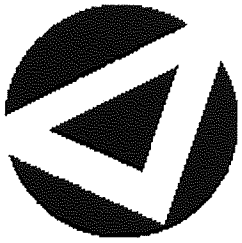
Digital Image Watermarking

Stéphane Bounkong

Doctor of Philosophy



ASTON UNIVERSITY

April 2004

ASTON UNIVERSITY

Digital Image Watermarking

STÉPHANE BOUNKONG

Doctor of Philosophy, 2004

Thesis Summary

In recent years, interest in digital watermarking has grown significantly. Indeed, the use of digital watermarking techniques is seen as a promising mean to protect intellectual property rights of digital data and to ensure the authentication of digital data. Thus, a significant research effort has been devoted to the study of practical watermarking systems, in particular for digital images.

In this thesis, a practical and principled approach to the problem is adopted. Several aspects of practical watermarking schemes are investigated. First, a power constaint formulation of the problem is presented. Then, a new analysis of quantisation effects on the information rate of digital watermarking scheme is proposed and compared to other approaches suggested in the literature. Subsequently, a new information embedding technique, based on quantisation, is put forward and its performance evaluated. Finally, the influence of image data representation on the performance of practical scheme is studied along with a new representation based on independent component analysis.

Keywords: Watermarking, information hiding, digital image, ICA

To my parents, Elise and Francis.

Acknowledgements

First and foremost I am grateful to my supervisor Prof. David Saad and to my advisor Prof. David Lowe for their patience, their ability to put order into my sometimes chaotic ideas, the sparkling discussions we had together, and the theoretical explanations they gave me whenever I needed it. Being a member of Neural Computing Research Group (NCRG) was very stimulating and during the three years I spent in Aston, I profited a lot from the discussions I had with students and members of staff. With this regard, a special thanks to Borémi Toch with whom I have spent hours arguing about theories in watermarking. I also would like to acknowledge the financial support of NCRG for making my study at Aston possible.

My fellow MSc and PhD students helped me adjust to life in UK. The lively discussions we engaged into are not forgotten. I thank my fellow course-mates, Borémi Toch, Lehel Csató, Randa Herzallah, Wei Lee Woon and Sun Yi.

For their remote support from France, I thank my family and my friends without whom the completion of this thesis would have been much harder.

In preparing this thesis I have been helped a lot by Borémi Toch and Chantal Gagnon who gave helpful comments on the grammatical aspects of the thesis. And finally, special thanks to Vicky Bond for her prompt replies whenever it was needed.

Contents

List of Figures

List of Tables

Declaration

This thesis describes the work carried out between January 2001 and April 2004 in the Neural Computing Research Group at Aston University under the supervision of Prof. David Saad.

This thesis has been composed by myself and has not been submitted, nor any similar dissertation, in any previous application for a degree.

Chapter 1

Introduction

The aims of this introductory chapter are to provide a general background to the digital watermarking problem, the scope and outline of this thesis, a formalised definition of the problem investigated and hence a framework to our discussion. Moreover, digital watermarking concepts and common techniques are also briefly reviewed and a glossary of the most commonly used notations is given.

1.1 Background

Digital media have become very popular over the last decade, reshaping the habits of modern society. The entertainment industry, where digital data has become omnipresent, has been particularly affected by its fast development. Nowadays, music records, video (e.g. TV recording or movie), art books and many others are retailed as digital data, representing a profitable business. Digital data are not limited to the entertainment industry, they are also used in medical imaging (X-ray), astronomy (satellite images), geography (maps), marketing (advertisement), monitoring (CCTV) and so on.

The development of digital media is due to major revolutions in the industry. Indeed, the progress of electronic and micro-electronic fields has significantly reduced personal computer costs, making them affordable and commonly used in industrialised countries. Furthermore, the development of reliable high capacity mass storage devices and efficient compression algorithms, such as MPEG [ISO/IEC JTC1/SC29/WG11 1988], JPEG [Wallace 1992], or JPEG2000 [JPEG2000 Final committee 2000], has made them very useful for multimedia applications. Finally, the development of the Internet has freed digital data distribution of most physical limitations.

However, this popularity has come with an increasing number of threats. As a consequence of multimedia device improvements, digital data, which can be exactly reproduced at virtually no cost, has become more vulnerable to illicit distribution or retailing. For instance, piracy has caused massive losses to the entertainment industry for the last few years, as published by the International Property Alliance and reported in Tab. 1.1.

Moreover, since digital data are easily forged, their credibility may be significantly depreciated for

Year	Motion Pictures	Records and Music
2002	1357.0	2540.7
2001	1288.0	2034.7
2000	1221.0	1800.3
1999	1268.0	1723.5
1998	1406.5	1634.1
1997	1786.0	1317.3

Table 1.1: Estimated trade losses due to piracy (in millions of U.S. dollars), published in the International Intellectual Property Alliance special 301 annual report, http://www.iipa.com.

sensitive matters. For instance, photos and CCTV records are often used as evidence in court and need to be authenticated. Due to the imperfective nature of their physical media, digital data may become corrupted. Moreover, for some application domains such as medical imaging, these defects can be critical, for example in diseases diagnosis based on X-ray. Thus, at times, digital data need to be proved uncorrupted.

Steganography, or the art of hiding information, has been used for centuries. But it is only recently that it has attracted the interest of the research community [Katzenbeisser and Petitcolas 2000]. In particular, digital watermarking, a branch of this art, has been the focus of many research activities.

By means of an imperceptible "watermark", digital watermarking aims at embedding information in a digital data about the data itself. The embedded information and the digital data in which it is embedded are also referred to as the "watermark information" and the "host data", respectively. Thus, the "watermarked data" is ideally imperceptibly different from the "original data". The digital watermark can be seen as an imperceptible perturbation of the original data and the difference between the watermarked and the original data is termed "watermark distortion".

Hence, digital watermarking techniques are considered useful for intellectual property rights protection, data authentication, data integrity protection, fingerprinting and monitoring [Cox, Miller, and Bloom 2002].

In intellectual property rights protection context, watermarking techniques may be used to prove digital data ownership [Craver 1996; Craver, Memon, Yeo, and Yeung 1998] by embedding the legitimate rights holder details or some sort of identification. This does not prevent the theft itself but provides a tool and a legal mechanism to protect the legitimate rights holder or customer.

Depending on the embedded information, digital watermarking techniques can also be used to identify not only the data rights holder but also its customer, allowing to trace back illicit distribution or retailing sources. The watermark and the watermarking technique, which help identify the customer rather than the data rights holder, are referred to as fingerprint and fingerprinting technique, respectively.

Directly embedded into the host data, the watermark information may also be used to ensure its integrity or authentication [Lin, Podilchuk, and Delp 2000; Martinian 2000]. Any attempt to tamper with a watermarked data will also affect the watermark information to some extent and may enable the detection of such attempt. For instance, a watermark whose information cannot be recovered after a

small perturbation of the watermarked data can be used to detect any distortion. Such watermarking scheme is referred to as a fragile watermarking scheme. Some techniques may even enable to locate the modification.

Another interesting application domain concerns digital data enhancement. In this context, the watermark is used to encode extra-data in the media itself. For instance, for a digitised song recording, the watermark information may consist in lyrics, artist biographies, related song names, or future album announcements.

Motivated by the industry emerging needs, digital watermarking has been the focus of many studies in the research community [Anderson and Petitcolas 1999] over the past few years. At first, watermarking techniques relied on intuitive principles or on methods from other fields. With the increasing number of publications on watermarking schemes, the need for a fair and standard framework to evaluate the proposed techniques has become necessary, but has not been fully answered yet.

Some efforts have been made towards this direction, resulting in various benchmarking tools [Petitcolas, Anderson, and Kuhn 1998; Petitcolas 2000; Petitcolas 2002; Voloshynovskiy, Pereira, Iquise, and Pun 2001; Pereira, Voloshynovskiy, Madueño, Marchand-Maillet, and Pun 2001]. However, if standard testing tools are now available, they do not put forward any evaluation framework to compare techniques, and they are merely a collection of attacks.

Until recently, the watermarking research field was missing the theoretical background and a consistent framework for making the switch from art to science. In the past few years, theoretical analysis [Moulin and O'Sullivan 2003; Somekh-Baruch and Merhav 2003; Cohen and Lapidoth 2002] have appeared, providing pieces of a general watermarking theory.

Based on theoretical results, more principled techniques [Chen 2000; Chen and Wornell 2001; Eggers, Bauml, Tzschoppe, and Girod 2003; Ramkumar 2000] were devised and investigated. They showed promising results relating watermarking theory to watermarking techniques. However, these recent studies have two common shortcomings. First, the analyses are often based on continuous values and it is hardly the case for digital data. Second, the typical statistical models chosen are often inadequate, sometimes resulting in misleading conclusions as shown in [Bounkong, Saad, and Lowe 2002b].

Moreover, research devoted to digital watermarking has seen a significant broadening of its range of investigation. Watermarking schemes can be improved in many ways. First theoretical studies aiming at the channel capacity [Moulin 2001] for different noise models may provide a better understanding of the communication limits. In the past few years, several studies have proposed an information-theoretic framework to the watermarking problem and investigated the capacity channel notably for Gaussian [Cohen and Lapidoth 2002; Karakos 2002] and quantisation [Bounkong, Saad, and Lowe 2002b; Eggers and Girod 2001] noise models.

Research in the design of practical schemes has also been carried out. Various directions have been investigated comprising the human perceptual system modelling, the host data representation or data embedding techniques. For instance, an appropriate perceptual model [Bartolini, Barni, Cappellini,

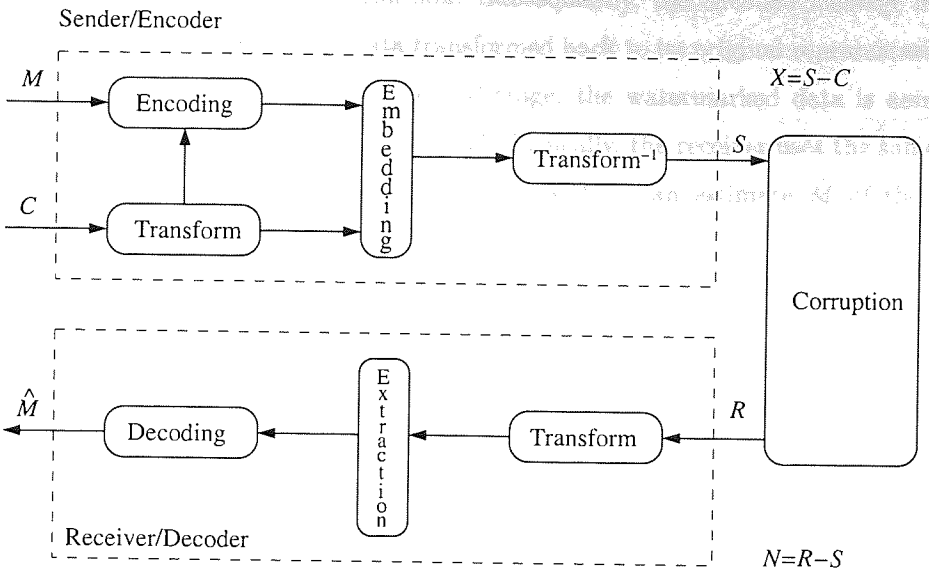


Figure 1.1: Watermarking framework block diagram.

and Piva 1998; Kutter and Winkler 2002] may be used to modulate the embedding process according to the impact on the watermark perceptibility. A suitable data representation may enable the scheme to be more robust against some modifications or gain some desirable properties [Malvar and Florêncio 2003]. In the same way, more robust embedding [Bounkong, Toch, Saad, and Lowe 2003a; Bounkong, Toch, Saad, and Lowe 2003b] and decoding [Barni, Bartolini, Cappellini, Piva, and Rigacci 1998] techniques may improve the retrieval of the watermark information. Intensive research has also been devoted to provide standard benchmarking packages [Petitcolas 2002] and standard testing procedures.

1.2 Scope and Outline of the Thesis

Digital watermarking is a very broad research topic which is related to various types of data such as image [Langelaar, Setyawan, and Lagendijk 2000], music [Toch, Lowe, and Saad 2003] or video [Wolfgang, Podilchuk, and Delp 1999]. From a theoretical point of view, the problem is fairly similar. However, in practice, many aspects may tear them apart: presence or absence of the time or spatial dimension, perceptual distortion constraints, or data size are a few examples among others.

Furthermore, in the past few years, digital watermarking study has significantly broaden its range from pragmatic watermarking schemes [Meerwald 2001a], secure communication protocol [Cox, Miller, and Bloom 2002], scheme performance evaluation [Petitcolas 2000] to channel capacity theoretical investigation [Karakos 2002]. Although they are all related, one may argue that they are also different areas of research.

In this thesis, our study will focus on techniques used for digital image watermarking, which can be described by the block diagram depicted in Fig. 1.1. The host data C is first represented in some transform domain, then the watermark information or message M is encoded using eventually a secret

key, the host data or any other side information. Subsequently, the encoded message is embedded into the host data and the modified host data transformed back to its original representation domain, resulting in the watermarked data S. In a second stage, the watermarked data is assumed to be corrupted during the transmission and received as R. Typically, the receiver uses the same transform as the sender before extraction and decoding, which results in an estimate $\hat{M}$ of the watermark information M.

In the present work, three different aspects of the framework described in Fig. 1.1 will be dealt with. They can be listed as follows:

Quantisation effects on digital watermarking – quantisation and digital data are closely related. Indeed, quantisation enables continuous data to be represented as digital data, but at the expenses of some loss of information. Similarly, quantised transmission reduces the amount of watermark information, or requires the distortion introduced by the watermark to be increased to maintain the amount of transmitted watermark information.

Information embedding techniques – they are arguably the central part of most watermarking schemes. The correct retrieval of the watermark information depends significantly on their performance.

Image data representations – they are the main tool to reduce the perceptibility of the distortion introduced by the watermark or to improve the recoverability of the watermark information.

Then, the thesis is organised as follows:

Chapter 1 is this introductory chapter.

Chapter 2 studies the relationship between clipping, quantisation and digital watermarking. In particular, two analyses of quantisation effects on digital watermarking are considered. The first one, which is generally adopted in the literature, consists of a stochastic modelling of the quantisation process, while the second one puts forward a deterministic modelling for the same process. Both aim at estimating the minimum watermark distortion required to transmit a given watermark information. These approaches are discussed and carried out on practical cases.

Chapter 3 is devoted to information embedding techniques. The main techniques used in the field are first reviewed. Then, an algorithm which derives the optimal embedding function for a given host, noise and decoding model is proposed. The characteristics of the optimal embedding function are discussed for specific noise and decoding models, and compared to techniques proposed in the literature.

Chapter 4 investigates image data representation effects on the perceptibility of the watermark. The most commonly used representations are first reviewed. Subsequently, a new and principled representation based on independent component analysis is proposed for digital image watermarking purposes. Then, the influence of image data representation on watermarking scheme

robustness is briefly discussed. Finally, two visual perceptual distortion measures are presented and compared with two image data representations based on two different block transforms.

Chapter 5 concludes this thesis. A summary of the achievements and contributions to the state of the art is presented. Further directions of studies are presented and possible solutions are outlined.

1.3 The Watermarking Problem

In this section, definitions of the digital image watermarking problem and of its related concepts are given.

1.3.1 Definitions

Definition 1 *In this thesis, the term "image" or "digital image" refers to a digital greyscale image. An image is a two dimensional matrix of size $h \times w$ which corresponds respectively to the height and width of the image in pixels. Such matrix is composed of integers which encode the luminosity or intensity of the pixel from 0 for black to 255 for white.*

Digital colour images have a similar structure but are composed of three colour planes. Thus, each pixel information is spread on three components. Greyscale image watermarking may be related to the watermarking of the luminance component of an image.

In this work, the cover image in which a message is embedded is termed "host data" or "cover image", and it is denoted by the $h \times w$ matrix C. An individual element or pixel of the image is referred to by the notation c_{ij}, where the matrix indices $i \in [1, h]$ and $j \in [1, w]$ correspond to the spatial location of the pixel. For instance, $c_{1,1}$ is the top left corner pixel and $c_{h,w}$ is the bottom right corner pixel. Finally, let us define $\mathscr{C}$, the alphabet of C, as the set of all possible images that can be represented by the matrix C. The cardinal of $\mathscr{C}$ is $8hw$.

Definition 2 *In this work, the watermark information or message which is embedded in the watermarking process is denoted by M. The latter is an integer comprised between 0 and $2^l - 1$ where l is a strictly positive integer.*

Also, M can be expressed as a sequence of l bits corresponding to its binary representation. In that case, each bit of the sequence is referred by m_i where i denotes the position of the bit in the sequence. For instance, if $l = 4$ and $M = 4$, then M can be expressed in binary representation as $M = 0100$ with $m_1 = 0$, $m_2 = 1$, $m_3 = 0$ and $m_4 = 0$.

Furthermore, as in the majority of the studies devoted to digital watermarking, it will be assumed that M is uniformly distributed over the 2^l possible messages. The binary entropy of M is therefore given by l, in bits.

Definition 3 *A digital watermarking scheme is a set of two processes: an encoding/embedding and an extracting/decoding process. Both are parts of a general communication framework in which the sender aims at transmitting the watermark information to the receiver under some constraints.*

Both encoding/embedding Q and extracting/decoding W processes can be either stochastic or deterministic, although deterministic processes are predominant in practice. In the context of digital watermarking, the channel [Cover and Thomas 1991][1] consists of the input/output alphabet $\mathscr{C}$ representing all possible digital images C and the probability $\mathrm{A}(R|S)$ of observing the received image $R \in \mathscr{C}$ given the watermarked image $S \in \mathscr{C}$. Such channel will be denoted ($\mathscr{C}$,A).

Thus, the watermarking process may, under some constraints, modify the original image C into a new image S in order to embed the watermark information M. At the extracting stage, the receiver is given the corrupted image R from which the message M has to be retrieved. The decoded message is denoted $\hat{M}$ and each bit of its binary expression is given by $\hat{m}_i$. Besides, the watermarked image S can be considered and referred to as the "codeword" of M conditioned by the host image C.

Depending on the availability of the original data C to the receiver, the scheme is said to be blind, oblivious or public if C is not available. It is non-blind, non-oblivious or private if C is available. Finally, it is semi-blind or semi-public if only some references about C are available but not C itself [Cox, Miller, and Bloom 2002].

Definition 4 *The watermark embedded in a host data by a watermarking scheme is defined as the pixelwise difference between the watermarked and the original host image.*

Thus, the embedded watermark denoted X is given by

$$x_{ij} = s_{ij} - c_{ij} \,,$$

with $i \in [1, h]$ and $j \in [1, w]$. With this definition, the Euclidean distance between the original and watermarked image is also the 2-norm of X. Remark that this definition does not imply an additive watermarking scheme.

Definition 5 *The corruption noise which affects the watermarked data during transmission is defined as the pixelwise difference between the received and the watermarked image.*

Thus, the corruption noise denoted N is given by

$$n_{ij} = r_{ij} - s_{ij} \,,$$

with $i \in [1, h]$ and $j \in [1, w]$. As previously, the Euclidean distance between the watermarked and the corrupted data is equal to the 2-norm of N and such definition does not restrict the study to additive noise.

[1] p.184

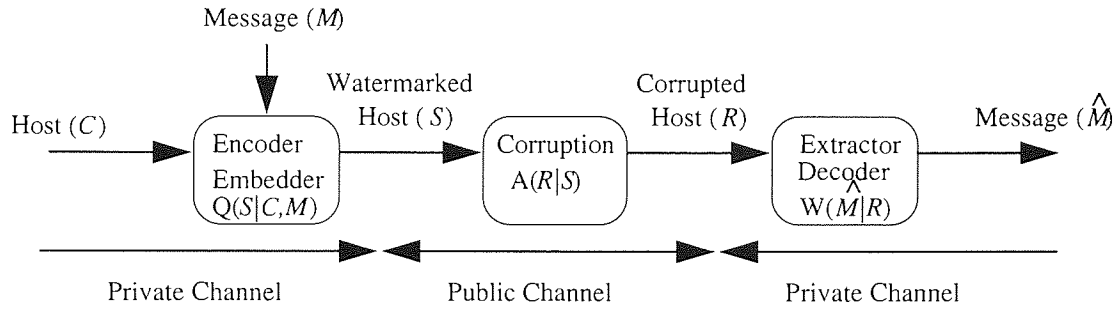


Figure 1.2: A general watermarking framework.

Definition 6 *The digital watermarking framework depicted in Fig. 1.2 can be described as a communication problem comprising a sender, a receiver and an attacker. In this problem, the sender communicates a message M through a public channel ($\mathscr{C}$,A) such that the presence of M cannot be perceived by an observer.*

As mentioned earlier, the watermarking scheme is an essential part of the framework as it provides the practical tools for a successful communication: the encoding $\mathrm{Q}(S|C,M)$ of the message M and the host data C into the watermarked data S and the decoding $\mathrm{W}(\hat{M}|R)$ from the received data R of the embedded message M or of related information. Thus, a watermarking scheme can be defined by the quadruplet ($\mathscr{C}$,Q,A,W).

Since the transmission over a communication channel may be noisy, the received data R and the sent data S are usually assumed to be different. The discrepancy between the sent and received data can be attributed to various sources such as random noise related to the physical transmission, quantisation noise due to lossy compression, or signal processing noise.

From this problem description follows further concepts such as the information rate, bit decoding error probability, bit good or correct decoding probability, bit error rate or channel capacity that we shall define below.

Definition 7 *The information rate of a watermarking scheme is defined as the ratio Ir between the binary length l of the message M and the number of bits ($8hw$) used to represent the host data C, and is given by*

$$Ir = \frac{l}{8\,h\,w}\,.$$

Definition 8 *The bit decoding error probability of a watermarking scheme is defined as the probability of having $\hat{m}_i$ different from m_i given R. The bit good decoding probability of a watermarking scheme is defined as the probability of having $\hat{m}_i$ equal to m_i given R. The bit decoding error probability and good decoding probability of a watermarking scheme are respectively denoted*

$$\mathrm{P}(\hat{m}_i \neq m_i|R) \qquad \text{and} \qquad \mathrm{P}(\hat{m}_i = m_i|R)\,.$$

Note that

$$\mathrm{P}(\hat{m}_i = m_i|R) = 1 - \mathrm{P}(\hat{m}_i \neq m_i|R)\,.$$

In some case, the decoding error probability cannot be derived and is hence approximated by the bit error rate defined as the average function of bit differences between M and $\hat{M}$, both expressed in binary and given by

$$Ber = \frac{1}{l}\sum_{i=1}^{l} m_i \oplus \hat{m}_i\,,$$

where $\oplus$ is the "exclusive or" or "xor" operator.

An alternative set of probabilities, composed of the detection and the false alarm probability, may be used to characterise watermarking scheme performance. Both can be related to the decoding error probability. For instance, the watermark detection may be characterised by a good decoding probability higher than a certain threshold, while the false alarm probability corresponds to the probability to detect the watermark in any data. In this work, we focus on watermarking scheme decoding error probability and good decoding probability.

Definition 9 *The watermarking channel capacity can be defined as the maximum mutual information between the message M and the received data R, where the maximum is taken over all possible joint distribution $\mathrm{P}(s, m|c)$. Since the message M is assumed uniformly distributed, the maximum is taken only over $\mathrm{P}(s|c, m)$.*

Thus, the watermarking channel capacity denoted $\overline{C}$ is given by

$$\overline{C} = \max_{\mathrm{P}(s|c,m)} \mathrm{I}(M, R)$$

where $\mathrm{I}(.,.)$ is the mutual information between two random variables. Due to the complexity of typical corruption processes, an analytical expression of $\overline{C}$ cannot usually be derived, even for simple host and noise models. In most practical scenarios, as $\overline{C}$ is unknown and its available bounds are quite loose, the decoding error probability and the information rate are important indications of the watermarking scheme performance.

Definition 10 *A watermark is said perceptible if one can tell the difference between the watermarked data and the original data in normal conditions of visualisation.*

The perceptibility is a measure based on a human physiological function, here in particular the human visual system (HVS) which is difficult to deal with as it is a highly subjective measure. Many researches [Nadenau, Winkler, Alleysson, and Kunt 2000; Avcibas, Sankur, and Sayood 2002; Taylor 1998] have focused on modelling the HVS and devising objective replacement for the human judgement. Although significant progresses have been achieved, no real consensus exists yet.

In some other frameworks, the watermark has to be undetectable, which is a characteristic more difficult to achieve [Cachin 1998] but also more objective, as it does not involve human judgement

directly. However, in this work and for most watermarking applications, indetectability is not required [Cox, Miller, and Bloom 2002] and most of the watermarking schemes are designed to be solely imperceptible.

1.3.2 Discussion

As most processes, watermarking schemes need to be evaluated with respect to some performance measures. From the previous description, at least three can be identified: the scheme information rate, its decoding error probability and the watermark perceptibility. The information rate and the decoding error probability can be directly related to the quantity and quality of information transmitted by the scheme over the channel, while the perceptibility appears as a constraint on the encoding scheme.

Intuitively, all three seem related and contradictory in the constraints imposed. For instance, embedding more information may introduce more distortion and presumably increase the perceptibility of the watermark if the same decoding error probability has to be preserved. In the same way, decreasing the decoding error probability will require either to reduce the information rate or increase the visibility of the watermark.

The relationship between these three criteria may vary depending on the nature of the corruption process and the watermarking technique used. As searching for global rules appears hopeless, studies are usually limited to a specific scheme and noise model [Eggers and Girod 1999; Voloshynovskiy, Herrigel, Iquise, and Pun 2001; Wong and Au 2003].

Note that perceptibility is not a stable measure as it may depend on the individual or on the perceptibility model used. For instance, human evaluation is time consuming, costly and can be influenced. Furthermore, since no consensus has been found yet on a model for the human visual system, a measure composed of several models could be more desirable.

Although the information rate and the decoding error probability are well defined, their evaluation for watermarking schemes may also be a difficult task. Indeed, the information rate has to be chosen beforehand and the decoding error probability depends strongly on the channel characteristics. However, taking into consideration the use of error correcting code may help modulating the relation between the decoding error probability and the information rate.

Now that the criteria to evaluate watermarking scheme performance has been established, some additional constraints should be described explicitly. The watermarking scheme and the noise process need to be bounded for any evaluation to be relevant. For instance, if the noise is not limited, nothing can be transmitted reliably, while an unlimited strong watermark will be completely robust.

As the information transmission needs to be imperceptible, it is tempting to use an imperceptibility measure with two pre-defined thresholds to constraint both watermarking scheme and noise process. However, as it has been pointed out previously, there is neither a principled agreed metric yet, nor a consensus about a good practical model. Furthermore, using a measure based on a model of the human visual system typically prevents further mathematical analysis.

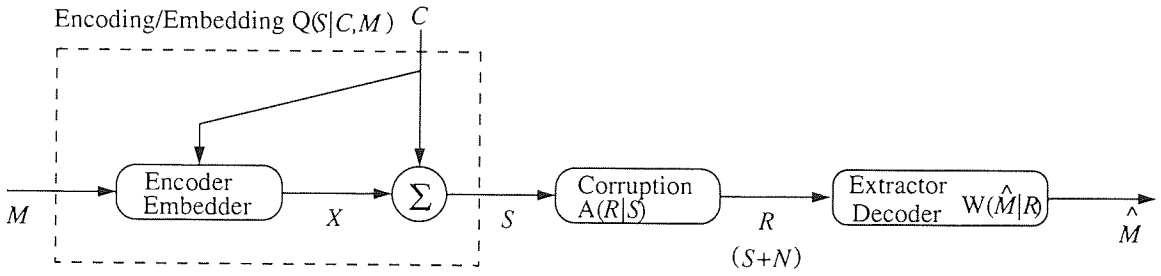


Figure 1.3: Watermarking problem as a power constrained communication problem.

Alternatively, a mathematically simpler constraint based on the Euclidean distance is generally used in theoretical studies. In this work, we choose to adopt a similar set of constraints based on the mean squared error (MSE) measure as it appears as the most principled approach. Thus, if σ_X^2 and σ_N^2 are the two thresholds related to the watermark X (Def. 4) and noise N (Def. 5) mean squared error, respectively, the problem constraints can be defined as

$$\frac{1}{hw}\sum_{ij} x_{ij}^2 \leq \sigma_X^2 \qquad \text{and} \qquad \frac{1}{hw}\sum_{ij} n_{ij}^2 \leq \sigma_N^2 \,.$$

In this thesis, these constraints are referred to as the watermark power constraint and noise power constraints.

Note that the Euclidean distance is not a good perceptibility measure. As shown in several studies [Girod 1993], it is quite poorly correlated to human visual system perception. Thus, more complicated models may be required for the evaluation of watermarking scheme performance.

1.3.3 Problem Formulation

In this section, the watermarking problem is defined as a power constrained communication problem.

Definition 11 *The digital watermarking framework depicted in Fig. 1.3 can be described as a power constrained communication problem, comprising an encoding/embedding process Q, a corruption process A and an extracting/decoding process W. In this problem, the sender communicates a message M to the receiver through a public channel $(\mathscr{C},A)$ such that the mean squared error introduced to the host image C by the watermark X resulting in the watermarked data S is upper bounded by a given constant*

$$\frac{1}{hw}\sum_{ij} x_{ij}^2 \leq \sigma_X^2 \,. \tag{1.1}$$

During the transmission, the watermarked image S may be changed by some power constrained corruption process A, such that the average squared distortion of the corruption noise is upper bounded by another constant

$$\frac{1}{hw}\sum_{ij} n_{ij}^2 \leq \sigma_N^2 \,. \tag{1.2}$$

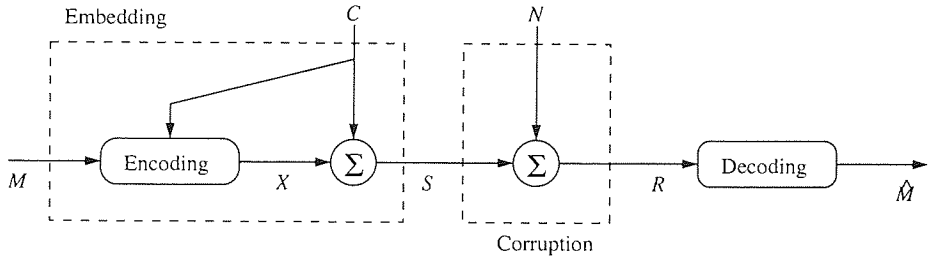


Figure 1.4: Costa's communication model.

In this context, the sender would like to maximise the information rate, while minimising both decoding error probability and perceptibility of the transmission. Thus, this problem can be summarised by the sextuplet $(\mathscr{C}, Q, A, W, \sigma_X^2, \sigma_N^2)$.

1.3.4 Related Problem

In [Costa 1983], a similar but not completely identical communication problem has been investigated. In Costa's framework depicted in Fig. 1.4, the host data C is continuous, following a normal law $C \sim \mathcal{N}(0, \sigma_C^2)$. The corruption noise N is additive, Gaussian and independent of the host data C, $N \sim \mathcal{N}(0, \sigma_N^2)$. The watermark is power constrained as in Eq. 1.1. For such a problem, Costa showed that the channel capacity denoted $\overline{C}_{Costa}$ is given by

$$\overline{C}_{Costa} = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_X^2}{\sigma_N^2} \right) \text{, in bits,} \tag{1.3}$$

where the capacity $\overline{C}_{Costa}$ is independent of the host data C and is similar to the capacity of the corresponding Gaussian channel. An optimal embedding scheme was also proposed, but the required random codebook makes it unfeasible for practical applications.

1.4 Watermarking Concepts Review

In this section, a brief review of techniques and concepts related to digital watermarking is carried out comprising:

- intellectual property rights protection, fingerprinting, integrity protection, authentication and data enhancement,
- data representation and domain transform,
- spread spectrum technique,
- global and spatially localised embedding,
- human visual system modelling and watermarking,
- encoding, embedding and decoding techniques.

1.4.1 Digital Watermarking and Applications

Among the application domains of digital watermarking, intellectual property rights protection [Craver, Memon, Yeo, and Yeung 1998; Langelaar, Setyawan, and Lagendijk 2000] has arguably attracted the most attention due to the tremendous commercial interest involved. To meet the expectations of the industrial community, watermarking schemes devoted to intellectual property rights protection must fulfil some requirements which consist of a very low decoding error probability or very high probability of detection, a very low probability of "false alarm[2]", a high level of imperceptibility and a reasonable information rate.

Fingerprinting [Karakos and Papamarcou 2000; Langelaar, Setyawan, and Lagendijk 2000] is another application which aims at identifying the receiver rather than the sender. One of its potential purpose lies in "traitor tracking", the goal being to find out the original source of the illicit redistribution of data. Schemes of this type normally need to have a very low decoding error probability or very high probability of detection, and a high information rate. Depending on the action taken when the source of illicit distribution is found, the probability of false alarm may be required to range between reasonably low and very low.

Data integrity and authentication [Xie and Arce 1998; Langelaar, Setyawan, and Lagendijk 2000] are two other promising domains of application for digital watermarking. These domains are currently dealt with by hash function and public encryption. As the watermark is embedded in the data itself, it is not lost when the data are printed out. Furthermore, watermarking for authentication and integrity protection should be seen as a complementary tool to existing methods and not as a replacement. Indeed, the encrypted digital signature produced by a hash function could be embedded as a watermark. In this context, watermarking schemes need to have a high decoding error probability[3], a very low perceptibility and a reasonable information rate level.

A last possible application is related to data enhancement. In this case, the watermark aims at making the data more informative and attractive by providing further information related to the data itself. This could comprise the author biography, comments, analysis and references. For such applications, the watermarking schemes is required to have a very high information rate and reasonable levels of decoding error probability and imperceptibility.

1.4.2 Data Representation and Domain Transform

For many image processing applications, a good data representation is critical [Jain 1989; Jung, Miltra, and Mukherjee 1996]. For instance, lossy compression standards have focused on the discrete wavelet transform (DWT) and the block discrete cosine transform (DCT), while image filters process the data in the discrete Fourier transform (DFT) or in the spatial domain.

Digital image watermarking techniques do not evade such practice. Indeed, some robustness

[2]a false alarm is raised when a watermark is detected in a received data which was not watermarked in the first place.

[3]Note that if the data is tampered with, the hash will not correspond to the data anymore.

(lower decoding error probability) and imperceptibility property are expected to be easier to achieve in a transform domain. However, all these principles and beliefs are merely commonly accepted assumptions or are only based on numerical simulations. In Chapter 4, image data representation influence on digital watermarking scheme performance is investigated.

The two most commonly used domain transforms are the DCT and the DWT. Interest in these domains have been motivated mostly by the JPEG [Wallace 1992] and the JPEG2000 [JPEG2000 Final committee 2000] lossy compression standards which are based on these transforms. Among several incentives to use them are their very good energy[4] compaction properties [Jain 1989] and a good decorrelation of the obtained coefficients [Jain 1989].

Other domain transforms have also been investigated such as the DFT which provides not only a good energy compaction and data decorrelation [Jain 1989], but also provides some invariance property to translation [Pereira and Pun 1999] and even rotation and scaling when used in conjunction with another Mellin Fourier [Ruanaidh and Pun 1997] transform. Independent component analysis (ICA) has also been the object of some studies [Bounkong, Saad, and Lowe 2002a; González-Serrano, Molina-Bulla, and Murillo-Fuentes 2001]. Indeed, ICA provides a representation comprising uncorrelated coefficients that are also as independent as possible [Hyvarinen, Karhunen, and Oja 2001]. Such property is usually considered desirable in information theory [Moulin and O'Sullivan 2003]. In addition, analyses are more easily carried out in this case which may help in optimising the watermarking scheme.

1.4.3 Spread Spectrum Technique

Spread spectrum (SS) communication aims at transmitting a narrowband signal over a much larger bandwidth such that the signal energy, in terms of the variance present in any single frequency, is undetectable. Such a method is commonly used in military communication in order to gain advantages in interference rejection (anti-jam), message privacy/security, or low probability of interception.

Such a technique has been popularised in the digital watermarking research field by Cox's seminal paper [Cox, Kilian, Leighton, and Shamoon 1997]. In the proposed scheme, the embedding is achieved by a random Gaussian noise sequence, representing the message information, which is used to modulate some frequencies of the host data signal. Note that the message as defined in Def. 2 can be used as a seed for the random Gaussian noise generator.

Spread spectrum watermarking is strongly related to frequency representation of the data but not to a particular decomposition, also it can be applied to the DCT, DFT or any other frequency decomposition method. Further references on SS based watermarking scheme can be found in [Kutter and Winkler 2002; Malvar and Florêncio 2003; Fridrich 1998; F.H. Hartung 1999].

[4] in the sense of the variance.

1.4.4 Global and Spatially Localised Embedding

The embedding process can be viewed from two opposed angles, as a global[5] process or as a local and spatially localised process. Both approaches have their benefits and drawbacks.

Global embedding techniques scatter each information bit of the message over the whole[6] data. This makes the embedding more robust, particularly against spatially localised corruptions. However, such method cannot adapt to local information of the host data. Due to the imperceptibility requirement, such embedding may not be optimal [Bartolini, Barni, Cappellini, and Piva 1998].

Localised embedding techniques on the contrary can take advantage of local data information and target particular regions [Lu and Liao 2000; P.-C. Su and Kuo 1999]. However, they are more sensitive to local distortions and may lead to the loss of some information bits, since they are not encoded throughout the whole image. One way to take advantage of the local nature of the embedding is to use the human visual system characteristics as presented in the next section.

1.4.5 Human Visual System Modelling and Watermarking

It is well known that the human visual system sensitivity is not consistent and depends on the distance of vision, lighting, or background colour [Cornsweet 1970]. Also, the contrast [Legge and Foley 1980; Peli 1990], luminosity, and therefore perceptibility [Cornsweet 1970] of an object may not be perceived in the same way depending on the context. Furthermore, high frequency pattern looks "messier" than low frequency pattern and changes made to them will be more easily noticeable in the latter than in the former. The human visual system is also more sensitive to stimulus in the front rather than on the side [Cornsweet 1970].

Commonly accepted assumption, confirmed by physiological experiments, is that the human visual system is highly inefficient in the sense that it cannot generally distinguish between the $2^{8h\,w}$ different images that can be encoded in a 8 bits $h \times w$ matrix. Actually, in normal conditions of observation, two images have to be quite different to be detected as such. In order to achieve high compression rates, lossy compression algorithms can afford to discard a significantly large amount of information without degrading too much the perceptual quality of the data.

Many practical watermarking schemes use the human visual system characteristics or defects to embed more robust or more informative watermark. Taking advantage of research carried out in other fields, imported results, such as quantisation tables derived for lossy compression, were exploited in practical watermarking schemes.

In the past few years, many of these human visual system based watermarking techniques [Wolfgang, Podilchuk, and Delp 1999; Delaigle, Vleeschouwer, and Macq 1998; Podilchuck and Zeng 1998; Bartolini, Barni, Cappellini, and Piva 1998; Kutter and Winkler 2002] were suggested, claiming improved results. However, considering the lack of consensus on what is a fair evaluation on these, the actual achievements are difficult to measure.

[5] in the sense of the host data.
[6] in the spatial sense.

1.4.6 Information Embedding Techniques and Watermarking

In the context of information embedding, a plethora of techniques exist but most of them are based on the two same principles: modulation and quantisation. A detailed analysis and discussion of both systems is carried out in Chapter 3.

In the context of watermarking, many practical schemes are based on these two principles. In the beginning, they were put forward as heuristic methods, and the investigation of their intrinsic properties were often neglected. However, with the increase of interest in watermarking research, more principled information embedding and decoding techniques have been devised and used in modern schemes.

1.5 Notations Summary

In this work, the following convention is adopted: functions are denoted with normal upper case letter such as "H" for the entropy of a random variable, matrices or vectors variables are denoted with italic upper cases letters such as "C" for the host data, while scalar variable are denoted by italic lower case letters such as "l" for the length of the message. For vectors and matrices elements the following conventions apply : if A is a matrix or a vector, then

a – is the generic term of A,

a_{ij} – is the element of the matrix A at the ith row and at the jth column,

a_k – is the kth element of the vector A or the a_k element of a sequence of elements $\{a_{ij}\}$ of A,

$a_{.j}$ **or** $a_{i.}$ – are respectively, the jth column and the ith row of the matrix A.

1.5.1 Variables and Functions

C – refers to the digital watermarking problem host data, which is a digital image of size $h \times w$ pixels as defined in Def. 1. The notation c_{ij} normally refers to an element in the spatial domain of the data, but may also be used for transform domain if there is no ambiguity; C is a matrix of size $h \times w$ pixels.

