be built. However, the smart cards also have problems which undoubtedly bring serious concerns.

The smart card ICCs are subject to non-invasive, semi-invasive and invasive attacks. The ICCs and the memories could be disabled. Some cases, the security and sensitive information stored in the chip could be compromised (Barker et al., 2005). Through these attacks, the criminals can also tamper, forge or reproduce false identity documents. Particularly for contactless smart cards, two types of attacks threaten throughout the entire life of the cards, namely skimming and eavesdropping. Skimming is an attack by “creating an unauthorised connection with a readable chip in order to gain

---

\(^8\)ISO/IEC 14443, “Identification Cards - Contactless integrated circuit(s) cards - Proximity cards”

\(^9\)i.d.
access to the data contained therein" (U.S. Office of Federal Register, 2005). Eavesdropping is the interception of the legitimate communication session between an ICCs and an authorised reader.

Apart from these classic attacks, recently, an unpublished survey report by British Identity and Passport Service has raised the question about the smart cards’ durability. Ten year durability of smart cards are considered possible conditionally. Particularly for contactless smart cards the life time is subject to flexibility of the base card, because the connections between the inlaid antenna and ICs can be damaged. Another recent report published by the British National Audit Office (NAO) brings a question mark to the undergoing British “biometrics passport”. In the report (Nation Audit Office, UK, 2007), the chips supplied by the manufactures can only be offered limited warranty, which is significantly shorter than the designated life time of the travel document. Another problem is that the intellectual propriety normally belongs to the chip manufactures, thus the actual user who concerns about security and continuous issuance may not have exclusive use of the same type of chip and the supply may be discontinued.

The smart card itself does not have any direct link between the card and the card presenter, when a verification process occurs. Unless the card is enabled with biometric authentication programs. So the authenticity of the smart card does not exist, and it has to depend on the effectiveness of one or more on card biometric schemes. In the following section, some of the sophisticated biometric schemes are introduced.

A.2 Biometrics

A.2.1 Background

Biometrics by definition is “the automated means of recognising a living person through measurements of physiological or behavioural traits” (NTWG, 2004, page 8). Fingerprint is the most widely used biometric system. The first publication was in 1684, describing systematic study of ridge, pore and furrow structure in fingerprints (Castle, 2001). Now automatic fingerprint recognition is used by law enforcement agencies in front line of forensic identification.

Classifications of biometrics may be defined as physiological and behavioural biometrics. Physiological biometrics are derived from the measurement of a part of human body’s anatomy, such as hands, fingers, face and eyes. Behavioural biometrics are derived from the measurement of a person’s action performance, between the beginning, progressing and end.
Example of behavioural biometrics are signature, voice and keystrokes.

A.2.2 Pros and cons

Distinctiveness of a certain biometric type is profound. The best commercialised type is the fingerprint scan. The chance that two persons have identical fingerprints is estimated “one in a hundred billion” (U.S. General Service Administration, 2004). Balanced between performance and cost, fingerprint is the most considered and the widest available biometric scheme. Another well established robust form of biometric is the iris scan. It has six times more distinct, identifiable features than fingerprint, and unlike fingerprints, iris does not subject to wear and injury. Similar to iris scan, retina scan is also very distinctive and robust. But it is the most expensive form of biometric implementation.

With recently established or amended international standards, biometrics are possible for international interoperability. For example, under the ISO/IEC 19794 umbrella, namely “Information technology – Biometric data interchange formats”, a dozen documents are available to developments of wide range of products. Among them, seven established international standards are:

- ISO/IEC 19794-1:2006 Framework
- ISO/IEC 19794-2:2005 Finger minutiae data
- ISO/IEC 19794-3:2006 Finger pattern spectral data
- ISO/IEC 19794-4:2005 Finger image data
- ISO/IEC 19794-5:2005 Face image data
- ISO/IEC 19794-6:2005 Iris image data
- ISO/IEC 19794-8:2006 Finger pattern skeletal data

Three dozens of relevant international standards regarding frameworks, testing and more types of biometrics are under development or review.

