Wavelet Packets-based Digital Watermarking for Image Authentication

by

Alexandre Paquet

B.Eng, Université Laval, 2000

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE STUDIES

(Department of Electrical & Computer Engineering)

We accept this thesis as conforming
to the required standard

The University of British Columbia

July 2002

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical & Comp Engineering

The University of British Columbia
Vancouver, Canada

Date August 1, 2002

DE-6 (2/88)

Abstract

The rapid expansion of the Internet and the overall development of digital technologies in the past years have sharply increased the availability of digital media. Digital contents can be reproduced without loss of quality, but they may also be easily modified, and sometimes, imperceptibly. In many contexts, any alteration of image, video or audio data must be detected. Therefore, some work needs to be done to develop security systems to protect the content of digital data. Watermarking is accepted as a plausible candidate for such an application as it allows for the invisible insertion of information in a host by its imperceptible modification.

This thesis is concerned with the protection of information contained in digital images. A novel, semi-fragile watermarking technique for the authentication of images is developed. Image protection is achieved by the insertion of a secret author's identification key in an image's wavelet packet (WP) decomposition. Rounding the mean of selected regions of WP coefficients embeds the binary key. To take maximum advantage of the host image's characteristics in the embedding process, an optimal quantization protocol is formulated. The image's verification is done without the use of the original *unmarked* image. The detection of unauthorized frequency or spatial tampering with the image is performed by a combined interband/intraband verification protocol. This new technique can detect malicious tampering with images, but stays unaffected by high quality JPEG compression.

Contents

List of Tables

List of Figures

List of Abbreviations

DCT	Discrete Cosine Transform
DVD	Digital Versatile Disk
DW	Digital Watermarking
DWT	Discrete Wavelet Transform
FB	Filter Bank
FT	Fourier Transform
HVS	Human Visual System
JPEG	Joint Photographic Experts Group
LP	Low Pass
MPEG	Moving Picture Experts Group
MSQE	Mean Square Quantization Error
QF	Quality Factor
QMF	Quadrature Mirror Filter
WFA	Windowed Fourier Analysis
WP	Wavelet Packets
WPC	Wavelet Packet Coefficients
WT	Wavelet Transform

Glossary

Cover Work:	Media content that is to be watermarked. Synonym: *Host.*
Decomposition:	Fragmentation of a signal into a known arrangement of frequency band by the subsequent application of an analysis filter bank or by the repetitive dilation of a wavelet basis. Synonym: *Wavelet Decomposition.*
Decomposition Level:	State of the wavelet decomposition in terms of the stage in a filter bank tree.
Frequency Band:	Intervals of given width in the frequency spectrum. Synonym: *Band.*
Level of Detail:	Group of wavelet coefficients belonging to the same frequency band and forming a category of details in the signal reconstruction.
Resolution:	Number of levels applied to an original signal in its wavelet decomposition. Synonym: *Scale.*
Subband:	Specific region of a frequency band resulting from a signal decomposition.
Telltaling:	Used in the context of content authentication to identify the specific tampering process applied to an original host.

Acknowledgements

I would like to acknowledge the precious contribution of several people who have, knowingly or not, made the present thesis possible. My utmost debt of gratitude is to my supervisor, Professor Rabab K. Ward, for her continuous and kind support throughout my master's degree. She was always generous with her time and provided me with invaluable guidance in technical and professional matters. She has been particularly important in different writing stages and preparation of presentations, always providing me with advice, imprinted with wisdom earned from years in the field of image processing. I am also thankful to a fellow member of the image processing laboratory, Mehran Azimi, for his help on practical implications of our field of study and to Professor Ioannis Pitas, who provided me with insightful comments on my research. In addition, I want to express my gratitude to Professor Saif Zahir who directed the early stage of my work, and who gave me the organizational skills needed for its completion. I would also like to thank Dr. Hoss Ahmadi who was the first to introduce me to the world of wavelets. Through fruitful discussions on the subject, he is the one who, ultimately, got me interested in the field of digital watermarking. Finally, I have to thank the Natural Sciences and Engineering Research Council of Canada for its financial support, without which the completion of this degree would have been much more difficult.
To all, thank you.

ALEXANDRE PAQUET

The University of British Columbia

July 2002

"There is no such thing as a long piece of work, except one that you dare not start."

-Charles Baudelaire, writer and poet (1821-1867)

Chapter 1

Introduction

"There is a single light of science, and to brighten it anywhere is to brighten it everywhere."

-Isaac Asimov, writer (1920-1992)

1.1 Digital Watermarking

In the past decade, the apparition of digital cameras, both photographic and video, as well as CD-ROM and DVDs, has eased the creation, storage and visualization of digital multimedia. In addition, the developments of faster computers, combined with the augmentation of storage capacities and transmission speed, facilitate the overall utilization of digital technologies. This has created a real explosion in the use of digital data, and at least in terms of entertainment and media, we are now clearly living in a digital world.

Digital contents show great advantages in terms of storage and processing. Furthermore, they can be reproduced without loss of quality, and allow

for easy and imperceptible modifications. This permits the wide distribution of high quality music and video contents, and the production of incredibly real visual animations by the film industry. However, it also brings some problems: intellectual properties are harder to protect and so are original contents. Therefore, new practices must be developed in order to enhance the characteristics, to guard intellectual properties and to secure the content of digital data.

Digital watermarking is a relatively new technology that embeds hidden information in image, music, video or audio data by their imperceptible modification. It differs from cryptography since watermarking is about concealing the existence of secret information, while the former tries to protect it. Although insertion procedures are designed so that humans do not notice the marks inserted, computer programs can be created to extract the original marks easily. Afterwards, they can be used for copyright protection, broadcast monitoring, or even, in relation with an embedding pattern, for content authentication purposes.

Due to watermarking's wide range of applications and high potential, this sub-discipline of communication security has attracted a lot of interest in the last eight years. It has now evolved as an established candidate for copyright protection, ownership identification and fingerprinting systems. Moreover, several commercial applications of watermarking for copy control devices are planned, or are already implemented. For all these contexts, a lot of effort is dedicated to the development of *robust watermarking* schemes that

permanently mark the works. On the other hand, the use of *fragile* embedding schemes-ones where the embedded key, that is, the mark, is destroyed by the modification of the work-is much less investigated. Nevertheless, this kind of system shows great promise for content authentication as it allows for the validation of digital data, thus giving it legal value. As digital media are now widely employed and commonly accepted as official documents, protection of their informative content will grow as an important issue, as with the protection of intellectual property in the past years.

1.2 Problem Definition

In many applications, such as courtroom evidence and video security systems, any modification of image, video or audio data must be detected if it cannot be prevented. As digital images are widely available, online or elsewhere, and because they are so easy to modify, some work needs to be done to protect the information they contain. As the number of images increases, the direct storage of unique *reference* patterns becomes impractical. Moreover, as some images need to be slightly compressed in order to be efficiently stored, authentication systems need to offer flexibility. Unfortunately, many of the approaches previously proposed lack this characteristic, while others require too much user interaction to be truly considered secure for commercial applications.

In this dissertation, we introduce a novel technique for the content authentication of digital images. The new approach is able to detect, as well as

localize, malicious image alterations, while offering robustness to high quality image compression. Our method is based on semi-fragile watermarking technology. It uses the knowledge of characteristics of the human visual system to round discrete *wavelet packet* coefficients from an images' decomposition to optimal quantization levels. Tampering detection is performed using *intra* and *inter* frequency band verifications. Combined together, they allow for the detection of possible alterations done either in the frequency or spatial domain, while rejecting low-level perturbation resulting from storage operations.

1.3 Organization of the Thesis

The goal of the thesis is to find a watermarking method that can detect, as well as localize, tampering in digital images. We first review earlier work on digital watermarking, and then develop a semi-fragile watermarking scheme for image authentication. In order to make the thesis complete, we give an overview of image processing techniques used, and a more extensive background on watermarking technologies. The thesis is organized as follows.