M – represents the message to be embedded in C as defined in Def. 2. Each element m_i of the message corresponds to the ith bit of M binary representation. The message length is of l bits.

X – is the watermark as defined in Def. 4. For the sake of simplicity, the same notation X may be used to denote the watermark in a transform domain when there is no ambiguity.

N – is the corruption noise of the watermarking problem and is defined in Def. 5. As previously, the notation N may be used to represent the noise in a transform domain.

S – is the watermarked host data.

R – is the received host data.

$\hat{M}$ – is the estimate of the message M given the received data R.

P – is the probability function of a random variable X and $\mathrm{P}(x)$ is the probability of a particular realisation x of the random variable X.

Ir – is the information rate in bit(s) of a watermarking scheme as defined in Def. 7.

Ber – is the bit error rate ($0 \leq Ber \leq 1$) of the watermarking scheme subject to a corruption noise as defined in Def. 8.

G – is the Gaussian function of the variable x, of mean μ and variance σ^2 given by

$$\mathrm{G}(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right).$$

H – is the entropy function of a random variable. For a continuous variable X, $\mathrm{H}(X)$ can be expressed as

$$\mathrm{H}(X) = -\int_x \mathrm{P}(x) \log(\mathrm{P}(x)) \mathrm{d}x\,;$$

for a discrete variable X, $\mathrm{H}(X)$ is given by

$$\mathrm{H}(X) = \sum_x \mathrm{P}(x) \log(\mathrm{P}(x))\,.$$

I – is the mutual information between two random variables. In the continuous case it is given by

$$\mathrm{I}(X, Y) = -\int_{x,y} \mathrm{P}(x, y) \log\left(\frac{\mathrm{P}(x, y)}{\mathrm{P}(x)\mathrm{P}(y)} \right) \mathrm{d}x\mathrm{d}y\,;$$

in the discrete case, we have

$$\mathrm{I}(X, Y) = -\sum_{x,y} \mathrm{P}(x, y) \log\left(\frac{\mathrm{P}(x, y)}{\mathrm{P}(x)\mathrm{P}(y)} \right).$$

$\overline{C}$ – is the watermarking channel capacity as given in Def. 9.

1.5.2 Acronyms

BER – bit error rate,

DC-QIM – distortion compensated QIM,

DCT – discrete cosine transform,

DEP – decoding error probability,

DES – discrete embedding scheme,

DFT – discrete Fourier transform,

DWT – discrete wavelet transform,

GDP – good decoding probability,

HVS – human visual system,

ICA – independent component analysis,

IR – information rate,

MSSIM – mean SSIM,

PCA – principal component analysis,

SCS – scalar Costa scheme,

SSIM – structural similarity index measure,

SS – spread spectrum,

QIM – quantisation index modulation.

Chapter 2

Digital Image, Quantisation and Watermarking

In this chapter, the relation between clipping, quantisation and digital watermarking is studied. In particular, two analyses of quantisation effects on digital image watermarking are considered. The first one, which is generally adopted in the literature, consists of a stochastic modelling of the quantisation process, while the second one puts forward a deterministic modelling for the same process. Both aim at estimating the minimum watermark mean squared error distortion required to transmit a given message. These approaches are demonstrated for practical cases and discussed.

2.1 Digital Images

A digital image C represented by a $h \times w$ matrix of 8 bits integers can take up to 2^{8hw} states. In contrast to a "real" image, such as a photograph or any analog two dimensional static visual object, a digital image is limited in resolution and tone.

Having discrete representations, digital images may be represented as isolated points among real images, which can be depicted as a two dimensional lattice in Fig. 2.1. Thus, if the watermark X is power constrained as $\|X\|^2 \leq d^2$, there are actually very few images to which the original image C can be modified to. The digital watermarking problem is about how to make use of these "available" or "accessible" images to convey some information to a receiver.

With such watermark power constraint $\|X\|^2 \leq d^2$, if n denotes the number of images at a Euclidean distance lower than d, it follows by definition that the information rate is upper bounded

$$Ir \leq \frac{n}{8hw}.$$

Moreover, as shown later in this chapter, n depends not only on the watermark power constraint threshold d^2 but also on the original image C.

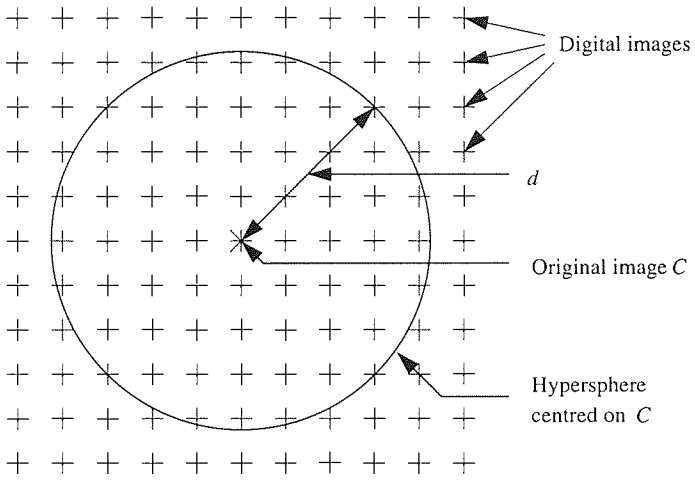


Figure 2.1: Two dimensional representation of the digital image space. In this figure, each "+" represents a digital image. Two neighbouring "+" differ only from a level of pixel intensity value. This figure illustrates that from a given digital image C, only a reduced number of digital images are at a Euclidean distance lower than d.

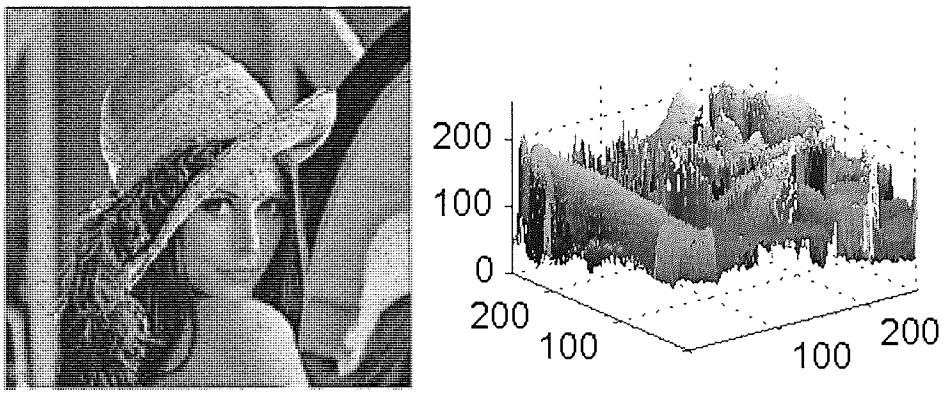


Figure 2.2: The Lena image of size 256×256 pixels and its relief map.

2.1.1 Clipping, Quantisation and Digital Images

Most digital processing techniques result in images, which cannot be represented within the discrete space of digital images $\mathscr{C}$, hence, they are followed by clipping or quantisation, which are used to find the closest digital representation. As a matter of fact, the majority of the watermarking practical schemes fall into this category. Thus, even for a noiseless transmission, the received data R is rarely equal to the watermarked data S.

Clipping is the process of cutting down to the desired size or shape. This pixelwise process affects each pixel whose value is greater than 255 or smaller than 0, by setting them to 255 and 0, respectively. Thus, an image can be seen as a relief map which height is bounded by 0 and 255 as in Fig. 2.2.

Image processing techniques are more likely to be followed by clipping when the pixel values of the original image C are close to 0 or 255. Since the histogram of an image pixel intensity values is usually spread over the whole range of possible values as in Fig. 2.3, some pixel values of a processed

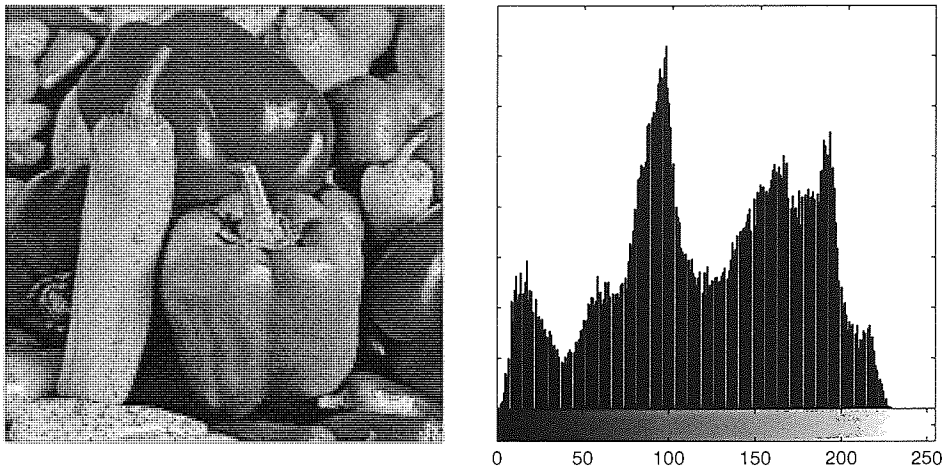


Figure 2.3: The Peppers image and its histogram.

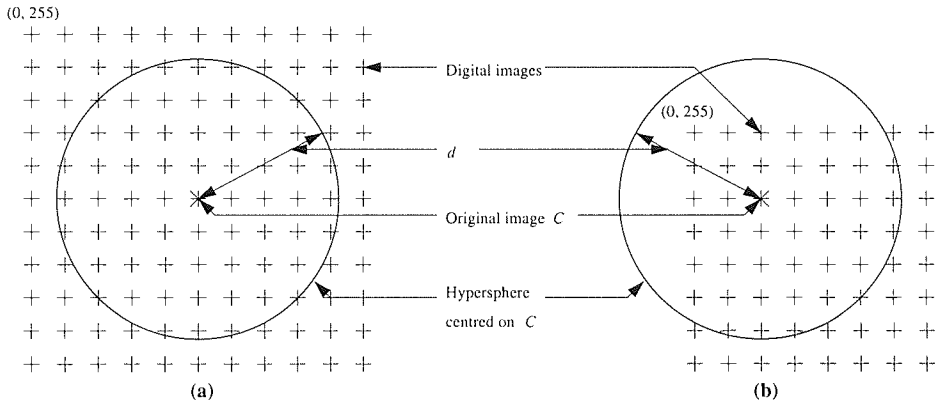


Figure 2.4: Influence of the digital data values on the number of neighbouring data. This figure illustrates the variation of the number of neighbouring digital data for two different data values at a Euclidean distance lower than d. Digital images composed by two pixels are represented by "+" symbols, and the pixel values of the top left "+" are given in brackets.

image have generally to be clipped to 0 or 255.

Thus, the original pixel values of the image C are relevant to the information rate of digital watermarking schemes. Indeed, depending on them, the number of accessible digital images n at a Euclidean distance lower than d from the original host image C may change considerably. For instance, let us consider a digital image composed of two pixels represented by integers between 0 and 255 and depicted as a two dimensional lattice in Fig. 2.4.

In Fig. 2.4a, 61 images are at a Euclidean distance lower than d of the original image C represented by a "*", while we have only 42 in Fig. 2.4b. Clearly, more information can be transmitted in the case (a) or a more robust embedding scheme can be devised.

In order to demonstrate how clipping can affect the performance of a watermarking scheme, let us consider the following scenario. The watermarking technique studied embeds a binary message M using a quantisation based method. Along a pre-defined axis Δ, the original image C is mapped to the closest watermarked image S, also termed codeword, which encodes the message M. Moreover,

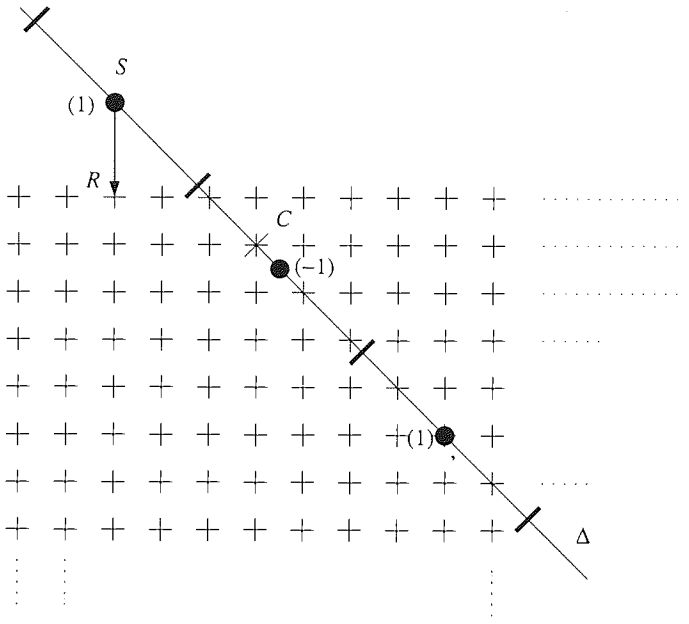


Figure 2.5: An example of clipping effects on a practical watermarking scheme.

the watermarking scheme is oblivious of the digital limitation of the image representation. Finally, the decoding is carried out "to the nearest" in the sense of the Euclidean distance. This implies that the message encoded in the closest codeword is the one which is decoded.

In Fig. 2.5, a two dimensional representation of the problem is depicted. Digital images are represented by "+", while the original digital image C is represented by "*". All possible codewords are symbolised by black dots, with, in brackets, the message they encode: -1 or 1. Black dashes delimit the message decoding areas.

Now, let us assume that the message to encode is $M = 1$, thus the original image C will be mapped to the watermarked image S. Subsequently, since S is not a representable digital image, it is clipped to the closest representable image R. Although the decoding to the nearest will result in the correct message $M = 1$, it can be observed that the other codeword encoding $M = 1$, on the bottom right of the figure, would be preferable to R since it has a lower Euclidean distance to the closest codeword encoding $M = 1$. Therefore, it will be more "robust" to further attack than R.

Clipping effects on watermarking techniques will greatly differ from an image to another, since clipping depends on the pixel values of the original image. In Fig. 2.6, the mean squared error distortion due to clipping is measured with respect to the mean squared error introduced by the watermark for the watermarking scheme described in [Cox, Kilian, Leighton, and Shamoon 1997]. The original host image used for this experiment is the Peppers image which can be found in Fig. 2.3.

The original image C is watermarked with ten different messages[1] for five different strengths[2] giving us 50 watermarked images. Subsequently each watermarked data S is clipped to find the

[1] random sequences [Cox, Kilian, Leighton, and Shamoon 1997]
[2] values of α [Cox, Kilian, Leighton, and Shamoon 1997]

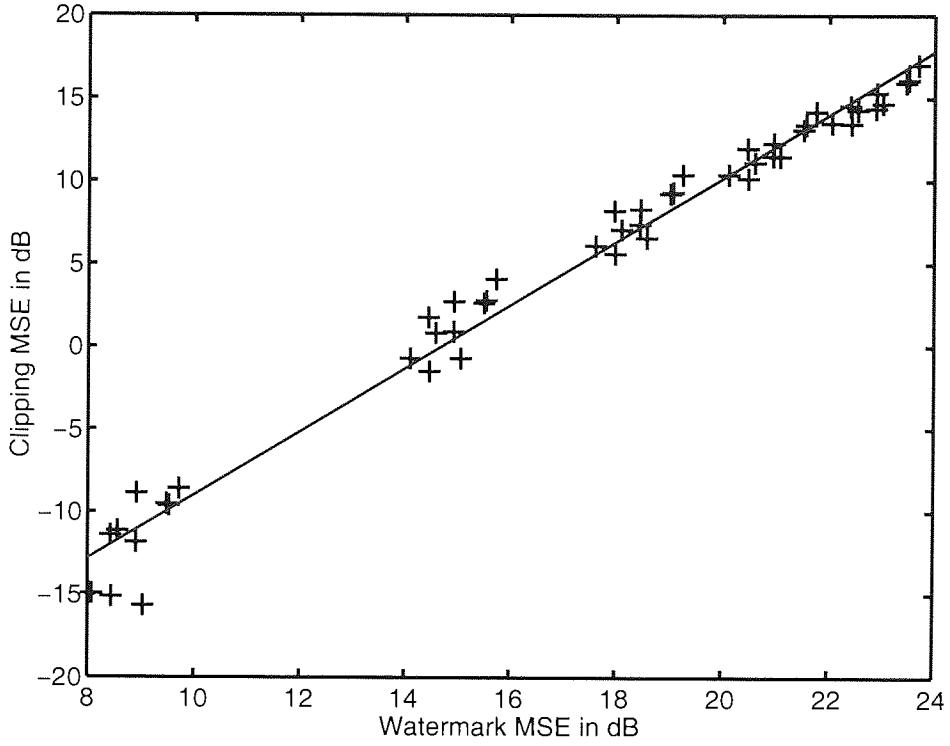


Figure 2.6: Clipping mean squared error (MSE) with respect to the watermark mean squared error.

closest digital representation denoted R. The distortion introduced by such process[3] is represented with respect to the watermark distortion[4] and shown as a cloud of points "+" on Fig. 2.6. A linear regression of these data points is also plotted with a plain line.

Figure 2.6 shows that clipping effects become more important when the watermark distortion increases. When raising the distortion introduced by the watermark, one usually aims at achieving a higher information rate or a lower decoding error probability. However, part of this effort is defeated by quantisation and clipping effects if the limitations of digital images are not taken into account.

2.1.2 Quantisation

Definition 12 *Quantisation is the mapping of a possibly continuous data (scalar, vector, matrix) to a neighbouring discrete data from a fixed set $\mathscr{Q}$. The data to be quantised and the fixed set $\mathscr{Q}$ are termed source data and quantisation grid or lattice, respectively.*

Quantisation is typically used to represent data using a finite number of bits. For instance, most signal processing techniques result in data which need to be quantised to the fixed set of representable data $\mathscr{C}$. In the context of lossy compression, it is also used to reduce the number of bits required to represent data at the expense of some precision. If $\mathscr{Q}$ is equal to the set of all positive and negative integers $\mathbb{Z}$, four basic quantisation mappings exist:

[3] measured by the mean squared error between R and S
[4] measured by the mean squared error between S and C

round – represented by $[.]$, maps the data to the nearest integer,

ceil – represented by $\lceil . \rceil$, maps the data to the nearest higher integer,

floor – represented by $\lfloor . \rfloor$, maps the data to the nearest lower integer,

fix – maps the data to the nearest integer which is smaller than the data in absolute value; the fix(.) operator is equivalent to $\lfloor . \rfloor$ for positive data and to $\lceil . \rceil$ for negative data.

Similar quantisers, for $\mathscr{U} = \delta\mathbb{Z}$ with $\delta \in \mathbb{R}_+^*$, can be easily derived from the integer quantisers ($\delta = 1$). For instance, the rounding mapping can be expressed as

$$R = \left[\frac{S}{\delta}\right]\delta \,, \tag{2.1}$$

where $S \in \mathbb{R}$ is the data to quantise, δ the quantisation step and $R \in \mathscr{U}$ is the quantised data.

For most image processing techniques, quantisation to the set of representable data $\mathscr{C}$, which is done with the "fix" mapping with a quantisation step $\delta = 1$, has little or no consequence on processed image. The quantisation mean squared error distortion, which can be derived from Eq. 2.3, is limited to $\frac{1}{3}$ for uniformly distributed data and is upper bounded by 1 for any other distribution model. Note that this estimation does not allow for clipping, whose effects may be more significant.

However, in the context of watermarking for data authentication and data integrity protection, it may represent a significant difficulty. Indeed, the watermarked data is not necessarily representable and is typically quantised to the closest digital data. Thus, the resulting data may not pass the integrity or authentication test. Several techniques can be considered to overcome this problem. For instance, one may use a watermark composed of discrete values resulting in watermarked data which are unchanged by the quantisation stage.

Quantisation is highly relevant to lossy compression [ISO/IEC JTC1/SC29/WG11 1988; Wallace 1992; JPEG2000 Final committee 2000; Shapiro 1993; Said and Pearlman 1996], as it reduces the number of bits required to represent information. Moreover, as lossy compression is essential to digital image transmission and storage, quantisation study in the context of watermarking is crucial. Due to the significant role they play in image processing, quantisation and lossy compression are the subject of many studies [Wong and Au 2003; Meerwald 2001b; Meerwald 2002] in the context of digital watermarking. The problem has been investigated both probabilistically [Eggers and Girod 1999; Eggers and Girod 2001] and deterministically [Bounkong, Saad, and Lowe 2002b] in the literature.

As seen in Chapter 1, digital watermarking is governed by three parameters: the information rate, the decoding error probability and the mean squared error distortion introduced by the watermark. Typically, the information rate is optimised with respect to the two other parameters as in [Eggers and Girod 2002]. Although this choice seems natural, it is nonetheless arbitrary and other choices are also possible.

In practice, the only truly fixed parameter is the information rate as both message and host data have fixed lengths. Depending on the watermarking technique used, the mean squared error distortion

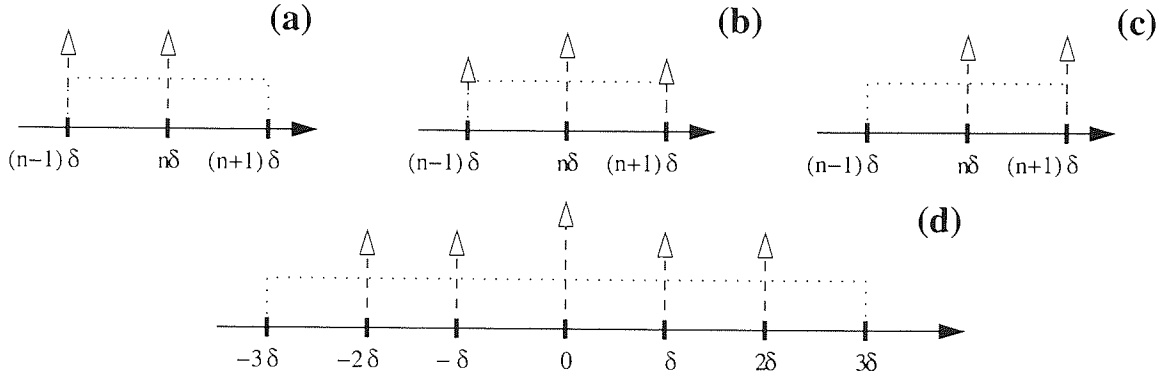


Figure 2.7: Data probability density after quantisation. The dotted lines represent the original probability density, while the dashed line arrows represent the probability density after quantisation for the different mapping: (a) floor, (b) round, (c) ceil and (d) fix.

introduced by the watermark may vary with the particular message and host data, while the decoding error probability will depend on the noise and the watermarked data.

In the following sections, the lowest mean squared error introduced by the watermark to embed information at a given information rate and quantisation noise distortion level is investigated. For the study, both probabilistic and deterministic approaches are considered in the context of digital watermarking.

2.1.3 Quantisation Mean Squared Error

Quantisation, as all images processing techniques, introduces some mean squared error distortion which depends on the following: the probability density of the source data, the quantisation step δ and the type of mapping (e.g. "round", "ceil", "floor" or "fix"). As shown in Fig. 2.7, "ceil" (c), "floor" (a) and "fix" (d) map the source data to the edge of the bin while "round" (b) maps it to the bin centre.

Given a uniformly distributed source, the mean squared error distortion for "round" denoted $< \text{round}^2 >$ and the mean squared error distortion for "ceil", "floor" and "fix" denoted $< \text{fix}^2 >$ can be expressed as

$$< \text{round}^2 > = \frac{1}{\delta} \int_{-\delta/2}^{\delta/2} t^2 \, \mathrm{d}t = \frac{\delta^2}{12}, \tag{2.2}$$

$$< \text{fix}^2 > = \frac{1}{\delta} \int_{0}^{\delta} t^2 \, \mathrm{d}t = \frac{\delta^2}{3}. \tag{2.3}$$

2.1.4 Probabilistic Model for Quantisation

In the probabilistic approach [Eggers and Girod 2001], a framework with continuous quantities is usually adopted and quantisation is typically modelled by an additive centred Gaussian noise N. Since quantisation may not operate uniformly over the whole space, such as in the case of lossy

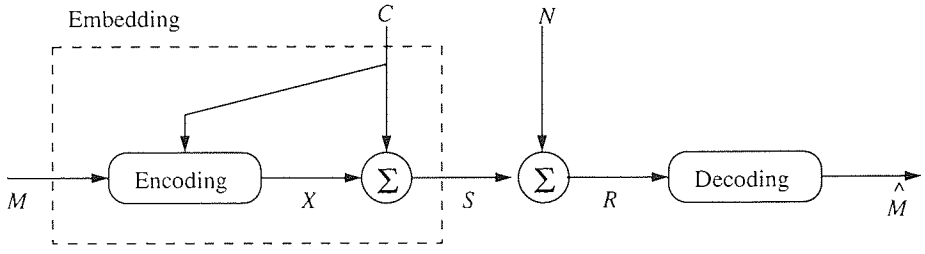


Figure 2.8: Watermarking framework subject to quantisation modelled as an additive Gaussian noise.

compression [Wallace 1992], a communication model as depicted in Fig. 2.8 is considered comprising n parallel independent channels.

Each channel is parameterised by a watermark variance σ^2_{Xi} and a Gaussian noise variance σ^2_{Ni} with $i \in \{1..n\}$. For a given ith channel, the noise variance σ^2_{Ni} is typically set to be equal to the actual quantisation distortion variance [Eggers and Girod 2001], which is obtained by numerical simulations or analytical derivation assuming a distribution model for the source data.

For instance, if the source data S is assumed uniformly distributed and the quantisation mapping is "fix", then the Gaussian noise variance σ^2_{Ni} is set to be equal to $\delta_i^2/3$, where δ_i is the size of the quantisation bin.

Moreover, for each of these Gaussian channels, the information rate is upper bounded [Cover and Thomas 1991] by the channel capacity $\overline{C}_i$ given by

$$\overline{C}_i = \frac{1}{2}\log_2\left(1 + \frac{\sigma^2_{Xi}}{\sigma^2_{Ni}}\right). \tag{2.4}$$

From now on, the following case will be assumed: with a uniformly distributed data and "fix" quantisation mapping characterised by the quantisation step δ_i for the channel of index i, the rough Gaussian approximation commonly used in the literature gives

$$\overline{C}_i = \frac{1}{2}\log_2\left(1 + \frac{3\sigma^2_{Xi}}{\delta_i^2}\right). \tag{2.5}$$

For the whole problem considering n parallel channels, the information rate is upper bounded by $\overline{C} = \sum_{i=1}^{n} \overline{C}_i$. Then, let us denote σ^2_X the global watermark power constraint parameter over the n channels, given by

$$\sigma^2_X = \sum_{i=1}^{n} \sigma^2_{Xi}. \tag{2.6}$$

So, given $\overline{C}$ and $\{\sigma^2_{Ni} | i \in [1,n]\}$, the problem can be summarised as finding a set of σ^2_{Xi} which minimises σ^2_X. Hence, let us optimise σ^2_X with respect to σ^2_{Xi} for given $\overline{C}$ and $\{\sigma^2_{Ni} | i \in [1,n]\}$. By introducing a Lagrange multiplier λ in Eq. 2.6, we have

$$\mathscr{L}(\{\sigma^2_{Xi}\}_{i=1..n}, \lambda) = \sum_{i=1}^{n} \sigma^2_{Xi} + \lambda\left(\overline{C} - \sum_{i=1}^{n} \overline{C}_i\right). \tag{2.7}$$

Then, solving the system of equations given by the first order conditions of optimality

$$\begin{cases} \frac{\partial}{\partial \sigma^2_{Xi}} \mathscr{L}(\{\sigma^2_{Xi}\}_{i=1..n}, \lambda) &= 0, \quad \text{for } i = 1..n, \\ \frac{\partial}{\partial \lambda} \mathscr{L}(\{\sigma^2_{Xi}\}_{i=1..n}, \lambda) &= 0, \end{cases} \tag{2.8}$$

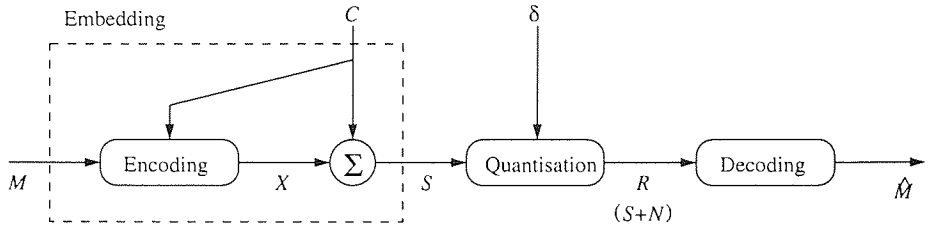


Figure 2.9: Watermarking framework subject to quantisation.

we get

$$\sigma_{Xi}^2 = \frac{1}{3}\left\{2^{\frac{2\overline{C}}{n}}\left(\prod_j \delta_j^2\right)^{\frac{1}{n}} - \delta_i^2\right\}. \tag{2.9}$$

Then, the positivity of the obtained solutions σ_{Xi}^2 has to be verified for each channel. In the two following paragraphs, we first suggest a procedure to cope with negative solutions, then we prove that this procedure results in the optimal solution.

If a negative solution σ_{Xi}^2 is found, then the channel corresponding to the largest quantisation step δ_j is removed from the optimisation process and the watermark variances σ_{Xi}^2 are recalculated. If a negative solution is found again, then the channel corresponding to the new largest quantisation step δ_j is also removed and the watermark variances σ_{Xi}^2 are recomputed. This procedure is repeated until all the solutions found for σ_{Xi}^2 are positive. The watermark variances σ_{Xi}^2 corresponding to the channels removed during the procedure are set to 0.

First, if a negative solution σ_{Xi}^2 is found, we have from Eq. 2.9 that the solution σ_{Xj}^2 found for the channel corresponding to the largest quantisation step δ_j is negative since $\delta_j^2 \geq \delta_i^2$. Second, if for a solution set of σ_{Xi}^2, there is a σ_{Xi}^2 equal to 0 and the watermark variance σ_{Xj}^2 corresponding to the largest quantisation step δ_j is strictly positive, then it can be easily shown that the proposed set of solutions σ_{Xi}^2 is not optimal since setting σ_{Xj}^2 to 0 and using the channel whose σ_{Xi}^2 is equal to 0 leads to a strictly lower global squared distortion σ_X^2. Hence, the procedure we suggested above results in the optimal solution. Note that the positivity constraints correspond to Kuhn-Tucker conditions. The proposed procedure provides a significant reduction of the number of solutions to compute: n cases at most instead of 2^n cases.

2.2 Deterministic Model for Quantisation: A New Approach

In the deterministic approach [Bounkong, Saad, and Lowe 2002b], quantisation is the mapping of a value to a fixed and reduced set of possible values. From a watermarking point of view, quantisation reduces the number of usable codewords in the neighbourhood of the host data and hence the achievable information rate.

In this section, two different scenarios are considered: 1) the quantisation step δ is known at the encoding and decoding, 2) only an upper bound $\overline{\delta}$ of the actual quantisation step δ is known. In

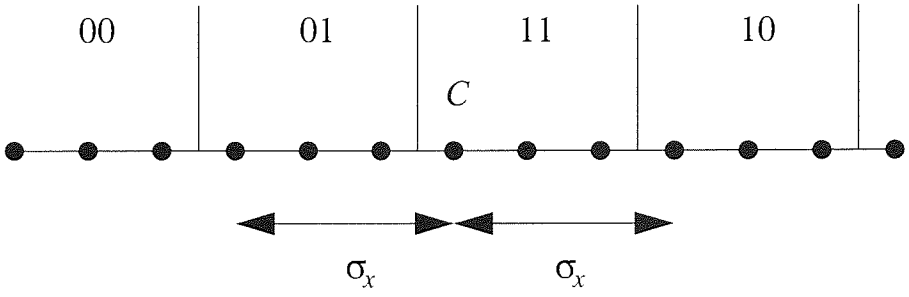


Figure 2.10: An example of quantisation effects on the achievable information rate for a one dimensional data.

the first case, the whole quantisation process is known and the received data R can be determined beforehand, while in the second case, only the maximum distortion level $\overline{\delta}$ is known, which means that the actual quantisation process may use a lower quantisation step $\delta \leq \overline{\delta}$. Both scenarios are described in Fig. 2.9. From now on, the quantisation mapping considered is the "fix" mapping.

2.2.1 Distortion, Information Embedding and Quantisation

For a given 2-norm $||X||$ of the watermark X, the amount of information that can be correctly decoded at the receiving end of the channel is limited by the number of quantisation bins at a lower Euclidean distance than $||X||$ from the original host data C as illustrated for the one dimensional case in Fig. 2.10.

In Fig. 2.10, digital data are represented as black dots, the original data point is labelled C, the quantisation bins are delimited by vertical lines and labelled with '00', '01', '11' and '10'. With the watermark power constraint $||X||^2 \leq \sigma_X^2$, the host data C can be modified into data belonging to three different quantisation bins labelled '01', '11' and '10'. Although the host data C can be modified into seven different codewords S, only three different informations can be decoded from the received data R.

Thus, given the quantisation bin size δ, also termed quantisation step, the relation between the information rate and the watermark power constraint parameter σ_X^2 may be derived. Note that the watermark power constraint parameter σ_X^2 and the bin size δ are related by $\sigma_X^2 = \delta^2/12$ as derived in Eq. 2.2 and Eq. 2.3 for uniformly distributed data.

Let us consider the lowest squared error distortion e^2 introduced to a scalar, subject to quantisation ("fix"), in order to transmit a given amount of information of l bits. Let us also assume the scenario depicted in Fig. 2.11, where each quantisation bin of width δ or 2δ represents a different information denoted by a letter from 'A' to 'F'; for each bin, an arrow indicates the corresponding quantised point which is marked by vertical lines.

The original data C is in the bin labelled 'A'. Its value is set to $C = \epsilon \simeq 0$, which corresponds to the worst case for the embedding. Now, let us defined q and r as

$$q = \text{fix}\left(\frac{C}{\delta}\right) \qquad \text{and} \qquad r = C - q\,\delta\,. \tag{2.10}$$

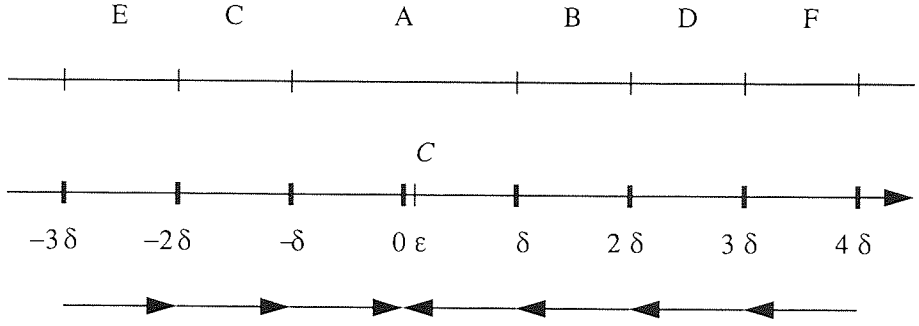


Figure 2.11: Embedding distortion and quantisation. The arrows indicate how the quantisation operates within each bin of width δ. Each quantisation bin is labelled by a letter. The original data is denoted C.

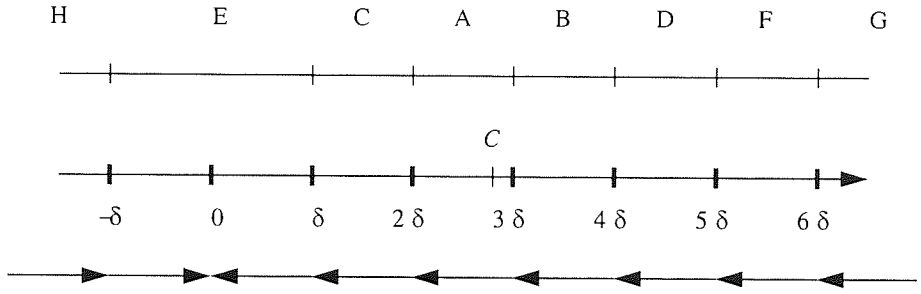


Figure 2.12: Embedding distortion and quantisation. The arrows indicate how the quantisation operate within each bin of width δ. Each quantisation bin is labelled by a letter. The original data is denoted C.

To embed 1 bit, the embedder needs to be able to map the host data C to two codewords of two different quantisation bins. For instance, in Fig. 2.11, the two closest bins to the host data C are labelled 'A' and 'B'. Thus, the squared error is given by $e^2 = \delta^2$ (for simplicity, the ϵ term is neglected as $\epsilon \simeq 0$). The quantisation bins labelled 'A' and 'B' are said "accessible". Similarly, to embed 2 bits, the quantisation bins labelled 'A', 'B', 'C' and 'D' have to be accessible, and the squared error is $e^2 = 4\delta^2$. Thus, to embed l bits, the lowest squared error e_l^2 is given by

$$e_l^2 = 2^{2(l-1)}\delta^2 \,. \tag{2.11}$$

In the previous example, one can notice that the host data C was in the bin labelled 'A', which is twice as large as the others. This case is also referred to as the worst case since the squared error distortion to introduce in order to embed l bits is maximum. Moreover, because of its 'double' width, this bin induces some irregularities in the computation of e_l^2 when the original host data C is not in the central double width bin as in Fig. 2.11.

For instance, let us consider the case depicted in Fig. 2.12 to illustrate the general case and the irregularities in the computation of e_l^2. To embed 1 bit, the bins labelled 'A' and 'B' have to be accessible resulting in squared error $e_1^2 = (\delta - |r|)^2$. To embed 2 bits, the bins labelled 'A', 'B', 'C' and 'D' have to be accessible, and the squared error is $e_2^2 = (2\delta - |r|)^2$. To embed 3 bits, we notice

first that the bin labelled 'E' has a width of 2δ, thus the squared error is $e_3^2 = (7\delta + |r|)^2$. Similarly, to embed l bits, the lowest resulting squared error e_n^2 is given by

$$e_l^2 = \begin{cases} (2^{l-1}\delta - k)^2 , & \text{if } |q| - l < 0 , \\ \left((2^{l-1} - 1)\,\delta + k\right)^2 , & \text{if } |q| - l \geq 0 , \end{cases} \tag{2.12}$$

$$k = \begin{cases} \delta - |r| , & \text{if } q = 0 , \\ \min(|r|, \delta - |r|) , & \text{if } q \neq 0 . \end{cases} \tag{2.13}$$

Note that if $q = 0$ and $r = \epsilon \simeq 0$, we get the same formula as in Eq. 2.11. Moreover, let us define the "additional squared distortion" d_l^2 to embed an extra bit, thus $l + 1$ instead of l, as

$$d_l^2 = e_{l+1}^2 - e_l^2 . \tag{2.14}$$

2.2.2 State of the Problem and Worst Case Solution for Known Quantisation Step

In this section, as in the stochastic approach described in Sec. 2.1.4, the communication model considered is composed by n independent parallel channels. Each channel is subject to a quantisation process ("fix") parameterised by δ_i with $i \in [1, n]$, known at the encoding. The watermark power constraint parameter over the n channels is denoted σ_X^2 and given by

$$\sigma_X^2 = \sum_{i=1}^{n} \sigma_{Xi}^2 , \tag{2.15}$$

where σ_{Xi}^2 is the power constraint parameter of the ith channel. Moreover, the global information rate Ir is given by $Ir = \sum_i Ir_i$ where Ir_i is the information rate of the ith channel.

So, given the global information rate Ir and the bin sizes of the n channels $\{\delta_i | i \in [1, n]\}$, the problem can be summarised as finding the set of watermark variances σ_{Xi}^2 or information rates Ir_i which minimises σ_X^2. Thus, let us optimise the watermark distortion σ_X^2 with respect to the information rates Ir_i for a given global information rate Ir. In the worst case, where $C = \epsilon \simeq 0$, the global power constraint parameter σ_X^2 can be expressed from Eq. 2.11 and Eq. 2.3 as

$$\sigma_X^2 = \frac{1}{3} \sum_{i | Ir_i > 0} 2^{2(Ir_i - 1)} \delta_i^2 , \quad \text{with} \quad Ir = \sum_i Ir_i . \tag{2.16}$$

Then, introducing a Lagrange multiplier in Eq. 2.16 as in Eq. 2.7, and solving the system of equations given by the first order condition of optimality, we get

$$Ir_i = \frac{Ir}{n} + \frac{1}{n} \sum_j \log_2 \delta_j - \log_2 \delta_i . \tag{2.17}$$

Then, the positivity of the obtained information rates Ir_i has to be verified for each channel. If a negative solution Ir_i is found, then the procedure described in Sec. 2.1.4 is carried out. This means that the channels corresponding to the largest quantisation steps δ_j are successively removed from the optimisation and the information rates Ir_i are recomputed, until all the solutions Ir_i are positive. All the information rates Ir_j corresponding to the channels removed during the procedure are set to 0.