Despite the profound accuracy represented by biometrics, as its matching depends on the development of pattern recognition technologies, the program is restricted by the designer's knowledge, methods of processing the templates and methods of capturing. The biometrics are always inexact (Callas, 2001). All systems have to balance between the number of the wrongly accepted and the number of wrongly rejected, called false positive rate and false negative
rate (McClimans, 2001; Matthews, 2002), respectively. For example, “The false negative rate range was zero to 44% while the false positive rate range was zero to 0.4%” (Pankanti et al., 2000). That indicates, in order to block all intruders, about a half of the users must be “sacrificed”. The biometric systems tend to work 99.99% of the time. Thus they are about as good as a 4-digit pin system (Callas, 2001).

In their current shape, biometrics technologies are not reliable to maintain a stable system. As pointed out by a government expert, “the use of current facial recognition technology with two dimensional images of limited resolution is not sufficiently reliable to enable fully automated searches, even in a relatively small database \(^{10}\), and the performance is known to decline as the database size increases” (Nation Audit Office, UK, 2007).

Cost issue is still remain a prime reason why the biometrics can not be deployed as a national or even international identification standard. This cost, according to a British governmental report (Nation Audit Office, UK, 2007), mainly spreads over additional equipments, personnel training and buying intellectual properties. UK government, recently reported, has dropped iris scan as a part of its biometrics scheme for National Identity Register. James Hall, the chief executive of the Identity and Passport Service has told the press “the decision was down to cost. Collecting every biometric involves significant extra cost and I believe we can achieve the objective - securing people’s identities - without irises” \(^{11}\).

Perhaps the most worrying concern about the biometrics is that no matter what biometric measurement uses, there always are 1% to 2% of population will fall outside the model (Callas, 2001), and the alternatives are also limited for individuals.

### A.3 Digital Watermarks

#### A.3.1 Background

Among those newly developed security technologies, digital watermark has unique specifications. The digital watermark is noise-like pixels, printed together with the photograph on physical media. Perceptibly, it has little effect on the image quality. Under high resolution scanners or cameras, however, a watermark can be detected and read. Traditionally digital watermarks are

\(^{10}\) Approximately 100,000 individuals

commonly used to identify the rightful owner or buyer of an imaging art piece. When a digital watermark is printed on an identification document, it can not only verify the card's user, but also authenticates the photograph on the card therefore the authenticity of the card.

The watermark is designed to be imperceptible. It is hidden in the photograph image. Not like smart card and magnetic strip, the watermark cannot be located precisely, so that removal and alter the watermark without causing visual distortion is incredibly hard due to the physical cover is the photograph pixels. Modern watermarking schemes use spreading spectrum communication technology (Cox et al., 1997), which was designed for secure communication (Proakis, 2001). The watermark is securely coded, without the knowledge of the spreading sequence, to detect watermark is nearly impossible.

A.3.2 Pros and cons

The digital watermarking technology has many unbeatable advantages to be used as a strong identification measurement. Also as a new member of the technology family which is considered to be a tool of authentification, it has some limitations requiring further research and development. In this section the pros and cons are explained.

Digital watermarks are designed to hide in the photograph of the identity document’s bearer, so it must be produced at the time of issuance. It combines facial biometric characteristics with an extraordinary large code, normally millions bit long. Unlike other security measures, watermarks does not rely on any other physical template to authenticate, so that it is not necessary to produce large volume of high valued blank templates prior to issuance. One figure shows that in the case of without contingency plan, store blank UK’s biometric passport could cost £60 million over five years (Nation Audit Office, UK, 2007). In the same report, it indicates that in the UK, passport book production cost increased from paper book’s £5.00 to ePassport’s £12.25. The extra 145% increase are the cost of additional authentification devices inlaid the new passport. On the contrary, watermarks can be printed on any material which is the best suitable, technically or economically. Deploy and retrieve requires no additional device, other than existing printers and scanners/cameras.

The watermarks have the same life time as the holder’s photographs. They will complete the designed life time of the document itself. If the photograph is protected against wear, tear, folding and colour fading, the watermarks embedded are also protected.