In Chapter 2, we go over the basis of wavelets. We first give the historical background and explain how wavelets came to be. The importance of multiresolution concepts for wavelets is then demonstrated. Starting from series expansion principles, and using quadrature mirror filter banks implementations, we explain how wavelets can be used for multiresolution decomposition of one-dimensional signals. *Wavelet packets*, a particular kind of wavelet decomposition that separates signals in symmetrical levels of detail, and that is

used in our watermarking system, are then described. Finally, we extend the use of wavelet and wavelet packets to multidimensional signals using separable transform concepts.

In Chapter 3, the bases of watermarking technologies are given. In the first section, we draw a portrait of the concepts leading to the use of watermarking for digital media by retracing its history. Then, we give a background on watermarking by explaining the general concepts. In that sense, generic classifications are given based on the host media, the application intended as well as the embedding domain, and the decoding procedure used. Specific requirements of different systems are also listed. To conclude, we review cornerstone papers on digital watermarking that have first laid the conceptual bases of the technology.

In the following chapter, specificities of content authentication systems are described, and our *WP-based* technique is introduced. To start, we come up with a classification of authentication approaches considered. From this, we draw the requirements that such systems should fulfill in order to be effective and efficient. Then, we detail specific methods introduced that have served as bases in the development of our own system. We emphasize two different families of embedding protocols: those acting in the spatial domains, and the others, acting in some transform domain. The pros and cons of each are highlighted through the examination of published work. Finally, a novel image authentication approach based on the quantization of wavelet packet coefficients is introduced in the last section of Chapter 4.

Afterwards, we present experimental results in Chapter 5. We first confirm the invisibility of the embedded marks, as well as the authentication capabilities of our system. Then, we prove its ability to detect and localize tampering, both in space and frequency, even in the presence of compression. Afterwards, in order to have more objective evaluation criteria, we compare our system with a commercially available watermarking tool in terms of tampering localization aptitude and resistance to attacks, and demonstrate that our system outperforms the commercial software. Finally, different strategies meant to increase the robustness of the previously presented system are examined in Section 5.5.

To finish, a summary of our work and the major results obtained, as well as future possible research work in the field of digital watermarking for image authentication, are presented in Chapter 6.

Chapter 2

Wavelet Analysis

"The problems that exist in the world today cannot be solved by the level of thinking that created them."

-Albert Einstein, physicist (1879-1955)

2.1 Introduction

Although the average person probably knows very little about wavelets, their impact on today's technological world is phenomenal. They represent a very powerful mathematical tool commonly used by scientists and engineers, and are currently applied in fields such as signal processing, computer vision and data compression. Several new applications of wavelets are discovered every year and will continue to be in the future.

The purpose of this chapter is to provide a solid understanding of wavelet transforms. Since we wish to make the present thesis readable, we avoid going too deep into the mathematical details of the approaches developed

throughout the years. We first highlight the cornerstones leading to modern wavelet theory. Then, in Section 2.3 we present the theoretical fundamentals of wavelet analysis for multiresolution decomposition. Finally, we introduce the concepts of wavelet packet decomposition used in our watermarking system before we extend the use of wavelets to two-dimensional signals.

2.2 Historical Perspective

The first known step toward the development of a unified wavelet theory occurred when a Hungarian mathematician named Alfred Haar completed his work on the orthogonal systems of functions. In 1910, he proposed the use of piecewise constant functions to form an orthogonal basis. His system uses a basis function (now referred to as the scaling function ϕ) as a starting point. Then, the mother's (ψ), daughters', sons', granddaughters' and grandsons' (and so on) functions are obtained by the subsequent scaling and translation of the basis, or father wavelet. Haar proved that the obtained set of functions can be used to represent a signal at different levels of detail [31]. Furthermore, he demonstrated that a decomposed signal can be reconstructed using the reverse operations. Although it was not called "*wavelets*" back then, the simplest of the wavelet families was nonetheless born, and is now named the Haar wavelet.

Half a century later, some interest was devoted to the study of Windowed Fourier Analysis (WFA) in order to achieve both spatial *and* frequency localization in signal decomposition. As it decomposes a signal into a sum of

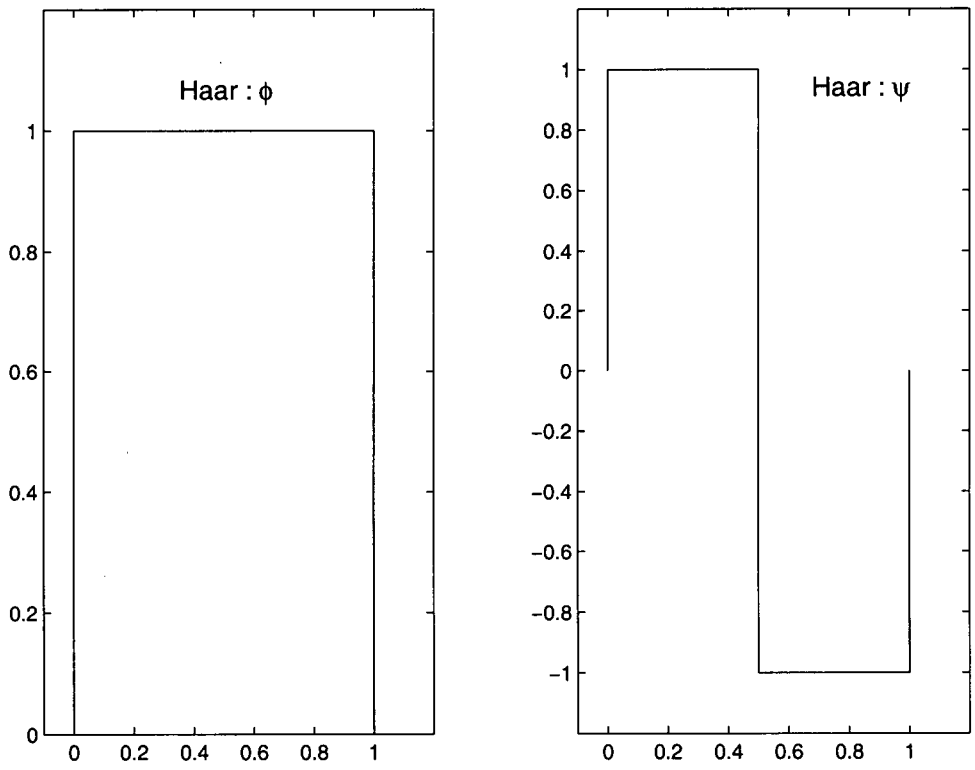


Figure 2.1: Haar Scaling (ϕ) and Wavelet (ψ) Functions

sines and cosines, which are both infinite in time, the standard Fourier decomposition lacks the time[1] localization necessary for the accurate analysis of several real signals. The idea of WFA is to study the frequencies of a signal for time-limited windows. This allows for some time localization of the frequency characteristics of a given signal. WFA concept finally allowed the examination of things in terms of both time and frequency.

Nonetheless, Haar remained the only example of a wavelet, and the next major advancements did not come until later in the 1980s. Jean Morlet and Alex Grossman teamed up in 1981. Together, they discovered that a signal could be transformed into wavelet form, and then synthesized back into the original signal without any loss of information. Then, in 1984, they were the first to use the term *wavelet* to describe their functions [30]. More specifically, they were called *Wavelets of Constant Slope.* Other researchers had used the term wavelets for different signal processing applications (see [61] for example) but Morlet and Grossman were the first to use it as it is now currently referred to, which is as follows:

> a wavelet is a unique function, limited in time and frequency, that can be translated and dilated to form multiresolution basis used to decompose a signal at different levels.

In addition, their major contribution was the finding of a simple signal recomposition method from its wavelet coefficients. They also discovered an-

[1]Time and space will be used alternatively throughout this thesis as time is, in fact, but only one possible space representation. However, since a lot of concepts used have first been developed for time-dependant signals, we find it helpful to use the same notation.

other interesting thing that is now commonly used in wavelet-based coding: *a small modification in the wavelet coefficients only causes a small change in the original signal.* This might not have appeared to be especially meaningful at the time, but when considering that modern wavelet-based compression schemes quantize wavelet coefficients, if it had been otherwise, data compression would be a much more difficult task today.