2.2.3 Maximum Distortion Algorithm

In the general case, for a given host data C, the best set of information rates Ir_i for each channel varies depending on the values of q_i and r_i with $i \in [1, n]$, which are given by Eq. 2.10. Simple analytical derivation can only be obtained for the worst case as in Eq. 2.17. Thus, we put forward an algorithmic way to tackle the general case using an iterative greedy algorithm.

Outline

Firstly, the distortion to embed one bit is computed for each channel $i \in [1, n]$. The number of embedded bits in each channel is denoted Ir_i and set to 0, the total number of bits to embed is denoted Ir. Then, the following operations are repeated until $Ir = \sum_i Ir_i$. The channel of minimal embedding distortion indexed j is selected. The quantity Ir_j is incremented by one and the distortion to embed an additionnal bit using this channel is computed. The algorithm stops when $Ir = \sum_i Ir_i$.

Algorithm

1. First, the information rate Ir_i of each channel and the global squared distortion σ_X^2 are set to 0, then the distortion K_i to embed one bit in the ith channel is evaluated for each channel as

$$K_i = e_{1,i}^2 \,. \tag{2.18}$$

2. The lowest distortion $K_j = \min_i K_i$ is selected, σ_X^2 is increased by K_j and Ir_j increased by one.

3. Then, the selected K_j is then updated using

$$K_j = d_{Ir_j,i}^2 \,. \tag{2.19}$$

4. Steps 2 and 3 are repeated until $Ir = \sum_i Ir_i$.

Discussion

Since at each iteration $\sum_i Ir_i$ increases by 1, the convergence of this algorithm terminates is trivial and achieved in Ir steps.

In the previous analysis and in the presented algorithm, the quantisation steps δ_i are assumed to be known at the encoding, which might not be the case in most practical cases. However, it is reasonable to assume that the quantisation steps δ_i are upper bounded. In the two following sections, the difference for the embedding between knowing the actual δ_i and only an upper bound $\overline{\delta}_i$ is investigated.

2.2.4 Known Quantisation Step and Known Maximum Quantisation Step

In this section, we demonstrate how an unknown lower quantisation step can be more disruptive than a known higher quantisation step.

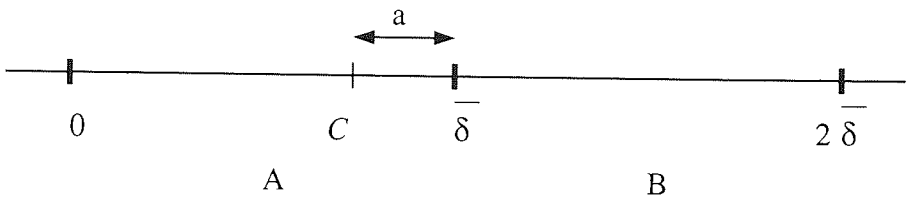


Figure 2.13: Quantisation bins for a given quantisation step of $\overline{\delta}$.

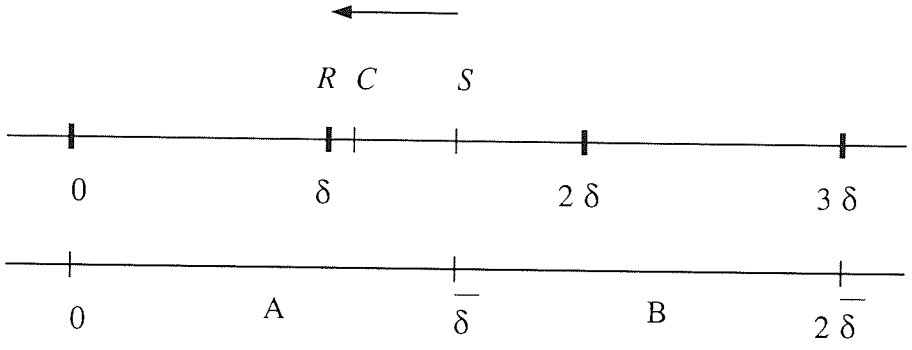


Figure 2.14: Information embedding assuming a quantisation step of $\overline{\delta}$ but actually corrupted with a quantisation step of $2/3\overline{\delta}$.

Let us assume a quantisation step $\overline{\delta}$ and the corresponding labelled lattice as a host data C depicted in Fig. 2.13. As in Sec. 2.2.1, to embed one bit of information, the bins labelled 'A' and 'B' have to be accessible, thus the squared distortion required is a^2.

Now, let us assume that we want to embed the information 'B' and therefore set $S = \overline{\delta}$. Besides, let us assume that the watermarked data S is quantised by a quantiser using a step $\delta = 2\overline{\delta}/3$ and not $\overline{\delta}$ as expected by the sender and the receiver (Fig. 2.14). The received data is therefore $R = \delta = 2\overline{\delta}/3$ and not the expected $R = \overline{\delta}$. Furthermore, the quantisation process with $\delta = 2\overline{\delta}/3$ has changed the watermarked data S from a quantisation bin labelled 'B' to the one labelled 'A'. Thus, the decoded message is 'A' and not 'B'.

Hence, the encoding and decoding areas used in Sec. 2.2.1 have to be changed. Notice that decoded data in the bin labelled 'A' cannot be from any other bins than the ones labelled 'A' or 'B' if $\delta \leq \overline{\delta}$. Then, remark that for all $\delta \leq \overline{\delta}$, data lying between 0 and $\overline{\delta}/2$ are genuinely encoding 'A' as no quantisation process could change data originally in the bin labelled 'B' to this area.

Thus, the encoding and decoding of the symbol 'A' is restricted to the area $\mathscr{E}_A = [-\overline{\delta}/2, \overline{\delta}/2]$. In such a case, there will be no false decoding of 'A'. Similar encoding and decoding areas can be derived for the other symbols as presented in the next section.

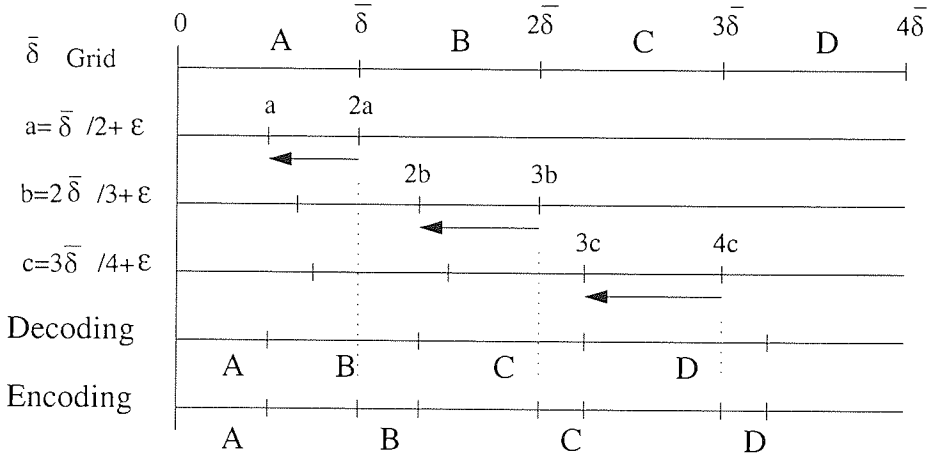


Figure 2.15: Uncertainty areas generated by multiple potential quantisers. Encoding and decoding strategy.

2.2.5 Modified Maximum Distortion Algorithm

We have seen that the framework proposed in Sec. 2.2.2 cannot be applied when only the upper bound $\overline{\delta}$ of the quantisation step is known. When the quantisation step δ is unknown, the quantisation bins defined by $\overline{\delta}$ cannot be directly used as decoding areas like in Sec. 2.2.2. Indeed, the quantisation bins boundaries, defined by the quantisation steps $\overline{\delta}$ and δ with $\overline{\delta} > \delta$, do not match and a bin defined by δ may overlap two bins defined by $\overline{\delta}$ resulting in ambiguous decoding regions as in Fig. 2.14.

However, as pointed out in the previous section, such ambiguity can be avoided by changing the encoding and decoding areas. Furthermore, since such ambiguity exists only between two contiguous bins, the encoding and decoding areas can be modified sequentially from the central bin at '0' to the outer bins. Since the problem is symmetric with respect to 0, the study can be restricted to the positive axis.

In Fig. 2.15, the intervals of ambiguity are underlined by arrows. They correspond to the ambiguity between 'A' and 'B', 'B' and 'C', and 'C' and 'D', respectively. A general formulation of these intervals can be obtained by computing the maximum remainder $R_{\overline{\delta}}$ over δ of the Euclidean division of $i\,\overline{\delta}$ by δ (Fig. 2.15, line 2 to 4), which can be expressed formally as follows, where $i \in \mathbb{N}$ denotes the index of the bin from the bin centre at '0',

$$R_{\overline{\delta}}(i) = \max_{\delta} \left\{ i\,\overline{\delta} - \delta \left[\frac{i\,\overline{\delta}}{\delta} \right] \right\}, \tag{2.20}$$

$$= \frac{i\,\overline{\delta}}{i+1} + \epsilon\,,\ 0 < \epsilon \ll \overline{\delta}\,. \tag{2.21}$$

Then, the intervals of ambiguity $I(i)$ can be expressed as

$$I(i) = [iR_{\overline{\delta}}(i),\ (i+1)R_{\overline{\delta}}(i)]\,. \tag{2.22}$$

To take them into account, the encoding and decoding bins boundaries considered in the algorithm have to be amended. The interval $[a, 2a]$ (Fig. 2.15, line 2) should be associated to 'B' at the decoding

and not used at the encoding. Similarly, the interval $[2b, 3b]$ (Fig. 2.15, line 3) has to be associated to 'C' at the decoding and not used at the encoding, and so on.

This also defines the encoding bounds (Fig. 2.15, line 6); for example to encode B, the watermarked data has to lie in $[\overline{\delta}, 2b]$ (Fig. 2.15, line 3). If the modified data is smaller than $\overline{\delta}$, using a quantiser with $\delta = R_{\overline{\delta}}(1) - \epsilon$ with an appropriate value ϵ will automatically bring it below a and lead to a bad decoding. If the data is greater than $2b$, no quantisation will automatically lead to a bad decoding corresponding to the message C. Once these boundaries are established, a similar algorithm to the one described in Sec. 2.2.2 can be applied.

2.2.6 Practical Applications of the Algorithms

From a practical point of view, the maximum distortion algorithms enable the computation of a relation between the watermark distortion and the number of bits embedded. However, they may not be adapted into practical techniques since the watermak distortion computed depends on the original cover data which is not available to the decoder. Therefore, the results obtained may only be used as bounds.

However, the use of a combinaison of the worst case analysis (Sec. 2.2.2) and of the maximum distortion algorithm with known maximum quantisation step may be adapted into a practical watermarking technique. This technique would achieve the performance of the worst case (Fig. 2.17). The implementation details of such technique is left for future investigation.

2.3 Case Study: JPEG Lossy Compression

In this section, both stochastic and deterministic approaches of quantisation effects to watermarking are demonstrated. The JPEG lossy compression standard has been chosen for this demonstration as it is arguably the most commonly used image file format over the Internet. The standard principles are briefly presented, followed by the numerical simulation details and their results analysis.

2.3.1 Background

The JPEG standard [Wallace 1992] is based on block discrete cosine transform (DCT) decomposition and quantisation. Since image data are strongly locally correlated and DCT is a good approximation of principal component analysis for such data [Jain 1989], very few obtained coefficients gather most of the image information. One achieves the JPEG compression with some quantisers tuned according to the human visual system combined with a lossless entropic encoder.

Figure 2.16 depicts JPEG compression for still images. The image is first divided into contiguous 8×8 pixels patches. Then, while taking the two dimensional discrete cosine transform of all patches, the obtained coefficients are quantised according to

$$R = \mathrm{fix}\left(\frac{S}{\delta}\right)\delta\,, \tag{2.23}$$

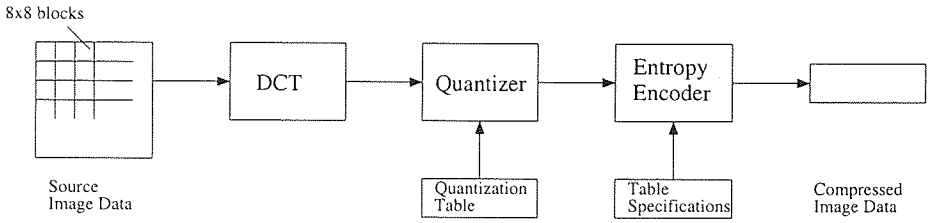


Figure 2.16: JPEG lossy compression block diagram.

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	89	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Table 2.1: JPEG quantisation table: Q.

where R is the quantised data, S the data before quantisation, δ is the quantisation step and where 'fix' is one of the integer quantiser defined in Sec. 2.1.2.

The quantisation step δ is computed from the JPEG quality factor parameter QF (Eq. 2.24 and Eq. 2.25) and the predefined quantisation table Q given in Tab. 2.1, which provides the different values for each coefficient in a patch. Finally, the quantised coefficients are reordered and encoded using a lossless compressor.

$$\delta = kQ\,, \tag{2.24}$$

$$k = \begin{cases} 50/QF & \text{if } QF < 50\,, \\ \frac{(200-2QF)}{100} & \text{if } QF \geq 50\,. \end{cases} \tag{2.25}$$

2.3.2 Numerical Studies

In this section, simulation results based on the statistical and determistic approaches described in Sec. 2.1.4-2.2.5 are reported. All methods were used to evaluate the watermark distortion σ_X^2 required to embed reliably a given number of bits, when the watermarked data S is subject to JPEG lossy compression. The results are given for various quality factor (and therefore for various quantisation steps δ_i). The deterministic approach presented in Sec. 2.2.3 and Sec. 2.2.5 are demonstrated on the well-known Lena image and on the worst case (monochrome black image).

Figure 2.17a and Fig. 2.17b represent the minimum watermark distortion σ_X^2 required as a function of the number of bits embedded in a 8×8 pixels patch. Each line corresponds to a different JPEG compression level which is characterised by its quality factor QF. These functions are estimated by the approach described in Sec. 2.2.2 and in Sec. 2.2.3 for a black image ($C \simeq 0$, worst case) and for the Lena image. In both cases, quantisation is treated as a deterministic process. The experiment

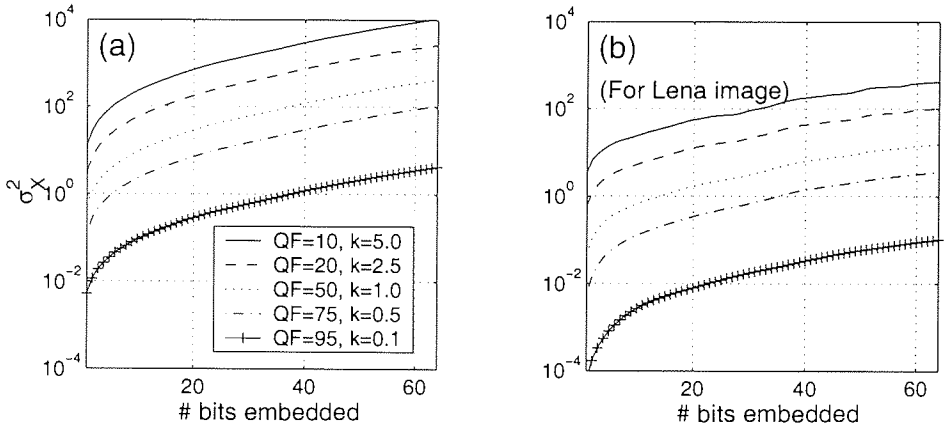


Figure 2.17: Minimum watermark distortion as a function of the number of bits embedded assuming a JPEG compression of the watermarked data. This compression is modelled as a quantisation corruption of known quantisation step. Each line of this figure represents the minimum watermark distortion (σ_X^2) required as a function of the number of bits embedded in a 8×8 pixels patch for a given JPEG quality factor (QF). These results are obtained by the deterministic modelling presented in Sec. 2.2.2 and Sec. 2.2.3. On the left, the host data is (a) a monochrome black image, while (b) the Lena picture is used on the right.

reported in Fig. 2.18a is similar the one reported in Fig. 2.17a except that the quantisation process is modelled by an additive centred Gaussian noise of same variance.

The results show that the Gaussian model approach overestimates the watermark distortion σ_X^2 needed in all studied ranges of quality factor QF and of number of bits embedded. This also explains why some schemes in the literature designed with the full knowledge of the JPEG compression standard achieve better results than expected by the Gaussian model.

This also shows that the stochastic noise is more efficient than quantisation when the δ_i are known at the encoding. However, the knowledge of the noise variance σ_{Ni}^2 is not as informative as the knowledge of the quantisation step δ_i. Indeed, having the latter is more useful, since it allows the computation of the received data R beforehand as a function of the watermarked data S. Knowing the noise variance σ_{Ni}^2 is actually more comparable to the knowledge of an upper bound $\overline{\delta}$ to the actual quantisation step δ (Sec. 2.2.5).

Figure 2.18b shows the watermark distortion σ_X^2 required to transmit Ir bits estimated by the approach described in Sec. 2.2.5 when only the upper bound $\overline{\delta}$ is available to the encoder through the quality factor lower bound $\overline{QF}$. The watermark distortion σ_X^2 required has increased significantly for low $\overline{QF}$ and is almost comparable to the results obtained for the Gaussian model.

Furthermore, for strong quantisation, knowing that the actual quantisation step used δ instead of an upper bound $\overline{\delta}$ is not as helpful as one would expect and it only shows little improvement for the watermark distortion σ_X^2 required. When the information rate Ir increases, the difference with the case in which the quality factor QF is known at the encoding disappears. Indeed, when a channel is used to embed more than one bit, the ratio between the distortion to embed the last bit and the total distortion converges to the same value in both cases (known quantisation step or known upper bound only).

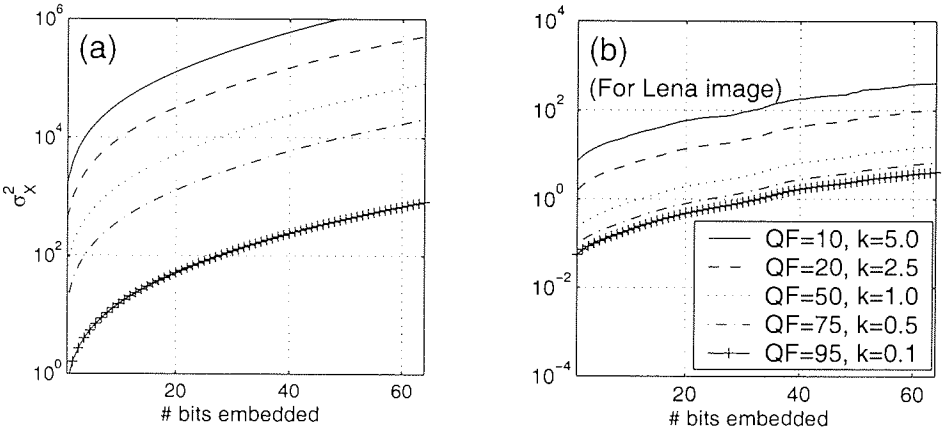


Figure 2.18: Minimum watermark distortion as a function of the number of bits embedded assuming a JPEG compression of the watermarked data. This compression is modelled as (a) an additive Gaussian noise or as (b) a quantisation corruption of unknown quantisation step. Each line of this figure represents the minimum watermark distortion (σ^2_X) required as a function of the number of bits embedded in a 8×8 pixels patch for a given JPEG quality factor (QF). In case (a), the Gaussian noise variance is computed from the compression distortion. In case (b), the approach described in Sec. 2.2.5 is used and the maximum quantisation step $\overline{\delta}$ is given by the step used for the quality factor QF.

2.4 Summary

In this chapter, clipping and quantisation effects on digital watermarking have been investigated. Clipping is shown to defeat part of the watermarking effort when ignored. Its corruptive power has been demonstrated for spread spectrum [Cox, Kilian, Leighton, and Shamoon 1997] technique. Then, quantisation effects on watermarking was studied. In particular, we put forward a new model from which we derived a relation between the watermark distortion and its information rate. The bounds we obtained were closer to the practical limits than the ones of previous studies. Thus, they provide better guidelines to watermarking technique designer. Furthermore, the analysis carried out could also be helpful to improve current techniques which are fairly suboptimal in terms of information rate with respect to the distortion they introduce.

Chapter 3

Information Embedding Techniques for Digital Watermarking

In this chapter, information embedding techniques for digital watermarking are investigated. Following a brief presentation of the information embedding problem framework in the context of digital watermarking, we review the main techniques used in the watermarking field. We also put forward a family of optimised embedding processes obtained from a general optimisation procedure. The characteristics of the optimised embedding processes are discussed for specific noise and decoding models and compared to techniques put forward in the literature.

3.1 Introduction

In the context of digital watermarking, the information embedding problem can be summarised as finding the best codeword S for the message M given the host data C. The codeword S is conditioned by the host data C. In this problem, the host data C and the codeword S are digital data. Furthermore, the watermark X defined as the difference between the codeword S and the host data C is constrained as

$$\|X\|^2 = \|S - C\|^2 \leq hw\,\sigma_X^2\,, \tag{3.1}$$

where σ_X^2 is the mean squared error of the watermark X, h and w are respectively the height and width of X. In the present work, we assume that the number of digital data at an Euclidean distance lower than $\sqrt{hw}\,\sigma_X$ from C is much greater than the number of different messages M given by 2^l where l is the length of M in bits.

In spite of the great number of practical schemes, only few techniques are used to embed and extract information in the context of digital watermarking. In general, their performance is investigated through numerical simulations of the global watermarking scheme. Here, we argue that such approach is not sufficient to assess accurately the contribution of the embedding method.

The two most popular methods are 1) modulation embedding used with a correlation based extractor and 2) quantisation embedding used with a decoding to the nearest codeword message. Both will be discussed in Sec. 3.2 and Sec. 3.3, respectively.

In Sec. 3.4, for clarity and simplicity reasons, the host data C and the codeword S are assumed to be scalar values, while the message M is assumed to be binary. Extensions to more general cases are easily obtained and outlined in the conclusion of this chapter. In Sec. 3.4, embedding processes of the form

$$S = q(C, M) + f(r, M)\,, \tag{3.2}$$

with

$$\begin{aligned} q(C, M) &= \left\{ \text{round}\left(\frac{C}{\Delta} - \frac{M}{4} \right) + \frac{M}{4} \right\} \Delta\,, & (3.3)\\ r &= C - q(C, M)\,, & (3.4)\\ |f(r, M)| &\leq \Delta/2\,, & (3.5)\\ \Delta &> 0\,, & (3.6) \end{aligned}$$

are focussed on. These techniques are tightly related to quantisation embedding by the term $q(C, M)$. Indeed, an example of quantisation embedding investigated in Sec. 3.3 can be expressed as

$$S = q(C, M)\,. \tag{3.7}$$

In Sec. 3.4, an optimisation procedure for the process f is put forward. Its goal is to maximise the good decoding probability at the receiving end, given a host, noise and decoding models. In Sec. 3.5, the properties and performance of the optimal embedding process are discussed with respect to the techniques suggested in the literature. Finally, a summary of the contributions and extensions of the optimisation algorithm presented conclude this chapter.

3.2 Modulation and Correlation Techniques

In communication and transmission, modulation is generally used to embed information in a host signal by modifying its phase or amplitude. For a digital image C, a phase shift in its frequency representation results in a translation in the spatial domain. Since such modification typically has a significant impact on the perceptual quality of the resulting image, most of the research effort on digital image watermarking has been devoted to amplitude modulation.

At the decoding stage, two approaches are usually considered. In the first one, the original data C and the embedded message M are assumed available to the decoder. The original data C is typically subtracted from the received data R resulting in a corrupted watermark $\tilde{X} = R - C$. Then, a correlation based measure between the watermark X and the corrupted watermark $\tilde{X}$ is computed and compared to a pre-defined threshold.

In the second one, only the received data R is required. In this approach, the message M is estimated by computing the first order statistics of different sets of received data coefficients. Although

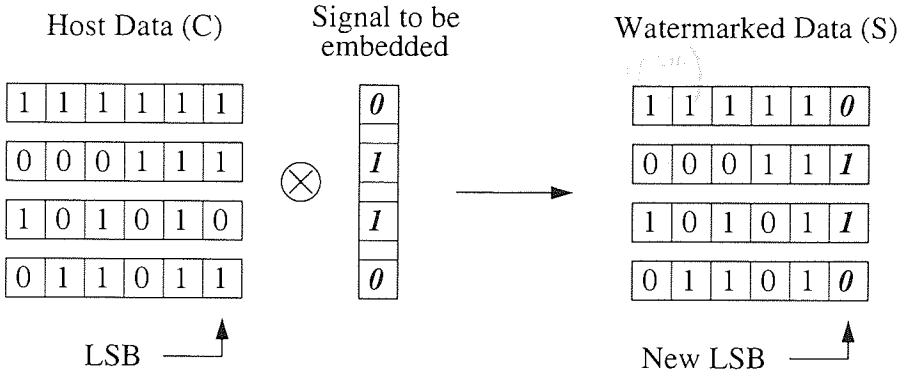


Figure 3.1: Least significant bit modulation technique: an example of embedding.

such method usually suffers from the original host data interference, recent research has shown that they can be reduced by taking into account the host data characteristics.

3.2.1 Embedding and Correlation Based Detection

Least Significant Bits Modulation Based Watermarking

Among the first techniques proposed for digital watermarking, least significant bits modulation consists in the replacement of the least significant bits of each element of the host data C by the element of a noise pattern M as described in Fig. 3.1. The noise pattern M corresponds to the encoded message.

In the following, it will be assumed that C_{LSB} is the matrix of the least significant bit of each element of C and that M is a binary noise pattern. Note that the distortion introduced is usually low and imperceptible. Assuming that the element of C_{LSB} and M are uniformly distributed, the expectation of mean squared error introduced by the replacement of C_{LSB} by M is given by $\frac{1}{2}$.

At the decoding stage, the binary noise pattern M and the received data least significant bits $\hat{M}$ are compared. The ratio r between the number of identical bits and the size of the noise pattern M is computed. Assuming that the transmission channel is a memoryless binary symmetric channel with the bit flip probability p for the watermarked data least significant bit, we have

$$\mathrm{E}[r] = 1 - p, \tag{3.8}$$

$$\mathrm{E}[r^2] = 1 - p, \tag{3.9}$$

$$\mathrm{Var}(r) = p(1-p), \tag{3.10}$$

where $\mathrm{E}[r]$ and $\mathrm{Var}(r)$ are the expectation and the variance of the ratio r.

The ratio r can therefore be compared to a pre-selected threshold t in $[0, 1]$, which controls the probabilities of detection $\mathrm{P}(r \geq t)$ and of false alarm $\mathrm{P}(s \geq t)$, where s is the ratio between the number of identical bits between the noise pattern M and the host data least significant bit C_{LSB}. Given the probability p_c for a bit of the noise pattern M and of the host data least significant bit C_{LSB} to be

identical, we have

$$P(r \geq t) = \sum_{i=\lceil hwt \rceil}^{hw} p^i (1-p)^{(hw-i)} \binom{hw}{i}, \tag{3.11}$$

$$P(s \geq t) = \sum_{i=\lceil hwt \rceil}^{hw} p_c^i (1-p_c)^{(hw-i)} \binom{hw}{i}, \tag{3.12}$$

with

$$\binom{hw}{i} = \frac{(hw)!}{(hw-i)!i!}. \tag{3.13}$$

Note that if the bit flip probability p of the least significant bit is $\frac{1}{2}$ during the transmission, the distortion introduced would remain imperceptible with a mean squared error of $\frac{1}{2}$, while the expectation of the ratio r would be as low as if the host data C was not watermarked. Least significant bits modulation techniques are very simple and introduce low distortion. However, mild attacks in the sense of the mean squared error can defeat them easily.

Spread Spectrum Based Watermarking

In spread spectrum based schemes, the watermark variance is spread through a large range of frequencies unlike in least significant bits modulation techniques based schemes which have an influence on a narrow band of high frequencies.

In spread spectrum based schemes, modulation is applied to the frequency representation of the host data, such as the Fourier or cosine representation, resulting in the dissemination of the watermark information throughout the whole image.

Typically, k frequency coefficients denoted C are selected to carry the watermark information which is encoded as a centred Gaussian random sequence denoted M. In Cox's seminal paper [Cox, Kilian, Leighton, and Shamoon 1997], few information embedding techniques were suggested; they can be summarised as

$$\text{(a)} \quad S = C + M, \qquad \text{(b)} \quad S = C(1+M), \qquad \text{and} \quad \text{(c)} \quad S = Ce^M, \tag{3.14}$$

where S is the vector of modified frequency coefficients.

Among the formulae proposed, Eq. 3.14b has arguably attracted most of the research interest and will be assumed for the rest of the discussion. If the frequency transform related to the k coefficients C is a combination of rotations and reflections such as the discrete cosine transform, and if $m_i \sim \mathcal{N}(0, \sigma^2)$, then the mean squared error introduced by the modulation, and its expectation are respectively given by

$$\frac{1}{hw}\|C - S\|^2 = \frac{1}{hw} C \cdot M \qquad \text{and} \tag{3.15}$$

$$\frac{1}{hw}\mathrm{E}\left[\|C - S\|^2\right] = \frac{k\,\sigma^2}{hw}\|C\|^2. \tag{3.16}$$

One observes that the distortion introduced to a coefficient depends on its own intensity. Such practice is partly motivated by the principle put forward in [Cox, Kilian, Leighton, and Shamoon 1997],

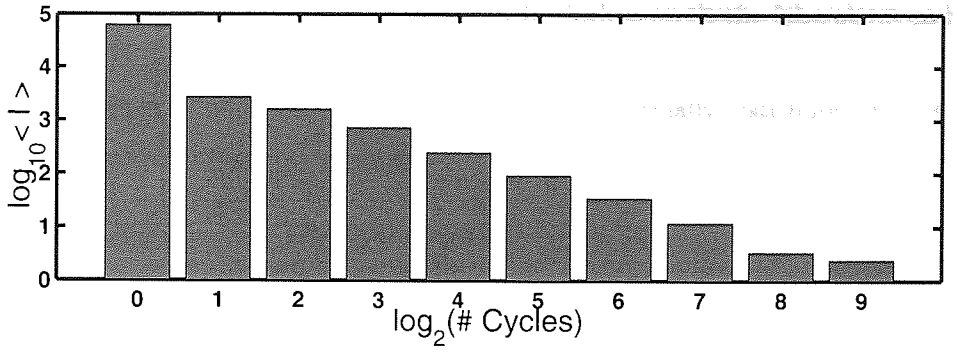


Figure 3.2: DCT frequency subband average component intensity.

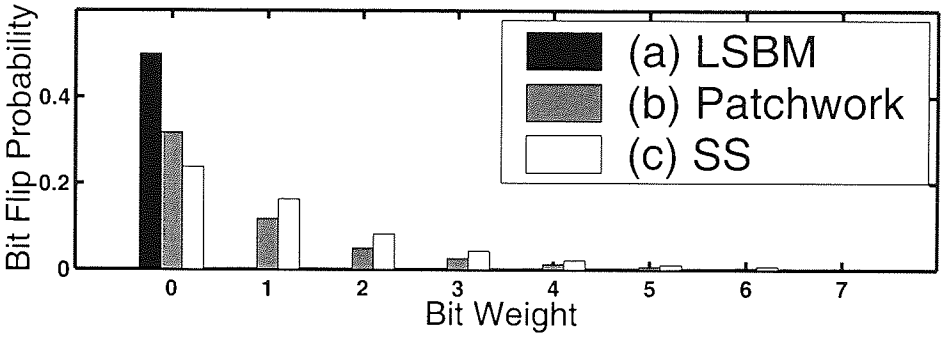


Figure 3.3: Bit flip probability due to (a) LSBM, (b) Patchwork and (c) SS embedding techniques for a 512×512 pixels image. For Patchwork technique, the number of selected pairs is $\frac{1}{4}\,512 \times 512 = 65536$. For all techniques, $D(C,S) = 0.5$ bits^2/pixel.

which can be summarised as follows: the watermark information should be embedded in perceptually significant features.

In the context of frequency watermarking, this consists in selecting coefficients with high intensity values. These are typically concentrated in low frequencies for digital images as shown in Fig. 3.2 for the discrete cosine transform coefficients, which also let the highly locally correlated nature of images show through.

Furthermore, spread spectrum based techniques embed usually more information in higher significance bits than least significant bits modulation or Patchwork techniques [Bender, Gruhl, Morimoto, and Lu 1996], as shown in Fig. 3.3. Hence, spread spectrum based techniques are more robust to common signal processing techniques and more attractive for practical schemes.

At the decoding stage, the original host data C is assumed available and the estimate $\hat{M}$ of the embedded Gaussian sequence M is given by

$$\hat{M} = \frac{R}{C} - 1\,, \tag{3.17}$$

where R is the received vector of selected coefficients. Then, the correlation r between M and $\hat{M}$ defined as

$$r = M \cdot \hat{M} \tag{3.18}$$

is computed and compared to a pre-defined threshold t, which controls the false alarm and detection probability.

If the two sequences M and $\hat{M}$ are independent, then r is normally distributed, $r \sim \mathcal{N}(0, \|\hat{M}\|^2)$ resulting in a false alarm probability given by

$$\mathrm{P}(r \geq t) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{t}{\|\hat{M}\|\sqrt{2}}\right), \tag{3.19}$$

where $\mathrm{erfc}(\cdot)$ is the complementary normal error function.

Assuming that the watermarked data S is corrupted such that the correlation r between M and $\hat{M}$ is given by $r = \|M\|^2 + n$, where n is an independent additive white Gaussian noise with $n \sim \mathcal{N}(0, \sigma_n^2)$, we have

$$\mathrm{E}[r] = k^2\sigma^2\,, \qquad \mathrm{E}[r^2] = 3k^2\sigma^4 + \sigma_n^2\,, \qquad \text{and} \quad \mathrm{Var}(E) = 2k^2\sigma^4 + \sigma_n^2\,. \tag{3.20}$$

Moreover, the probability of detection $\mathrm{P}(r \geq t)$ is given by

$$\mathrm{P}(r \geq t) = \int_{n=-\infty}^{+\infty}\int_{u=t-n}^{+\infty} \mathrm{G}(n, 0, \sigma_n^2)\mathrm{P}(\|M\|^2 = u)\,\mathrm{d}u\,\mathrm{d}n\,, \tag{3.21}$$

$$= \int_{n=-\infty}^{t} \mathrm{G}(n, 0, \sigma_n^2)\,\mathrm{erfc}\left(\frac{\sqrt{t-n}}{k\,\sigma\sqrt{2}}\right)\mathrm{d}n + \frac{1}{2}\,\mathrm{erfc}\left(\frac{t}{\sigma_n\sqrt{2}}\right), \tag{3.22}$$

where $\mathrm{G}(x, \mu, \sigma^2)$ is the Gaussian probability density function of the variable x, of mean μ and of variance σ^2.

Moreover, investigations concerning more complex noise models such as lossy compression are usually carried out numerically as they are typically intractable analytically.

Numerical results published in the literature showed generally good performance for spread spectrum based techniques when the original host data is available to the decoder [Cox, Kilian, Leighton, and Shamoon 1997]. However, these performances are significantly degraded for blind schemes. Also, further schemes based on modulation have been developed to overcome this problem as we shall see in the next section.

3.2.2 Modulation and Message Estimation

Basic Modulation Technique for Blind Decoding

In the case where the original host data is not available to the decoder, modulation can also be used as an information embedding technique. For instance, let us assume that C is a vector of selected frequency coefficients as in the previous section, m a binary message in $\{-1; 1\}$, and U a vector of length k such that $u_i = \pm\,\sigma_U$, with $i \in [1; k]$ and $\sigma_U \in \mathbb{R}+$. Then, the modified frequency coefficients vector S is given by

$$S = C + mU\,, \tag{3.23}$$

and the mean squared error introduced can be expressed as

$$\frac{1}{hw}\|S - C\|^2 = \frac{k}{hw}\,\sigma_U^2\,. \tag{3.24}$$

From the received frequency coefficients vector R and the vector U which is available to the decoder, the estimate $\hat{m}$ of the embedded message m is computed as

$$\hat{m} = \begin{cases} 1 & \text{if } \tilde{m} > 0\,, \\ -1 & \text{otherwise}\,, \end{cases} \tag{3.25}$$

with

$$\tilde{m} = \frac{R \cdot U}{\sigma_U^2}\,. \tag{3.26}$$

Assuming that the coefficients vector S is corrupted by an additive noise N, we have

$$\tilde{m} = m + \frac{C \cdot U}{\sigma_U^2} + \frac{N \cdot U}{\sigma_U^2}\,. \tag{3.27}$$

Moreover, if the host data C and the noise N are independent and normally distributed as

$$c_i \sim \mathcal{N}(0, \sigma_C^2) \qquad \text{and} \qquad n_i \sim \mathcal{N}(0, \sigma_N^2)\,, \tag{3.28}$$

we have

$$\tilde{m} \sim \mathcal{N}\left(m, \frac{\sigma_C^2 + \sigma_N^2}{k\sigma_U^2}\right), \tag{3.29}$$

which results in an error probability of

$$\mathrm{P}(\hat{m} \neq m) = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{k\sigma_U^2}{2(\sigma_C^2+\sigma_N^2)}}\right). \tag{3.30}$$

As shown in the previous equation, both host data and corruption noise interfere with the decoding in the same way. Thus, the host data C can be considered as a source of noise. Moreover, since $\sigma_C^2 \gg \sigma_U^2$, the resulting performances are usually quite poor. One of the main reasons comes from the embedding formula which does not take into account the original host data.

Improved Modulation Technique for Blind Decoding

In [Malvar and Florêncio 2003], an improved modulation technique is proposed. In order to alleviate the interference due to the frequency coefficients vector C, the new embedding process changes the amplitude of the sequence U by introducing a function f in the embedding process

$$S = C + \mathrm{f}(C, U, m)U\,. \tag{3.31}$$

For simplicity reasons, the function f is here approximated by a linear function as put forward in [Malvar and Florêncio 2003]. Hence, it follows that f is expressed as

$$\mathrm{f}(C, U, m) = \left(\alpha m - \lambda \frac{C \cdot U}{\sigma_U^2}\right), \tag{3.32}$$

where the two parameters α and λ are related by

$$\alpha = \sqrt{\frac{k\sigma_U^2 - \lambda^2\sigma_C^2}{k\sigma_U^2}}\,, \tag{3.33}$$

to maintain the same expected mean squared error $\frac{k}{hw}\sigma_U^2$ as in Eq. 3.24. Note that the embedding function in Eq. 3.23 corresponds to the particular case where $\alpha = 1$ and $\lambda = 0$.

As previously, the decoding process rely on the computation of the inner product between R and U normalised by σ_U^2 which gives

$$\tilde{m} \sim \mathcal{N}(\mu_{\tilde{m}}, \sigma_{\tilde{m}}^2) \qquad \text{with} \quad \mu_{\tilde{m}} = \alpha\, m \quad \text{and} \quad \sigma_{\tilde{m}}^2 = \frac{\sigma_N^2 + (1-\lambda^2)\,\sigma_C^2}{k\sigma_U^2}\,. \tag{3.34}$$

The resulting decoding error probability $\mathrm{P}(\hat{m} \neq m)$, which can be optimised with respect to λ, is given by

$$\mathrm{P}(\hat{m} \neq m) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\mu_{\tilde{m}}}{\sigma_{\tilde{m}}\sqrt{2}}\right), \tag{3.35}$$

$$= \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{k\sigma_U^2 - \lambda^2\sigma_C^2}{2(\sigma_N^2 + (1-\lambda)^2\sigma_C^2)}}\right), \tag{3.36}$$

$$= \frac{1}{2}\,\mathrm{erfc}\left(\frac{1}{\sqrt{2}}\sqrt{\frac{\frac{k\sigma_U^2}{\sigma_C^2} - \lambda^2}{\frac{\sigma_N^2}{\sigma_C^2} + (1-\lambda)^2}}\right), \tag{3.37}$$

with

$$\lambda = \frac{1}{2}\left(\gamma - \sqrt{\gamma^2 - 4\frac{k\sigma_U^2}{\sigma_C^2}}\right) \qquad \text{and} \qquad \gamma = 1 + \frac{\sigma_N^2}{\sigma_U^2} + \frac{k\sigma_C^2}{\sigma_U^2}\,. \tag{3.38}$$

In [Malvar and Florêncio 2003], it was shown that the approximated improved modulation technique given by Eq. 3.31 and Eq. 3.32 performs significantly better than the original modulation embedding technique given in Eq. 3.23, and as well as the quantisation index modulation technique which is described in the next section.