Watermarks can only be embedded at the issuance. Without authorised
access to the sequence generator, it is impossible to enrol a new user and assign a new document. Successful forgery is almost impossible. By auditing access to the issuance counters, watermarks can also stop fraud.

Sadly, at the present form, watermarking is not yet the perfect solution. Because there is no international standard defining watermark technology to provide international interoperability, even nationally. Lack of supplier and competition strain the technical progress toward practical mass deployment.

Technically, the watermarking algorithms are also marginally behind the competitive technologies. For example, on the storage side, contactless smart cards normally have several tens of thousands bytes of capacity, the reported highest digital watermarking capacity in the digital domain is one of sixty-fourth of the number of total pixels (Miller et al., 2004). For another example, digital watermarks use a single form of biometrics, the facial imaging, however, the smart card is capable to use a variety of biometrics. In the case of reading time, it only takes 8 seconds to read a contactless smart card specified in the ePassport standards, but between algorithms, reading time of the digital watermarks from a printed photograph varies.

A.4 Discussions

Security issues in present days are commonly linked with identity authentications. The human societies around globe have encountered an identity crisis never has happened. The conventional identity systems are desperately required improvement. Conventional technologies commonly applied to verify identity are the printed ID cards and passports, but their authenticity is increasingly doubted. Many modern technologies are being or about to be deployed to protect humanity from the enemies come from inside.

Smart cards with contact or contactless interface, containing one or more micro-control units and a number of memory modules, provide a robust platform to implement a number of applications, which are mainly focused on fast payment and identity. In reality this robust platform is subject to even more sophisticated attacks, physically and logically. Also, without a strong authentification application, the platform is valueless, especially to prove the authenticity of the identity document and it’s bearer.

Biometrics characteristically are such an application that can be built upon a robust platform like smart cards such that the combination could provide indubitable authenticity and may ultimately realise solid identity. However, at the present shape, biometrics, despite the international support, are still not so accurate to realise the ideal.

An ancient concept - watermarks - derived into a sophisticated technol-
ogy, which was developed to protect the intellectual property of digitally formed artworks, is given a new life to protect the authenticity of the identity document. Watermarks possess unique self-authentification property at an extremely low cost and live as long as the identity document itself. Digital watermarks are secured by the modern communication and cryptography theories. Digital processing techniques ensure its stealth. However, as all new technologies, digital watermarking algorithms are infantile. Many improvement must be made through large amount of research and development. For example the number of information bits which is able to hide in forms of digital watermarks is still limited and not sufficient. The high capacity hiding techniques are still causing perceptible distortions. The robustness of the watermarks requires further improvement.
Appendix B

2D filter design

Filtering is a major signal processing topic in watermarking system design. It is important to review the concept of LTI filter design. A digital filter system, used in image processing, is linear time-invariant system (Wade, 1994). This system is essentially an algorithm for converting one sequence into another (Rabiner and Gold, 1975). The linear property is defined as if $x_1(n)$, $x_2(n)$ are the inputs of the system, respectively, and $y_1(n)$, $y_2(n)$ are the coordinated outputs. If $ax_1(n) + bx_2(n)$ is applied as the input, $ay_1(n) + by_2(n)$ is obtained at the output (Rabiner and Gold, 1975). The time-invariant property is defined as: if $x(n)$ applies at the input, and $y(n)$ is obtained at output; then input sequence $x(n - m)$ produces the output $y(n - m)$. Then an LTI system is described as

$$y(n) = \sum_{m=-\infty}^{\infty} h(m)x(n - m)$$  \hspace{1cm} (B.1)

where $h(m)$ is the impulse response of a LTI system.