The real breakthrough in wavelets analysis, however, happened in the late 1980's when a lot of papers now considered classic were published. Yves Meyer and Stéphane Mallat were two important contributors to this newborn field. Investigating the use of wavelets in many different applied fields, they were amongst the first to develop the concept of multiresolution analysis for wavelets [49]. This was an important step for the advancement of research on wavelets. As a result, multiresolution is now an extensively used signal decomposition approach. Mallat and Meyer were the first to mention scaling functions of wavelets, which allow researchers and mathematicians to construct their own wavelets using established criteria [80].

Around the same time, a Belgian physicist named Ingrid Daubechies employed multiresolution analysis to create her own family of wavelets. Using construction methods related to filter banks, she introduced in [23] a family of compactly supported orthogonal wavelet systems with arbitrarily high, but fixed regularity. These wavelets offer a number of desirable properties (such as compact support, orthogonality, regularity, and continuity) that make them trully attractive[2]. This is why the *Daubechies Wavelets* are now some of the

[2]More on this in the next section.

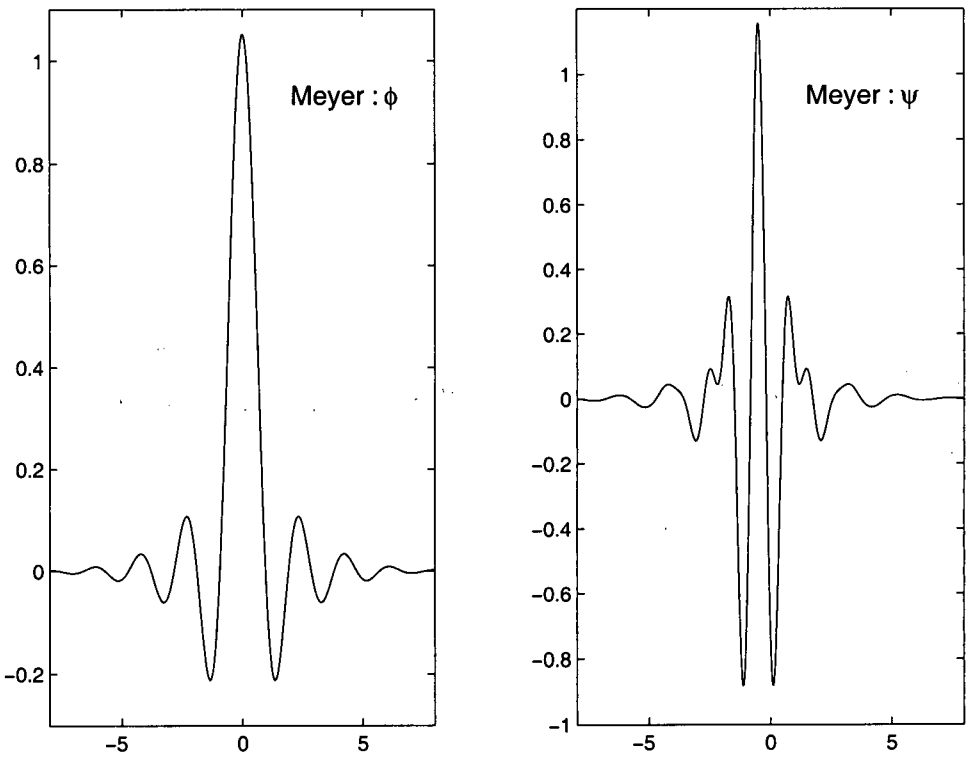


Figure 2.2: Meyer Scaling (ϕ) and Wavelet (ψ) Functions

most common ones today.

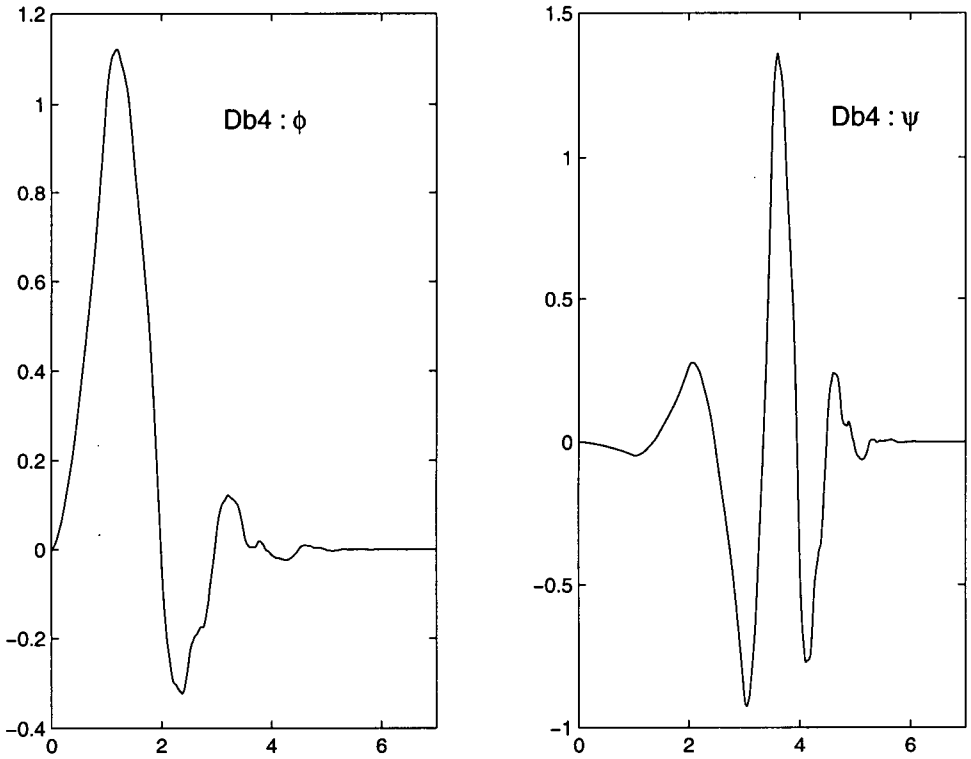


Figure 2.3: Daubechies-4 Scaling (ϕ) and Wavelet (ψ) Functions

Daubechies' work was probably the starting point of much focused research on wavelets that has lead to their acceptance as a modern mathematical tool and their wide use in sciences and engineering. Of course, many other researchers have contributed to the advancement of the field in the last decade, and several applications have been found. In particular, wavelet transforms prove to be extremely effective for image coding, and upcoming image compression standards-such as JPEG-2000-make use of them. From this, it is clear that wavelets are definitely a tool for the future, and this is why the knowledge

of their historical and theoretical bases is of great interest.

2.3 Wavelets and Multiresolution Analysis

In her now classic book [24], Ingrid Daubechies defines the wavelet transform as the following:

> a tool that cuts up data or functions or operators into different frequency components and then studies each component with a resolution matched to its scale.

The wavelet transform of a signal evolving in time depends on two variables: frequency and time. Therefore, these transforms provide an accurate tool for time-frequency localization. This is the most important factor that explains why wavelet transforms have already attracted so much attention.

Extensive publications on the general theory of wavelets are found in [24, 69, 80], while [68] details their specific applications to image processing. We now briefly examine the concepts linked with series expansion, and subsequently, the theory behind filter bank analysis. Finally, we introduce the idea of multiresolution and its application to wavelet decomposition.