3.3 Quantisation Based Techniques

As seen in the previous section, one of the main problem in digital watermarking consists in the interference due to the host data itself. As the latter is usually not available to the decoder, it may constitute an additional source of noise and hence reduce the information rate achieved. As we shall see here, using a quantisation based embedding method solves part of the problem.

The rest of this section is organised as follows: quantisation for information embedding is first presented. Then, two embedding techniques based on quantisation used in the watermarking field are reviewed, leading to a discussion on quantiser structures for information embedding and digital watermarking. Finally, we introduce the improvement related to our research.

3.3.1 Quantisation: an Information Embedding Technique

In the context of digital watermarking, some noise may contaminate the transmission, thus the watermarked data can easily change into a neighbouring codeword. This implies that adjacent codewords should usually be mapped to the same decoded message. Thus, codewords encoding different messages should be as far as possible from each other.

One can notice that this description is pretty similar to the well-known sphere packing problem. In this context, the sphere centres could represent the different codewords and the minimum distance between two centres encoding different information, the embedding robustness. Therefore, an information embedding technique based on quantisation can be considered.

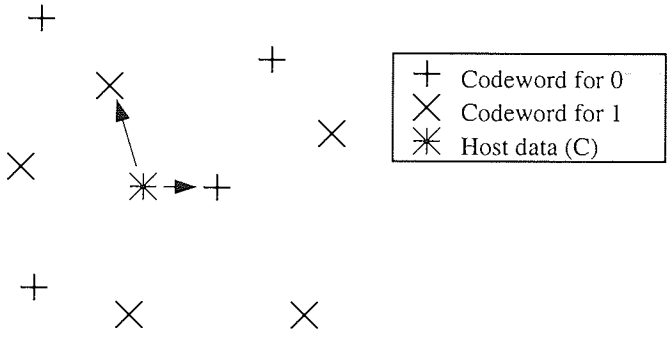


Figure 3.4: Graphical example of quantisation embedding.

The latter may make use different quantisation grids or lattices depending on the information to encode. For instance, let us consider two quantisation grids associated with the information 0 and 1, respectively. At the embedding, if the message M to embed is 0, then the corresponding codeword S will be the closest lattice point of the grid associated with 0.

At the decoding, assuming the availability of the lattices, the decoder has to find the most probable encoding lattice and hence the most probable encoded message. Moreover, the most commonly used approaches are the decoding to the nearest and the maximum a posteriori decoding. Note that the host data interference has been reduced to the uncertainty in the lattice point chosen to encode the message.

Note also that the quantisation grids used are usually regular such that they can be factorised to a simple mathematical expression and easily transmitted to the decoder. Then, in the next section, we shall present two information embedding methods based on quantisation.

3.3.2 Quantisation Based Techniques Examples

In the context of digital watermarking, the most famous quantisation based information embedding techniques are arguably quantisation index modulation (QIM) [Chen and Wornell 2001] and scalar Costa scheme [Eggers, Bauml, Tzschoppe, and Girod 2003]. Both techniques originally operate on scalar and rely on a set of uniform quantisers. As each quantiser is associated with a symbol to be encoded, the number of required quantisers depends on the size of the alphabet, which is assumed binary $M \in \{-1, 1\}$ for the rest of this section. The host data C is also assumed to be a scalar.

Definition 13 *The characteristic distance of a uniform quantiser denoted Δ is defined as the minimum distance between two points of the quantisation grid. Note this is also the distance between two consecutive grid points.*

Thus, the uniform quantisation grids can be expressed as the set

$$\mathscr{U}_M = \left\{ a_M \pm k\Delta \,\middle|\, a_M \in \mathbb{R},\, j \in \mathbb{Z}\,, M \in \{-1, 1\} \right\},$$

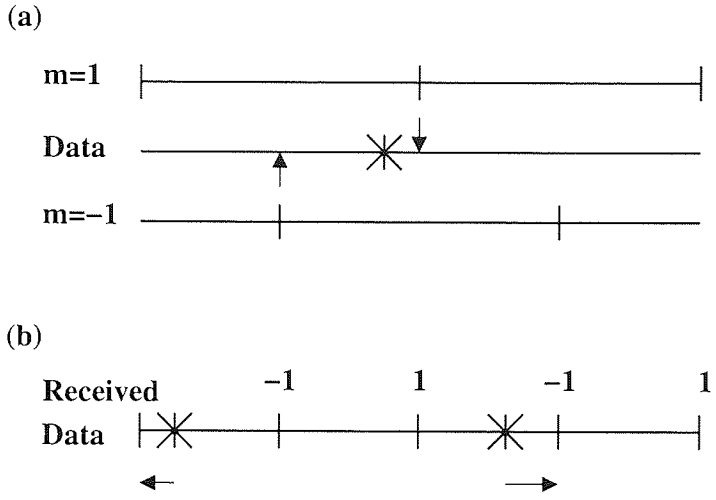


Figure 3.5: QIM embedding (a) and decoding (b). In (a), (*) denotes the original host data C and the arrows indicate the modified points, corresponding to different message bits $M = \pm 1$. In (b), the (*) denotes two possible received values and the arrows indicate their respective decoded messages.

where a_{-1} and a_1 have to be chosen such that the distance between codewords is maximum. Since this only implies that

$$|a_1 - a_{-1}| = \left(k + \frac{1}{2}\right)\Delta \qquad \text{with} \qquad k \in \mathbb{N}, \tag{3.39}$$

we set arbitrarily $a_{-1} = -\Delta/4$ and $a_1 = \Delta/4$.

Thus, the QIM embedding function can be expressed as

$$S = \mathcal{Q}(C, M, \Delta) = \left\{\text{round}\left(\frac{C}{\Delta} - \frac{M}{4}\right) + \frac{M}{4}\right\}\Delta\,, \tag{3.40}$$

and depicted by Fig. 3.5. Furthermore, if C is assumed to be uniformly distributed, the expected mean squared error introduced can be directly given as a function of Δ,

$$\mathrm{E}[(S - C)^2] = \frac{\Delta^2}{12}\,. \tag{3.41}$$

As shown in Fig. 3.7, the probability density function (PDF) of the watermark data is concentrated on the lattices points.

Assuming a decoding to the nearest lattice point and a corruption modelled by a centred Gaussian noise, the decoding error probability is given by

$$\mathrm{P}(\hat{M} \neq M|R) = \sum_{i=0}^{n}(-1)^i \,\mathrm{erf}\left(\frac{(i+1/2)\Delta}{\sigma_N\sqrt{2}}\right) + \mathcal{O}\left(\mathrm{erfc}\left(\frac{(n+1/2)\Delta}{\sigma_N\sqrt{2}}\right)\right), \tag{3.42}$$

where $\mathcal{O}(X)$ stands for a quantity smaller than $|X|$.

From Eq. 3.41 and Eq. 3.42, one can notice that a low mean squared error requires a low quantisation step Δ, while a low decoding error probability requires a big Δ. As low distortion and low error rate are both desirable but unfortunately contradicting, a compromise has to be found.

Besides, Chen and then Eggers suggested to introduce a parameter $\alpha \in [0, 1]$ into the QIM technique resulting in a distortion compensated QIM (DC-QIM) technique. For the parameter α which

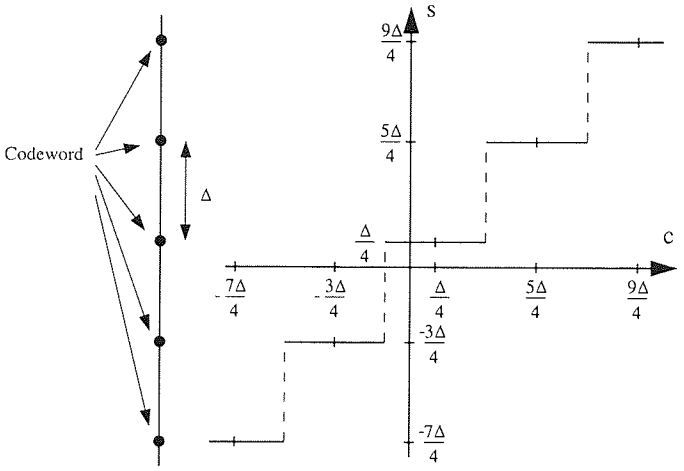


Figure 3.6: An example of QIM embedding function for scalar data and a message $M = 1$.

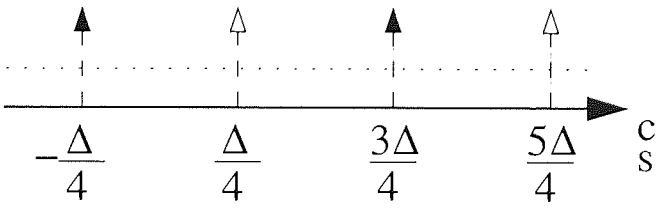


Figure 3.7: Host data probability density function after QIM embedding. White headed arrows represent the contribution due to a message $M = 1$, while black headed arrows correspond to a message $M = -1$. The dotted line corresponds to the original host data probability density function.

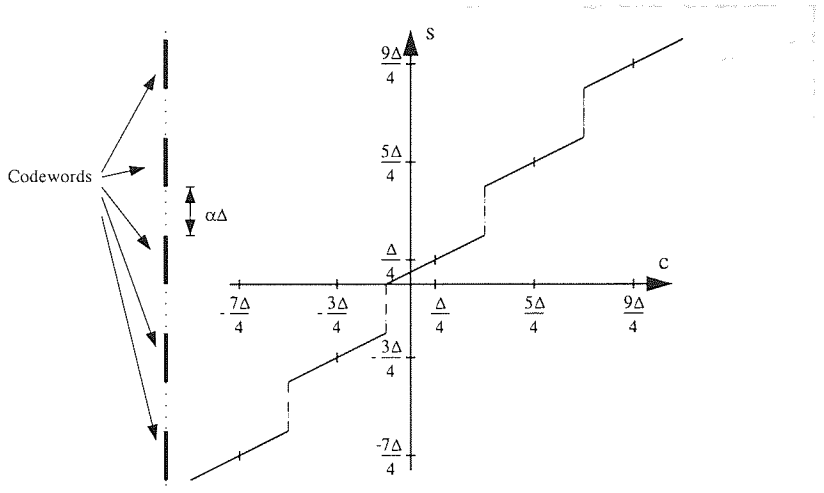


Figure 3.8: An example of DC-QIM embedding function for scalar data and $M = 1$.

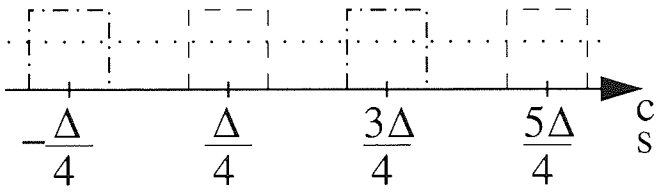


Figure 3.9: Probability density function of the codeword S after DC-QIM embedding. Dashed lines represent the contribution due to $M = 1$, while dashed dotted lines correspond to $M = -1$. The dotted line corresponds to the original host data C probability density function.

gives the maximum information rate, the technique is also named scalar Costa scheme (SCS). The codeword S resulting from DC-QIM embedding function can be seen as the average between the original host data point C weighted with $(1 - \alpha)$ and the lattice point or QIM codeword, weighted with α. Thus, the codeword expression and the mean squared error introduced are given by

$$S = (1 - \alpha)C + \alpha\mathcal{Q}(C, M, \Delta)\,, \tag{3.43}$$

and

$$\mathrm{E}[(s_i - c_i)^2] = \frac{\alpha^2\Delta^2}{12}\,. \tag{3.44}$$

Hence, DC-QIM embedding function does not map the different host data C within a bin to a single codeword S as QIM embedding function in Fig. 3.7. As shown in Fig. 3.9, for a particular bin and message to encode, the probability density function within the bin is rescaled around the bin centre with a factor $(1 - \alpha)$.

The parameter α can be chosen such that the decoding error probability is minimised for a given mean squared error or conversely. Note that the QIM embedding function is a particular case of the DC-QIM embedding function for $\alpha = 1$. As shown in [Chen 2000], significant improvement can be achieved, in particular when the signal to noise ratio (SNR) of the watermark X to the corruption N

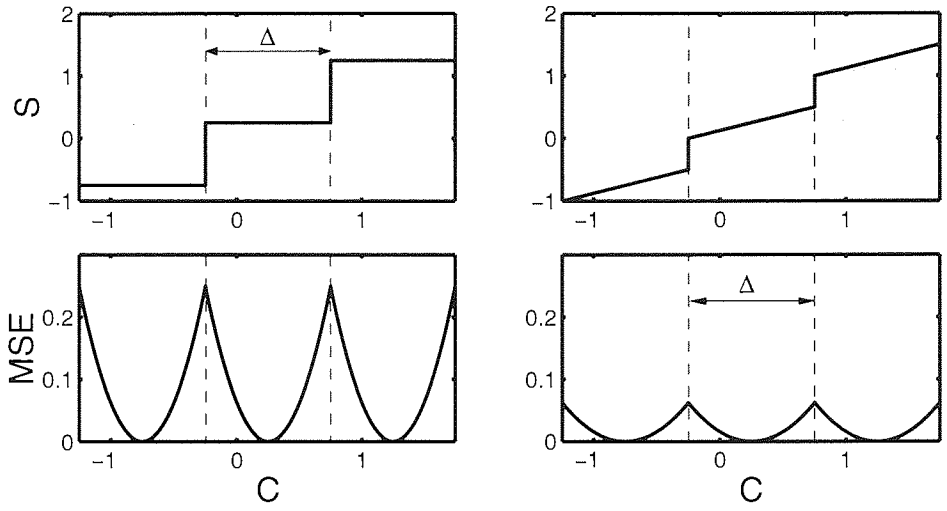


Figure 3.10: Embedding distortion for QIM and DC-QIM embedding. In the top row, QIM and DC-QIM embedding functions are represented, while their respective squared error is depicted on the bottom row as a function of the original host data value C.

is negative (in dB).

From QIM and DC-QIM embedding functions representation in Fig. 3.10, one notices that the squared error introduced at the edges of the quantisation bin has the most significant contribution to the overall mean squared error. However, for certain noise models such as the white Gaussian noise, the reduction of decoding error probability obtained for a host data at the edges is typically very low considering the mean squared error introduced.

For instance, in Fig. 3.11, let us consider the points A and B on the line (a). The point A is at the border of a bin encoding 1 and a bin encoding 0 and the point B is at the centre of a bin encoding 0. After white Gaussian noise corruption, line (b), the probability to be decoded as 1 is represented by the area in grey, which is greater for the point originally at A. Now, let us assume that both points are moved towards the centre of the bin encoding 1 by the same distance and let us observe the probability of being decoded 1 for the new points after a white Gaussian noise corruption, line (c). One notices that in both cases the grey areas have increased. However, it can also be noticed that the point A originally closer from the bin centre benefits more from being moved toward the bin centre than the point B. As shown in line (d), the increase of the grey surface on line (c) is represented on line (d) by the difference between the dark grey and light grey areas. Thus, here we conjecture and show in Sec. 3.4 that a better embedding technique can be devised.

3.3.3 Analysis

In this section, two non-optimality sources of quantisation based embedding schemes are discussed. We focus on the structure of the embedding lattice and on the embedding function based on quantisation grids.

As mentioned previously, QIM and DC-QIM embedding are primarily scalar embedding techniques.

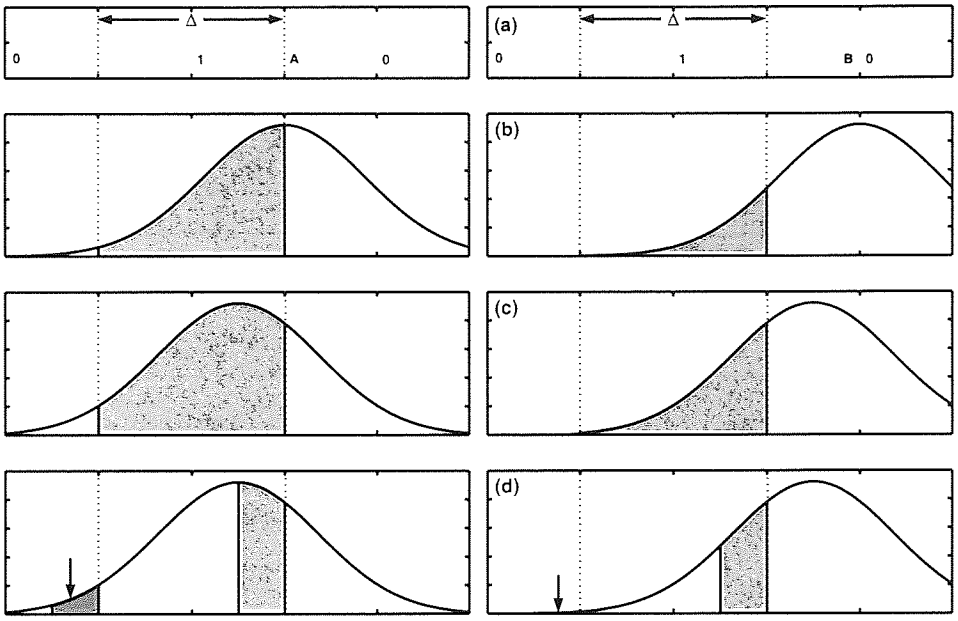


Figure 3.11: Gain in good decoding probability for two different points. On the first row (a), decoding areas delimited by dotted lines and labelled 0 or 1, while the original host data values are denoted 'A' and 'B'. On the second row (b), their good decoding probabilities are coloured in grey. On the third row (c), assuming that the data values have been changed for new values closer to the bin centre corresponding to 1, their new good decoding probabilities are also coloured in grey. Finally, on the bottom row (d), dark grey areas (pointed by an arrow) represent the decrease in good decoding probability from the line (b) to the line (c), while the light grey corresponds to the increase.

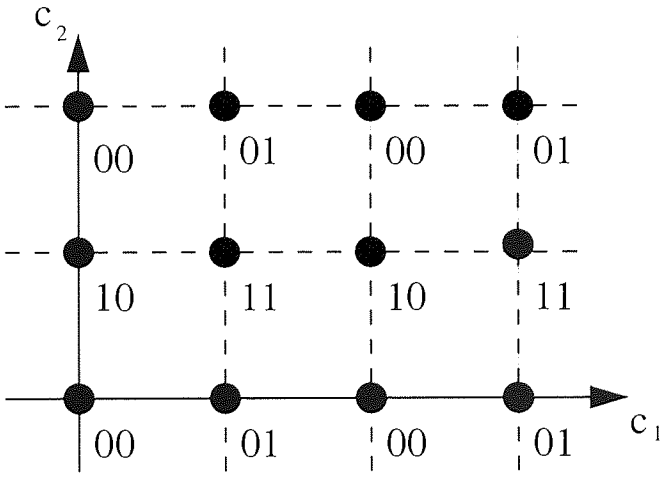


Figure 3.12: Two dimensional embedding lattice resulting for QIM or DC-QIM embedding. The codewords '00', '01', '10' and '11' are represented at the lattice nodes by black dots.

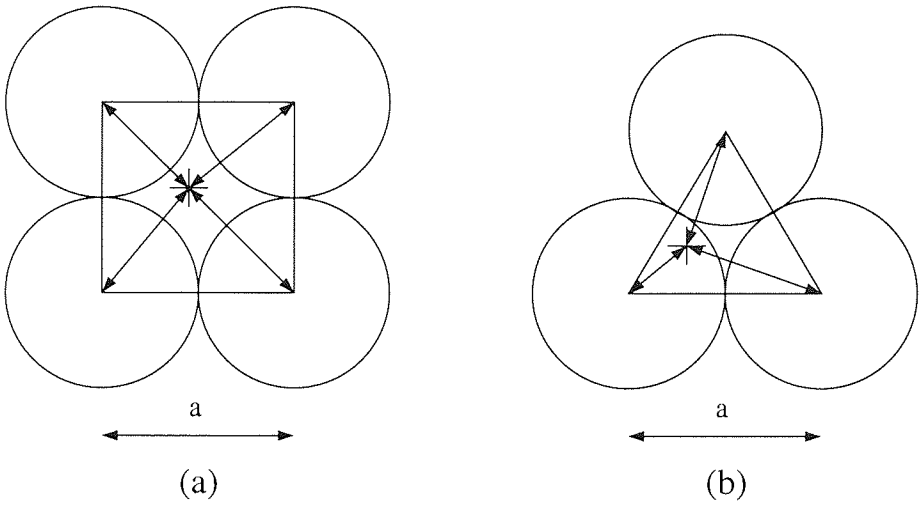


Figure 3.13: Examples of regular lattice and sphere packing in two dimensions: (a) square lattice and (b) hexagonal lattice.

Assuming that originally one bit was embedded per scalar, to double the information rate, one should either double the size of the alphabet or the number of scalars used.

In the latter case, if two scalars are considered to encode two bits, the embedding lattice of codewords may be depicted as in Fig. 3.12. For QIM technique, the decoding error probability and the mean squared error of such encoding will solely depends on the quantisation step Δ. However, it is a well-known fact that the square regular lattice is not optimal.

Let us consider an alternative lattice as depicted in Fig. 3.13b. The characteristic distance of both lattices or the minimum distance between two sphere centres is a. Hence, the robustness of the embedding is unchanged. Now, let us compare the two average squared distortions introduced,

$$\text{For Fig. 3.13a,} \quad \mathrm{E}[(C-S)^2] = \frac{2}{3}a^2\,, \tag{3.45}$$

$$\text{For Fig. 3.13b,} \quad \mathrm{E}[(C-S)^2] = \frac{5}{12}a^2\,. \tag{3.46}$$

This result gives a squared average distortion per bit of $a^2/3$ for the square lattice against $5a^2/(12\log_2(3)) \simeq 0.26a^2$ for the triangle lattice. Note that in two dimensional space this triangle or cubic face centre lattice is optimal. As it is beyond the scope of this work, for further references on higher dimensionality data quantisation and sphere packing, one can refer to [Sloane 1981; Conway and Sloane 1982; G.D. Forney 1988].

Then, as demonstrated here, some improvement can be obtained beyond scalar embedding with vector quantisation. However, as the dimensionality of the data increases, the marginal gain decreases and the embedding and decoding techniques become more and more complex. Indeed, spheres are not space filling object. Also, polytopes such as square or hexagon[1] are typically used as replacement, as depicted in Fig. 3.14. Thus, practical systems usually use low dimensionality quantisers as speed and simplicity are more desirable than a small marginal gain.

[1] in the two dimensional case.

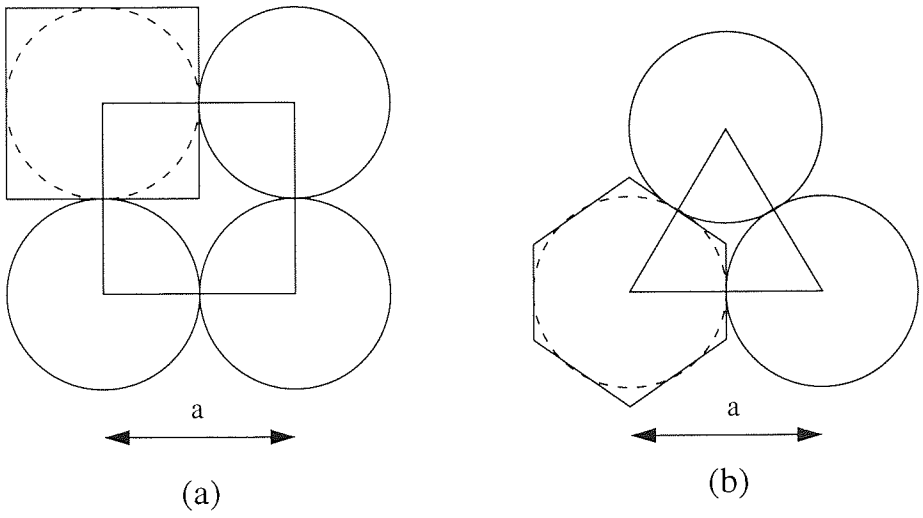


Figure 3.14: Examples of regular lattice based decoding associated with the embedding depicted in Fig. 3.13.

As mentioned in a previous section, QIM codewords are reduced to $\mathscr{U}_M$, which corresponds to the lattice points. However, further researches has pointed out that for the Gaussian noise model studied, such policy may not be optimal especially when the signal to noise ratio is negative (in dB). Hence, the DC-QIM and SCS optimised embedding techniques were proposed but it was also shown to be non-optimal [Bounkong, Toch, Saad, and Lowe 2003b].

3.4 Optimal Discrete Embedding Process

In this section, for clarity and simplicity reasons, the host data C and the codeword S are assumed to be digital scalar values, while the message M is taken from $\{-1, 1\}$. Extensions to more general cases are easily obtained and outlined in the conclusion of this chapter. This section focuses on embedding processes of the form

$$S = q(C, M) + f(r, M)\,, \tag{3.47}$$

with

$$q(C, M) = \left\{\text{round}\left(\frac{C}{\Delta} - \frac{M}{4}\right) + \frac{M}{4}\right\}\Delta\,, \tag{3.48}$$

$$r = C - q(C, M) \qquad \text{and} \tag{3.49}$$

$$|f(r, M)| \leq \Delta/2\,, \tag{3.50}$$

where $\Delta \in \mathbb{R}+$ is the quantisation step. These embedding processes are tightly related to quantisation embedding. Indeed, if $f(r, M) = (1 - \alpha)r$, one obtains the DC-QIM embedding technique, and thus the QIM technique for $\alpha = 1$.

Due to the symmetry of the problem with respect to the message value -1 or 1, it is assumed that $M = 1$. Furthermore, by definition of r in Eq. 3.50, we have $|r| \leq \frac{\Delta}{2}$. For sake of simplicity, let us

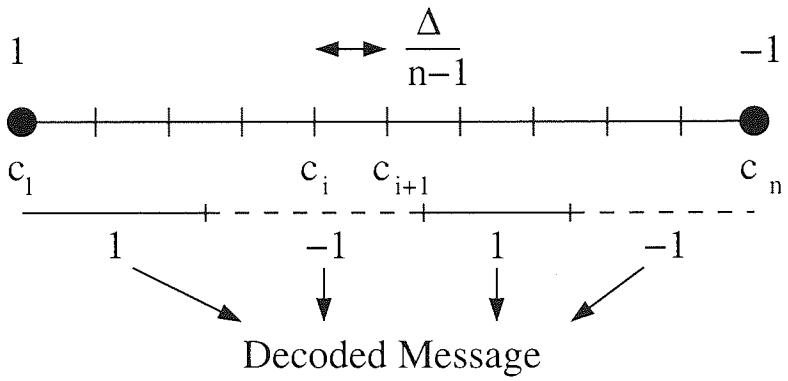


Figure 3.15: An example of decoding areas. The elements of the alphabet $\mathscr{C}$ are represented by vertical lines or black dots. Decoding areas are underlined by dashed lines for $\hat{M} = 1$ and dotted lines for $\hat{M} = -1$ are given. The data points are denoted c_i and the distance between two adjacent points is $\Delta/(n-1)$ where $\Delta \in \mathbb{R}+$. Furthermore, $c_1 = \frac{-\Delta}{2}$ and $c_n = \frac{\Delta}{2}$.

requalify r as the host data C and redefine the watermarked data S or codeword as $S = f(C, M = 1)$, with $|C| \leq \frac{\Delta}{2}$, $|S| \leq \frac{\Delta}{2}$ and $\Delta \in \mathbb{R}+$.

In this section, an optimisation procedure for the process f is put forward. Its goal is to maximise the good decoding probability at the receiving stage given a host, noise and decoding models.

3.4.1 Definitions

Definition 14 *Let us define the alphabet $\mathscr{C}$ of the host data $C \in \mathscr{C}$ as*

$$\mathscr{C} = \left\{ c_i \,\middle|\, c_i = \left(\frac{(i-1)}{n-1} - \frac{1}{2} \right) \Delta, i \in [1, n] \right\}, \tag{3.51}$$

where $n \in \mathbb{N}$ is the number of digital data and $\Delta \in \mathbb{R}+$ is the width of the quantisation bin. Then, let us denote p_i the probability of having $C = c_i$, $p_i = P(c_i)$.

From the constraint defined in Eq. 3.50, we have that $\mathscr{C}$ is also the alphabet of the watermarked data S. Besides, let us assume that the number of possible host data points n is reasonably large.

Definition 15 *Decoding is characterised by the partition of $\mathscr{C}$ into two non-overlapping subsets denoted $\overline{\mathscr{C}}$ and $\underline{\mathscr{C}}$ as in Fig. 3.15. These two subsets are associated with the decoded message $\hat{M} = 1$ and $\hat{M} = -1$, respectively.*

Definition 16 *A discrete embedding process f is a mapping of $\mathscr{C}$ into $\mathscr{C}$ given the message M such that*

$$\forall i \in [1, n], \exists j \in [1, n], c_j = f(c_i)\,. \tag{3.52}$$

Note that although it is generally true in this section, the embedding process f may not be deterministic. Thus, let us define the probability p_{ij} of having the watermarked data value c_j given the original host data value c_i for an embedding process as

$$\forall (i, j) \in [1, n]^2, p_{ij} = \mathrm{P}(c_j | c_i)\,. \tag{3.53}$$

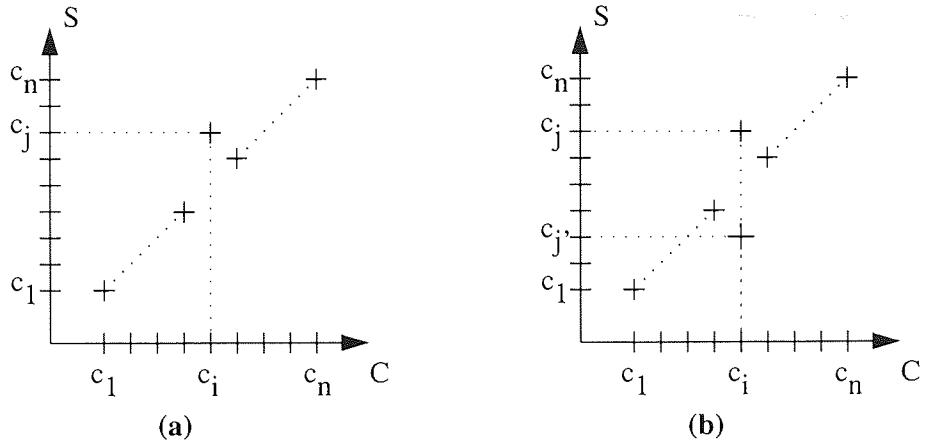


Figure 3.16: Examples of (a) mapping function $F_{ij}(c_k)$ and of (b) stochastic mapping process $F_{i\sim}(c_k)$. In (a), the original data c_i is mapped to the data c_j, while in (2), the original data c_i may be mapped either to the data c_j or to $c_{j'}$.

Definition 17 *The mapping function F_{ij} with $(i,j) \in [1,n]^2$ of $\mathscr{C}$ in $\mathscr{C}$ depicted in Fig. 3.16a is defined as*

$$\forall k \in [1,n], F_{ij}(c_k) = \begin{cases} c_j & \text{if } k = i\,, \\ c_k & \text{otherwise.} \end{cases} \tag{3.54}$$

Let us also define the stochastic mapping process $F_{i\sim}$ with $i \in [1,n]$ of $\mathscr{C}$ into $\mathscr{C}$, depicted in Fig. 3.16b, as

$$\forall k \in [1,n], F_{i\sim}(c_k) = \begin{cases} c_j & \text{if } k = i\,, \text{ with } j \in [1,n] \text{ and } p_{ij}\,, \\ c_k & \text{otherwise.} \end{cases} \tag{3.55}$$

Note that a deterministic mapping F_{ij} is a stochastic mapping $F_{i\sim}$ such that $\exists j \in [1,n], p_{ij} = 1$.

Theorem 1 *Any embedding process f can be expressed as a composition of mapping processes as follows*

$$f = \bigcirc_{i=1}^{n} F_{i\sim}\,, \tag{3.56}$$

where $\bigcirc$ is the generalised function composition operator.

Proof: Since $\{F_{i\sim}\}_i$ partitions the space of all mappings of $\mathscr{C}$ into $\mathscr{C}$ and $F_{i\sim}$ can represent all mapping process for the host data c_i where $i \in [1,n]$, we have that any embedding process f is a composition of $F_{i\sim}$.

If the received data R is equal to the sent data S, then one should retrieve the embedded message M at the decoding stage. However, at the embedding, the nearest codeword S encoding the message $M = 1$ may be far from the original data C and hence the power constraint as defined in Eq. 3.1 may prevent the embedding of correct message $M = 1$ in the watermarked data S. Thus, the codeword S may actually be decoded as $\hat{M} = -1$ as in Fig. 3.18.

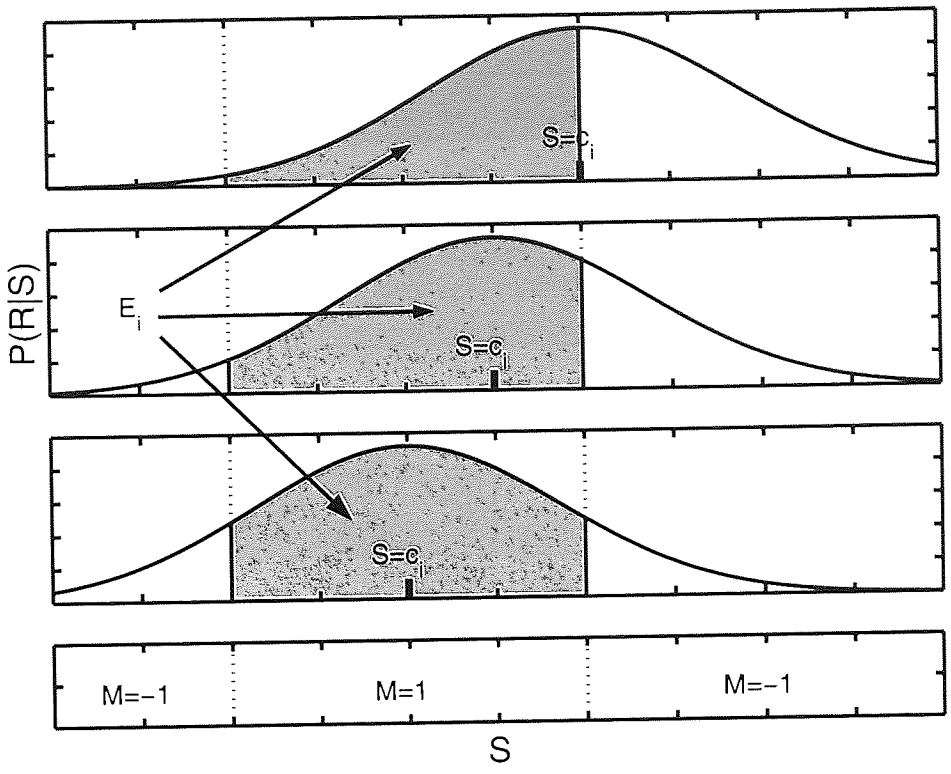


Figure 3.17: An example of good decoding probability for a Gaussian noise model for different codeword S. On the bottom line, the decoding areas are labelled by the message decoded, $M = -1$ or $M = 1$. The grey areas represent the good decoding probability assuming $M = 1$.

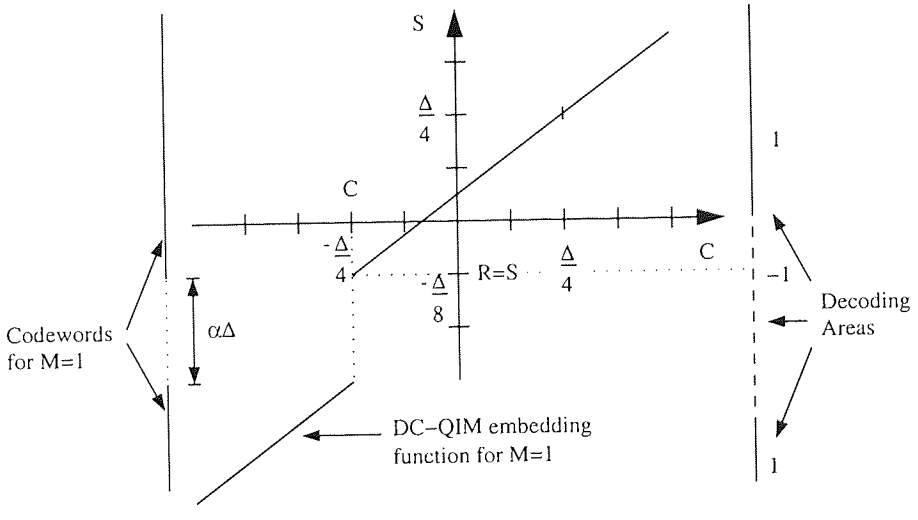


Figure 3.18: An example of bad decoding for DC-QIM embedding in a noiseless transmission, with $\alpha = 1/4$. The host data value is $C = -\Delta/4$ and the message to embed is $M = 1$. Using the embedding function for $M = -1$ given in Eq. 3.43, the watermarked data obtained is $S = -\Delta/8$. Since the transmission is noiseless, the received data is $R = S = -\Delta/8$. Using the decoding areas indicated on the right, the extracted message is $\hat{M} = -1$, resulting in an error as $\hat{M} \neq M$.

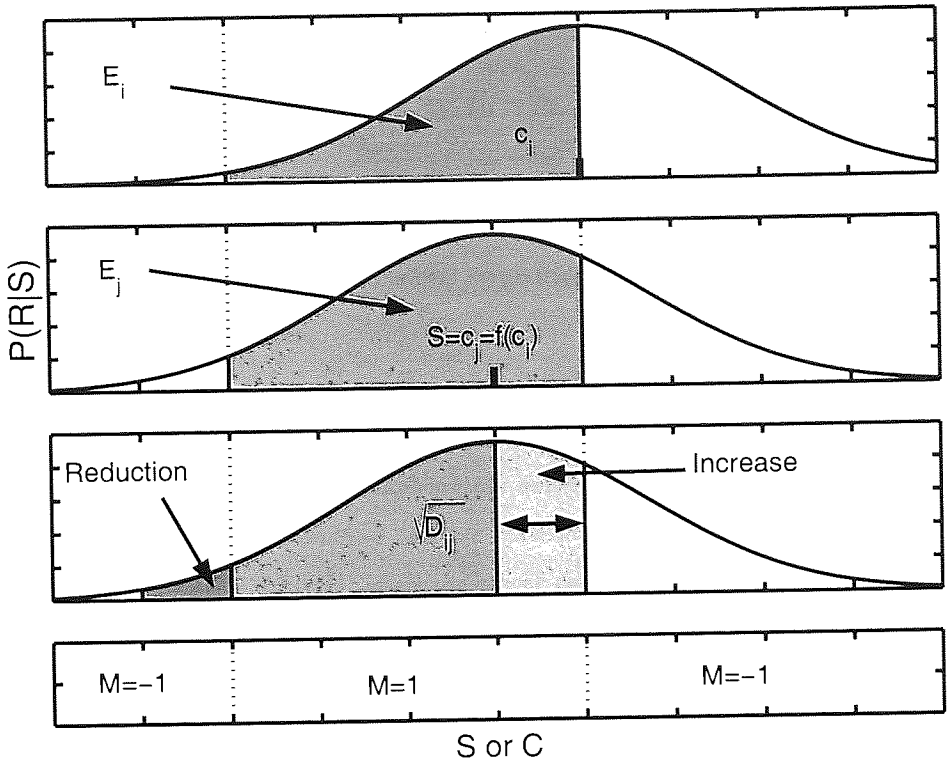


Figure 3.19: An example of good decoding probability variation for a Gaussian noise model. On the bottom line, decoding areas are labelled by the message decoded, $\hat{M} = -1$ or $\hat{M} = 1$. On the first and second lines, the grey areas represent the good decoding probability assuming $M = 1$. On the third line, the grey and light grey areas represent the good decoding probability. Furthermore, the light grey areas correspond to an increase of good decoding probability, while the dark grey areas correspond to a decrease.