If a special input sequence

$$x(n) = e^{j\omega n} \quad -\infty < n < \infty$$

is applied to an LTI system input, the Equation B.1 gives the output as

$$y(n) = \sum_{m=-\infty}^{\infty} h(m)e^{j\omega(n-m)}$$

$$= e^{j\omega n} \sum_{m=-\infty}^{\infty} h(m)e^{-j\omega m}$$

$$= x(n)H(e^{j\omega})$$  \hspace{1cm} (B.2)
\( H(e^{j\omega}) \) is called the frequency response of the system (Rabiner and Gold, 1975), and its impulse response can be described in form of Fourier series pair

\[
h(n) = \sum_{m=-\infty}^{\infty} H(e^{j\omega}) e^{j\omega m} \tag{B.3}
\]

In the Equation B.3, \( h(n) \) can be considered to be essentially a superposition of sinusoid \( e^{j\omega} \) of amplitude \( H(e^{j\omega}) \) (Rabiner and Gold, 1975).

Similarly, for arbitrary input sequence \( x(n) \)

\[
x(n) = \sum_{m=-\infty}^{\infty} X(e^{j\omega}) e^{j\omega m} \tag{B.4}
\]

where

\[
X(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m) e^{-j\omega m} \tag{B.5}
\]

So, in frequency domain, when \( x(n) \) inputs the system \( h(n) \), it yields the following output,

\[
Y(e^{j\omega}) = X(e^{j\omega}) H(e^{j\omega}) \tag{B.6}
\]

The Equation B.6 describes an essential property of convolution summary, which is that convolution in the time domain is converted to multiplication in the frequency domain (Rabiner and Gold, 1975).

There are many ways to design digital frequency selective filter. Between two major types of filters, Finite Impulse Response (FIR) and Infinite Impulse Response (IIR), the FIR filter has guaranteed linear phase response and its structurally simple. Thus this work focuses on design work of FIR filters. The design of a FIR filter can start with windowing, frequency sampling or optimal methods. The method used here is using the frequency domain sinusoidal (Ifeachor and Jervis, 2002). The straightforward way is to specify the ideal frequency response of the desired frequency band to be selected by this filter. The specified the frequency response of the filter on the ideal basis is:

\[
H(e^{j\omega}) = \begin{cases} 
1, & -\omega_c \leq \omega \leq \omega_c \\
0, & \text{otherwise} 
\end{cases} \tag{B.7}
\]
From B.7, the impulse response can be obtained as

\[ h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega n} d\omega = \frac{1}{2\pi} \frac{2\sin(\omega_c n)}{n} = 2f_c \frac{\sin(\omega_c n)}{\omega_c n} = 2f_c \text{sinc}(\omega_c n) \]  \hspace{1cm} (B.8)

where

\[ \text{sinc}(x) = \begin{cases} 1 & \text{when } x = 0 \\ \frac{\sin(x)}{x} & \text{when } x \neq 0 \end{cases} \]  \hspace{1cm} (B.9)

according to L'Hôpital rule (James et al., 1992).

The Figure B.1 shows the three examples of high-pass, band-pass and band-stop filter’s frequency responses. (The spectrum is normalised against the sampling frequency.)

![Figure B.1](image)

Figure B.1: Ideal filter response in frequency domain

For a band-pass filter, its frequency response might be described as follow,
if $\omega_1 < \omega_2$

$$H(e^{j\omega}) = \begin{cases} 1 & \omega_1 < |\omega| < \omega_2 \\ 0 & \text{otherwise} \end{cases} \quad (B.10)$$

The impulse response can be obtained from

$$h(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\omega})e^{jn\omega}d\omega + \int_{\pi}^{2\pi} H(e^{j\omega})e^{jn\omega}d\omega \quad (B.11)$$

$$= 2f_{c_2}\text{sinc}(\omega_{c_2}n) - 2f_{c_1}\text{sinc}(\omega_{c_1}n) \quad (B.12)$$

The Equation B.11 shows the method which combines two pass-bands, on each side. On the contrast, the Equation B.12 is composed by subtracting one low pass filter on frequency band from $-\omega_{c_2}$ to $\omega_{c_2}$ by another that is from $-\omega_{c_1}$ to $\omega_{c_1}$.