2.3.1 Series Expansion

The goal of series expansion is to represent a signal or function as a combination of bases. Essentially, it means that we want to find a set of elementary signals $\{\varphi_i\}_{i\in\mathcal{Z}}$ so that we can write an original signal x, as a linear combination

of the following basis:

$$x = \sum_i \alpha_i \varphi_i \tag{2.1}$$

where the expansion coefficients α_i's can be obtained by the computation of the inner product of the basis dual set $\{\tilde{\varphi}_i\}$ with the signal x as follows:

$$\alpha_i = \sum_n \tilde{\varphi}_i[n]x[n] \tag{2.2}$$

When the set $\{\varphi_i\}$ is orthonormal and complete[3], we have an *orthonormal basis* and the basis, and its dual are the same, that is, $\varphi_i = \tilde{\varphi}_i$. Therefore, it means that the expansion coefficients can be found directly with the basis coefficients, as follows:

$$\alpha_i = \sum_n \varphi_i[n]x[n] \tag{2.3}$$

This principle is used in Fourier series decomposition in order to describe periodic signals by the combination of harmonically related sinusoids. This gives a perception of a signal in the frequency domain, in terms of their frequency content (see A.1). Fourier series representation is among the most popular series expansion techniques. Two main reasons motivate its use: one, it is easily implemented; and two, it yields the smallest mean square error (*MSE*) in power between the signal x and its series representation [3]. However, Fourier series are limited to periodic signals, a fact that prevents its use in several signal-processing applications. Nevertheless, the harmonic decomposition principle is used to compute the Fourier Transform (FT), which extracts the frequency content of a signal, periodic or not (Equations A.4 and

[3]A set $\{\varphi_i\}$ is considered complete if all the signals x in the representation space $\mathcal{S}$ can be expanded as in Equation 2.2.

A.5). The FT allows us to go from the time domain to the frequency domain and back, without the loss of information[4].

From this, signal decomposition using series expansion (and associated transforms) presents great advantages. It allows for the extraction of information-in the FT example about frequency content-that would not be available from the direct examination of the signal. The choice of the basis $\{\varphi_i\}$ used determines what characteristics are scrutinized in the analysis. The FT is a very powerful tool when we are interested in the frequency spectrum of a signal. However, it lacks the time localization needed in a lot of applications. This explains the amount of work done to develop other bases. In this context, it is important to note that wavelets are a specific kind of basis that allows for signal decomposition. In the next subsection, we introduce filter banks, a concept that simplifies the notions of multiresolution presented in Subsection 2.3.3.

2.3.2 Filter Banks

A multirate *filter bank* is a set of (M) parallel filters having either the same input or output. When the filters are used to split a common input (x), it is referred to as an *analysis bank*. On the other hand, it is called a *synthesis bank* when it is used to recombine split inputs to form one common output ($\hat{x}$ or x_r). An important feature of multirate filter banks is that they split a signal into different frequency bands. Moreover, perfect reconstruction from the de-

[4]This holds only as long as the entire signal is known, a fact that can be highly problematic for real-time applications.

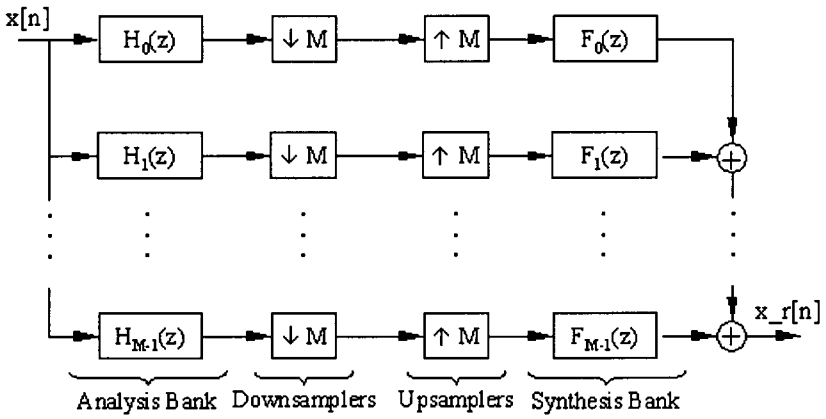


Figure 2.4: Multirate Quadrature Mirror Filter bank

composition is maintained as long as specific filter requirements are fulfilled. Therefore, filter banks (FB) play an important role in the study of wavelets. In fact, wavelet transforms can be constructed and easily implemented by the use of a particular class of filter bank systems developed by Croisier et al. in [22] and known as *quadrature mirror filter* (QMF) banks (Figure 2.4). Thorough design of QMF is outside the scope of our project, but it is important to have a general comprehension of filter banks in order to understand how wavelets allow the decomposition and the perfect reconstruction of signals. For that reason, we highlight the main steps leading to the construction of wavelets through the design of filter banks.

Since wavelets rely on octave-based decomposition, we are interested only in two-channel filter banks, hence dropping the concept of multirate. However, the reader must understand that the concepts explained in this section can be generalized to multirate systems as well. Several books are pub-

lished on the subject, and we invite interested people to consult [80] for more theoretical information. In our development, we first apply the concepts developed in 2.3.1 to Haar decomposition, and explain how a filter bank can be obtained with this particular basis. Then, we present the filter's requirements for the development of other bases.

$$\varphi_{2k}[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } n = 2k,\ 2k+1 \\ 0 & \text{otherwise} \end{cases}$$

$$\varphi_{2k+1}[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } n = 2k \\ -\frac{1}{\sqrt{2}} & \text{for } n = 2k+1 \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

$$\varphi_{2k}[n] = \varphi_0[n-2k] \quad \varphi_{2k+1}[n] = \varphi_1[n-2k] \tag{2.5}$$

Haar first came up with the bases presented in Equation 2.4, and corresponding to the scaling function (father wavelet) and mother wavelet respectively. This orthonormal basis (see A.2) has been shown to be suitable for wavelet decomposition as the basis functions are translates of each other (Equation 2.5). Particularly, those functions can easily apply to FB representation. First, the transform X associated with the basis has to be defined as follows:

$$\begin{aligned} X[2k] = \langle \varphi_{2k}, x \rangle &= \tfrac{1}{\sqrt{2}}(x[2k] + x[2k+1]) \\ X[2k+1] = \langle \varphi_{2k+1}, x \rangle &= \tfrac{1}{\sqrt{2}}(x[2k] - x[2k+1]) \end{aligned} \tag{2.6}$$

which allows the perfect reconstruction of the signal $x[n]$ from the following:

$$x[n] = \sum_{k \epsilon \mathcal{Z}} X[k]\varphi_k[n] \tag{2.7}$$

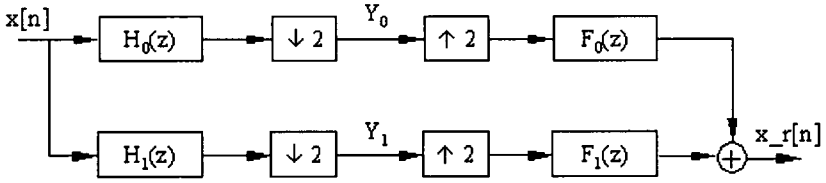


Figure 2.5: Two Channels QMF Bank

Once we have an appropriate basis for signal decomposition, we need to get the proper filters to obtain the simplified Figure 2.5. From it, we have that $y_0[k] = h_0[n] * x[n]|_{n=2k}$[5] and $y_1[k] = h_1[n] * x[n]|_{n=2k}$. Since the transform is already defined for the Haar basis as a function of $X[2k]$ and $X[2k+1]$, we want $y_0[k] = X[2k]$ and $y_1[k] = X[2k+1]$. Fortunately, the extraction of the corresponding filters $h_0[n]$ and $h_1[n]$ is pretty straightforward from the following:

$$X[2k] = h_0[n] * x[n]|_{n=2k} = \sum_{l \epsilon \mathcal{Z}} h_0[2k-l]x[l] = \frac{1}{\sqrt{2}}x[2k] + \frac{1}{\sqrt{2}}x[2k+1] \tag{2.8}$$

$$X[2k+1] = h_1[n] * x[n]|_{n=2k} = \sum_{l \epsilon \mathcal{Z}} h_1[2k-l]x[l] = \frac{1}{\sqrt{2}}x[2k] - \frac{1}{\sqrt{2}}x[2k+1] \tag{2.9}$$

and so, the analysis filter bank is defined with the impulse response of the following associated filters:

$$h_0[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } n = -1,\ 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.10}$$