Definition 18 *The good decoding probability E_i for a codeword $S = c_i$ with $i \in [1, n]$ given a noise model and a message $M = 1$ is defined (Fig. 3.17) as*

$$E_i = \mathrm{P}(R \in \overline{\mathscr{C}} | S = c_i, M = 1)\,. \tag{3.57}$$

Thus, the global good decoding probability for the identity embedding function $S = C$ given the received data R can be defined as the p_i weighted average of good decoding probability over all values $c_i \in \mathscr{C}$ of the host data C and expressed as

$$\mathrm{P}(R \in \overline{\mathscr{C}} | S = C, M = 1) = \sum_{i=1}^{n} p_i E_i\,. \tag{3.58}$$

Definition 19 *From Def. 16 and Def. 18, the global good decoding probability for an embedding process f can be defined as the weighted average of good decoding probability E_j over the original data values $C = c_i$ and the codewords $S = c_j$ used, expressed as*

$$\mathrm{P}(R \in \overline{\mathscr{C}} | S = f(C), M = 1) = \sum_{i=1}^{n} p_i \Big(\sum_{j=1}^{n} p_{ij} E_j \Big)\,. \tag{3.59}$$

Definition 20 *Let us define the difference of good decoding probability E_{ij} and the squared error D_{ij} resulting from the mapping of c_i to c_j with $(i, j) \in [1, n]^2$ given a noise model and as depicted in*

Fig. 3.19,

$$E_{ij} = E_j - E_i\,, \tag{3.60}$$

$$D_{ij} = (c_j - c_i)^2\,. \tag{3.61}$$

Definition 21 *As in Def. 19, the mean squared error denoted D introduced by an embedding process f can be defined as the weighted average squared error over the original data values $C = c_i$ and the codewords $S = c_j$ used, expressed as*

$$D = \sum_{i=1}^{n} p_i \Big(\sum_{j=1}^{n} p_{ij} D_{ij} \Big)\,. \tag{3.62}$$

Definition 22 *Let us define the discrete information embedding problem as finding the set of n mapping processes $F_{i\sim}$ such that the resulting discrete embedding process has the highest good decoding probability, $\mathrm{P}(R \in \overline{\mathscr{C}} | S = f(C), M = 1)$, while the power constraint defined below is enforced:*

$$D \leq \sigma_X^2\,. \tag{3.63}$$

Let us assume that for an embedding process f associated with a mean squared error D and another mean squared error D', if $D' > D$, then there exists an embedding process f', such that $\mathrm{P}(R \in \overline{\mathscr{C}} | S = f'(C), M = 1) > \mathrm{P}(R \in \overline{\mathscr{C}} | S = f(C), M = 1)$.

As n is finite, there is only a finite number of possible embedding functions ($n!$), hence there is at least one optimal function in the sense of the good decoding probability. Moreover, as it is a finite problem, an exhaustive search is possible in theory. However, for large values of n, such search may be computationally too demanding or even intractable. Thus, a constructive and iterative algorithm to find an optimal embedding process is proposed.

3.4.2 Embedding Function Properties

In this section, some properties of embedding processes are stated and proved. Embedding processes are used to embed information, and thus to reduce the related decoding error probability. However, this may also introduce some distortion to the original data, which is here measured by the mean squared error.

From the previous section definitions, for a given original data value $C = c_i$ and noise model, all possible codewords $S = c_j$ can be represented on a plane with coordinates D_{ij} for the x-axis and E_{ij} for the y-axis as in Fig. 3.20. One notices that some codewords c_j are more relevant than others. For instance, the codewords $S = c_4$ and $S = c_6$ should not be considered since $S = c_4$ reduces the good decoding probability, and a linear average between the codewords $S = c_7$ and $S = c_9$ will result in a higher good decoding probability.

Thus, for a given original host data value $C = c_i$ with $i \in [1, n]$ and noise model, let us build the ordered sequence $\{\mathscr{H}^i\}_k$ of relevant codewords S whose elements are denoted h_k^i. The subscript i denotes the original host data value $C = c_i$, while k is the sequence index. The last index of

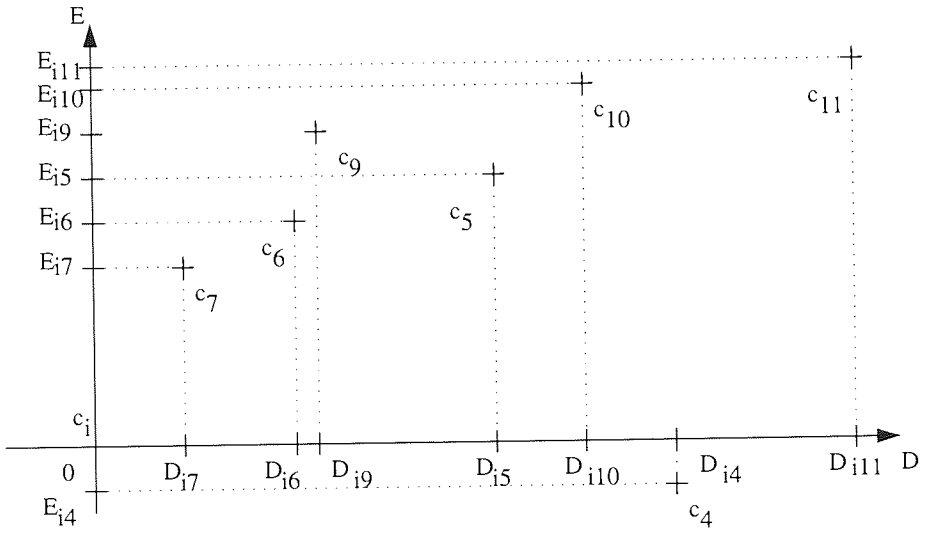


Figure 3.20: An example of the representation of codewords c_j on a plan for the original data c_i. The x-axis and y-axis represent respectively the D_{ij} and E_{ij} related to each codeword for the original data c_i.

the sequence is denoted $\overline{k}^i$. Since irrelevant codewords, such as c_4 in the previous example, are not element of the sequence $\{\mathscr{H}^i\}_k$, the last index of the sequence $\overline{k}^i$ is lower than the number of possible codewords n. Following a necessary definition, the sequence $\{\mathscr{H}^i\}_k$ is defined by recurrence in Def. 24.

Definition 23 *Let us define $\mathscr{F}^i_{kl}$ as the mapping function from the codeword h^i_k to h^i_l, then $\mathscr{E}^i_{kl}$ and $\mathscr{D}^i_{kl}$, as the difference of good decoding probability and the squared error associated with $\mathscr{F}^i_{kl}$, respectively, expressed as*

$$\forall m \in [1,n], \mathscr{F}^i_{kl}(h^i_m) = \begin{cases} h^i_l & \text{if } m = k\,, \\ h^i_m & \text{otherwise,} \end{cases} \tag{3.64}$$

$$\mathscr{E}^i_{kl} = \mathscr{E}^i_l - \mathscr{E}^i_k\,, \tag{3.65}$$

$$\mathscr{D}^i_{kl} = (h^i_l - h^i_i)^2 - (h^i_k - h^i_i)^2\,. \tag{3.66}$$

In the previous definition, $\mathscr{E}^i_{kl}$ represents the improvement in terms of decoding for using the codeword h^i_l instead of the codeword h^i_k, while $\mathscr{D}^i_{kl}$ represents the additional distortion to introduce. The function $\mathscr{F}^i_{kl}$ is the function which changes the codeword h^i_k into the codeword h^i_l and which does not change any codeword which is not h^i_k.

The aim of the next definition is to build the ordered sequence of all potentially codewords. A codeword is considered useful if its improve the decoding performance (Eq. 3.68, first condition) and using it offers the best decoding improvement over introduced distortion ratio (Eq. 3.68, second condition).

Definition 24 *For a given original host data value $C = c_i$ and noise model, let us define the ordered*

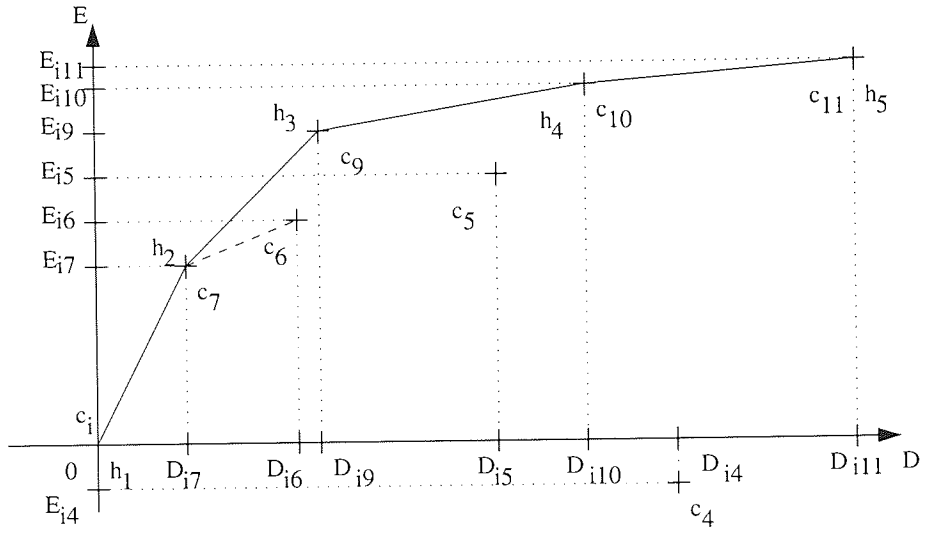


Figure 3.21: Construction of the sequence $\{\mathscr{H}^i\}_k$ from the representation of codewords c_j on a plan for the original data c_i. The x-axis and y-axis represent respectively the D_{ij} and E_{ij} related to each codeword for the original data c_i.

sequence $\{\mathscr{H}^i\}_k$ *of relevant codewords* $S = h_k^i$ *as follows*

$$h_1^i = c_i \tag{3.67}$$

$$\forall k > 1, h_{k+1}^i = c_j \quad \begin{cases} \text{if} & (E_{ij} - \mathscr{E}_{1k}) > 0 \\ \text{and} & \frac{E_{ij} - \mathscr{E}_{1k}}{D_{ij} - \mathscr{D}_{1k}} = \max\limits_{j' \in [1,n]} \frac{E_{ij'} - \mathscr{E}_{1k}}{D_{ij'} - \mathscr{D}_{1k}}, \end{cases} \tag{3.68}$$

where the two conditions in Eq. 3.68 aim at excluding codewords like $S = c_4$ *and* $S = c_6$*, respectively.*

If for a given index k*, two or more codewords* $S = c_j$ *with* $j \in [1, n]$ *are possible, the one resulting in the lowest* $D_{1j} - \mathscr{D}_{1k}$ *is taken. In the case of equality, both*[2] *points are selected and considered as a double point for the sequence. Furthermore, let us define the good decoding probability* $\mathscr{E}_k^i$ *for the codeword* $S = h_k^i$ *as*

$$\mathscr{E}_k^i = \mathrm{P}(R \in \overline{\mathscr{C}} \mid S = h_k^i). \tag{3.69}$$

Finally, let us denote $\overline{k}^i$ *as the last index value of the sequence* $\{\mathscr{H}^i\}_k$*.*

In our previous example in Fig. 3.21, the first elements of the sequence $\{\mathscr{H}^i\}_k$ are given by $h_1 = c_i$, then $h_2 = c_7$, $h_3 = c_9$, $h_4 = c_{10}$, and $h_5 = c_{11}$. Some codewords $S = c_j$ with $j \in [1, n]$ have not been taken to be an element of $\{\mathscr{H}^i\}_k$, as for instance, c_4, c_5 and c_6.

For instance, for the codeword c_4, as $\forall k \in [2, \overline{k}_i], \mathscr{E}_{ik} > 0$ and since $E_{i4} < 0$, we have that $\forall k \in [2, \overline{k}_i], E_{i4} - \mathscr{E}_{ik} < 0$, thus the codeword c_4 never complies with the first selection rule in Eq. 3.68.

The codeword c_6 can be considered for two indices in the sequence $\{\mathscr{H}^i\}_k$ corresponding to the elements h_2 and h_3, since $E_{i6} > 0$ and $E_{i6} - \mathscr{E}_{i7} > 0$. However, in both cases the maximality condition

[2] On a line segment, there are at most two points at a given distance of a third one.

of Eq. 3.68 is not satisfied by the codeword c_6, but by the codeword c_7 and then c_9, respectively. Then, $\forall k \in [4, \overline{k}_i], E_{i6} - \mathscr{E}_{ik} < 0$, thus the codeword c_6 never satisfies the first selection rule. The same situation arises for the codeword c_5 which can be considered for h_2, h_3 and h_4 but never after.

Property 1 *For a given host data value c_i, by construction of $\{\mathscr{H}^i\}_k$ and $\{\mathscr{E}^i\}_k$, we have,*

$$\forall k \in [2, \overline{k}^i], \quad \mathscr{E}^i_k > 0\,, \tag{3.70}$$

$$\forall (k,k') \in [1, \overline{k}^i]^2, \quad k < k' \Rightarrow (\mathscr{E}^i_{kk'} > 0) \wedge (\mathscr{D}^i_{kk'} > 0)\,, \tag{3.71}$$

$$\forall k \in [1, \overline{k}^i - 1], \quad \mathscr{G}^i_k > 0\,, \tag{3.72}$$

$$\forall k \in [1, \overline{k}^i - 1], \quad k < k+1 \Rightarrow \mathscr{G}^i_k \geq \mathscr{G}^i_{k+1}\,, \tag{3.73}$$

$$\forall (k,k') \in [1, \overline{k}^i - 1]^2, \quad k < k' \Rightarrow \mathscr{G}^i_k \geq \mathscr{G}^i_{k'}\,, \tag{3.74}$$

where $\mathscr{G}^i_k$ is the gradient between the points h_k and h_{k+1}, which is given by

$$\mathscr{G}^i_k = \frac{\mathscr{E}^i_{k+1} - \mathscr{E}^i_k}{\mathscr{D}^i_{i,k+1} - \mathscr{D}^i_{ik}} = \frac{\mathscr{E}^i_{k,k+1}}{\mathscr{D}^i_{k,k+1}}\,. \tag{3.75}$$

Furthermore, it can be easily shown that

$$\forall j \in [1,n], \exists (k,\lambda) \in [1, \overline{k}_i] \times [0,1], \\ \left(\lambda \mathscr{D}^i_k + (1-\lambda)\mathscr{D}^i_{k+1} \leq D_{ij}\right) \wedge \left(\lambda \mathscr{E}^i_k + (1-\lambda)\mathscr{E}^i_{k+1} \geq E_{ij}\right). \tag{3.76}$$

Proof Prop. 1: See Appendix A.

Definition 25 *Let us define the stochastic mapping $\mathscr{F}^i_\sim$ associated to the host data value $C = c_i$ as*

$$\forall k \in [1, \overline{k}^i], \mathscr{F}^i_\sim(c_k) = \begin{cases} h^i_l & \text{if } k = i, \text{ with } p^i_l \text{ and } l \in [1, \overline{k}^i]\,, \\ c_k & \text{otherwise,} \end{cases} \tag{3.77}$$

where p^i_l is the probability of having the codeword $S = h^i_l$ given an original data value $C = c_i$.

Furthermore, let us define $k^i \in [1, \overline{k}^i]$ as the maximum index l of $\mathscr{F}^i_\sim$ such that $p^i_l > 0$, and given by

$$k^i = \max_{l \in [1, \overline{k}^i]} \left\{ l \,|\, p^i_l > 0 \right\}. \tag{3.78}$$

Theorem 2 *If $F_{i\sim}$ is a stochastic mapping process with $i \in [1,n]$, then one of the two following statements are true:*

1. *the mean squared error $D_{i\sim}$ introduced by $F_{i\sim}$ is greater than the mean squared error $\mathscr{D}^i_{1\overline{k}^i}$ introduced by $\mathscr{F}^i_{1\overline{k}^i}$, and the increase of good decoding probability $E_{i\sim}$ resulting from $F_{i\sim}$ is lower than the increase of good decoding probability $\mathscr{E}^i_{1\overline{k}^i}$ resulting from $\mathscr{F}^i_{1\overline{k}^i}$;*

2. *otherwise, there is a unique couple (k^i, λ) in $[1, \overline{k}^i] \times [0,1[$, such that the positively weighted average $\lambda \mathscr{D}^i_{1,k^i-1} + (1-\lambda)\mathscr{D}^i_{1,k^i}$ is equal to $D_{i\sim}$, and that the increase of good decoding probability $E_{i\sim}$ resulting from $F_{i\sim}$ is lower than the one resulting from $\mathscr{F}^i_\sim$ composed of the two mappings $\mathscr{F}^i_{1,k^i-1}$ and $\mathscr{F}^i_{1,k^i}$ with the probabilities $p^i_{1,k^i-1} = \lambda$ and $p^i_{1k^i} = 1-\lambda$.*

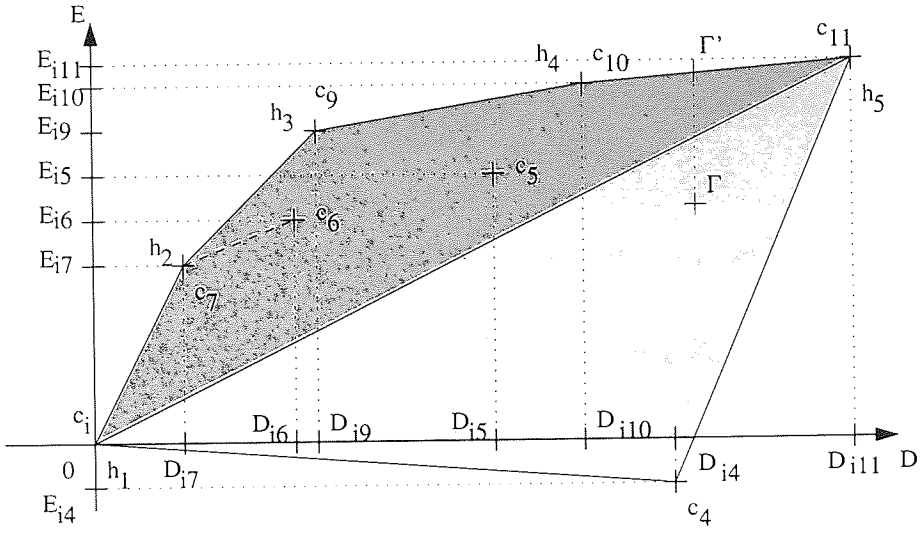


Figure 3.22: Graphical representation of stochastic mapping processes $F_{i\sim}$ and $\mathscr{F}^i_{\sim}$. Dark grey areas are associated with both types of mappings, while light grey is associated with $F_{i\sim}$ processes only.

This can be formally expressed as

$$\left(D_{i\sim} \geq \mathscr{D}^i_{1\overline{k}^i} \Rightarrow E_{i\sim} \leq \mathscr{E}^i_{1\overline{k}^i}\right) \vee \left(\exists!(k^i,\lambda) \in [1,\overline{k}^i] \times [0,1[, \right.$$
$$\left. \left(\lambda \mathscr{D}^i_{1,k^i-1} + (1-\lambda)\mathscr{D}^i_{1k^i} = D_{i\sim}\right) \wedge \left(E_{i\sim} \leq \lambda \mathscr{E}^i_{1,k^i-1} + (1-\lambda)\mathscr{E}^i_{1k^i}\right)\right), \quad (3.79)$$

with

$$D_{i\sim} = \sum_{j=1}^{n} p_{ij} D_{ij} \qquad \text{and} \qquad E_{i\sim} = \sum_{j=1}^{n} p_{ij} E_{ij}\,. \quad (3.80)$$

Note that the inequality on the good decoding probability is strict if

$$\exists j \in [1,n], (c_j \notin \{\mathscr{H}^i\}) \wedge (p_{ij} > 0)\,. \quad (3.81)$$

Proof Theorem 2: see Appendix A.

In short, Eq. 3.79 re-expressed the fact that if a codeword does not belong to $\mathscr{H}^i$ then a better codeword in terms of decoding with respect to the distortion introduced can be found.

From a geometric point of view (Fig. 3.22), it means that any codeword which is not in $\mathscr{H}^i$ is below the curve outlined by the elements of $\mathscr{H}^i$. And if Γ is an average of positively weighted codewords $S \in \mathscr{C}$, then either: 1) there is average Γ' of two positively weighted points $h^i_{k^i-1}$ and $h^i_{k^i}$ with $k^i \in [1,\overline{k}^i]$, such that the y-coordinate of Γ' is greater than the y-coordinate of Γ or 2) in the case where $\sum_{j=1}^{n} p_{ij} D_{ij} > \mathscr{D}^i_{1\overline{k}^i}$, then Γ is below $h^i_{\overline{k}^i}$. Note that this theorem is similar to Prop. 1, Eq. 3.74.

In Fig. 3.22, Γ and Γ' are the averages of (c_i, c_4, c_{11}) and of (h^i_4, h^i_5), respectively. Both have the same x-coordinate but the y-coordinate of Γ' is greater than the one of Γ.

Property 2 *For a given noise model and an embedding process f, then f satisfies the Prop. 2 if f can be expressed as a composition of stochastic mapping processes $\mathscr{F}^i_{\sim}$ with $i \in [1,n]$ such that each $\mathscr{F}^i_{\sim}$*

results in the same good decoding probability as a composition of two deterministic mapping processes $\mathscr{F}^i_{1,k^i-1}$ *and* $\mathscr{F}^i_{1,k^i}$ *with* $\forall i \in [1,n], k^i \in [1,\overline{k}^i]$. *Then, we can write*

$$f = \bigcirc_{i=1}^{n} \mathscr{F}^i_{\sim}\,, \tag{3.82}$$

and

$$\begin{aligned} \mathrm{P}(R \in \overline{\mathscr{C}} \,|\, S = f(C), M = 1) &= \sum_{i=1}^{n} p_i \left(p^i_{k^i-1} \mathscr{E}^i_{k^i-1} + p^i_{k^i} \mathscr{E}^i_{k^i} \right), &(3.83)\\ &= \sum_{i=1}^{n} p_i \left(\mathscr{E}^i_{k_i-1} + p^i_{1,k_i} \mathscr{D}^i_{k_i,k_i-1} \mathscr{G}^i_{k_i-1} \right), &(3.84) \end{aligned}$$

where the transition from Eq. 3.83 to Eq. 3.84 follows from the definition of $\mathscr{G}^i_{k_i-1}$ *given in Eq. 3.75.*

Theorem 3 *Given a noise model, if* f *is an optimal embedding process in the sense of the good decoding probability, then* f *satisfies Prop. 2.*

Proof Theorem 3: This theorem is given by Theorem 1 and then by successive applications of Theorem 2 to each $\mathscr{F}^i_{\sim}$ *with* $i \in [1,n]$. *For instance, if for a given* $i \in [1,n]$, *the property is not true, it contradicts the optimality of* f. *Therefore, we derive Theorem 3.* □

Property 3 *Given a noise model and an embedding process* f, *then* f *satisfies Prop. 3 if it satisfies Prop. 2 and*

$$\forall i \in [1,n], (p^i_{k^i} < 1) \Rightarrow \mathscr{G}^i_{k^i-1} = \mathscr{G}\,, \qquad \text{with} \qquad \mathscr{G} = \min_{j \in [1,n]} \mathscr{G}^j_{k^j-1}\,. \tag{3.85}$$

Theorem 4 *Given a noise model, if* f *is an optimal embedding process in the sense of the good decoding probability, then* f *satisfies Prop. 3.*

Proof Theorem 4: see Appendix A.

Property 4 *Given a noise model and an embedding process* f, *then* f *satisfies Prop. 4 if it satisfies Prop. 3 and*

$$\forall (i,k) \in [1,n] \times [1,\overline{k}^i],\ (k > k^i) \Rightarrow \left(\mathscr{G}^i_{k-1} \leq \mathscr{G}\right). \tag{3.86}$$

Theorem 5 *Given a noise model, if* f *is an optimal embedding process in the sense of the good decoding probability, then* f *satisfies Prop. 4.*

Proof Theorem 5: see Appendix A.

Theorem 6 *Given a noise model and an embedding process* f, *if* f *satisfies Prop. 4 and* $D = \sigma_X^2$, *then* f *is an optimal embedding process in the sense of the good decoding probability.*

Proof Theorem 6: see Appendix A.

3.4.3 Maximum Good Decoding Probability Algorithm

In this section, a constructive algorithm for an optimal embedding function in the sense of the good decoding probability is presented. The optimality of the obtained function follows from the properties stated and proved in the previous section.

The algorithm is iterative and starts from the identity embedding function which introduces no distortion. At each iteration, the current embedding function is optimised in the sense of the good decoding probability while introducing an additional squared error.

The notations used here have been defined in Sec. 3.4.1 and Sec. 3.4.2. Then, the proposed algorithm proceeds as follows:

1. First, the sequences $\mathscr{E}^i$ and $\mathscr{D}^i$ are constructed. The good decoding probability resulting from the identity embedding function is computed as $P = \mathrm{P}(R \in \overline{\mathscr{C}}|S = C, M = 1) = \sum_{i=1}^{n} p_i E_i$, the mean squared error introduced set as $D = 0$ and all the k^i with $i \in [1, n]$ are set to 0.

2. Then, a modification of the current embedding function is selected according to the greatest increase of the good decoding probability given by

$$\exists j \in [1, n], \mathscr{G} = \mathscr{G}^j_{k^j} = \max_{i \in [1,n]} \mathscr{G}^i_{k^i} \,. \tag{3.87}$$

   Hence, the conditions of optimality defined in Eq. 3.82 and Eq. 3.86 are enforced.

3. So, if

$$D + p_j \mathscr{D}^j_{k_j, k_j+1} \leq \sigma_X^2 \,, \tag{3.88}$$

   then D is increased by $p_j \mathscr{D}_{k_j,k_j+1}$, P by $p_j \mathscr{E}^j_{k^j,k^j+1}$ and k^j by one, otherwise in order to enforce the condition of optimality expressed in Eq. 3.85, we actually use the stochastic embedding for c_j defined as

$$p^j_{k_j} = 1 - \lambda \qquad \text{and} \qquad p^j_{k_j+1} = \lambda \,, \tag{3.89}$$

   with

$$\lambda = \frac{\sigma_X^2 - D}{p_j \mathscr{D}^j_{k^j,k^j+1}} \,, \tag{3.90}$$

   then D is set to σ_X^2, P is increased by $p_j \lambda \mathscr{E}^j_{k^j,k^j+1}$ and k^j by one.

4. Steps 2 to 3 are repeated as long as $D < \sigma_X^2$.

As the algorithm results in an embedding process f, which satisfies Prop. 4, we have from Theorem 6 that f is also optimal in the sense of the good decoding probability. Furthermore, as at each iteration D increases by at least

$$d = \min_{\substack{i \in [1,n] \\ p_i \neq 0}} p_i \frac{\Delta^2}{n^2} \,, \tag{3.91}$$

we have that the algorithm converges trivially in at most σ_X^2 / d iterations. For Gaussian noise, uniformly distributed host data and decoding to the nearest models, the resulting optimal embedding function is denoted DES, standing for *discrete embedding scheme*.

3.5 Numerical Studies

In this section, we investigate numerically the performance of the different schemes discussed in previous sections. First of all, the elements of comparison between the different techniques is laid out. Then, few embedding processes resulting from the proposed algorithm are demonstrated. Finally, a set of selected numerical simulations are presented and their results discussed.

3.5.1 Element of Comparison

For embedding techniques, two performance measures, the maximum information rate or the decoding error probability, measured both with respect to the signal to noise ratio (SNR) between the watermark and noise variances, are typically used. As the power constrained formulation of the problem has strongly related them to each other, it has become relevant to investigate the relation between these measures in order to evaluate the performance of an embedding scheme.

Definition 26 *The maximum information rate of a scheme $\overline{I}_{scheme}$ is defined as the maximum mutual information between the message M and the received data R over the parameters of the scheme and is given by*

$$\overline{I}_{scheme} = \max_{\alpha} I(R, M)\,, \tag{3.92}$$

where α symbolises the scheme parameters if any.

Definition 27 *The signal to noise ratio between the watermark and the noise, also termed watermark to noise ratio (WNR), is given by*

$$SNR = 10\log_{10}\left(\frac{\sigma_X^2}{\sigma_N^2}\right), \tag{3.93}$$

where σ_X^2 and σ_N^2 are the watermark and noise variances, respectively.

Definition 28 *The decoding error probability is defined as one minus the good decoding probability and is given by*

$$\mathrm{P}(\hat{M} \neq M|R) = 1 - \mathrm{P}(\hat{M} = M|R)\,. \tag{3.94}$$

3.5.2 Examples of Optimal Embedding Processes

As seen previously, the derivation of an optimal embedding process is done only for a given noise and host data models. Thus, from now on, we assume a uniformly distributed host data model and a centred Gaussian noise model. Furthermore, the decoding to the nearest grid point is also assumed.

As seen previously, the DC-QIM technique is a parameterised version of the QIM embedding technique. For a fixed watermark to noise ratio, there is an optimal bin size Δ resulting in the scalar Costa scheme. Thus, the QIM technique can be seen as a particular case of the DC-QIM embedding technique. Similarly, the approach we put forward here can be seen as an even more general case.

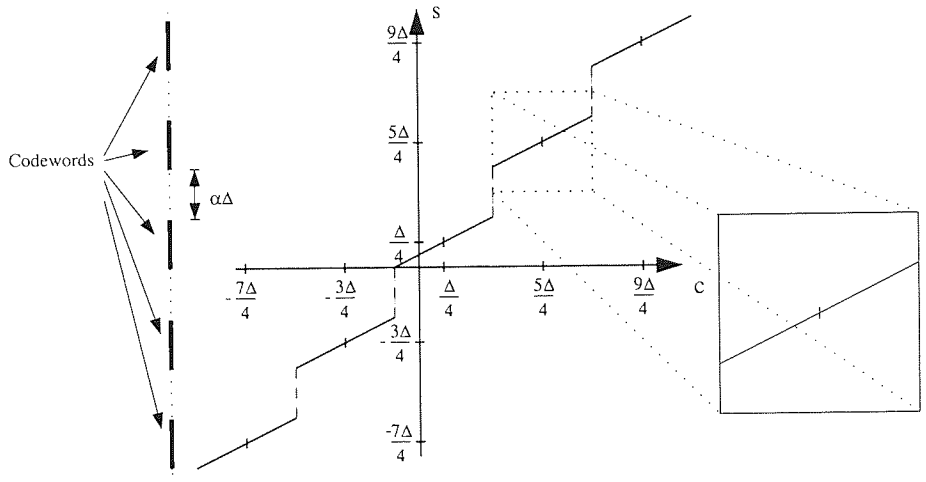


Figure 3.23: An example of DC-QIM embedding function for $M = 1$. As the following embedding processes are similarly regular, only the part framed by dotted lines shall be represented for them.

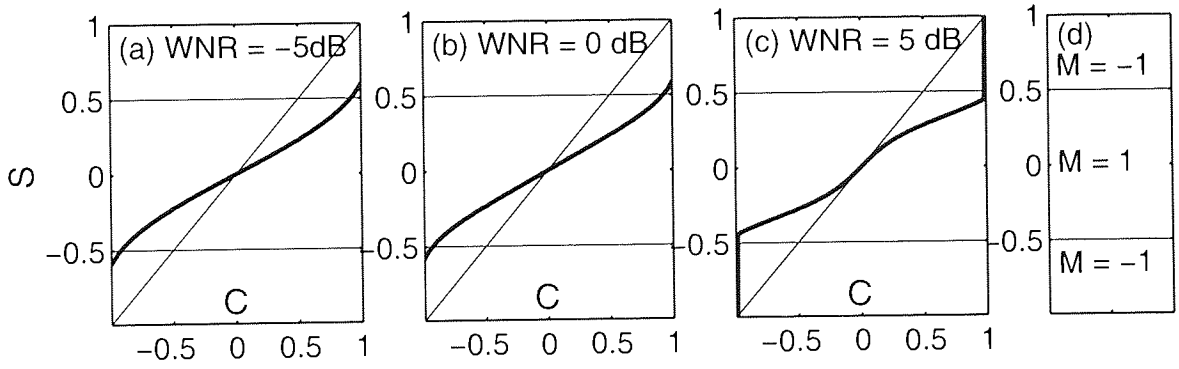


Figure 3.24: DES embedding for the message $M = 1$, the bin size $\Delta = 2$ and the watermark variance $\sigma_X^2 = 1/12$. The x-axis represents the host data C original value, while the y-axis corresponds to the watermarked data S. The plain thick line represents the embedding function while the identity function is given as reference by the plain thin line.

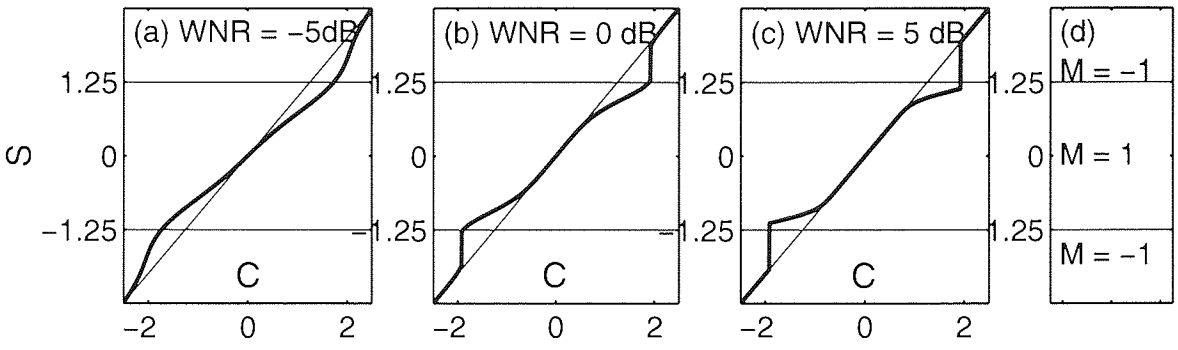


Figure 3.25: DES embedding for the message $M = 1$, the quantisation step $\Delta = 5$ and the watermark variance $\sigma_X^2 = 1/12$. The x-axis represents the original host data C, while the y-axis corresponds to the watermarked data S. The plain thick line represents the embedding function while the identity function is given as reference by the plain thin line.

Indeed, the embedding process derived by our algorithm takes into account not only the bin size Δ but also the noise probability density function and the watermark to noise ratio.

For the DC-QIM scheme, the quantisation step Δ and the watermark variance σ_X^2 are sufficient to determine completely the resulting embedding function as in Fig. 3.23. As all the following presented embedding processes have the same regular character as the one in Fig. 3.23, we shall plot only a reduced part as framed in Fig. 3.23. In the case of our approach, for a given quantisation step Δ and watermark variance σ_X^2, there is still a very large number of possible embedding processes. The optimal embedding process will be completely determined only given the noise model, hence the watermark to noise ratio in our examples as in Fig. 3.24.

In Fig. 3.24, few optimal embedding functions derived with our algorithm are represented for the same power constraint and bin size, but for different watermark to noise ratio values. In Fig. 3.24d, the decoding areas for -1 and 1 are indicated. Although Fig. 3.24a and Fig. 3.24b look pretty similar, we would like to point out that the embedding processes derived are generally different and depend on the watermark to noise ratio.

For instance, in Fig. 3.24c, one notices that for a fairly high watermark to noise ratio of 5dB, the host data close to the centre of the bin is barely modified during the embedding process unlike the ones at the edges. Data close to the centre have little chance to be badly decoded in presence of weak Gaussian noise, thus the distortion allowance is mainly used at the edges.

On the contrary, in Fig 3.24a and Fig. 3.24b, the distortion is more evenly spread throughout the bin. For some data which are originally located at the edges, in the case of a noiseless transmission, the decoding would surprisingly result in decoding error. Such behaviour was also reported for DC-QIM in Fig. 3.18 and is not so awkward. Indeed, although the decoding for a noiseless transmission results in decoding error, the embedding reduces nonetheless the decoding error probability under the assumption of a Gaussian noise model.

In Fig. 3.25, the bin size has been increased compared to Fig. 3.24, from $\Delta = 2$ to $\Delta = 5$. Note that to achieve the QIM embedding method, a watermark distortion of $\sigma_X^2 = 4/12$ and $\sigma_X^2 = 25/12$

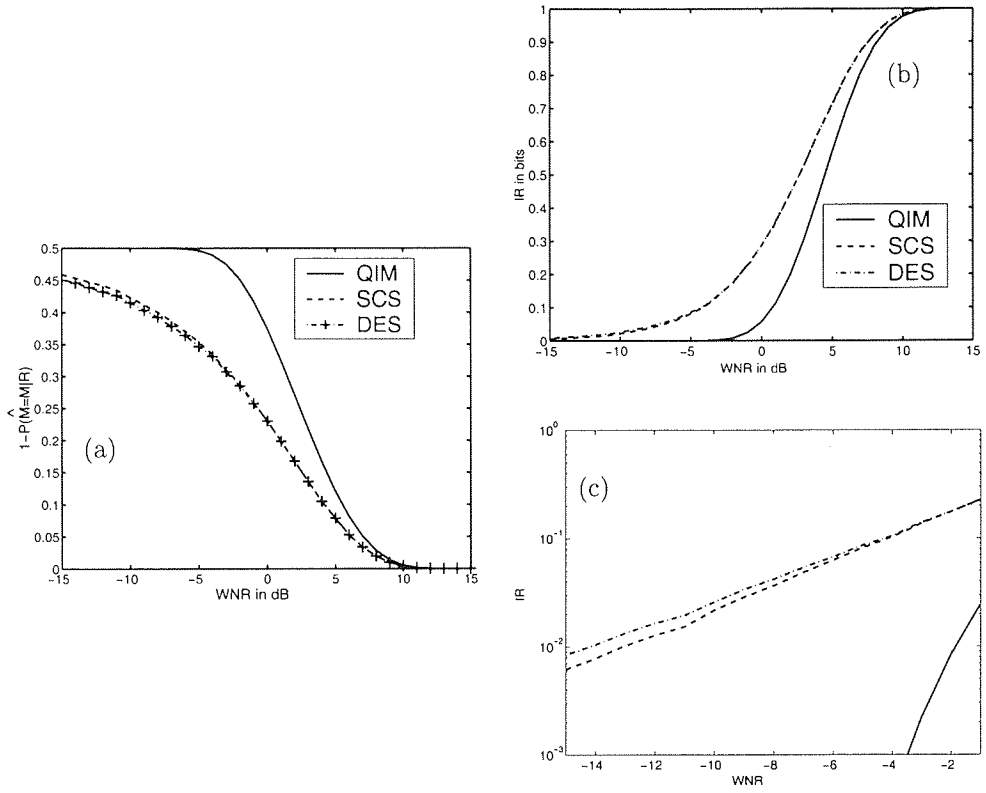


Figure 3.26: Decoding error probability and information rate (IR) against the watermark to noise ratio (WNR) for QIM, SCS and DES embedding techniques. Plot (c) corresponds to a zoomed part of plot (b) using a log scale.

would be required respectively, while here it is $\sigma_X^2 = 1/12$. Clearly, the watermark power constraint does not enable all the data to be correctly decoded for a noiseless transmission.

Then, as it can be seen from the plots, the embedding function distortion is mainly concentrated at the borders of the decoding areas from which the best improvement in terms of decoding probability are obtained. This concentration is due to the bell shape of the Gaussian noise probability density function. As it can be noticed, the distortion density along the decoding edges is smoother in Fig. 3.25a than in Fig. 3.25b and Fig. 3.25c. This can be interpreted as a direct consequence of the larger standard deviation of the noise and therefore of its smoother bell shape.

Hence, the optimal embedding process does not depend solely on the noise standard deviation, but on all characteristics of the noise probability density which are taken into account in the proposed algorithm.

3.5.3 Simulation Results

In this section, we compare the performance of the QIM, DC-QIM, SCS and DES embedding techniques for a centred Gaussian noise model with a decoding to the nearest for various parameter values.