If $\omega_1 = 0$, the second term of the Equation B.12 is always zero as shown in Equation B.8. The filter is a low-pass filter with a cut-off frequency at $f_{c_2}$. If the the higher band $\omega_{c_2}$ equals to $2\pi$, the first term in the Equation B.12 is always zero. The resulted filter is a high-pass filter. For band stop filters.

$$h(n) = 2f_{c_1}\text{sinc}(\omega_{c_1}n) - 2f_{c_2}\text{sinc}(\omega_{c_2}n) \quad (B.13)$$

By applying inverse Fourier transform, the desired frequency response is sampled at a large number of points, even infinite (Wade, 1994). Such a system is non-causal. To realise a practical filter, the number of coefficients has to be truncated to a finite number. This direct truncation leads to the well known Gibbs phenomenon. A finite weighting sequence $w(n)$, called a window, can be used to modify the coefficients $h(n)$, to control the convergence of the Fourier series (Rabiner and Gold, 1975).

Some window functions that approximate the desired characteristics.

$$w_{\text{Hamming}}(n) = (0.54 + 0.46 \cos(2\pi n/N)) \quad (B.14)$$

$$w_{\text{Blackman}}(n) = 0.42 - 0.50 \cos(2\pi n/N) + 0.08 \cos(4\pi n/N) \quad (B.15)$$

$$w_{\text{Blackman-Harris}}(n) = 0.42323 + 0.49755 \cos(2\pi n/N) + 0.07922 \cos(4\pi n/N) \quad (B.16)$$

"For linear time-invariant systems the convolution theorem is valid" (Rabiner and Gold, 1975). So, if a two-dimensional system with finite impulse response $h(n_1, n_2)$ has an input $x(n_1, n_2)$, the output $y(n_1, n_2)$ is determined as

$$y(n_1, n_2) = x(n_1, n_2) * h(n_1, n_2)$$

$$= \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} x(m_1, m_2) h(n_1 - m_1, n_2 - m_2) \quad (B.17)$$
Compare the impact on filters response with different windows


g(Nx, Ny) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathcal{H}(e^{j(\omega_x)}e^{j(\omega_y)}) d\omega_x d\omega_y

= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} H(e^{j\omega_x}) e^{j\omega_x} d\omega_x \int_{-\pi}^{\pi} H(e^{j\omega_y}) e^{j\omega_y} d\omega_y

= g(Nx)f(Ny)

---

Figure B.2: Frequency response of a high-pass filter with 125 coefficients applied with Rectangular window, Hamming and Blackman-Harris

where $M_1, M_2$ are the length of filter each dimension. As defined in (Rabiner and Gold, 1975), a two-dimensional sinusoid sequence on the input of the system is

$$x(n_1, n_2) = e^{j(\omega_1 n_1 + \omega_2 n_2)} \quad -\infty < n_1 < \infty; \quad -\infty < n_2 < \infty$$ (B.18)

Substitute Equation B.18 into Equation B.17 and expand during time period $0 \leq m_1 < M_2, 0 \leq m_2 < M_2$

$$\mathcal{H}(e^{j\omega_1}, e^{j\omega_2}) = \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} h(m_1, m_2) e^{-j\omega_1 m_1} e^{-j\omega_2 m_2}$$ (B.19)

The filter coefficients can be derived from applying the inverse Fourier transform with ideal frequency specifications, for example a 2-D low pass filter has the impulse response as:

$$h(n_x, n_y) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} H(e^{j(\omega_x)}e^{j\omega_y}) e^{j(\omega_x n_x + \omega_y n_y)} d\omega_x d\omega_y$$

= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} H(e^{j\omega_x}) e^{j\omega_x} d\omega_x \int_{-\pi}^{\pi} H(e^{j\omega_y}) e^{j\omega_y} d\omega_y

= g(n_x)f(n_y)$$ (B.20)
Shown in Equation B.20, a 2-D filter’s impulse response is obtained through the product of two separately designed 1-D filter coefficients (Rabiner and Gold, 1975), respectively to satisfy the specifications on horizontal \( H(e^{j\omega_x}) \) and vertical \( H(e^{j\omega_y}) \) directions respectively. The product can be made by forming the \( H(e^{j\omega_y}) \) into a diagonal matrix and repeating \( H(e^{j\omega_x}) \) for every row.