[5]First, the $n = 2k$ results from the downsampling operation. Second, the filters are identified $H_0(z)$ and $H_1(z)$ in the figure. $\mathcal{Z}$-domain representation is used simply by convention.

$$h_1[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } n = 0 \\ -\frac{1}{\sqrt{2}} & \text{for } n = -1 \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

It is clear that the impulse responses found are time-reversed versions of the basis functions (which correspond to the (Haar) scaling (ϕ) and wavelet (ψ) functions of Figure 2.1). This comes from the fact that convolution is an inner product with time reversal, giving the following:

$$h_0[n] = \varphi_0[-n] \quad \text{and} \quad h_1[n] = \varphi_1[-n] \tag{2.12}$$

At this point, we are able to decompose the original signal $x[n]$ in two half length signals $y_0[n]$ and $y_1[n]$. However, this is only half the way since our goal is to reconstruct the signal $\hat{x}$ using synthesis filters $f_0[n]$ and $f_1[n]$ (see Figure 2.5). Keeping in mind that the downsampling operation has to be reversed (by upsampling) in order to restore the original length of the signal, and using the series expansion procedure given in 2.2, we directly have the following:

$$\hat{x}[n] = \sum_{k\epsilon\mathcal{Z}} y_0[k]\varphi_{2k}[n] + \sum_{k\epsilon\mathcal{Z}} y_1[k]\varphi_{2k+1}[n] \tag{2.13}$$

For QMF, 2.13 yields perfect reconstruction ($\hat{x}[n] = x[n]$), and the y's are transforms of x using Haar's basis set (i.e. $y_0[k] = X[2k]$ and $y_1[k] = X[2k+1]$). Equation 2.13 thus becomes the following:

$$x[n] = \sum_{k\epsilon\mathcal{Z}} X[2k]\varphi_{2k}[n] + \sum_{k\epsilon\mathcal{Z}} X[2k+1]\varphi_{2k+1}[n] \tag{2.14}$$

Keeping in mind that the reconstruction operation includes upsampling and convolution, the previous equation (2.14) allows the extraction of the

following:

$$\varphi_{2k}[n] = f_0[n-2k] \quad \text{and} \quad \varphi_{2k+1}[n] = f_1[n-2k] \tag{2.15}$$

Therefore, the synthesis filter bank is defined by the following filters' impulse responses:

$$f_0[n] = \varphi_0[n] \quad \text{and} \quad f_1[n] = \varphi_1[n] \tag{2.16}$$

To sum up, we have shown that it is possible to use a simple basis in order to construct a complete two-channel filter bank, which allows decomposition and perfect restoration of the original signal. Furthermore, we have found that, for the Haar basis, Equations 2.12 and 2.16 describe the system completely. Those equations can be made more general in order to fit other systems. In fact, the two following generalizations[6] (from [80]) are used as requirements for the definition of another set of basis $\{\varphi_i\}$.

1. The impulse responses of the synthesis filters equal the first set of basis functions, as follows:

$$f_i[n] = \varphi_i[n] \quad i = 0, 1. \tag{2.17}$$

2. The impulse response of the analysis filters are the time-reversed versions of the synthesis ones, as follows:

$$h_i[n] = f_i[-n] \quad i = 0, 1. \tag{2.18}$$

[6] Which, in fact, represent the first prerequisites of QMF.

Equations 2.17 and 2.18 can be relaxed and rewritten in the $\mathcal{Z}$ domain to form the perfect reconstruction condition on the following filters (see A.3):

$$F_0(z)H_0(z) + F_1(z)H_1(z) = I \tag{2.19}$$

In retrospect, we have made the link between series expansion and filter bank analysis. From this, we are now ready to define more two-channel filter banks, and extend the principles to general multiresolution wavelet analysis. Nonetheless, one more aspect needs to be covered in order to have a better understanding of the physical implication of the signal decomposition. We need to know what the characteristics of the signals $y_0[k]$ and $y_1[k]$ are in relation to the original signal $x[n]$. Taking the Haar decomposition case again, we have the first intermediate signal as follows:

$$y_0[k] = X[2k] = \frac{1}{\sqrt{2}}(x[2k] + x[2k+1]) \tag{2.20}$$

which corresponds to the output of an averaging-therefore low pass-filter. On the other hand, the second intermediate signal $y_1[k]$ is as follows:

$$y_1[k] = X[2k+1] = \frac{1}{\sqrt{2}}(x[2k] - x[2k+1]) \tag{2.21}$$

The later gives the difference between two successive samples of the original signal $x[n]$. Therefore, it is a high pass filter. It means that the outputs of the analysis stage of a two channel filter bank are one, a low pass filtered version of the original signal and two, a high pass filtered version of the original signal.

To summarize, two channel filter banks allow for the decomposition of a signal into two different *subband* signals at each stage: one that describes

the coarse behaviour of the original signal (low frequency), while the other contains the details of the signal (high frequency). Furthermore, it allows for the reconstruction of the original signal without the loss of information. As first Haar, then Morlet, Grossman et al. have noticed, this offers very promising capacities. In fact, these are exploited in the design and use of multiresolution systems, the subject of the following section.

2.3.3 Multiresolution and Wavelet Theory

At this point, we want to use the concepts developed in 2.3.1 and 2.3.2 to extend the bases of wavelet analysis. Once again, our goal is not to go too deep in the theoretical details. A thorough mathematical approach to filter bank analysis and wavelet decomposition is available in [1]. The purpose of this section, and in fact of the entire chapter, is to make the present thesis self-sustained; that is, we want to be sure that the reader has a general comprehension of wavelets and their applications in order to understand their use in our digital watermarking scheme.

The first thing to be comfortable with is the concept of multiresolution. As seen in the previous section, a signal can be represented by a coarse approximation y_0, plus added details y_1. Mallat [48] and Meyer [53] first showed that the detail signal $y_1[n]$ is the difference between the original one (i.e. $x[n]$) and its coarse representation $y_0[n]$. From this, the coarse and detail subspaces are orthogonal to each other. This permits the recursive application of successive decomposition at all resolutions, which results in the division of the

original signal in narrower levels of detail. A multiresolution analysis consists of a sequence of embedded closed subspace V_i, such as the following[7]:

$$... \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2}... \tag{2.22}$$

Next, we need to know what multiresolution means in terms of filter banks. From Equation 2.22 and the arguments around it, there is nothing to prevent the use of additional two-channel QMF banks to further decompose some or all of the subband signals y_i, and then again decompose the resulting signals. The idea can be applied recursively to form a filter bank with a tree structure. If only the approximation signal is decomposed into two subbands at any given resolution (using the same basic QMF and downsamplers), it yields an octave band filter bank. Furthermore, it has been shown that, if the analysis filters $H_0(z)$ and $H_1(z)$ satisfy certain regularity conditions [48], the filter bank representation can be used to compute a wavelet decomposition (and reconstruction) of the original signal $x[n]$.

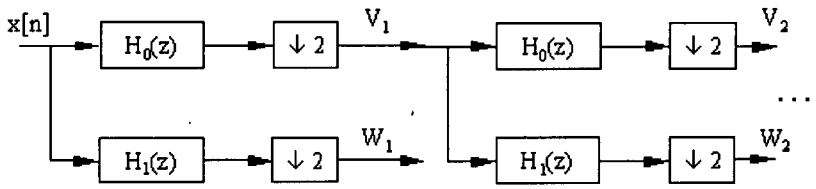


Figure 2.6: Two Levels of Wavelet Decomposition using Filter Bank Representation

The multiresolution decomposition is described in terms of subspaces V_j and W_j, which relate to the intermediate signals y_0 and y_1, as seen in

[7] A formal definition of multiresolution is found in [80] (See also A.4).