As it can be seen from Fig. 3.26, DES technique shows better performance than QIM and SCS techniques for the whole range of watermark to noise ratios studied. However, the difference is relatively small especially for positive watermark to noise ratio. Nonetheless, the DES scheme provides

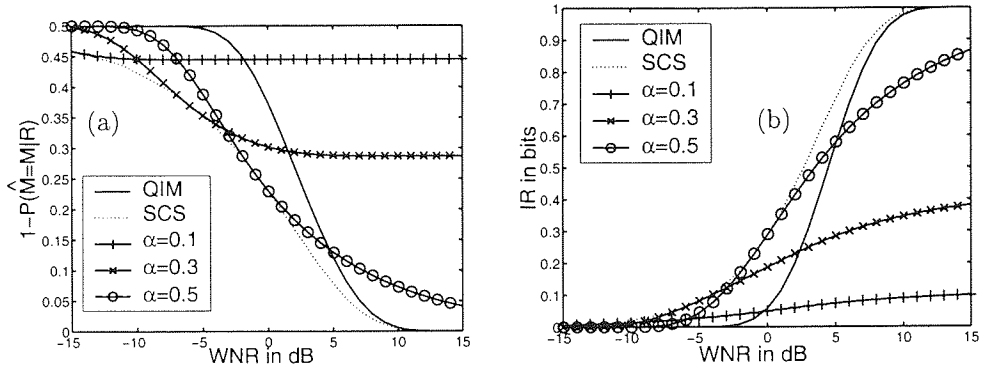


Figure 3.27: Decoding error probability and information rate (IR) against the watermark to noise ratio (WNR) for the DC-QIM embedding technique for various values of α. The dotted line represents the SCS technique limit performance.

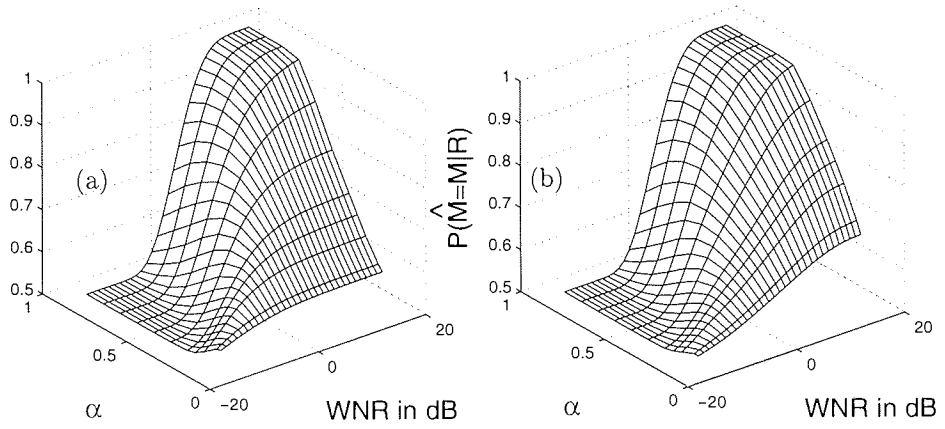


Figure 3.28: Good decoding probability against the watermark to noise ratio (WNR) for the DES embedding technique with a fixed parameter α and expected watermark to noise ratio. On plot (a), the WNR expected is -10 dB, while on plot (b), it is 10 dB.

a significant improvement in terms of information rate for negative watermark to noise ratio (in dB) as shown in Fig. 3.26c. Indeed, on the range of watermark to noise ratio studied, it may represent up to 40 percent of increase in information rate compared to the SCS scheme for a watermark to noise ratio of -15dB as depicted in Fig. 3.26c.

However, both SCS and DES schemes cannot be realistically considered as practical schemes. Indeed, both are optimised with respect to the watermark to noise ratio which might not be known in practice. Furthermore, they do not perform so well if the actual watermark to noise ratio is different from the expected one, hence, practical applications have to rely on schemes like DC-QIM or DES techniques with fixed parameters.

As it can be seen from Fig. 3.27, while decreasing the parameter α, the DC-QIM scheme performances are increasing for negative watermark to noise ratio (in dB). However, this improvement is achieved at the expense of the performance at positive watermark to noise ratio. Similar behaviour can be observed for the decoding error probability of the DES scheme in Fig. 3.28 and for the information rate in Fig. 3.29.

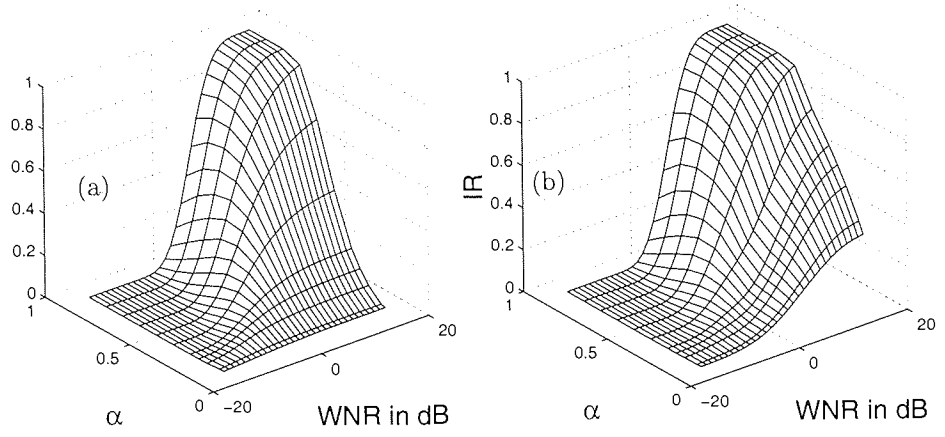


Figure 3.29: Information rate (IR) against the watermark to noise ratio (WNR) for the DES embedding technique with a fixed parameter α and expected watermark to noise ratio. On plot (a), the WNR expected is -10dB, while on plot (b), it is 10dB.

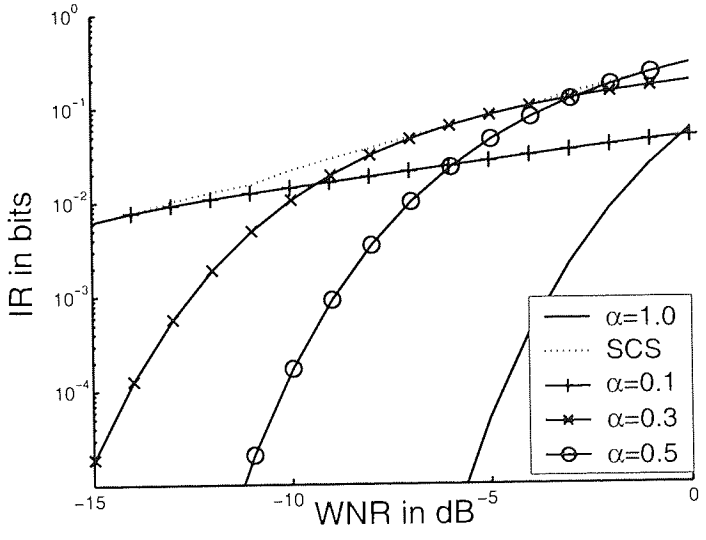


Figure 3.30: Log scaled representation of the information rate (IR) against the watermark to noise ratio (WNR) for DC-QIM embedding technique for various values of α. The dotted line represents the SCS technique limit performance.

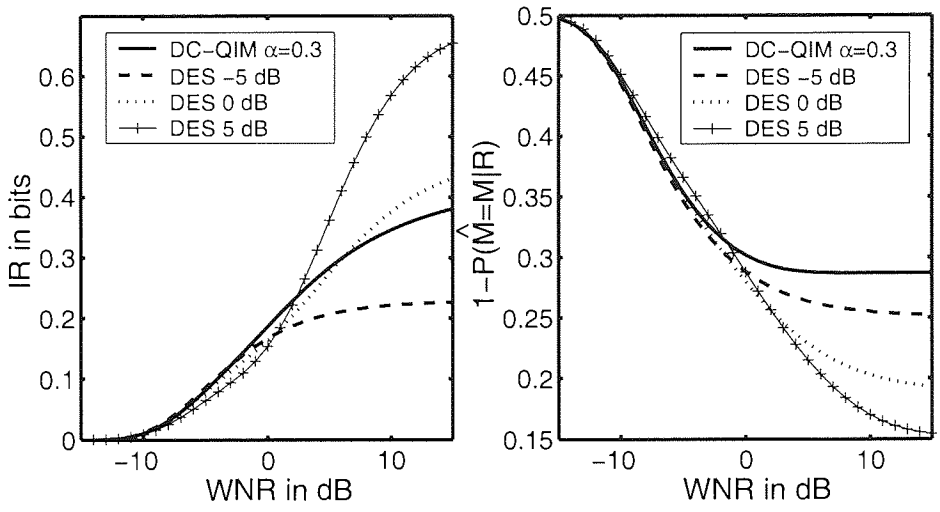


Figure 3.31: Comparison of the DC-QIM and DES schemes information rates (IR) and decoding error probabilities against the watermark to noise ratio (WNR) for $\alpha = 0.3$ and various value of expected WNR (for DES only).

Thus, for the DC-QIM scheme, small values of α are necessary to achieve 'reasonable' decoding error probability and information rate at negative watermark to noise ratio (in dB), as demonstrated in Fig. 3.27 and Fig. 3.30, but degrade significantly the performance at positive watermark to noise ratio (in dB). For instance, an α value of 0.3 is close to optimal for a watermark to noise ratio between -7 dB and -3 dB, but results in a decoding error probability close to 0.3 at 15 dB. Similarly, for watermark to noise ratios below -10 dB, an α value of 0.1 provides good information rates, but is then upper bounded by an information rate of 0.1 bit/element for the whole range of watermark to noise ratio studied.

However, in the case of the DES scheme, the problem encountered by the DC-QIM technique is slightly alleviated by the fact that for a given bin size, several DES embedding processes are possible. Indeed, for each bin size Δ, the embedding process can be optimised with respect to the expected watermark to noise ratio, choosing the bin size is not everything as it is for the DC-QIM scheme.

Thus, as shown in Fig. 3.31, for an α value of 0.3, the DES scheme enables the decoding error probability to be as low as 0.15 for a watermark to noise ratio of 15 dB, while the decoding error probability at negative watermark to noise ratio (in dB) are comparable to the DC-QIM scheme performance. These results are obtained when the DES scheme is optimised with an expected watermark to noise ratio between 0 dB and 5 dB as demonstrated in Fig. 3.31. Thus, using a large bin size Δ provides the robustness at negative watermark to noise ratio (in dB), while the optimisation enables a low decoding error probability at positive watermark to noise ratio (in dB). Similar discrepancy can be observed for larger bin size Δ as in Fig. 3.32.

Hence, although the performances of the DES scheme are significantly better only on a limited range of watermark to noise ratio than the ones achieved by the SCS scheme, the DES technique with fixed parameters is in practice superior to DC-QIM technique for the whole range of watermark to

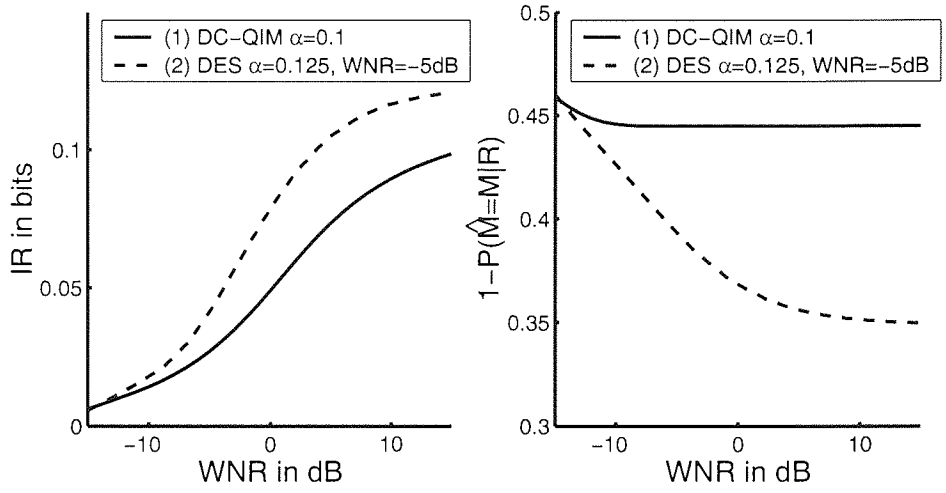


Figure 3.32: Comparison of the DC-QIM and DES schemes information rate (IR) and decoding error probability against the watermark to noise ratio (WNR).

noise ratio studied.

3.6 Contributions Summary and Future Research

In this chapter, we proposed an information embedding framework for digital data. Our approach is based on a regular lattice structured codebook and on trying to achieve low decoding error probability. Both are suitable for practical schemes and are superior to random codebooks, which are theoretically optimal, but also impractical.

Analysing different embedding schemes, as well as our optimised embedding technique, reveals that for a lattice based embedding processes, the lattice characteristic size Δ has a predominant role for low watermark to noise ratio. However, we have also pointed out that efficient practical schemes cannot solely rely on this fact.

Indeed, the embedding process has also to take into account the noise charateristics, such as its standard deviation or its shape. The algorithm proposed in this work takes all the characteristics of the noise unlike previously proposed embedding techniques.

Thus, the embedding processes derived by the proposed optimisation algorithm outperformed all previously proposed embedding schemes for a wide range of watermark to noise ratio. Furthermore, the algorithm is quite general and can be applied easily to various noise and decoding models. Moreover, as stated previously and demonstrated below, it can be used for higher dimensionality data and messages taken from larger alphabet.

For instance, let us assume a two data dimensional space with an hexagonal lattice and three different messages as depicted in Fig. 3.33. As shown in the figure, the working space is now a triangle instead of a line segment, and the data points can be labelled from c_1 to c_n. The same algorithm as presented in Sec. 3.4.3 is applicable in this case.

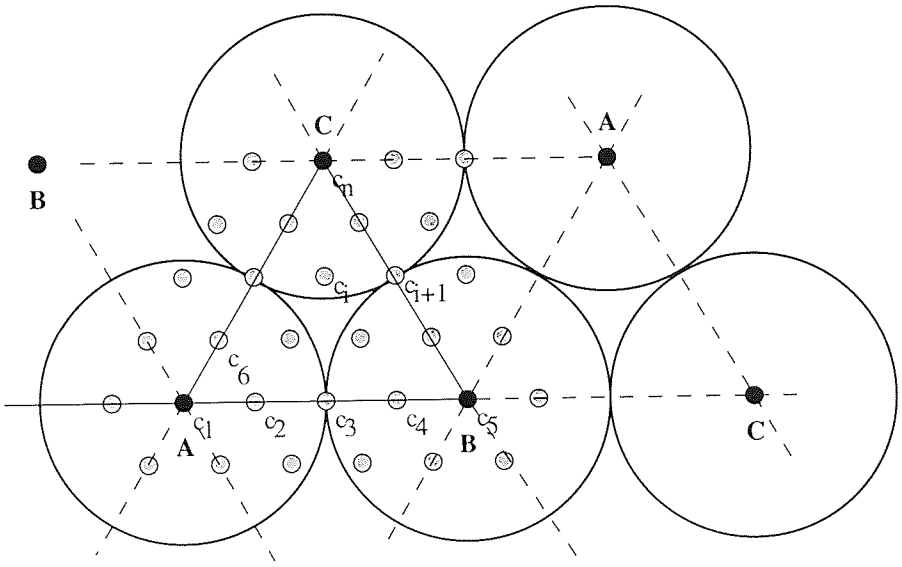


Figure 3.33: Two dimensional data space with an hexagonal lattice and three different messages denoted 'A', 'B' and 'C'. The lattice points are marked with black dots, while normal data points labelled c_i are marked with grey dots.

Chapter 4

Image Representation in Practical Watermarking Schemes

This chapter investigates image data representation effects on the performance of digital watermarking schemes. The most commonly used representations are first reviewed. Subsequently, a new and principled representation based on independent component analysis (ICA) is proposed for digital image watermarking purposes [Bounkong, Toch, Saad, and Lowe 2003a]. Then, the influence of image data representation on the robustness of watermarking schemes is briefly discussed. Finally, two visual perceptual distortion measures are presented and used to compare two image data representations based on block transforms.

4.1 Introduction

A digital image C in its native representation can be seen as a $h \times w$ matrix. Typically, the histogram of image pixel values vary significantly from image to image. Moreover, to the best of our knowledge, for this representation, no statistical model has been devised to characterise image data. Thus, no realistic assumption about the data can actually be taken.

However, some well-known characteristics of image data and of the human visual system, such as the strong local[1] correlation between pixels values and the varying sensitivity to frequency, provide strong incentives to the use of a frequency representation. Thus, many image processing methods, such as lossy compression techniques [Wallace 1992; JPEG2000 Final committee 2000], use a frequency representation for image data. The most commonly used are the discrete Fourier transform (DFT), discrete cosine transform (DCT) and the discrete wavelet transform (DWT).

These allow to represent a typical image in a relatively small number of coefficients (respectively, low frequency [Jain 1989] and coarse level coefficients). Such representations simplify significantly the processing of image data and are adopted by most of the signal processing techniques.

[1] in the spatial sense

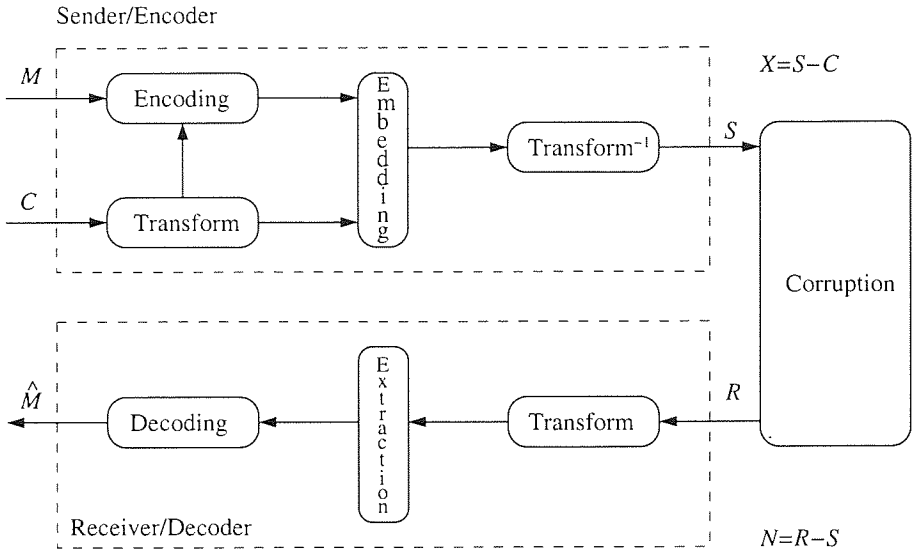


Figure 4.1: Watermarking framework block diagram.

Most practical image watermarking schemes, as described in Fig. 4.1, also use a transform domain to represent the image data C. Following the selection of preferred embedding sites among the transform domain coefficients, the watermark information M is embedded in the selected coefficients. Thus, given an embedding process, the domain transform and coefficient selection process may have significant effects on the watermark perceptibility or on the scheme decoding error probability. These effects are the focus of this chapter.

4.2 Image Data Representation

In the context of watermarking, many representations have been studied extensively, such as the cosine transform representation and the Fourier transform representation. More recently, applying the wavelet transform to the watermarking problem has been particularly popular in the watermarking literature.

A generally acknowledged reason for using a frequency representation is that it enables the spatial spreading of the watermark information M over the whole watermarked data S, although only few coefficients are actually modified. Moreover, such a modification, for well chosen frequencies, is typically smooth and does not introduce sharp or perceptible artefacts. Thus, a particular representation may be chosen because of the reduced perceptibility of the modification introduced.

Another common motivation for image transform is found in the simplification they may provide in deriving robust schemes against standard lossy image compression standards. Indeed, as lossy compression distortions are introduced by the quantisation of transform coefficients, embedding information in the same transformed space where the distortion is known may be more efficient as discussed in Chapter 2. Some transforms, such as a combination of Fourier and Mellin-Fourier transforms, may

also provide inherently some resistance against affine geometrical transformation as presented in [Ruanaidh and Pun 1997].

Besides these motivations, one may argue that since any of the commonly used transforms (i.e. Fourier, cosine and wavelet transform) are able to represent any image data C or watermarked data S, a watermark X obtained by some modifications in a transform space (e.g. Fourier) can as well in theory be obtained by some other modifications in another transform space (e.g. wavelet).

Thus, the actual motivations to use a transform space has to be found elsewhere. One of them is related to practical considerations. Indeed, the modification of one Fourier transform coefficient may correspond to the modification of countless wavelet coefficients and conversely. Local[2] modification may also be obtained more easily with block or wavelet transforms rather than with full frame transforms. The transform domain can also be used as the secret key as in [Bounkong, Toch, Saad, and Lowe 2003a].

Thus, the choice of a specific transform may significantly vary depending on the final purpose of the application or the designated role for the transform space.

4.2.1 Examples

DCT: was widely studied in the context of MPEG [ISO/IEC JTC1/SC29/WG11 1988] and JPEG [Wallace 1992] compression. Hence, this representation was naturally considered for digital watermarking schemes. Existing results on the perceptibility of modifications carried out in previous studies [Watson 1993] could also be reused, since imperceptibility of the watermark is a central requirement to the digital watermarking framework.

Two variations are predominant in the watermarking field: the full size DCT and the 8×8 block DCT. The first one provides a very compact representation of the image pixel value variance and the spreading the watermark information M over the whole[3] image. This characteristic, common to all full frame transforms, typically enables schemes to be fairly robust against localised corruption. However, this approach does not allow to take advantage of local information such as the presence of edges or smooth areas.

The use of the 8×8 block DCT for digital watermarking finds one of its main motivation in the JPEG standard, which uses this data representation. The consequences of the JPEG compression are fully known and can be taken into account for the embedding process [Bounkong, Saad, and Lowe 2002b]. Furthermore, the previously described characteristics of the full frame DCT are preserved to a reduced scale. However, as a drawback, this approach introduces blocking artefacts due to the image partition and the independent processing typical to each image block. Moreover, derived schemes tend to be more sensitive to local distortions, although this can be alleviated by the use of an appropriate error correcting code.

[2] in the spatial sense.
[3] in the spatial sense

DFT: is widely studied in signal processing and has been used since the beginning of digital watermarking. It is seen as a natural way to decompose the image in a frequency representation. Moreover, it provides some desirable properties such as the linear shift invariance of the frequency response. Besides, if one additionnaly uses a log polar mapping, it is possible to obtain a space insensitive to translation, rotation and scaling operations as described in [Ruanaidh and Pun 1997].

As a global transform, the DFT possesses the same benefits as the DCT: spatial spreading of the message information, robustness to local distortion and good variance compaction [Jain 1989].

DWT: was introduced after the two previous transformed spaces, but has also attracted a lot of interests in the past few years, as testified by the number of publications [Meerwald 2001a] related to its use for digital watermarking.

One of the main reason of such interest is arguably related to the new image lossy compression standard JPEG2000 [JPEG2000 Final committee 2000]. Hence, as for the 8×8 block DCT, previous studies results regarding the visual effects of wavelet coefficients modifications can be re-used.

Moreover, this data representation replaces the classical concepts of frequency and phase by the notions of multi-resolution and location, providing an elegant framework to image data processing. Thus, local information can be taken into account unlike the case of global transforms. Moreover, unlike the 8×8 block DCT, the DWT does not indirectly induce blocking artefacts but ringed artefacts which are arguably less perceptible.

4.2.2 Motivations for ICA based Watermarking

In this thesis, an ICA based representation of image data is suggested for digital watermarking applications. In this section, some motivations are given to this approach.

Applied to the watermarking problem, an ICA based representation has several benefits.

1. As an adaptive technique, it may provide a better[4] representation of image data as argued in [Field 1994; Bell and Sejnowski 1997; Simoncelli and Olshausen 2001].

2. The mixing and demixing matrices used are not only unknown to an outsider but also difficult to estimate as discussed in [Toch, Lowe, and Saad 2003].

3. Since the coefficients resulting from the ICA decomposition are independent, the values of a set of coefficients cannot be used to infer the value of one another, hence, a modification applied to one of them cannot be found using the others.

4. It facilitates the use of Bayesian decoding techniques based on statistical models that can be constructed due to the simple factorised statistics of the sources; principled Bayesian techniques

[4] in the sense that less coefficients are required to represent the image data.

are expected to improve the decoding performance in real systems. The topic is not investigated here, and is deferred for future research. However, some preliminary results can be found in [Bounkong, Toch, Saad, and Lowe 2003a].

5. Finally, the main justification for using an ICA based representation is coming from theoretical studies reported in the information theory literature. Assuming that the source (host data) is blockwise memoryless known to the decoder and the attack is also blockwise memoryless, the capacity of a watermarking scheme for constrained watermark and noise distortions is maximised when the block components are statistically independent [Moulin and O'Sullivan 2003]. Note that this property assumes that the sum over marginal entropies of all channels is the same whether the channels are statistically independent or not. Since, in general, only the marginal entropies can be accurately evaluated, an ICA based representation offers a significant advantage in the design of an optimal practical scheme.

4.2.3 Introduction to Independent Component Analysis

Independent component analysis (ICA) was introduced several years ago as a blind source separation technique, but since then has been used in a broad range of applications [Hyvarinen, Karhunen, and Oja 2001]. The ICA decomposition represents a given set of signals V as a mixture of statistically independent sources U. Although no prior constraint exists on the nature of the mixture, researches have mainly focused on the linear case [Hyvarinen 1999], which is arguably simpler and makes the interpretation of results easier. From an information theoretic point of view, the independence is usually a desired property. It allows efficient encoding, since no cross channel redundancy is present. Any distortion affecting one channel does not affect the others.

Now, we shall introduce the ICA problem. Let us denote by U an n-dimensional vector of independent random variables. This vector components are mixed linearly as in Fig 4.2 to an m-dimensional vector V. If the mixture coefficients a_{ij} are represented in a matrix form $A = (a_{ij})$, also termed the mixing matrix, the previous mixture can be expressed as

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}. \tag{4.1}$$

Thus, given a set of vectors V termed observed or sensored data, the problem is to find the corresponding set of source vectors or underlying data U and the mixing matrix A. This is actually carried out by finding a matrix W called the demixing matrix such that the estimate set of vectors U is maximising some measure of independence.

Some ambiguities arise from this definition. First, one can notice that the components of the vector U are defined only up to a scaling factor and permutation. Typically, the row of the demixing

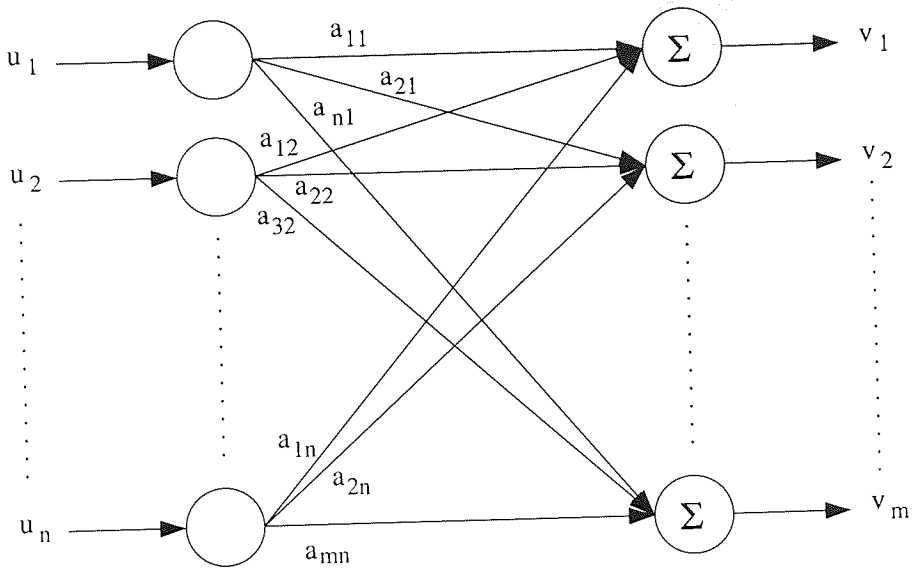


Figure 4.2: Independent component analysis framework.

matrix W are arbitrarily scaled such that each component of the vector U has a unitary variance. In the present work, such ambiguities have no consequence.

Another problem concerns the real number of sources n. Unlike for synthetic data, for real data such as MEG or EEG to which ICA has been applied to with some success, n is usually unknown. In some cases, n is often chosen according to what is usually accepted in the research field. Such problem also arise for digital images and will be discussed in the next section.

In order to estimate the independent components, different methods are available using various measures of independence such as the sum of cumulants, the non-gaussianity or the maximum likelihood. But all have been shown to be somehow equivalent [Lee, Girolami, Bell, and Sejnowski 2000]. For our studies, we use a few of the methods [Hyvarinen, Karhunen, and Oja 2001; Blaschke and Wiskott 2002] which have shown comparable results. Further details about ICA techniques can be found in [Hyvarinen, Karhunen, and Oja 2001; Lee, Girolami, Bell, and Sejnowski 2000].

4.2.4 ICA for Image Watermarking

The problem of encoding image information efficiently is relevant both to data transmission and to signal processing. Within the image processing community, much work has been done on image coding that uses a linear expansion [Jain 1989]. Typically, the bases are chosen for their low-entropy coding properties, or low computational cost. Rather than being adapted to the data, they are hand designed or derived from a mathematical model. Adaptive, linear transforms such as principal component analysis (PCA) or ICA have also been studied [Lewicki and Olshausen 1999] for images although they received less attention.

Adaptive decomposition such as PCA relies only on the data second order statistics, which are

Figure 4.3: An example of natural scene image.

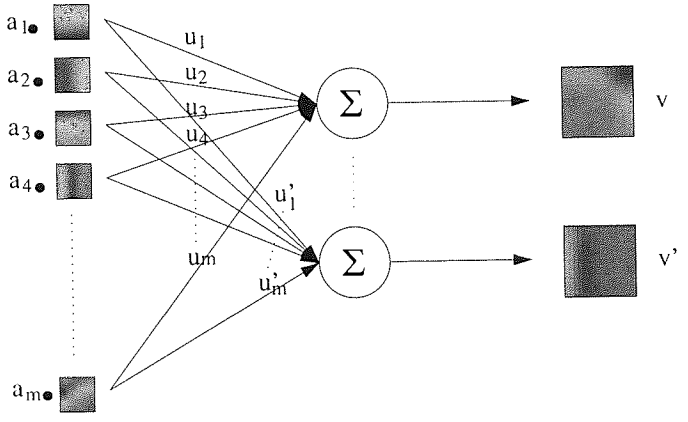


Figure 4.4: ICA for images.

gathered in the correlation matrix. However, it has been argued in [Field 1994; Simoncelli and Olshausen 2001] that most of the information in natural images (Fig. 4.3) is actually in higher-order statistics. It has also been suggested that images generally have "independent components" of sparse distributions [Field 1994; Bell and Sejnowski 1997]. Using the ICA approach, which is tightly related to sparse coding [Hoyer 1999], may give a better representation of the image data.

In the context of digital images and visual data, each image sample V can be considered as a superposition of basis vectors weighted by the vector U as depicted in Fig 4.4. The basis vectors, which correspond to the columns of A, can typically be interpreted as edges and bars composing the

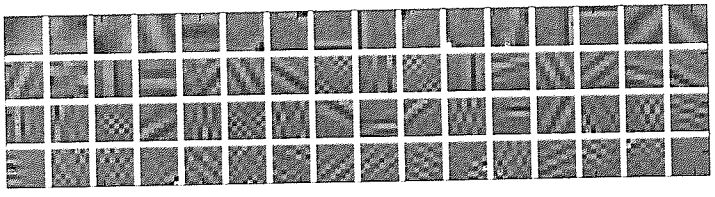


Figure 4.5: An example of ICA basis obtained from natural scene images.

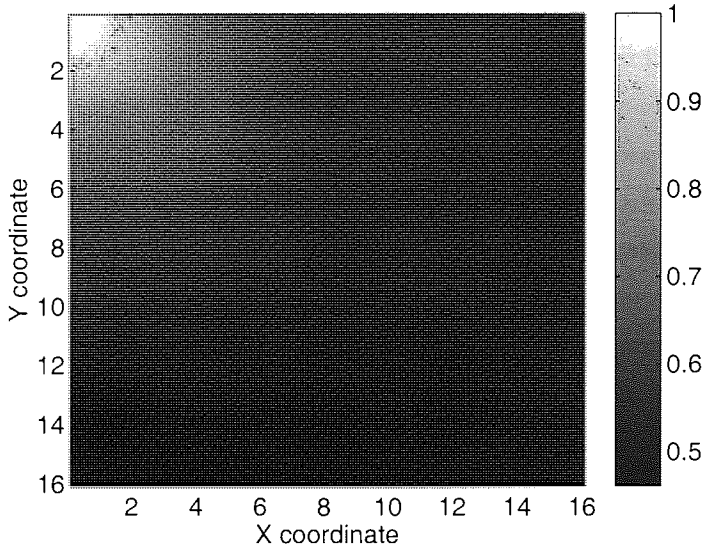


Figure 4.6: Normalised correlation between the top left pixel and the others.

image. As it can be seen in Fig. 4.5, they are both localised spatially and in frequency like Gabor filters [H.G. Feichtinger 1998]. As they represent more or less the same patterns at different scale, they can also be related to a wavelet decomposition. However, ICA bases possess more varied orientation and phase than typical wavelet bases.

As for PCA, ICA mixing and demixing matrices need to be estimated from some training data. Such estimation cannot usually be performed on a set of full size pictures as it would be too demanding computationally. Thus, a commonly accepted heuristic consists in taking small window samples of the full size pictures in order to compute the matrices. Such an approach can be partially justified by the fact that pixels usually have only short range correlation as it can be seen in Fig. 4.6.

From a practical point of view, the choice of n has two significant consequences. The first one is purely physical and is related to the practical estimation of the ICA mixing and demixing matrices. As n increases, the time and computational power required also increase and may make the matrix estimation infeasible. But, n is also tightly related to the spatial spreading of the information, the robustness of the embedding and the watermark perceptibility. Thus, a trade off between feasibility and performance has to be found.

Another important aspect related to the estimation of the ICA mixing and demixing matrices concerns the reduction of the data dimension using PCA. The PCA technique can be used to project the data onto a subspace, for example, of the most important eigenvectors. Such a reduction has the advantage of speeding up the basis vectors estimation, enabling the use of bigger windows. However, since the number m of latent sources u is unknown, it can be argued that such process may discard significant information. The evaluation of n is still an open problem, especially for real data like sound or images.

Here, we choose arbitrarily to use 8×8 block samples for our analysis. Hence, we have $n = 64$, but

another choice could be possible and provide better performance in practice. However, this relatively small number allows us first of all to have $m = n$. Besides, $n = 64$ matches the image block size used by the JPEG standard [Wallace 1992] and Watson's perceptual distortion measure [Watson 1993], which will be employed later. Thus, the analysis of the impact of JPEG lossy compression and of the perceptual distortion introduced by the watermark can be carried out in a more principled way.

Then, we propose to pre-process our sensored data, represented by image block samples, before the evaluation of the ICA matrices. For instance, for each image block sample, we suggest to remove its average luminance and to normalise its Euclidean norm. One of our main motivations is to give each sample the same weight. However, such practice may amplify the data noise, when the image sample is almost uniform, hence a selection of the data samples should be carried out in order to discard such data. As a result of the pre-processing, ICA operates on data samples which are on the unitary hypersphere.

Once ICA mixing and demixing matrices are obtained, they can be used to transform the data before the embedding stage as described in Sec. 4.1 and depicted in Fig. 4.1.

4.3 Image Data Representation and Noise Corruption

In this section, noise corruption effects are briefly discussed with respect to the image data representation. The aim is here to clarify in what context image transform domain can be helpful and to which extent. The noise processes discussed are among the most commonly studied in the digital image watermarking field and comprise random noise, quantisation noise, cropping and affine transforms.

4.3.1 Random Noise

Random noise effects can be represented as a cloud of possible received data R around the data actually sent S as depicted in a two dimensional representation in Fig. 4.7. The sent data S and the noisy data R are represented by '+' and '.', respectively. The noise model chosen for the examples is a centred Gaussian.

In Fig. 4.7a, the noise N does not privilege any direction, thus no image data representation can help in the design of a more robust scheme, which has therefore to rely only on the robustness of the embedding and decoding processes. For instance, the white Gaussian noise, which is commonly used in watermarking studies, is isotropic. Thus, watermarking schemes, which differ only in the image data representation used, should result in a similar decoding performance against this noise. Typical embedding process performance has been investigated and discussed in Chapter 3.

In Fig. 4.7c, the noise N is independent from the watermarked data S, but shows two privileged directions: the first one, denoted by dashed lines, corresponds to the maximum noise variance, while the second direction which is orthogonal to the dashed lines corresponds to the minimum noise variance. These directions can be used for the image data representation and help to design a robust scheme. For instance, the direction corresponding to the minimum noise variance can be used to convey the

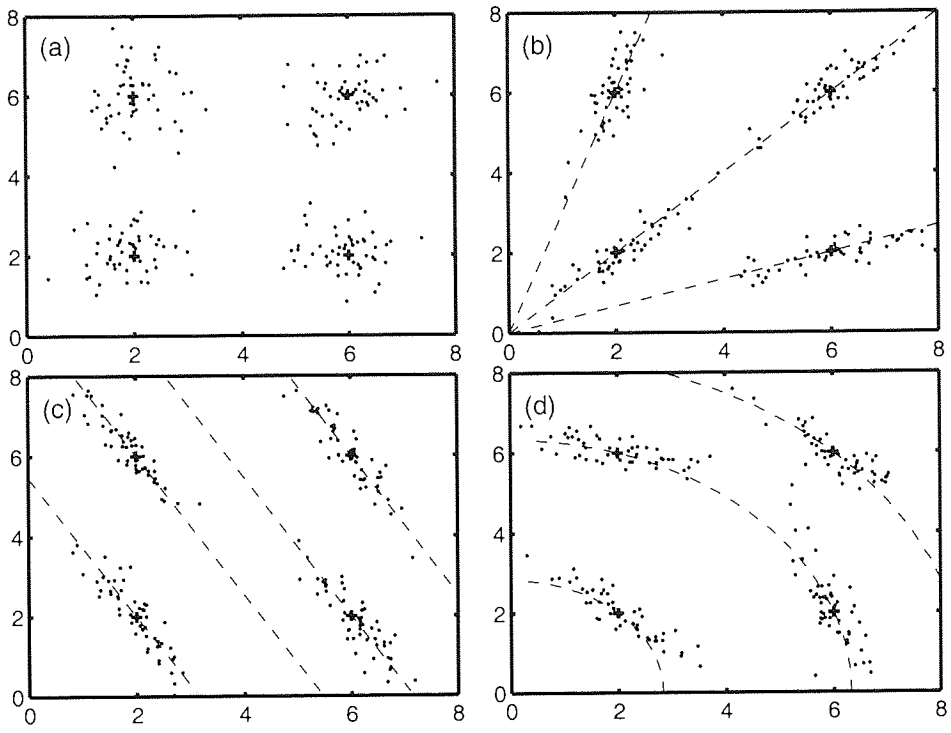


Figure 4.7: Examples of random noise. The watermarked data is represented by '+', while the received data is represented by '.'. Dashed lines represent the noise maximum variance directions.

watermark information.

In Fig. 4.7b and Fig. 4.7d, the noise characteristics, here the covariance matrix, depend on the watermarked data S, thus finding the exact relation between the noise N and the watermarked data S requires finding an appropriate representation for the image data C. In these examples, the noise variance is either maximal (b) or minimal (d) along the radial vector, thus polar coordinates system may be here very helpful to devise a robust scheme.

As seen in these examples, the knowledge of the noise characteristics is essential to the choice of a useful image data representation with respect to the noise robustness. As the noises sources vary and the data is watermarked only once, it may be useful to use several embedding directions or decide to make the watermark more robust only against a certain type of noise. Noise estimation, which is not investigated here, may also help greatly in improving the scheme performance, which is one of the possible future research directions.

4.3.2 Distortion Due to Quantisation

Distortion due to quantisation as typically encountered in image processing is depicted in the two dimensional representation of Fig. 4.8. Quantisation effects are represented by small arrows, quantisation points corresponding to the received data R by black dots. Each quantisation bin is delimited by dashed dotted lines. The original data axes P_1 and P_2 are represented by plain lines, while the quantisation axes Q_1 and Q_2 are represented by dashed and dash-dotted lines.

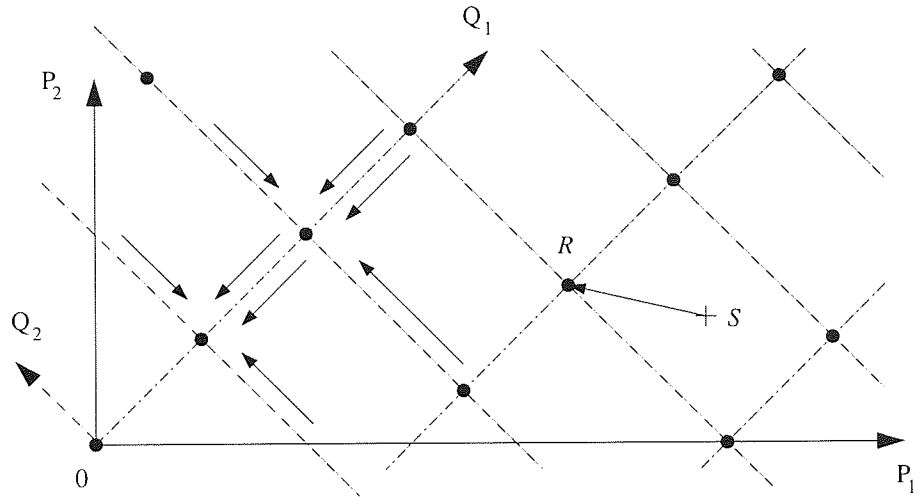


Figure 4.8: An example of data quantisation.