Same as the 1D filter design, a 2D FIR filter also has to truncate coefficients, therefore the 1D windowing function has a 2D derivation \( W(e^{j\omega_x}, e^{j\omega_y}) \). then,

\[
w(n_x, n_y) = w_x(n_x)w_y(n_y)
\]  

(B.21)

![Frequency response](image1.png)

![Impulse response](image2.png)

Figure B.3: Two Hamming 1-D filter multiplied to be a 2-D window function

The two-dimensional filter design, like 1D filter discussed previously, is generalised following. First use a rectangular bandpass filter as an example, because it is easily convertible to lowpass, highpass and bandstop filters.
Figure B.4: Ideal 2-D bandpass filter responses in frequency domain

The ideal passband frequency response is described as

\[
H_x(e^{j\omega_x}, e^{j\omega_y}) = \begin{cases} 
1, & |\omega_{c1x}| \leq |\omega_x| \leq |\omega_{c2x}| \quad \text{and} \\
|\omega_{c1y}| \leq |\omega_y| \leq |\omega_{c2y}| \\
0, & \text{otherwise}
\end{cases}
\]

In the Figure B.4, the shadowed area is the subtraction of two rectangles. The large one is spanned by \(\pm \omega_{c2x}\) and \(\pm \omega_{c2y}\). The small one is spanned by \(\pm \omega_{c1x}\) and \(\pm \omega_{c1y}\). The frequency response of the 2D bandpass filter is given as

\[
h(n_x, n_y) = \text{sinc}(\omega_{c2x} n_x) \cdot \text{sinc}(\omega_{c2y} n_y) - \text{sinc}(\omega_{c1x} n_x) \cdot \text{sinc}(\omega_{c1y} n_y)
\]

The generalised two-dimensional band-pass filter impulse response is given by

\[
h(n_x, n_y) = 2f_{c2x}2f_{c2y} \text{sinc}(\omega_{c2x} n_x) \cdot \text{sinc}(\omega_{c2y} n_y) - 2f_{c1x}2f_{c1y} \text{sinc}(\omega_{c1x} n_x) \cdot \text{sinc}(\omega_{c1y} n_y)
\]

(B.22)

where \(|\omega_{c2x}| > |\omega_{c1x}|\) and \(|\omega_{c2y}| > |\omega_{c1y}|\).
One example frequency response, which is presented in linear scale and log scale in Figure B.5.

In the above content, designing 2D filter using 2-Dimensional square ideal frequency response was reviewed. Alternative design routines are also available, for example Rabiner and Gold (1975) and Banks (1990) also introduced design methods based on circle ideal frequency response. Through some experiments it has been found out that the circled designing method does not supply significant improvement over its complexity and also, the cutoff frequency is only normalised along horizontal and vertical axes.
Bibliography

General Bibliography


nal dither modulation: a high-rate data-hiding method invariant to gain

2770/2003 of *Lecture Notes in Computer Science*, Springer Berlin / Hei-
delberg, chapter 6, pp. 81–92.

Inc.


Hill.

processing*, Prentice-Hall.

Salmon, R. A. (2001). Watermarking: what it is, why we do it, how can it
be used, *White Paper WHP009*, BBC Research & Development.

Schalkoff, R. J. (1989). *Digital image processing and computer vision*, John
Wiley & Sons, Inc.

in watermarking: A survey, in A. G. Tescher, B. Vasudev and V. M. Bove

soft-feedback equalization method for multilevel magnetic recording, *IEEE

second edn, Addison Wesley Longman.

Smith, J. R. and Comiskey, B. O. (1996). Modulation and information hid-
ing in images, *Proc. of the first information hiding workshop*, Vol. 1174 of
*Springer-Verlag Lecture Notes in Computer Science*, Isaac Newton Insti-
tute, Cambridge, U.K.


**Publications**


**Bibliography for security research**


MISSING PAGES REMOVED ON INSTRUCTION FROM THE UNIVERSITY