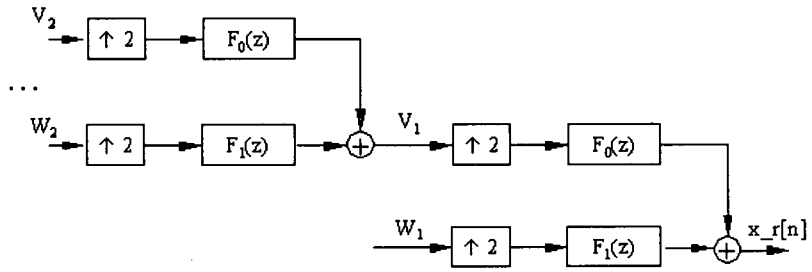


Figure 2.7: Two Levels of Wavelet Recomposition using Filter Bank Representation

Section 2.3.2, and, hence, to the scaling ($\phi(t)$) and wavelet ($\psi(t)$) functions. As the number of decomposition levels used increases, the subspace number j increases as well[8]. The wavelet space W_j corresponds to the difference between the present scaling space V_j and previous one V_{j-1}. It means that $V_j \oplus W_j = V_{j-1}$. This is shown in Figures 2.6 and 2.7, and graphically generalized in the frequency domain by Figure 2.8.

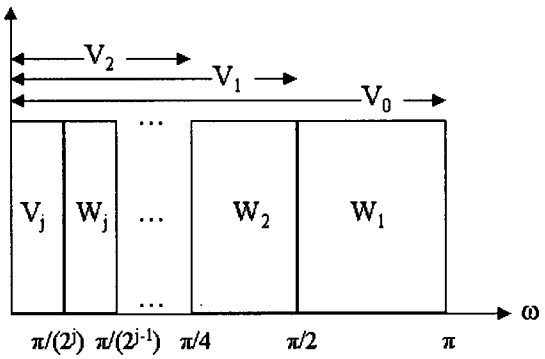


Figure 2.8: Ideal Spectrum Division from Wavelet Decomposition

Up to now, we have seen that the analysis side of the octave band filter

[8]We use the same notation as [24] and [69] but several others have been used; for example [80] employs exactly the opposite convention.

bank calculates the forward wavelet transform, while the synthesis side calculates the inverse wavelet transform. Therefore, a two-channel QMF bank can be directly used to form wavelet decomposition. This is probably the most frequently employed method of designing and implementing wavelet transforms. It is nonetheless important to summarize the more theoretical approach to wavelet decomposition, which is more related to series expansion than filter banks.

In order to account for the fact that the signal has to be decomposed into two bands at each level, two basis functions have to be defined. At each scale, a scaling function $\phi(t)$ is defined in addition to the wavelet function $\psi(t)$. The first one acting like the low pass filter H_0, while the second one is linked to the high pass filter H_1. Furthermore, in order to keep the total length of the decomposed signals equal to the length of the original signal-a fact that is taken care of by the downsampling operation in the filter bank representation-and to increase the definition at each level, a dilation operation needs to be performed (see Equation A.13) on the original basis. Finally, given the original basis ϕ, the scaling and wavelet functions at level j are as follows:

$$\phi^{(j)}(t) = 2 \sum_{k=-\infty}^{\infty} h_0(k)\phi(2^j t - k) \tag{2.23}$$

$$\psi^{(j)}(t) = 2 \sum_{k=-\infty}^{\infty} h_1(k)\phi(2^j t - k) \tag{2.24}$$

Consequently, the appropriate initial filter h_0 and scaling function $\phi^{(0)}$, make possible the definition of different scaling ($\phi(t)$) and wavelet ($\psi(t)$) functions by iteration of the dilation equations. Of course, the choice of the original

basis is of primordial importance, but as stated earlier, this is not the concern of the present thesis. In addition, a lot of work already exists in the field and that gives us enough material to work with. We introduce, in A.5, the six steps towards multiresolution, as proposed in [69]. Here, we present the Haar scaling and wavelet functions for V_0 and W_0 as they are shown in Figure 2.1:

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & \text{otherwise} \end{cases} \tag{2.25}$$

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise} \end{cases} \tag{2.26}$$

One final point needs to be made in order to complete our discussion about multiresolution wavelet transform. What is the advantage of this specific form of signal decomposition over others? Figure 2.9 clearly shows that wavelet decomposition allows both frequency and time localization, while Fourier Transform does not. For FT, Δf and Δt are fixed, even if the signal is first windowed in time. On the other hand, wavelet transform gives scalable time/frequency resolution as $\Delta\omega \propto 2^j$ and $\Delta t \propto 2^{-j}$, thus allowing the choice of more or less levels of decomposition in accordance with the importance of each resolution. Finally, as different bases exist, it is also possible to choose and/or adapt a particular basis according to the application considered.

This section reviewed the fundamentals of multiresolution analysis and its application to wavelet theory. We explained how multiresolution concepts are applied to octave band filter bank systems for the implementation of

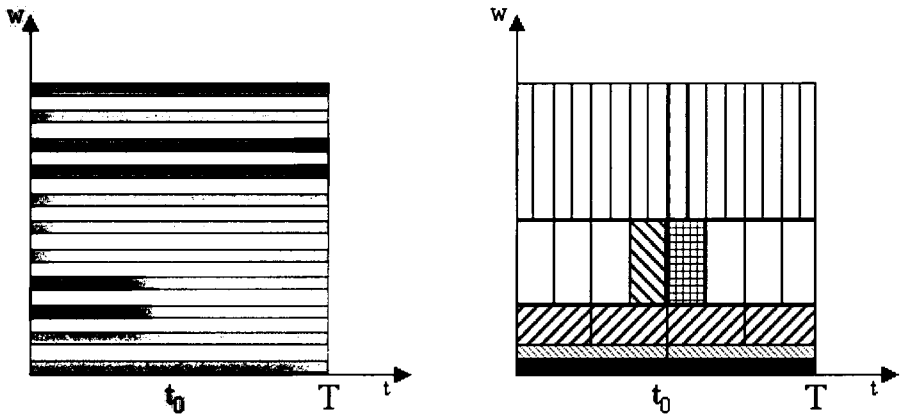


Figure 2.9: Frequency Tilling for Fourier and Wavelet Transforms

wavelet transforms. In addition, we highlighted the main steps leading to the computation of scaling and wavelet functions from a given basis by dilation operations. Finally, we showed the time-frequency localization allowed by wavelet transform, and the degree of freedom granted by the existence of a wide range of basis. As a result, the advantages provided by the relatively new wavelet transform over the more traditional Fourier analysis should be very clear.

2.4 Wavelet Packet Analysis

We just showed that the wavelet transforms offer several advantages in terms of localization, resolution and flexibility compared to the Fourier transform. In addition, we demonstrated that wavelet transforms can be easily implemented by the use of octave band QMF banks. This means that, by recursive filtering on the signal's coarse approximation, it is possible to achieve efficient and

simple standard wavelet analysis. This is of great importance as it allows for both efficient and accurate signal decomposition based on solid mathematical definitions.

Here, we are interested in seeing what would happen if we generalized the discussion of Subsection 2.3.3 to other, more arbitrary, tree structures. We are particularly interested in the full-tree decomposition scheme where all the outputs of the first stage, that is the high passed as well as the low passed ones, are further decomposed. This means that, starting from a single two-channel filter bank, we decompose a one-dimensional signal in 2^j bands at each j resolution level. Of course, this full tree decomposition, first proposed in [14] and known as *wavelet packets* (WP), can be implemented using the same filter bank-related approach [12, 13]. The wavelet packet library is produced by cascading filtering and downsampling operations in a tree-structure. Figures 2.6 and 2.7 thus become Figures 2.10 and 2.11.