In image processing, such as in JPEG lossy compression, quantisation is often done towards the zero of the quantisation axis as is Fig. 4.8. In addition, all watermarked data S within the same quantisation bin are mapped to a single point R. Thus, using the quantisation axes Q_1 and Q_2 for the image data representation helps in coping with this effective corruption. For instance, since the quantisation step is smaller along the Q_1 axis, this axis can be used to embed the watermark information resulting in a lower level of distortion. Further details on quantisation effects on digital watermarking can be found in Chapter 2.

4.3.3 Image Cropping

Corruption by cropping is a synchronisation attack and can be dealt with by carrying out exhaustive searches based on frequency analysis, feature detection or on a previously embedded synchronisation pattern. Such search may be computationally demanding and hence difficult to carry out online, but is nonetheless necessary to synchronise the received data R with the extractor.

Once the cropped data has been aligned, cropping can be considered from two points of view depending on the embedding and data representation chosen. In a global transform domain the cropping affects all transform coefficients to some extent. In this case, the consequences have to be examined case by case depending on the coefficients in which the watermark information was embedded.

For a local transform domain, such as a block transform, cropping can be seen as corruption by an erasure channel as defined in [Cover and Thomas 1991] and depicted for a binary channel in Fig. 4.9. Indeed, the remaining data is typically unmodified and hence the problem can be dealt using error correcting code and in particular erasure correcting codes [Luby, Mitzenmachen, Shokrollahi, and Spielman 2001] or low density parity check codes [MacKay 1999].

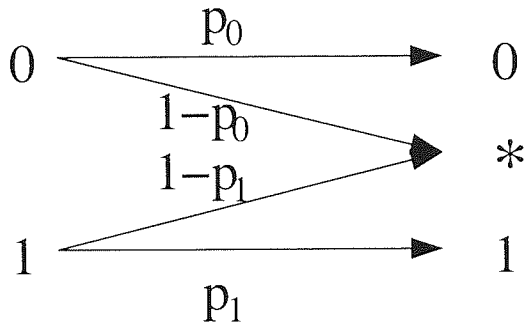


Figure 4.9: Binary erasure channel.

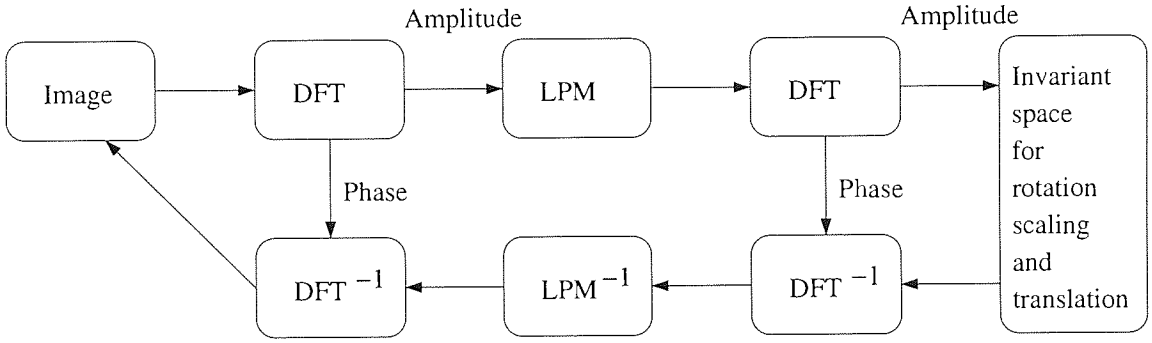


Figure 4.10: Rotation, scaling, and translation invariant scheme block diagram. (DFT: discrete Fourier transform; LPM: log polar map)

4.3.4 Geometrical Transform

Geometrical transforms such as rotation or translation can be classified as synchronisation attacks. As for cropping corruption, the main problem consists in aligning the decoder with the received data R as the embedded information M is usually unmodified. Thus, the problem can be handled, as mentionned previously, by exhaustive searches which may be extremely demanding and time consuming.

One can also use invariant transform domain as suggested in [Ruanaidh and Pun 1997]. However, in order to get such domain, one has to accept a significant reduction in the number of elements in which the message information can be embedded. For instance, in the framework of [Ruanaidh and Pun 1997] depicted in Fig. 4.10, the phases of two of the three successive transforms are not used. As the phase represents each time half of the coefficients, it means that the invariant domain comprises only one forth of the original coefficients used to encode the image data.

4.4 Watermark Perceptual Effects

In this section, the assessment of the watermark visibility and its relation with image data representation are discussed. The issue is indeed central to the digital watermarking problem and to the best

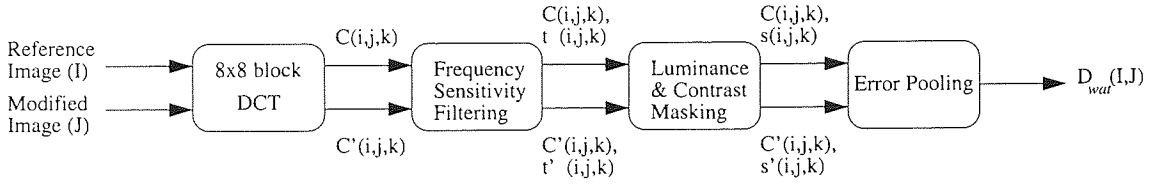


Figure 4.11: Watson perceptual metric block diagram.

of our knowledge, it is still an open problem. Early studies have relied on mathematical objective distortion metrics rather than on perceptual measures. Mathematical measures were quite appealing due to their computational simplicity.

With the emergence of perceptual measures in the field of digital watermarking, it became more and more difficult to compare results of new schemes as the watermark perceptibility requirements were either described by the word "imperceptible" or just given in terms of a parameter value used in the specific scheme. The mean squared error, the most commonly used mathematical metric, has been shown to be quite poorly correlated to perceptual distortions [Girod 1993]. However, as a well defined measure with clear physical meanings, the mean squared error can be used to define the watermark power constraint, unlike perceptual distortion measures which are not universal.

In this section, two image representations, 8×8 DCT and ICA block transforms, are investigated with respect to the mean squared error distortion introduced by the watermark X. Two perceptual distortion measures described in Sec. 4.4.1 are used to quantify the watermark perceptual distortion in Sec. 4.4.2.

4.4.1 Perceptual Measures

In this section, we present briefly two perceptual measures that we use in the context of practical watermarking scheme performance evaluation. Both assume that the image data are properly scaled and aligned.

The first one is a well-known error sensitivity and DCT based measures due to Watson [Watson 1993]. It is a *bottom-up* approach, simulating the different functions of the human visual system, such as luminosity sensitivity and contrast masking. The second approach called structural similarity [Wang, Bovik, Sheikh, and Simoncelli 2004] is on the contrary a *top-down* approach, which emphases the image data structural similarity and intends to mimic the overall functionality of the human visual system.

Watson's Perceptual Metric

Figure 4.11 depicts Watson's perceptual metric. First, both the reference I and the modified J image are divided into contiguous 8×8 pixels patches. The DCT of all patches is taken, the obtained coefficients $C(i,j,k)$, where $(i,j) \in [1,8]^2$ denotes the position of the coefficient in a given patch and

1.40	1.01	1.16	1.66	2.40	3.43	4.79	6.56
1.01	1.45	1.32	1.52	2.00	2.71	3.67	4.93
1.16	1.32	2.24	2.59	2.98	3.64	4.60	5.88
1.66	1.52	2.59	3.77	4.55	5.30	6.28	7.60
2.40	2.00	2.98	4.55	6.15	7.46	8.71	10.17
3.43	2.71	3.64	5.30	7.46	9.62	11.58	13.51
4.79	3.67	4.60	6.28	8.71	11.58	14.50	17.29
6.56	4.93	5.88	7.60	10.17	13.51	17.29	21.15

Table 4.1: DCT frequency sensitivity table t.

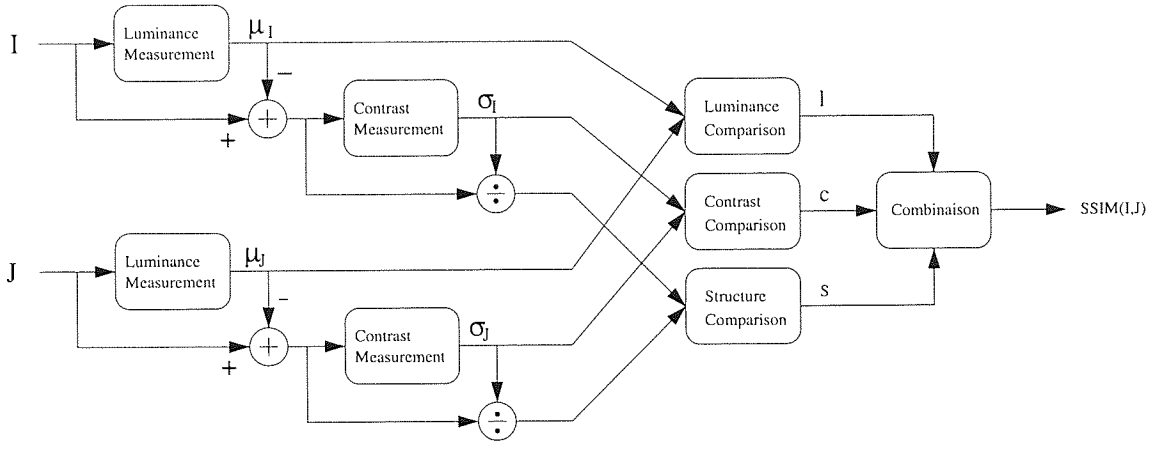


Figure 4.12: Structural similarity block diagram.

k the patch index, are used to compute the luminance-masked threshold $t_L(i,j,k)$ as follows

$$t_L(i,j,k) = t(i,j)\left(\frac{C(0,0,k)}{C_{0,0}}\right)^a, \tag{4.2}$$

with $a = 0.649$, where $t(i,j)$ is given by the frequency sensitivity table Tab. 4.1 and $C_{0,0}$ is the full image DCT first coefficient. Then, from $t_L(i,j,k)$, the masking threshold $s(i,j,k)$ is computed to take into account the contrast masking phenomena as in

$$s(i,j,k) = \max\left\{t_L(i,j,k), |C(i,j,k)|^w t_L(i,j,k)^{(1-w)}\right\}, \tag{4.3}$$

with $w = 0.7$. Finally, the perceptual distance $d(i,j,k)$ between each reference and modified coefficients is computed and pooled into a single value $D_{wat}(I,J)$ as follows

$$d(i,j,k) = \frac{C(i,j,k) - C'(i,j,k)}{s(i,j,k)}, \tag{4.4}$$

$$D_{wat}(I,J) = \left(\sum_{i,j,k} |d(i,j,k)|^p\right)^{\frac{1}{p}} \quad \text{with } p = 4\,. \tag{4.5}$$

Further details and discussion can be found in [Watson 1993].

Structural Similarity

As depicted in Fig. 4.12, the structural similarity metric is based on three comparisons: luminance, contrast and structure. As these are usually varying across the image, we will always refer to the *local*

luminance, contrast or structure. The local luminance and contrast are computed as the mean and variance of the local pixels values, while the local similarity in structure is quantified by a correlation based measure. The details of the calculations can be summarised as follows.

First, the local mean intensities and standard deviations for each pixel of the reference image I and the modified image J, denoted $\mu_I(i,j)$, $\mu_J(i,j)$, $\sigma_I(i,j)$ and $\sigma_J(i,j)$, respectively, are computed as[5]

$$\mu_I(i,j) = \frac{1}{N}\sum_{k=1}^{N} I_k\,, \tag{4.6}$$

$$\sigma_I(i,j) = \left(\frac{1}{N-1}\sum_{k=1}^{N}(I_k - \mu_I(i,j))^2\right)^{1/2}, \tag{4.7}$$

where I_k denotes the intensity value of a pixel in the neighbourhood of the pixel indexed (i,j) and N is their number.

Then, the luminance, contrast and structure similarity measures between I and J, respectively $l(i,j)$, $c(i,j)$ and $s(i,j)$ are evaluated for each spatial location (i,j) as in

$$l(i,j) = \frac{2\mu_I(i,j)\mu_J(i,j) + K_1}{\mu_I(i,j)^2 + \mu_J(i,j)^2 + K_1}\,, \tag{4.8}$$

$$c(i,j) = \frac{2\sigma_I(i,j)\sigma_J(i,j) + K_2}{\sigma_I(i,j)^2 + \sigma_J(i,j)^2 + K_2}\,, \tag{4.9}$$

$$s(i,j) = \frac{\sigma_{IJ}(i,j) + K_3}{\sigma_I(i,j)\sigma_J(i,j) + K_3}\,, \tag{4.10}$$

with

$$\sigma_{IJ}(i,j) = \frac{1}{N-1}\sum_{k=1}^{N}(I_k - \mu_I(i,j))(J_k - \mu_J(i,j))\,. \tag{4.11}$$

These three similarities measures are symmetric, lower than one, and equal to one if and only if all $I_k = J_k$. The constants K_1, K_2 and K_3 have been introduced to avoid instability, when the denominators are closed to zero.

Finally, these similarity measures are combined in a multiplicative way, which for specific parameter values [Wang, Bovik, Sheikh, and Simoncelli 2004] gives[6]

$$\mathrm{SSIM}(i,j) = l(i,j)\,c(i,j)\,s(i,j)\,, \tag{4.12}$$

$$= \frac{(2\mu_I\mu_J + K_1)(2\sigma_{IJ} + K_2)}{(\mu_I^2 + \mu_J^2 + K_1)(\sigma_I^2 + \sigma_J^2 + K_2)}\,, \tag{4.13}$$

with $K_1 = 0.01$, $K_2 = 0.03$ and $K_3 = 2K_2$. The average over all spatial location (i,j), given by

$$MSSIM(I,J) = \frac{1}{hw}\sum_{i,j} SSIM(i,j)\,, \tag{4.14}$$

can be used to evaluate the overall image quality as suggested in [Wang, Bovik, Sheikh, and Simoncelli 2004], where further details and discussion can be found.

[5] expressions of $\mu_J(i,j)$ and $\sigma_J(i,j)$ are similarly obtained.

[6] in these equations, for sake of readibility $\mu_I(i,j)$ is denoted μ_I. Similarly we have $\mu_J = \mu_J(i,j)$, $\sigma_I = \sigma_I(i,j)$, $\sigma_J = \sigma_J(i,j)$ and $\sigma_{IJ} = \sigma_{IJ}(i,j)$.

4.4.2 Image Data Representation and Watermark Perceptibility

In the context of digital watermarking, various data representation spaces have been investigated. Since image data representations are typically based on frequency decompositions, they provide a natural way to take advantage of the human visual system non-uniform sensitivity over the different frequency bands. Furthermore, they allow the spatial spreading of the watermark.

Moreover, several studies on such feature spaces are available due to needs in the field of digital images compression. These studies have been exploited for improving digital watermarking techniques. For instance, the DCT and the DWT are the two main transforms to benefit from these studies and are also the two most commonly used in practical watermarking schemes.

In the following analysis, DCT and ICA 8×8 block transforms are investigated, and the study of other transforms is the subject of future research. The study is carried out for a given watermark mean squared error with respect to the Watson perceptual metric and the structural similarity previously presented.

Watermark Perceptual Distortion with Respect to Watson's Measure

In this section, we are interested in the minimum perceptual distortion measured by Watson's metric for a given mean squared error σ_X^2. Since the DCT is merely a rotation, the watermark mean squared error σ_X^2 can be expressed as a function of the original and modified DCT coefficients $C(i,j,k)$ and $C'(i,j,k)$ defined in Sec. 4.4.1

$$\sigma_X^2 = \frac{1}{hw} \sum_{i,j,k} \left(C(i,j,k) - C'(i,j,k) \right)^2 . \tag{4.15}$$

By introducing a Lagrange multiplier λ related to the watermark mean squared error in Watson's distortion expression given in Eq. 4.5, we have

$$\mathscr{L}(\{\delta(i,j,k)\}_{(i,j,k)}, \lambda) = \left(\sum_{i,j,k} \left(\frac{\delta(i,j,k)}{s(i,j,k)} \right)^4 \right)^{\frac{1}{4}} + \lambda \left(\sigma_X^2 - \frac{1}{hw} \sum_{i,j,k} \delta(i,j,k)^2 \right) , \tag{4.16}$$

with

$$\delta(i,j,k) = C(i,j,k) - C'(i,j,k) . \tag{4.17}$$

Solving the system of equations given by the first order conditions of optimality

$$\begin{cases} \frac{\partial}{\partial \sigma_{X_i}^2} \mathscr{L}(\{\delta(i,j,k)\}, \lambda) &= 0 , \quad \text{for } i = 1..8 , \ j = 1..8 , \text{ and } k = 1..(hw/64) , \\ \frac{\partial}{\partial \lambda} \mathscr{L}(\{\delta(i,j,k)\}, \lambda) &= 0 , \end{cases} \tag{4.18}$$

we obtain the least perceptually distorted data $C'(i,j,k)$ in the sense of Watson's metric

$$C'(i,j,k) = C(i,j,k) \pm u(i,j,k)\, \sigma_X \sqrt{hw} , \tag{4.19}$$

$$u(i,j,k) = \sqrt{\frac{s(i,j,k)^4}{\sum_{i,j,k} s(i,j,k)^4}} , \tag{4.20}$$

$$\underline{D}_{wat}(I, J^*) = \frac{\sigma_X \sqrt{hw}}{\left(\sum_{i,j,k} s(i,j,k)^4 \right)^{1/4}} , \tag{4.21}$$

where $\underline{D}_{wat}(I, J^*)$ is the minimum Watson's distortion for a watermark mean squared error of σ_X^2.

Consequently, for a given 8×8 block[7], there are 2^{64} vectors[8] of minimal perceptual distortion and they can be defined as

$$v^T = \left(\pm u_{11}, \pm u_{12}, \ldots, \pm u_{ij}, \ldots, \pm u_{i'j'}, \ldots, \pm u_{88} \right). \tag{4.22}$$

Minimum perceptual distortion is achieved when the distortion occurs along one of these axes.

In addition, it is well known that independent modifications of image block mean values, corresponding to the coefficient $C(1,1,k)$, introduce visible blocking artefacts, which are not taken into account in Watson's model. Thus, let us define the 2^{63} vectors of minimal perceptual distortion in the subspace excluding the image local mean $C(1,1,k)$ given by

$$w^T = \left(\pm u'_{12}, \pm u'_{13}, \ldots, \pm u'_{ij}, \ldots, \pm u'_{i'j'}, \ldots, \pm u'_{88} \right), \tag{4.23}$$

$$u'(i,j,k) = \sqrt{\frac{s(i,j,k)^4}{\sum_{k,(i,j)\neq(1,1)} s(i,j,k)^4}} \quad . \tag{4.24}$$

Assuming a watermarking scheme based on the quantisation of a single DCT coefficient value per image block and a uniformly distributed host data over the quantisation bin $[-\Delta/2; \Delta/2]$ where Δ is the quantisation step, the average perceptual distortion $< D_{wat}(I,J) >$ can be derived as

$$< D_{wat}(I,J) > = \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} D_{wat}(I, J(t)) \mathrm{d}\, t \,, \tag{4.25}$$

$$= \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \left(\sum_k \frac{t^4}{s_k^4} \right)^{1/4} \mathrm{d}\, t \,, \tag{4.26}$$

$$= \frac{\Delta}{4} \left(\sum_k \frac{1}{s_k^4} \right)^{1/4} \,, \tag{4.27}$$

where s_k is the selected masking threshold $s(i,j,k)$ for the kth block. In practice, $s(1,1,k)$, which corresponds to the luminance of the block, is rarely chosen as a modification of this term as it introduces noticeable blocking artefacts which are not taken into account in Watson's model.

Thus, the comparison between two watermarking feature spaces can be done through the comparison of their DCT representation. For instance, the average perceptual distortion for a QIM embedding occurring along a unit vector, which DCT coefficients are denoted α_{ij}, assuming a uniformly distributed host data over the quantisation bin $[-\Delta/2; \Delta/2]$ where Δ is the quantisation step given by

$$< D_{wat}(I,J) >= \frac{\Delta}{4} \left(\sum_{i,j,k} \frac{\alpha_{ij}^4}{s(i,j,k)^4} \right)^{1/4} . \tag{4.28}$$

Figure 4.14 shows the perceptual distortions introduced by watermarking schemes using different transform spaces for a unitary watermark mean squared error distortion. All are based on a QIM embedding applied to a single coefficient per 8 × 8 image block. Hence, the difference of perceptual

[7] for a fixed k, $u(i,j,k)$ is denoted u_{ij}.
[8] the 2^{64} is due to the ± sign in Eq. 4.19.

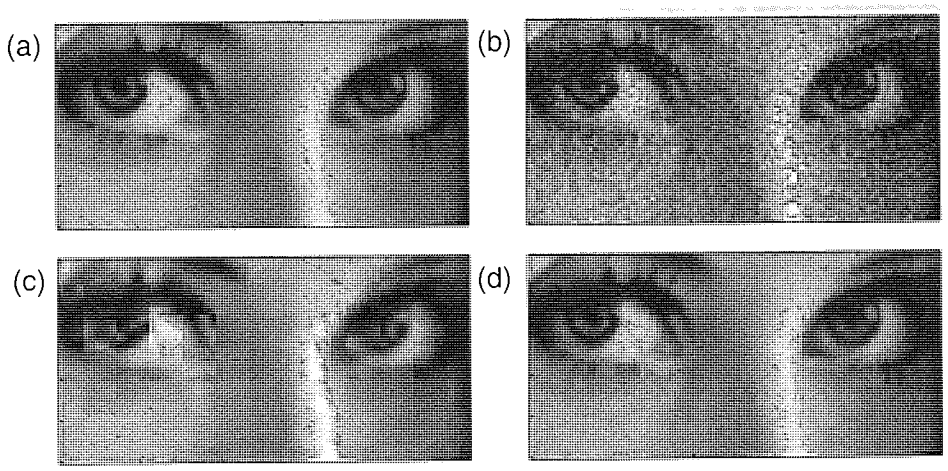


Figure 4.13: Examples of distortion measured by Watson's metric for some image modification. (a) Original image 50x100 pixels cropped from Lena picture; (b) centred Gaussian noise, MSE=220, $D_{wat} = 112$; (c) JPEG compression, MSE=60, $D_{wat} = 7.7$; (d) JPEG compression, MSE=5, $D_{wat} = 4.9$.

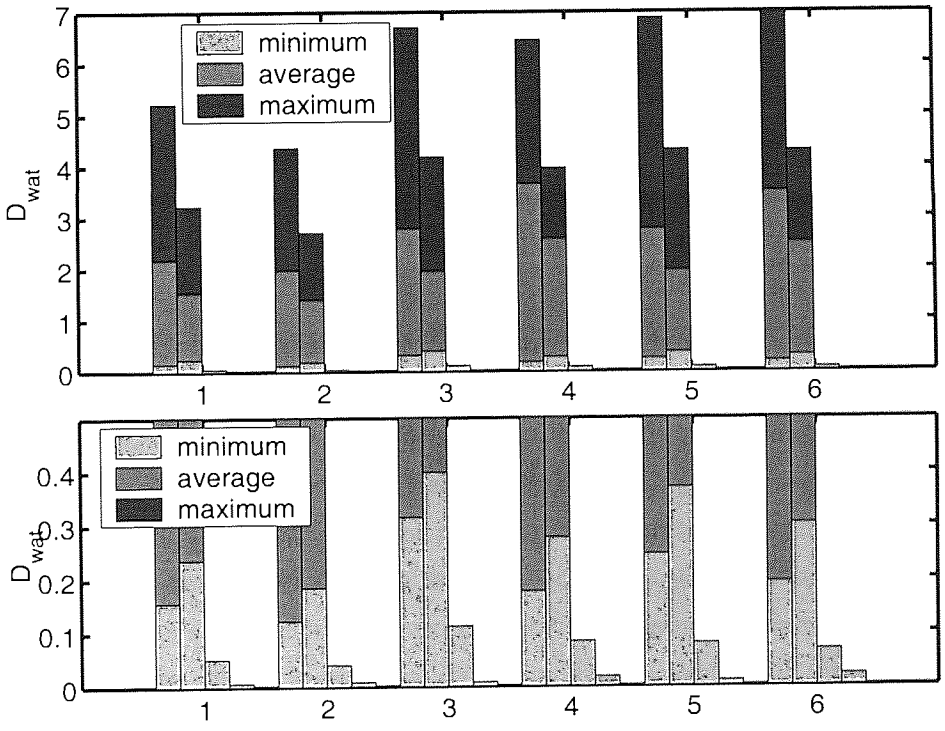


Figure 4.14: Watson perceptual distortion for various modified images. Top and bottom plots are the same at different scale. The x-axis labels correspond to the different standard images (Appendix B): (1) 3.2.25, (2) Baboon, (3) F16, (4) Fishingboat, (5) Lena and (6) Peppers. For each image, the bars correspond to Watson perceptual distortion measured when QIM is performed on one coefficient on each 8×8 patch for different transformation. From the left to right, each of the four bars corresponds to a different set of vectors defined by (i) the DCT basis vectors, (ii) the ICA transform vectors, the directions of minimal distortion (iii) in the subspace excluding the image mean component $(w_i)_{i\in[1:63]}$ (Eq. 4.23) and (iv) in the full space $(v_i)_{i\in[1:64]}$ (Eq. 4.22). The different grey for each bar represent the average minimum (light grey), mean (grey) and maximum (dark grey) perceptual distortion level assuming a MSE of $\sigma_X^2 = 1$.

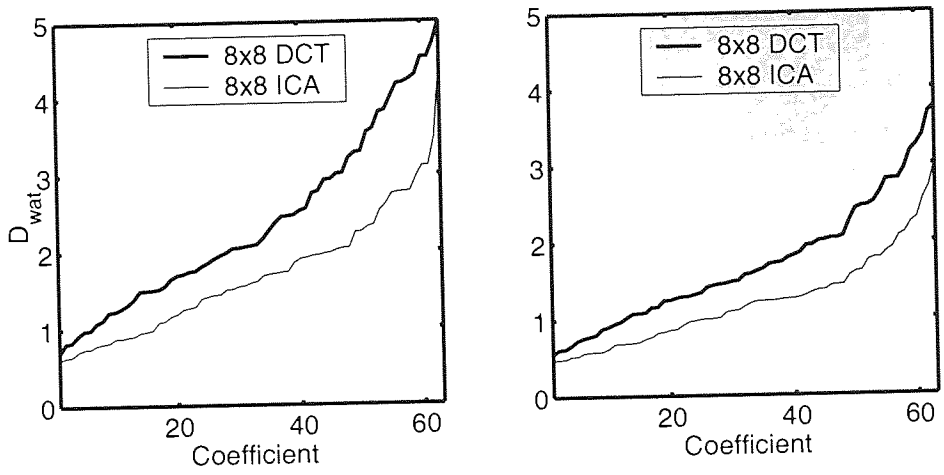


Figure 4.15: Average perceptual distortion resulting from QIM embedding with respect to the DCT/ICA transform coefficient for (a) Lena and (b) baboon images, with a mean squared error $\sigma_X^2 = 1$.

distortion is directly related to the transform domain used. The four transform spaces compared are DCT, ICA, and the optimal transforms defined by Eq. 4.22 and Eq. 4.23.

The distortion values reported are averaged measures for one 8×8 image block for various images labelled from 1 to 6 on the x-axis. Three values are marked for each technique: the minimum, the mean and the maximum. The minimum and maximum are obtained by assuming that the best or worst[9] basis vector is chosen for each image block at the embedding, while for the mean, it is assumed that the modified vector is randomly chosen. Note that for the two optimal spaces, there are neither mean nor maximum values.

As shown in Fig. 4.14, neither the DCT or ICA space performances are close to the optimal space performance represented by the third and fourth bars. Note that from Eq. 4.2 and Eq. 4.3, it can be shown that the predominant components of the optimal vectors v and w usually correspond to low frequency components[10] such as the mean. Indeed, for typical images, DCT low frequency coefficients are high compared to high frequency ones due to strong local correlation in images. However, as for the mean, independent modifications of low frequency coefficient values often lead to more visible block artefacts which are not taken into account in Watson perceptual measure.

The DCT space shows slightly lower minimal perceptual distortion than ICA transform space but also a significantly higher average mean perceptual distortion. Thus, if the set of coefficients to be quantised are selected randomly or independently of the perceptual distortion, using ICA feature space will lead to a lower perceptual distortion.

This result is confirmed in Fig. 4.15 where the average perceptual distortion related to a modification of the DCT or ICA transform coefficients are represented for (a) Lena and (b) Baboon images. This can easily be explained by the form of the pooling function (Eq. 4.5) used by Watson perceptual measure. For such a convex function, spreading the watermark mean squared error over many DCT

[9] in the sense of the perceptual distortion introduced.
[10] small values for the indices i and j in $C(i, j, k)$.

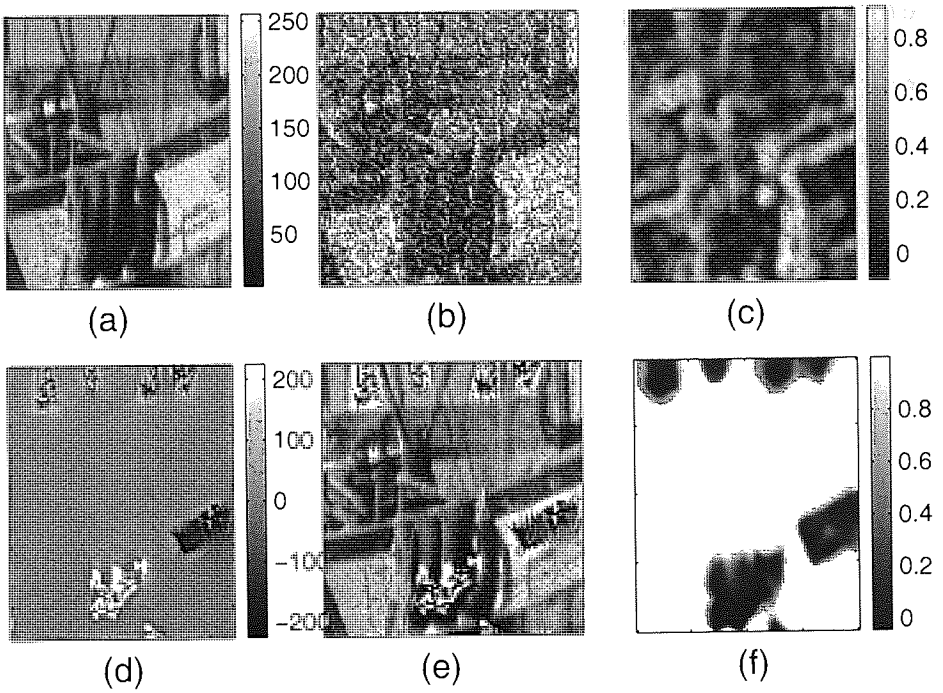


Figure 4.16: An example of optimisation of the MSSIM for a given mean squared error distortion level. (a) The original image (90×90, 8bits/pixel, cropped from the Fishingboat image); (b) the initial image corrupted by a centred Gaussian noise of variance 2500 (MSSIM=0.33); (c) SSIM map between (a) and (b); (d) image difference between (e) and (a); (e) MSSIM optimised image (MSSIM=0.79); (f) SSIM map between (a) and (e).

frequency coefficients will result on average in a lower value.

Watermark Perceptual Distortion with Respect to the Structural Similarity

Unlike Watson's metric, the SSIM does not allow any direct derivation of an optimal embedding space in the perceptual sense. It is therefore difficult to measure how close a transform is to the optimal one. Thus, the approach we put forward here will be based on a best case analysis as proposed in [Wang, Bovik, Sheikh, and Simoncelli 2004].

Starting from a distorted image, we ascent the gradient of MSSIM while maintaining the mean squared error at a constant level σ_X^2. Then, the MSSIM of the optimised image can be used as an approximation for the actual optimal MSSIM and to assess watermarking schemes imperceptibility. Moreover, carrying out such optimisation procedure for various images may enable us to infer some characteristics of the optimal transform.

In [Wang, Bovik, Sheikh, and Simoncelli 2004], the starting corrupted image used is arbitrarily set to be distorted by a centred Gaussian noise. However, in conjunction with the proposed pooling function given in Eq. 4.14, a sharp and visible localised distortion may result in the same MSSIM as many mild and imperceptible distortions. As each pixel influence is spatially localised, the optimisation may lead to a concentration of the distortion as illustrated in Fig. 4.16.

In Fig. 4.16e, a few regions with highly visible artefacts can be shown corresponding to the dark

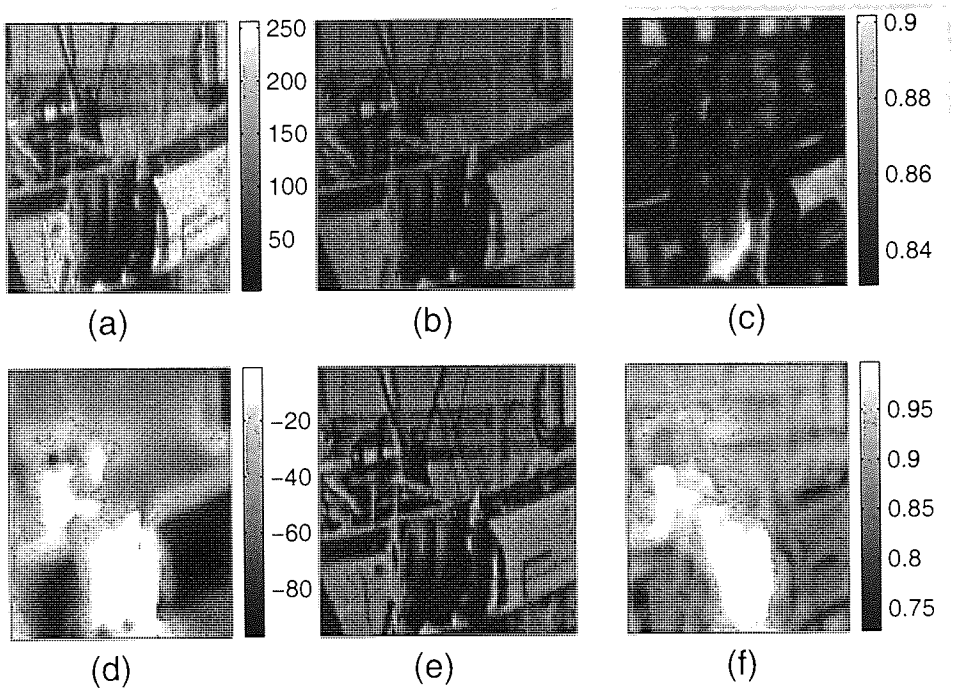


Figure 4.17: An example of optimisation of the MSSIM from a rescaled image for a given mean squared error distortion level. (a) The original image (90×90, 8bits/pixel, cropped from the Fishing-boat image); (b) the initial image rescaled with $k = 0.6394$ (MSE=2500, MSSIM=0.84); (c) SSIM map between (a) and (b); (d) image difference between (e) and (a); (e) MSSIM optimised image (MSSIM=0.91); (f) SSIM map between (a) and (e).

areas in Fig. 4.16f. In these regions, the pixel values are completely different from their original values found as shown in Fig. 4.16d and hence contribute significantly to the global MSE. In all other areas, the reported distortions are very mild ($\simeq$ 1 luminance intensity level). Comparing Fig. 4.16c and Fig. 4.16f reveals that in order to optimise the MSSIM, the optimisation procedure has increased the regions of high SSIM in Fig. 4.16c to the expenses of the areas of low SSIM. Clearly such optimised image cannot be taken as a reference for further investigations.

However, we found that pixel value rescaling can be used to find a better distorted image in the sense of the MSSIM. Indeed, it is well known that the perceptual impact of such distortion is rather low considering its mean squared error. Moreover, our experiments show that the MSSIM is pretty consistent with this fact. Furthermore, in the case of rescaling, the relation between the MSSIM and the mean squared error can be explicitly derived. Let us denote I and J, the original and the rescaled image, respectively, with the rescaling constant $k \in [0;1]$ such that $J = k\,I$. Then, the mean squared error and the MSSIM[11] between I and J can be derived as a function of k

$$MSE(I,J) = \frac{1}{hw}(1-k)^2\|I\|^2, \tag{4.29}$$

$$MSSIM(I,J) = \frac{4k^2}{(1+k^2)^2}. \tag{4.30}$$

This also provides a tighter lower bound for the optimal MSSIM at a given mean squared error distortion level. Furthermore, the rescaled image can be used as a starting point for the optimisation

[11] for simplification purposes, $K1$ and $K2$ are set to 0 in this derivation.

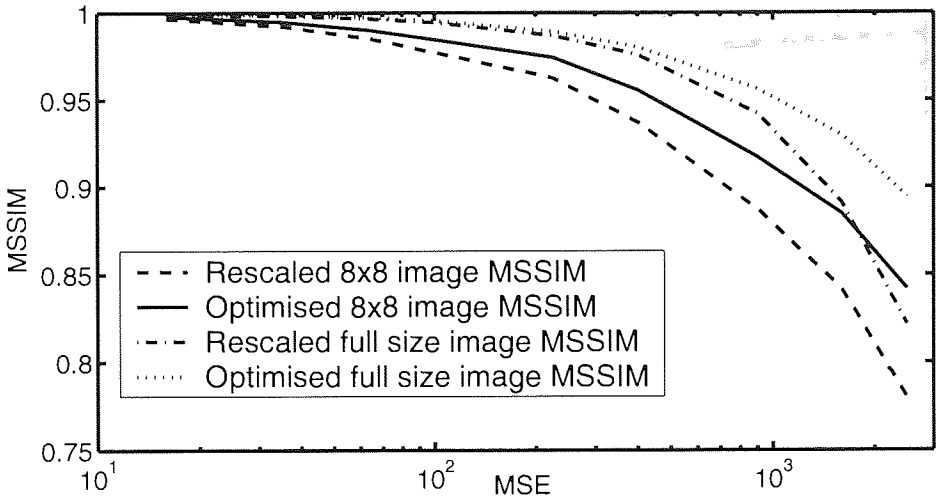


Figure 4.18: Approximation of the optimal MSSIM with respect to the MSE.

procedure proposed in [Wang, Bovik, Sheikh, and Simoncelli 2004]. An example is shown in Fig. 4.17. Figure 4.17c and Fig. 4.17f show a more evenly spread perceptual distortion than in Fig. 4.16f, which is corroborated by the absence of sharp artefact in Fig. 4.17b and Fig. 4.17e. It can also be noticed from the white zones in Fig. 4.17d that textured regions are less distorted. From the bottom part of Fig. 4.17f, it can also be seen that textured areas, when they are distorted, show a higher perceptual error than plain areas.

From this two steps optimisation procedure, approximate values of the optimal MSSIM are computed with respect to a given mean squared error value. These simulations, carried out for full frame images and 8×8 pixels patches, show a better optimised MSSIM for full size images than for the reduced size patches depicted in Fig. 4.18. Intuitively, this can be explained by the greater freedom for spreading the perceptual distortion in full size images.

Besides, even for very high mean squared error values, very low perceptual distortions, represented by high MSSIM values in Fig. 4.18, can be achieved. As for the optimal perceptual vector defined in Eq. 4.23 for the Watson model, the distortion of the mean value contributes for more than 95 % of the mean squared error. Thus, methods based on DCT or ICA block transforms, which do not modify the image patch mean intensity values, cannot be optimal with respect to the MSSIM.

Both studies based on Watson metric and the SSIM suggest that the variation of the image mean intensity value is the main ingredient in the least perceptible distortion for a given mean squared error level. However, such modification of the mean cannot be usually done at the block level, as independent modifications of block mean intensity values introduce sharp transitions from one block to another. Increasing the size of the image block for the transform reduces blocking artefacts, by decreasing the number of block transitions. Thus, this increase can also be used to improve the global imperceptibility of the watermark.