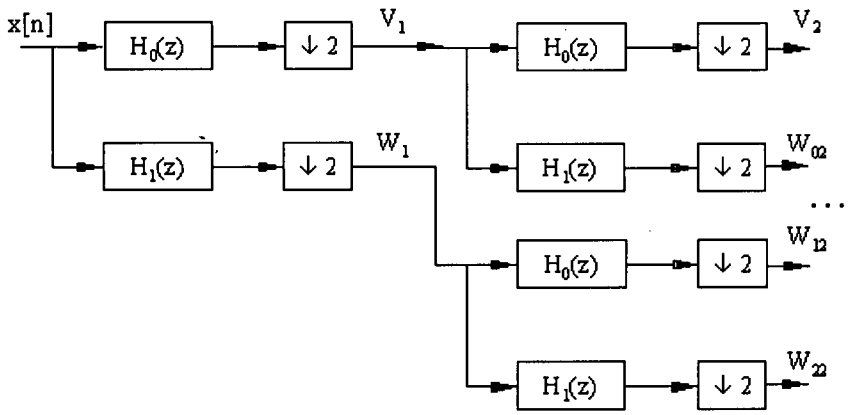


Figure 2.10: Two Levels of Wavelet Packet Decomposition using Filter Bank Representation

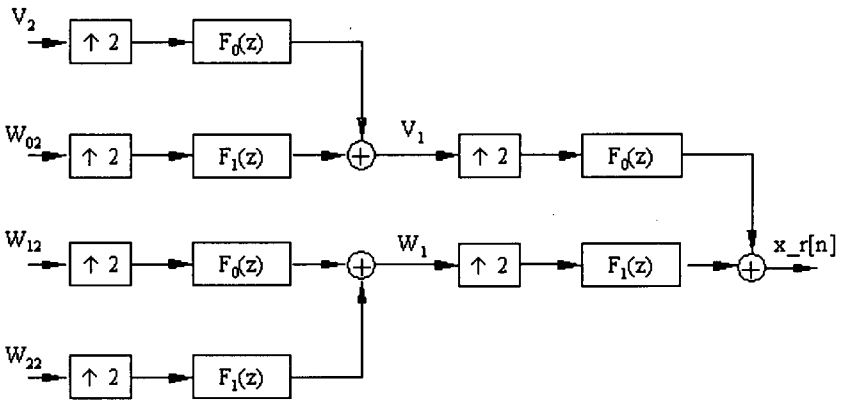


Figure 2.11: Two Levels of Wavelet Packet Recomposition using Filter Bank Representation

The wavelet packet algorithm generates a library of orthonormal functions that are derived from a single filter kernel. This kind of decomposition offers important advantages. In particular, wavelet packets are very promising for image compression. This comes from the fact that full tree decomposition allows for the selection of the best basis. As the complete tree structure yields *over defined* decomposition, not all the outputs have to be used in the analysis or for the reconstruction. For example, from Figure 2.10, W_{12}, W_{22} and V_1 can be selected as bases for the reconstruction of $x[n]$, leaving unused V_2 and W_{02}. Generally speaking, the WP algorithm searches through the library of bases V_i's and W_i's to find the least computationaly expensive set, which also provides the best compression[9].

In addition, from the examination of Figure 2.12, it is clear that WP

[9]Due to this capability, the FBI chose the use of a wavelet packets-based image compression scheme for their fingerprints databank as it allows for the best performance for images with important high frequency content.

increases the frequency resolution at higher frequencies. On the other hand, as the frequency resolution increases, the space resolution decreases. This is a problem that has to be kept in mind for the implementation of our watermarking scheme. There exists, however, one other advantage of wavelet packets that makes up, in our application, for the above problem: the implementation of wavelet packet decomposition assures the symmetry of the final decomposition bands. It means that all the bands are of the same size, and that the translation from frequency to time domain is much more straightforward. This proves very useful when the time comes to compute the space localization of tampering from the wavelet domain (Section 5.3).

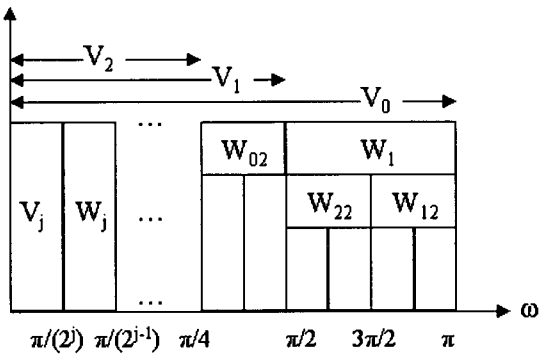


Figure 2.12: Ideal Spectrum Division from Wavelet Packet Decomposition

To sum up, this short section was designed to provide the concepts of *wavelet packets*. The principal differences as well as the advantages involved in WP (over *traditional* wavelets) signal decomposition are explained. We are now confident that interested persons will agree with our choice of wavelet packets for image decomposition. In any case, more detail on our implementation of wavelet packets is given in the description of our watermarking system

in Chapter 4.

2.5 Multidimensional Signals

Since we are particularly concerned with the utilization of wavelets for image decomposition, we need to generalize wavelets to two-dimensional signals. Filter bank concepts are easily applied to images by considering them as two one-dimensional signals: the rows and the columns. We can simply start by utilizing a one-dimensional transform on the rows before the same is done on the columns. This can be formulated in terms of the tensor product of a one-dimensional transform having analysis and synthesis basis functions $\tilde{\varphi}_i$ and φ_i. Therefore, the two-dimensional signal $I(x, y)$ can be expressed as a combination of the following bases:

$$I(x,y) = \sum_i \sum_k \langle I(x,y), \tilde{\varphi}_{i,k}(x,y) \rangle \; \varphi_{i,k}(x,y) \tag{2.27}$$

where the 2D basis functions are defined from the 1D ones as follows:

$$\tilde{\varphi}_{i,k}(x,y) = \tilde{\varphi}_i(x) \; \tilde{\varphi}_k(y) \quad \text{and} \quad \varphi_{i,k}(x,y) = \varphi_i(x) \; \varphi_k(y) \tag{2.28}$$

As the previous transform can always be expressed as two separate one-dimensional transforms, it is referred to as *separable*. While being constrained in their definition and construction, separable transforms are favored because of their computational efficiency with separable filters [80]. For this reason, separable transforms are the most commonly used today. For more information on non-separable systems used for image processing applications, the reader is invited to see [66, 79].

2.6 Summary

In this chapter, we reviewed well-known principles of signal decomposition, and particularly, the fundamentals of wavelet transform. We were specifically interested in the implementation of wavelets with octave band quadrature mirror filters. Although no new concepts were proposed, the originality of our presentation resides mainly in its simplicity. Even if wavelets are based on fairly complex mathematical theories, we strived to make this chapter readable. We first started from commonly known series expansion theorems and examined the implications for signal representation. Then, we explained the theories behind filter bank decomposition and described the implication of octave band QMF for the implementation of DWT. Afterwards, we expanded the use of filter banks to multiresolution analysis and focused on *wavelet packets*, a particular type of wavelet decomposition that is used in our digital watermarking algorithm. We concluded the chapter by broadening the use of wavelet transform to multidimensional signals.

Chapter 3

Digital Watermarking

"An invasion of armies can be resisted, but not an idea whose time has come."

-Victor Hugo, writer and poet (1802-1885)

3.1 Introduction

Digital watermarking is a relatively new technology that allows the imperceptible insertion of information into multimedia data. The supplementary information, called *watermark*, is embedded into the cover work through its slight modification. This mark is hidden from view during normal use and only becomes visible as a result of a special visualization process. An important point of watermarking techniques is that the embedded mark must carry information about the host in which it is hidden.

The protection of currency bills is the most widely known use of watermarking. For example, Benjamin Franklin's face is embedded on 100 US$ bills

and matches its printed portrait on the same bill, but is visible only when the bill is held up to a light. The watermark is used to make the illegal reproduction of bills detectable, if not impossible. In this context, the paper watermark carries information about the legitimacy of the currency bill. Authentication is only one possible application of watermarking and its use on digital work offers other great possibilities. It has been foreseen as a good candidate technology for enhancing multimedia data by the addition of information available to the users for content improvement, copyright protection, authentication, and so forth.

The purpose of the present chapter is to provide the reader with a solid basis on watermarking technologies. First, we define watermarking and draw a portrait of the concepts leading to its use for digital media by retracing its history. Then, we give a background on watermarking by highlighting its applications. We also give a generic classification of watermarking schemes and review cornerstone papers that explicate its bases.