In order to evaluate the 8×8 DCT and ICA block transforms with respect to the structural

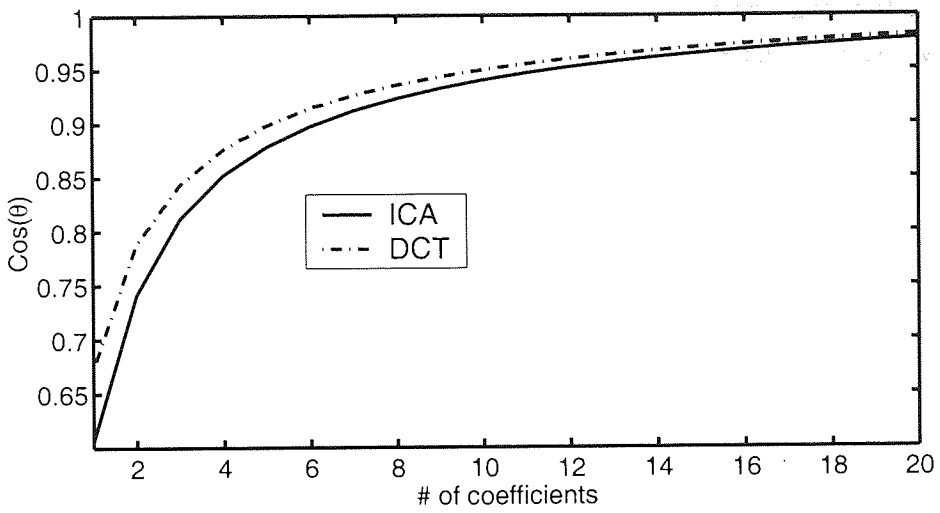


Figure 4.19: Optimised patches projection on the DCT and ICA bases.

similarity measure, we consider the projection on the DCT and ICA bases of the difference between the optimised and the original 8×8 pixels image patches. In Fig. 4.19, we report the average cosine between the perceptually least distortive vector and the closest hyperplan defined by a set of DCT or ICA basis vectors with respect to the size of the set. As shown in Fig. 4.19, neither DCT nor ICA provide a very good representation as 5 to 6 coefficients are needed on average to obtain a cosine of 0.9. However, it is expected from overcomplete ICA basis to provide better performance. The investigation of such ICA basis is not carried out here and will be the subject of future research.

4.5 Summary

In this chapter, we investigated image data representation influence on the performance of practical watermarking schemes. For instance, a general analysis of the relation between scheme robustness and image domain transform was carried out. Considering the high number of different types of attacks, no image transform can provide an optimal embedding space for all of them, thus image transforms have to be dedicated to specific corruption or serve other purposes, such as the watermark imperceptibility.

In this chapter, we also put forward a new image data representation for digital image watermarking. This approach based on ICA is mainly motivated by the properties resulting from such a transformation. For instance, the transform provides independent channels, thus the original value of a channel infer from the others after a modification. Moreover, ICA mixing and demixing matrices can be used as secret keys as they are obtained from a specific set of images.

The influence of this new domain transform, used as a block transform, was compared to the DCT block transform with respect to two perceptual measures proposed in [Watson 1993; Wang, Bovik, Sheikh, and Simoncelli 2004]. The analysis results show that ICA performance is competitive with respect to DCT performance. Future research on more advanced ICA approaches, comprising non-linear ICA and overcomplete ICA, are expected to show further improvement.

Further preliminary research on ICA domain transform based watermarking were also carried out in [Bounkong, Toch, Saad, and Lowe 2003a]. The topics investigated are related to fragile watermarking and maximum a posteriori decoding, which will be considered in future research.

Chapter 5

Conclusion

In this work, digital watermarking has been put forward as a mean to protect intellectual property rights of digital media content or to ensure their authentication. Digital watermarking schemes are composed of two elements: a message embedding process and a message retrieving process. The successful retrieval of the embedded message is used to prove the ownership of the data or to ensure its integrity.

The first principle of digital watermarking is for the watermark to be unobtrusive. The quality of the digital cover data must not be altered with respect to the original. Typically, the watermarked and original data have to be perceptually identical. Since this principle is quite subjective and hence very difficult to model, a simplistic mathematical approach is often used instead. For instance, we used a model based on the mean squared error measure. For data represented by a vector of digital values, we upper bound the mean squared difference between the original and watermarked vector.

Since this mathematical model bounds the watermarked data within an hypersphere centred at the original data, the amount of information that can be embedded is limited by the number n of digital data in this hypersphere. In the case of a noiseless transmission of the watermarked data, up to k bits (with $k = \log_2(n)$) of information can be transmitted. For a noisy transmission, the amount of information that can be reliably transmitted may be reduced significantly.

In the past few years, many studies have focused on finding how much information can be reliably transmitted within the watermarking scenario via noisy transmission channels. Another approach to this problem is to find the power constraint that is required to transmit reliably a given amount of information. For a quantised transmission channel, such investigation has been carried out in Chapter 2.

A quantised transmission channel reduces the number of different digital data that can be transmitted by replacing data by the nearest representative quantised value. For instance, for the n possible watermarked data, there will be only m different quantised data with $m \leq n$. The number of bits that can be transmitted is hence given by $\log_2(m)$.

Given the quantisation channel characteristics, we derived a relation between the size of the limiting

hypersphere and the quantity of information in bits that can be reliably transmitted. An analytical expression was derived for the worst case scenario, while a practical iterative algorithm has been suggested for the general case. Both analysis and algorithm are based on a deterministic quantisation model and are used to estimate the size of the limiting hypersphere given the quantity of information to transmit.

We compared these estimations with those obtained when a stochastic quantisation model is adopted. The study has pointed out the fact that a stochastic model, based solely on the second order statistics of the quantisation noise, leads to a significant overestimation of the size of the limiting hypersphere. It has also established that for the same noise variance, a centred Gaussian noise is a more efficient attack than JPEG lossy compression.

Moreover, we show that an efficient watermarking scheme, robust against quantisation attacks, can be derived by taking into account the quantisation bin limits. Each bin encodes at most a single message. Depending on the knowledge on the corruption quantisation step at the embedding stage, some digital data may not be used to encode any message. In addition, a general methodology has been developed that may enable the design of an optimal watermarking scheme against quantisation attack. Indeed, the watermark distortion bounds provided by the proposed algoritms can be used as guidelines to improve current method. The main remaining problem would be to find an optimal labelling of the quantisation bins within the set of all possible watermark messages.

For more general noisy transmission channels, we developed a new framework for information embedding based on quantisation. Quantisation based techniques assume that a set of data lattices corresponding to the different messages is available to the decoder. These lattices are used at the receiving end to decode the received values. Typically, the minimum distance between two lattice points Δ, also termed the characteristic distance of a lattice, is chosen according to the size of the limiting hypersphere (power constraint).

Further research has shown that increasing the minimum distance between two lattice points may reduce the decoding error probability. Given the size of the limiting hypersphere, if the lattice characteristic distance Δ used is large, it may not be always possible to map the original data to the closest lattice point. Thus, the embedding process consists in bringing closer the original data to the relevant lattice point.

We put forward an optimisation procedure that aimed at minimising the decoding error probability. The proposed algorithm is quite general and can be applied easily to various noise and decoding models as demonstrated in Chapter 3. Numerical results show that the embedding process derived by the proposed optimisation procedure outperforms previously suggested embedding schemes for a wide range of watermark to noise ratios.

The analysis of different embedding schemes, as well as our optimised embedding technique, reveals that for quantisation based embedding processes, the quantisation step Δ (minimum distance between two lattice points) has a predominant role in the scheme performance for low watermark to noise ratios. However, it was also pointed out that using large quantisation steps does not necessarily result in a

good performance.

In addition, this study has demonstrated the limitation of quantisation based embedding processes which use a pre-defined quantisation step Δ. Indeed, the optimal quantisation step Δ^* is typically different for each watermark to noise ratio and is rarely known at the embedding stage. When the quantisation step used Δ is different from Δ^*, the scheme performance may degrade significantly. Consequently, the main challenge is to design a practical watermarking scheme with good performance for both low and high watermark to noise ratios.

The proposed optimisation procedure has shown some promising results with this respect. Indeed, in the proposed framework, the embedding process can be based on a large quantisation step Δ and optimised with respect to a high watermark to noise ratio. Given the noise model, these two characteristics ensure a low decoding error probability at both low and high watermark to noise ratios providing an efficient embedding processes.

In our study, the embedding problem was presented as a one dimensional problem for simplicity, while application of the optimisation procedure to higher dimension data were outlined. Embedding processes based on vector quantisation have better performance but are also more complex. Thus, the improved performance has to be weighted against the increase in complexity.

The representation of high dimensionality data is another main issue of the digital watermarking problem. Its main purpose is to simplify the message embedding. Since the data perceptual distortion and the transmission corruption are typically not isotropic, privileged directions can be exploited for improving digital watermarking schemes.

We proposed a new image data representation based on independent component analysis (ICA) for digital watermarking (Chapter 4). Such a transform yields to independent channels, thus the original value of a channel cannot be inferred from other channels after a modification. Moreover, independent channels also provide a significant advantage in the design of efficient watermarking scheme in terms of capacity. Finally, ICA mixing and demixing matrices can be used as secret keys as they are obtained from a specific set of images.

After discussing the effects of image data representation on the decoding error probability, the proposed ICA-based representation was studied and compared to a scheme based on discrete cosine transform (DCT) using the watermark perceptual distortion as a measure. The analysis shows that the performance of ICA-based techniques is competitive and sometimes superior with respect to DCT based methods.

In addition, our study has shown that a very large image distortion, in the sense of the average squared difference, can be relatively imperceptible. Such imperceptible distortion mainly consists of a strong modification of the global image mean intensity value and of modifications preserving the image structure and contrast. That naturally excludes modifications based on block transforms, which typically induce sharp transitions from one block to another.

Considering the great variety of image data, finding a good representation for digital watermarking purposes remains a difficult task. Moreover, a formal statement of the problem has still to be found.

In our research, we expose the limitations of image block transform based representations and point to possible improvements.

Other future research directions include multi-resolution embedding processes or image data representation based on overcomplete bases. Image data based on overcomplete bases are expected to provide more flexibility in the representation of the data and to enable less noticeable distortion.

Using a multi-resolution embedding process would aim at achieving good information rates independently of the transmission noise level. The new scheme would be based on a multi-resolution quantisation grid. Depending on the estimated noise at the receiving end, either a coarse or fine grid would be used for decoding. In addition, the optimisation procedure presented in Chapter 3 could be extended to accomodate multi-resolution embedding process.

Bibliography

Anderson, R. and F. Petitcolas (1999, August). Information hiding and digital watermarking: an annotated bibliography.

Avcibas, I., B. Sankur, and K. Sayood (2002, April). Statistical evaluation of image quality measures. *Journal of Electronic Imaging* **11**, 206–223.

Barni, M., F. Bartolini, V. Cappellini, A. Piva, and F. Rigacci (1998, September). A map identification criterion for dct-based watermarking. In *Proceedings of European Signal Processing Conference*, Volume 1, pp. 17–20.

Bartolini, F., M. Barni, V. Cappellini, and A. Piva (1998, October). Mask building for perceptually hiding frequency embedded watermarks. In *Proceedings of the IEEE International Conference on Image Processing*, Volume 1, New York, pp. 450–454. IEEE.

Bell, A. and T. Sejnowski (1997, December). The independent components of natural scenes are edge filters. *Vision Research* **37**(23), 3327–3338.

Bender, W., D. Gruhl, N. Morimoto, and A. Lu (1996). Techniques for data hiding. *IBM Systems Journal* **35**(3/4), 313–336.

Blaschke, T. and L. Wiskott (2002, August). An improved cumulant based method for independent component analysis. In *Proc. Int. Conf. on Artificial Neural Networks, ICANN 02*, Volume 2415 of *Lecture Notes in Computer Science*, Berlin, pp. 1087–1093. Springer-Verlag.

Bounkong, S., D. Saad, and D. Lowe (2002a, August). Independent component analysis for domain independent watermarking. In *Artificial Neural Networks - ICANN 2002*, Volume 2415 of *Lecture Notes in Computer Science*, Berlin, pp. 510–515. Springer-Verlag.

Bounkong, S., D. Saad, and D. Lowe (2002b). Quantisation effects and watermarking capacity. In *Proceedings of the Seventh International Symposium on Communications Theory and Applications*, pp. 215–220.

Bounkong, S., B. Toch, D. Saad, and D. Lowe (2003a, December). ICA for watermarking digital images. *Journal of Machine Learning Research* **4**, 1471–1498.

Bounkong, S., B. Toch, D. Saad, and D. Lowe (2003b). Structured codebooks for SCS watermarking. In *Proceedings of the Signal Processing, Pattern Recognition, and Applications Conference*, pp. 77–81. ACTA Press.

Cachin, C. (1998, April). An information-theoretic model for steganography. In *Information Hiding*, Volume 1525 of *Lecture Notes in Computer Science*, Berlin, pp. 306–318. Springer-Verlag.

Chen, B. (2000, June). *Design and analysis of digital watermarking, information embedding, and data hiding systsems*. Ph. D. thesis, Massachusetts Institute of Technology.

Chen, B. and G. Wornell (2001, May). Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory* **47**(4), 1423–1443.

Cohen, A. and A. Lapidoth (2002, June). The gaussian watermarking game. *IEEE Transactions on Information Theory* **48**(6), 1639–1667.

Conway, J. and N. Sloane (1982, March). Fast quantizing and decoding algorithms for lattice quantizers and codes. *IEEE Transactions on Information Theory* **28**(2), 227–232.

Cornsweet, T. (1970). *Visual Perception.* New York: Academic Press.

Costa, M. (1983, May). Writing on dirty paper. *IEEE Transactions on Information Theory* **29**(3), 439–441.

Cover, T. and J. Thomas (1991). *Elements of information theory.* New York: John Wiley and Sons.

Cox, I., J. Kilian, T. Leighton, and T. Shamoon (1997, December). Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Information Theory* **6**(12), 1673–1687.

Cox, I., M. Miller, and J. Bloom (2002). *Digital Watermarking.* San Fransisco: Morgan Kaufmann.

Craver, S. (1996, July). Can invisible watermarks resolve rightful ownership? Research report RC 20509, IBM.

Craver, S., N. Memon, B.-L. Yeo, and M. Yeung (1998, May). Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications. *IEEE Journal on Selected Areas in Communications* **16**(4), 573–586.

Delaigle, J., C. D. Vleeschouwer, and B. Macq (1998, May). Watermarking algorithm based on a human visual model. *Signal Processing* **3**(66), 319–335.

Eggers, J., R. Bauml, R. Tzschoppe, and B. Girod (2003, April). Scalar costa scheme for information embedding. *IEEE Transactions on Signal Processing* **51**(4), 1003–1019.

Eggers, J. and B. Girod (1999, September). Watermark detection after quantization attacks. In *Information Hiding*, Volume 1768 of *Lecture Notes in Computer Science*, Berlin, pp. 172–186. Springer-Verlag.

Eggers, J. and B. Girod (2001, February). Quantization effects on digital watermarks. *Signal Processing* **81**(2), 239–263.

Eggers, J. and B. Girod (2002). *Informed Watermarking.* London: Kluver Academic Publishers.

F.H. Hartung, J.K. Su, B. G. (1999, April). Spread spectrum watermarking: malicious attacks and counterattacks. In *Security and Watermarking of Multimedia Contents*, Volume 3657 of *Proceedings of SPIE*, Bellingham, pp. 44–51. SPIE – The International Society for Optical Engineering.

Field, D. (1994, July). What is the goal of sensory coding? *Neural Computation* **6**(4), 559–601.

Fridrich, J. (1998, July). Combining low-frequency and spread spectrum watermarking. In *Mathematics of Data/Image Coding, Compression, and Encryption*, Volume 3456 of *Proceedings of SPIE*, Bellingham. SPIE – The International Society for Optical Engineering.

G.D. Forney, J. (1988, September). Coset codes - part I: Introduction and geometrical classification. *IEEE Transactions on Information Theory* **34**(5), 1123–1151.

Girod, B. (1993). What's wrong with mean-squared error? In A. Watson (Ed.), *Digital images and human vision*, pp. 207–220. Cambridge, Massachusetts: The MIT Press.

González-Serrano, F., H. Molina-Bulla, and J. Murillo-Fuentes (2001, May). Independent component analysis applied to digital watermarking. In *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, Volume 3, New York, pp. 1997–2000. IEEE.

H.G. Feichtinger, T. S. (1998). *Gabor analysis and algorithms, theory and applications.* Boston: Birkhuser.

Hoyer, P. (1999, April). *Independent Component Analysis in Image Denoising.* Ph. D. thesis, Helsinki University of Technology.

Hyvarinen, A. (1999). Survey on independent component analysis. *Neural Computing Surveys* **2**, 94–128.

Hyvarinen, A., J. Karhunen, and E. Oja (2001). *Independent Component Analysis.* New York: John Wiley and Sons.

ISO/IEC JTC1/SC29/WG11 (1988). Coding of moving pictures and audio. Standard ISO/IEC-11172 and ISO/IEC-13818 and ISO/IEC-14496, ISO/IEC. http://mpeg.telecomitalialab.com/standards.htm.

Jain, A. (1989). *Fundamentals of Digital Image Processing.* London: Prentice Hall.

JPEG2000 Final committee (2000, March). Jpeg2000 part 1 draft version 1.0. technical report FCD15444-1, ISO/IEC.

Jung, S.-H., S. Miltra, and D. Mukherjee (1996, June). Subband DCT: definition, analysis, and applications. *IEEE Transactions on Circuits and Systems for Video Technology* **6**(3), 273–286.

Karakos, D. (2002, May). *Digital watermarking, fingerprinting and compression: an information-theoretic perspective.* Ph. D. thesis, University of Maryland.

Karakos, D. and A. Papamarcou (2000, December). Relationship between quantization and distribution rates of digitally fingerprinted data. UMD TR 2000-51, Institute for Systems Research Technical Report.

Katzenbeisser, S. and F. Petitcolas (2000). *Information hiding techniques for steganography and digital watermarking.* London: Artech House.

Kutter, M. and S. Winkler (2002, January). A vision-based masking model for spread-spectrum image watermarking. *IEEE Transactions on Signal Processing* **11**(1), 16–25.

Langelaar, G., I. Setyawan, and R. Lagendijk (2000, September). Watermarking digital image and video data. a state-of-the-art overview. *IEEE Signal Processing Magazine* **17**(5), 20–46.

Lee, T., M. Girolami, A. Bell, and T. Sejnowski (2000). A unifying information-theoretic framework for independent component analysis. *Computers and mathematics with applications* **39**, 1–21.

Legge, G. and J. Foley (1980, December). Contrast masking in human vision. *Journal of the Optical Society of America* **70**(12), 1458–1471.

Lewicki, M. and B. Olshausen (1999, July). A probabilistic framework for the adaptation and comparison of images codes. *Journal of the Optical Society of America* **16**(7), 1587–1601.

Lin, E., C. Podilchuk, and E. Delp (2000, May). Detection of image alterations using semi-fragile watermarks. In *Security and Watermarking of Multimedia Contents II*, Volume 3971 of *Proceedings of the SPIE*, Bellingham, pp. 152–163. SPIE – The International Society for Optical Engineering.

Lu, C.-S. and H.-Y. Liao (2000). Oblivious coctail watermarking by sparse code shrinkage: a regional and global-based scheme. In *Proceedings of the IEEE International Conference on Image Processing*, Volume 3, New York, pp. 13–16. IEEE.

Luby, M., M. Mitzenmachen, M. Shokrollahi, and D. Spielman (2001, February). Efficient erasure correcting codes. *IEEE Transactions on Information Theory* **47**(2), 569–584.

MacKay, D. (1999, February). Good error correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory* **45**(2), 399–431.

Malvar, H. and D. Florêncio (2003, April). Improved spread spectrum: a new modulation technique for robust watermarking. *IEEE Transactions on Signal Processing* **51**(4), 898–905.

Martinian, E. (2000, May). *Authenticating multimedia in the presence of noise.* Ph. D. thesis, Massachusetts Institute of Technology.

Meerwald, P. (2001a). Digital image watermarking in the wavelet transform domain. Master's thesis, Universität Salzburg.

Meerwald, P. (2001b, May). Quantization watermarking in the JPEG2000 coding pipeline. In *Communications and Multimedia Security Issues of The New Century*, IFIP International Federation for Information Processing, London, pp. 69–79. Kluver Academic Publishers.

Meerwald, P. (2002, May). Authentication watermarking and JPEG2000. Workshop on Multimedia Security and Watermarking.

Moulin, P. (2001, June). The role of information theory in watermarking and its application to image watermarking. *Signal Processing* **81**(6), 1121–1139.

Moulin, P. and J. O'Sullivan (2003, March). Information-theoretic analysis of information hiding. *IEEE Transactions on Information Theory* **49**(3), 563–593.

Nadenau, M., S. Winkler, D. Alleysson, and M. Kunt (2000, September). Human vision models for perceptually optimized image processing - a review. Submitted to Proceedings of the IEEE.

P.-C. Su, H.-J. W. and C.-C. Kuo (1999, April). Digital watermarking in regions of interest. In *Proceedings of the Conference on Image Processing, Image Quality and Image Capture Systems*, Springfield, VA 22151, pp. 295–300. IS&T - The Society for Imaging Science and Technology.

Peli, E. (1990, October). Contrast in complex images. *Journal of the Optical Society of America A* **7**(10), 2030–2040.

Pereira, S. and T. Pun (1999, September). Fast robust template matching for affine resistant image watermarking. In *Information Hiding*, Volume 1768 of *Lecture Notes in Computer Science*, Berlin, pp. 207–218. Springer-Verlag.

Pereira, S., S. Voloshynovskiy, M. Madueño, S. Marchand-Maillet, and T. Pun (2001, April). Second generation benchmarking and application oriented evaluation. In *Information Hiding*, Volume 2137 of *Lecture Notes in Computer Science*, Berlin, pp. 340–353. Springer-Verlag.

Petitcolas, F. (2000, September). Evaluation of copyright marking systems. *IEEE Signal Processing Magazine* **17**(5), 58–64.

Petitcolas, F. (2002). Stirmark 4.0. http://www.petitcolas.net/fabien/watermarking/stirmark/index.html.

Petitcolas, F., R. Anderson, and M. Kuhn (1998, April). Attacks on copyright marking systems. In *Information Hiding*, Volume 1525 of *Lecture Notes in Computer Science*, Berlin, pp. 219–239. Springer-Verlag.

Podilchuck, C. and W. Zeng (1998, May). Image-adaptative watermarking using visual models. *IEEE Journal on Selected Areas in Communications* **16**(4), 525–539.

Ramkumar, M. (2000, January). *Data hiding in multimedia - theory and applications*. Ph. D. thesis, New Jersey Institute of Technology.

Ruanaidh, J. Ó. and T. Pun (1997, October). Rotation, scale and translation invariant digital image watermarking. In *Proceedings of the IEEE International Conference on Image Processing*, Volume 1, New York, pp. 536–539. IEEE.

Said, A. and W. Pearlman (1996, June). A new fast and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology* **6**, 243–250.

Shapiro, J. (1993, December). Embedded image coding using zerotrees of wavelet coefficient. *IEEE Transactions on Signal Processing* **41**, 3445–3462.

Simoncelli, E. and B. Olshausen (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience* **24**, 1193–1216.

Sloane, N. (1981, May). Tables of sphere packings and spherical codes. *IEEE Transactions on Information Theory* **27**(3), 327–338.

Somekh-Baruch, A. and N. Merhav (2003, March). On the error exponent and capacity games of private watermarking systems. *IEEE Transactions on Information Theory* **49**(3), 537–562.

Taylor, C. (1998, December). *Image quality assessement based on a human visual system model*. Ph. D. thesis, Purdue university.

Toch, B., D. Lowe, and D. Saad (2003). Watermarking of audio signals using ICA. In *Proceedings of the Third International Conference on Web Delivering of Music*, pp. 71–74.

Voloshynovskiy, S., A. Herrigel, V. Iquise, and T. Pun (2001, April). Blur/deblur attack against document protection systems based on digital watermarking. In *Information Hiding*, Volume 1768 of *Lecture Notes in Computer Science*, Berlin, pp. 330–339. Springer-Verlag.

Voloshynovskiy, S., S. Pereira, V. Iquise, and T. Pun (2001, June). Attack modelling: Towards a second generation benchmark. *Signal Processing* **81**(6), 1177–1214.

Wallace, G. (1992, February). The JPEG still picture compression standard. *IEEE Transactions on Consumer Electronics* **38**(1), 18–34.

Wang, Z., A. Bovik, A. Sheikh, and E. Simoncelli (2004, January). Image quality assessement: From error measurement to structural similarity. *IEEE Transactions on Image Processing* **1**(1), –. To appear.

Watson, A. (1993, September). DCT quantization matrices optimized for individual images. In *Human Vision, Visual Processing, and Digital Display IV*, Volume 1913 of *Proceedings of SPIE*, Bellingham, pp. 202–216. SPIE – The International Society for Optical Engineering.

Wolfgang, R., C. Podilchuk, and E. Delp (1999, April). Perceptual watermarks for digital images and video. In *Security and Watermarking of Multimedia Contents*, Volume 3657 of *Proceedings of SPIE*, Bellingham, pp. 44–51. SPIE – The International Society for Optical Engineering.

Wong, P. and O. Au (2003, August). A capacity estimation technique for JPEG-to-JPEG image watermarking. *IEEE Transactions on Circuits and Systems for Video Technology* **13**, 746–752.

Xie, L. and G. Arce (1998, October). Joint wavelet compression and authentication watermarking. In *Proceedings of the IEEE International Conference on Image Processing*, Volume 2, New York, pp. 427–431. IEEE.

Appendix A

Proofs

A.1 Proof Prop. 1

Proof Prop. 1, Eq. 3.70: As $\mathscr{E}_1^i = 0$, Eq. 3.68 of Def. 24 gives that $\mathscr{E}_2^i > 0$. Then, assuming that $\mathscr{E}_k^i > 0$ with $k \in [2, \overline{k}^i]$, Eq. 3.68 results in $\mathscr{E}_{k+1}^i > \mathscr{E}_k^i$ and therefore $\mathscr{E}_{k+1}^i > 0$, thus by recurrence, we have Prop. 1, Eq. 3.70.□

Proof Prop. 1, Eq. 3.71: It follows from Def. 23 that

$$\mathscr{E}_{kk'}^i = \sum_{l=k}^{k'-1} \mathscr{E}_{l,l+1}^i \,, \tag{A.1}$$

moreover by construction $\mathscr{E}_{l,l+1}^i > 0$ as in Eq. 3.68, thus $\mathscr{E}_{kk'}^i > 0$.

Similarly, we have

$$\mathscr{D}_{kk'}^i = \sum_{l=k}^{k'-1} \mathscr{D}_{l,l+1}^i \,, \tag{A.2}$$

thus

$$\mathscr{D}_{kk'}^i < 0 \Rightarrow (\exists l \in [k, k'-1], \mathscr{D}_{l,l+1}^i < 0) \wedge (\mathscr{E}_{l,l+1}^i > 0) \,, \tag{A.3}$$

such that

$$\mathscr{G}_l^i < \frac{\mathscr{E}_{l-1,l+1}^i}{\mathscr{D}_{l-1,l+1}^i} \,, \tag{A.4}$$

which contradicts the maximality condition in Eq. 3.68. Therefore,

$$\forall l \in [1, \overline{k}_i], \mathscr{D}_{l,l+1}^i > 0 \,, \tag{A.5}$$

hence $\mathscr{D}_{kk'}^i > 0$. Thus, we have Prop. 1, Eq. 3.71.□

Proof Prop. 1, Eq. 3.72: It follows from Eq. 3.75 and from the above properties, that $\mathscr{G}_k^i$ is the ratio of two positive numbers, hence $\mathscr{G}_k^i > 0$. Thus, we have Prop. 1, Eq. 3.72.□

Proof Prop. 1, Eq. 3.73: If $\mathscr{G}_k^i < \mathscr{G}_{k+1}^i$, then

$$\mathscr{G}_k^i < \frac{\mathscr{E}_{k-1,k+1}^i}{\mathscr{D}_{k-1,k+1}^i} \,, \tag{A.6}$$

which contradicts the maximality condition in Eq. 3.68, hence $\mathscr{G}_k^i \geq \mathscr{G}_{k+1}^i$. Thus, we have Prop. 1, Eq. 3.73.□

Proof Prop. 1, Eq. 3.74: Property 1, Eq. 3.73 is straightforward from Prop. 1, Eq. 3.73 and the transitivity of the inequality.□

Proof Prop. 1, Eq. 3.76: From a geometrical point of view, Prop. 1, Eq. 3.76 means that all codewords $S = c_j$ with $j \in [1, n]$ are below the curve consisting in the line segments joining h_k to h_{k+1} with $k \in [1, \overline{k}^i]$ or are below $h_{\overline{k}^i}$ if $D_{ij} > \mathscr{D}_{1\overline{k}^i}^i$.

Let us prove the latter case first. If

$$(D_{ij} > \mathscr{D}^i_{1\overline{k}_i}) \wedge (E_{ij} > \mathscr{E}^i_{1\overline{k}_i}), \tag{A.7}$$

then the codeword $S = c_j$ satisfies the condition of Eq. 3.68 and therefore should be in the sequence $\{\mathscr{H}^i\}_k$, which contradicts the fact that $D_{ij} > \mathscr{D}^i_{1\overline{k}^i}$, and hence, $E_{ij} \leq \mathscr{E}^i_{1\overline{k}^i}$.

Now, let us assume that $\mathscr{D}^i_{1k} \leq D_{ij} \leq \mathscr{D}^i_{1,k+1}$, then from Eq. 3.68, we have

$$\mathscr{G}^i_k \geq \frac{E_{ij} - \mathscr{E}^i_{1k}}{D_{ij} - \mathscr{D}^i_{1k}}, \tag{A.8}$$

which gives Eq. 3.76 with

$$\lambda = \frac{D_{ij} - \mathscr{D}^i_{1k}}{\mathscr{D}^i_{k,k+1}}. \tag{A.9}$$

Thus, we have Prop. 1, Eq. 3.76.□

A.2 Proof Theorem 2

First, let us assume that

$$\sum_{j=1}^{n} p_{ij} D_{ij} > \mathscr{D}^i_{1\overline{k}^i}. \tag{A.10}$$

If

$$\sum_{j=1}^{n} p_{ij} E_{ij} > \mathscr{E}^i_{1\overline{k}^i}, \tag{A.11}$$

then $\exists j \in [1,n], E_{ij} > \mathscr{E}^i_{1\overline{k}^i}$, which contradicts Prop. 1, Eq. 3.74, therefore

$$\sum_{j=1}^{n} p_{ij} E_{ij} \leq \mathscr{E}^i_{1\overline{k}^i}. \tag{A.12}$$

Now, let us assume that

$$\mathscr{D}^i_{1,k^i-1} \leq \sum_{j=1}^{n} p_{ij} D_{ij} \leq \mathscr{D}^i_{1,k^i}, \tag{A.13}$$

then from Eq. 3.68, we have

$$\forall j \in [1,n], \mathscr{G}^i_{k^i-1} \geq \frac{E_{ij} - \mathscr{E}^i_{1,k^i-1}}{D_{ij} - \mathscr{D}^i_{1,k^i-1}}, \tag{A.14}$$

and therefore

$$\mathscr{G}^i_{k^i-1} \geq \frac{\sum_{j=1}^{n} p_{ij} E_{ij} - \mathscr{E}^i_{1,k^i-1}}{\sum_{j=1}^{n} p_{ij} D_{ij} - \mathscr{D}^i_{1,k^i-1}}, \tag{A.15}$$

which gives Eq. 3.79 with

$$\lambda = \frac{\sum_{j=1}^{n} p_{ij} D_{ij} - \mathscr{D}^i_{1,k^i-1}}{\mathscr{D}^i_{k^i-1,k^i}}. \tag{A.16}$$

Thus, we have Theorem 2.□

A.3 Proof Theorem 4

From Theorem 3, we have the first part of the proof. Then, let us assume f such that Eq. 3.85 is not satisfied for a given $i \in [1,n]$ and that for a given $j \in [1,n]$, $\mathscr{G}^j_{k_j-1} = \mathscr{G}$, and hence $\mathscr{G}^i_{k_i-1} > \mathscr{G}^j_{k_j-1}$. Then, let us modify f to obtain f' in the following way and show that f' has a greater good decoding probability for the same mean squared error distortion introduced, and therefore that f is not optimal.

To obtain f', we modify $p^i_{k^i-1}$, $p^i_{k^i}$, $p^j_{k^j-1}$ and $p^j_{k^j}$ such that

$$\hat{p}^i_{k^i-1} = p^i_{k^i-1} - \lambda\,, \tag{A.17}$$
$$\hat{p}^i_{k^i} = p^i_{k^i} + \lambda\,, \tag{A.18}$$
$$\hat{p}^j_{k^j-1} = p^j_{k^j-1} + \mu\,, \tag{A.19}$$
$$\hat{p}^j_{k^j} = p^j_{k^j} - \mu\,, \tag{A.20}$$

with

$$\lambda = p^i_{k^i-1}\min(1,\alpha)\,, \quad \mu = p^j_{k^j}\min(1,\alpha^{-1})\,, \quad \text{and} \quad \alpha = \frac{p_j p^j_{k^j}\mathscr{D}^j_{k^j,k^j-1}}{p_i p^i_{k^i-1}\mathscr{D}^i_{k^i,k^i-1}}\,. \tag{A.21}$$

By construction, the mean squared error distortion introduced by the embedding processes f and f' are equal, but as $\mathscr{G}^i_{k^i-1} > \mathscr{G}^j_{k^j-1}$, we have that

$$\mathrm{P}(R \in \overline{C}\,|\,S = f'(C), M = 1) > \mathrm{P}(R \in \overline{C}\,|\,S = f(C), M = 1)\,, \tag{A.22}$$

therefore the embedding process f cannot be optimal if Eq. 3.85 is not satisfied. Hence, the Theorem 4 is verified. □

A.4 Proof Theorem 5

From Theorem 4, we have the first part of the proof. Then, let us assume an embedding process f such that Eq. 3.86 is not satisfied for a given $(i,k) \in [1,n] \times [1,\overline{k}^i]$ and that for a given $j \in [1,n]$, $\mathscr{G}^j_{k^j-1} = \mathscr{G}$, and hence $\mathscr{G}^i_{k^i} > \mathscr{G}^j_{k^j-1}$. Then, let us modify f to obtain f' in the following way and show that f' has a greater good decoding probability for the same mean squared error distortion introduced, and therefore that the embedding process f is not optimal.

First of all, note that as $\mathscr{G}^i_k$ is decreasing when k increases, we have

$$\mathscr{G}^i_k > \mathscr{G} \Rightarrow \mathscr{G}^i_{k^i} > \mathscr{G}\,. \tag{A.23}$$

To obtain f', we modify $p^i_{k^i}$, $p^i_{k^i+1}$, $p^j_{k^j-1}$ and $p^j_{k^j}$ such that

$$\hat{p}^i_{k^i} = p^i_{k^i-1} - \lambda\,, \tag{A.24}$$
$$\hat{p}^i_{k^i+1} = \lambda\,, \tag{A.25}$$
$$\hat{p}^j_{k^j-1} = p^j_{k^j-1} + \mu\,, \tag{A.26}$$
$$\hat{p}^j_{k^j} = p^j_{k^j} - \mu\,, \tag{A.27}$$

with

$$\lambda = p^i_{k^i}\min(1,\alpha)\,, \quad \mu = p^j_{k^j}\min(1,\alpha^{-1})\,, \quad \text{and} \quad \alpha = \frac{p_j p^j_{k^j}\mathscr{D}^j_{k^j,k^j-1}}{p_i p^i_{k^i}\mathscr{D}^i_{k^i+1,k^i}}\,. \tag{A.28}$$

By construction, the mean squared error introduced by the embedding processes f and f' are equal, but as $\mathscr{G}^i_{k^i} > \mathscr{G}^j_{k^j-1}$, we have that

$$\mathrm{P}(R \in \overline{C}\,|\,S = f'(C), M = 1) > \mathrm{P}(R \in \overline{C}\,|\,S = f(C), M = 1)\,, \tag{A.29}$$

therefore the embedding process f cannot be optimal if Eq. 3.86 is not satisfied. Hence, the Theorem 5 is verified. □

A.5 Proof Theorem 6

In order to prove the theorem above, we show that the good decoding probability of an optimal embedding process f' in the sense of the good decoding probability is equal to the good decoding probability for the embedding process f which satisfies Prop. 4 and $D = \sigma^2_X$. Then, from Prop. 2, the good decoding

probability and mean squared error for the embedding processes f and f' can be re-expressed similarly as in

$$\mathrm{P}(R \in \overline{\mathscr{C}}|S = f(C), M = 1) = \sum_{i=1}^{n} p_i \left(\mathscr{E}^i_{k^i-1} + p^i_{k^i} \mathscr{D}^i_{k^i-1,k^i} \mathscr{G}^i_{k^i-1}\right), \tag{A.30}$$

$$= \sum_{i=1}^{n} p_i \Big(\underbrace{p^i_{k^i} \mathscr{D}^i_{k^i-1,k^i} \mathscr{G}^i_{k^i-1}}_{(1)} + \sum_{k=1}^{k^i-2} \mathscr{D}^i_{k,k+1} \mathscr{G}^i_k \Big), \tag{A.31}$$

$$D = \sum_{i=1}^{n} p_i \left(\mathscr{D}^i_{k^i-1} + p^i_{k^i} \mathscr{D}^i_{k^i-1,k^i}\right), \tag{A.32}$$

$$= \sum_{i=1}^{n} p_i \Big(\underbrace{p^i_{k^i} \mathscr{D}^i_{k^i-1,k^i}}_{(1)} + \sum_{k=1}^{k^i-2} \mathscr{D}^i_{k,k+1} \Big). \tag{A.33}$$

Now, let us show that $\mathscr{G} = \mathscr{G}'$, where $\mathscr{G}$ and $\mathscr{G}'$ are respectively associated with the embedding processes f and f' and defined as

$$\mathscr{G} = \min_{i \in [1,n]} \mathscr{G}^i_{k^i-1} \qquad \text{and} \qquad \mathscr{G}' = \min_{i \in [1,n]} \mathscr{G}^i_{k'^i-1}. \tag{A.34}$$

If $\mathscr{G} < \mathscr{G}'$, then it can be derived from Eq. 3.86 that Eq. A.33 part (1)[1] for f is greater than D' for f', thus $D > D'$. Similarly, if $\mathscr{G} > \mathscr{G}'$, then $D < D'$. Since $D = D' = \sigma^2_X$, we have $\mathscr{G} = \mathscr{G}'$. Thus, we can re-write similarly the good decoding probability and mean squared error for the embedding processes f and f' as follows

$$\mathrm{P}(R \in \overline{\mathscr{C}}|S = f(C)) = \underbrace{\sum_{i=1}^{n} p_i \sum_{\substack{k=1 \\ \mathscr{G}^i_k \neq \mathscr{G}}}^{k^i-1} \mathscr{D}^i_{k-1,k} \mathscr{G}^i_k}_{(1)} + \mathscr{G} \underbrace{\sum_{\substack{i=1 \\ \mathscr{G}^i_{k^i} = \mathscr{G}}}^{n} p_i p^i_{k^i} \mathscr{D}^i_{k^i-1,k^i}}_{(2)}, \tag{A.35}$$

$$D = \underbrace{\sum_{i=1}^{n} p_i \sum_{\substack{k=1 \\ \mathscr{G}^i_k \neq \mathscr{G}}}^{k^i-1} \mathscr{D}^i_{k-1,k}}_{(1)} + \underbrace{\sum_{\substack{i=1 \\ \mathscr{G}^i_{k^i} = \mathscr{G}}}^{n} p_i p^i_{k^i} \mathscr{D}^i_{k^i-1,k^i}}_{(2)}. \tag{A.36}$$

As Eq. A.36 part (1) is identical for D and D', Eq. A.36 part (2) for D and D' have the same value since $D = D'$. Therefore, Eq. A.35 part (1) and part (2) have the same values for the embedding processes f and f', hence

$$\mathrm{P}(R \in \overline{\mathscr{C}}|S = f(C), M = 1) = \mathrm{P}(R \in \overline{\mathscr{C}}|S = f'(C), M = 1), \tag{A.37}$$

and the embedding process f is optimal in the sense of the good decoding probability. □

[1] associated with Eq. A.31 part (1).