3.1.1 Definition of Watermarking

First, we need to point out what differentiates digital watermarking from information hiding and steganography. In information hiding, the goal is to make the information imperceptible, or to keep the existence of the information secret [58]. Since it includes applications such as user anonymity in networks or database secrecy, information hiding is considered beyond the embedding of messages in content, and therefore, out of our scope of interest. Steganog-

raphy, on the other hand, is more related to the technology discussed here. The word steganography comes from two Greek words: *steganos*, which means covered; and *graphia*, which means writing. In [20], it is defined as *the art of concealed communication.* The hidden message, however, does not necessarily carry information about the cover work. For this reason, we consider steganography and watermarking as two different applications, and focus on the later for the remaining of the thesis.

3.2 Historical Perspective

Watermarking was first used in the thirteenth century in Fabriano, Italy, to label pieces of hand-made paper [50]. The inventors inserted designs in the paper sheets by thinning certain regions by placing wire in the mold. Afterwards, one could access the inserted design by holding the dry piece of marked paper up against a strong light. Watermarking was used to distinguish the mold used for fabrication, to identify the paper maker [20], or even simply for decorative purposes [36]. It was named watermarking because the patterns formed by the wires were perceived as watery areas on the marked articles [50].

This technique became accepted as a labeling tool for paper sheets. By the eighteenth century, papermakers were using watermarks to record information about produced paper. In that way, watermarks served-and still do-as a means of identifying paper with the members of the trade organization who manufactured it. At approximately the same time, the increasing number of

commercial exchanges and currency bill circulation boosted the problems of money counterfeiting. For that reason, watermarking quickly became an effective way to avoid the duplication of currency bills. As it has been proven to be effective, watermarking is still used as a currency protection technique.

In the mid 1950's, Emil Hembrooke, an engineer from the *Muzak* Corporation, was the first to extend the use of watermarking to other media. He filed up a patent for the *watermarking* of musical works. The insertion of an ownership identification key was designed to identify the work at hand. It was performed by the intermittent application of a narrow notch filter on the audio signal using *Morse*-based coding. In [34], the system is described as follow:

> The present invention makes possible the positive identification of the origin of a musical presentation and thereby constitutes an effective means of preventing such piracy, i.e. it may be likened to a watermark in paper.

The connection between the insertion of undetectable information in digital content and paper watermarking technology was thus made. Watermarking would, however, have to wait longer to attract enough attention to become an active field of research. In 1988, Komatsu and Tominaga were the first to use the term *digital watermarking* for their image authentication system [38]. Although there were several publications in the interval, a cornerstone paper by Cox et al. [15] was the starting point of more intensified research. Figure 3.1 shows[1] that the number of publications on watermarking

[1]Thanks to Peter Meerwald from the University of Salzburg for the data.

increased almost exponentially between 1995 and 1999. Of course, this was not only due to the paper by Cox et al., but mainly to the organization of the watermarking researchers. The first *Information Hiding Workshop* was held in 1996 and the *Society of Photo-Optical Instrumentation Engineers*, SPIE, started organizing conferences specifically on *Security and Watermarking of Multimedia Contents* in 1999. In addition to official efforts, individuals also contributed to the formation of a new research community. The work of Martin Kutter on the *Digital Watermarking World* is the first, and probably the best, example of personal efforts for the advancement of the technology. In the meantime, commercial use of digital watermarking (DW) interested companies and organizations. The music industry came up with the Secure Digital Music Initiative, *SDMI*, in 1999, in order to create an environment for the legitimate distribution of digital music. In addition, several companies (e.g. Digimarc Corporation, Alpvision and Alpha-Tec) specializing in digital watermarking have also been created. As a result, an increasing amount of effort and funding is dedicated for research in different areas of DW. It is therefore expected that a number of businesses will be created in the near future to deploy more applications based on this new technology [18].

3.3 Background on Watermarking

A lot of work has been done in order to develop reliable watermarking systems. Numerous digital media have been considered for the embedding of information to serve a wide range of applications. This has lead to the investigation of

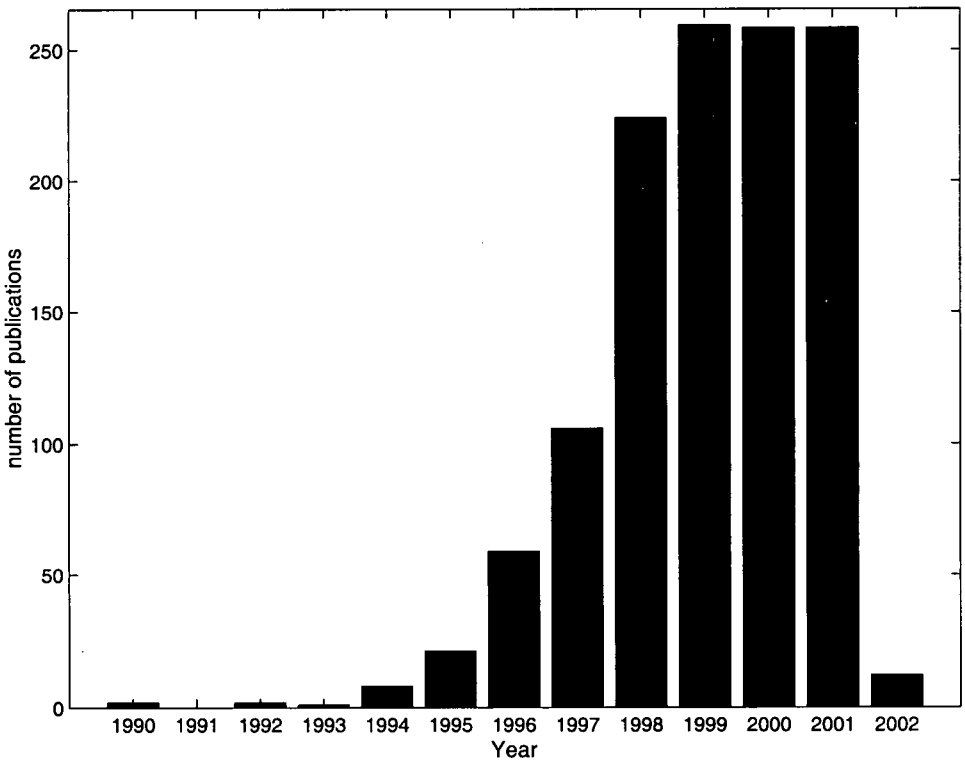


Figure 3.1: Publications on Digital Watermarking per Year

several approaches, both in terms of embedding and detection procedures. In this section, we review work done on different media as well as previously intended applications and general methodologies used for the watermarking of digital images.

3.3.1 Host Media for Watermarking

Since the first use of watermarking on audio signals for identification, watermarking has been tried on several other digital media. There are several papers on the insertion of information in gray scale [27, 39, 43, 50] or color [7, 29, 63, 87] digital images, while others exclusively investigate the use of compressed [40] or printed copies [25]. Watermarking of video has also attracted some attention [10, 26, 35, 70], especially for the protection of MPEG streams [27, 33]. Of course, more research has been done on the watermarking of audio signals [32, 46], and interesting applications have been found for speech in particular [11]. Text data have also been investigated by many [8, 9, 44]. Some more theoretical approaches of watermarking for combined media have been studied [45, 59, 81], but the resulting systems are limited since they do not take advantage of the specific type of host media involved. In order to summarize the relative importance of the investigation of each medium in terms of number of publications reviewed, we present Figure B.1 in Appendix B.

3.3.2 Applications of Watermarking

In addition to the exploitation of different hosts, researchers have been looking at different applications for digital watermarking. The application that attracts the most attention is *copyright protection* [27, 43, 44, 70, 71]. In this context, a watermark is permanently embedded in the work to identify its original owner. In order to be efficient, the embedded mark has to be robust, that is, it has to be detectable as long as the host carries its information, hence, the name of *robust watermarking.* Another use of robust watermarking is for the labeling or *fingerprinting* of digital media. This application is technically similar to the previous one except that, here, a different mark is embedded in each copy of the same work to allow its tracking.

As shown in Figure B.2, there are a number of other very interesting applications that require recognition. Watermarking technologies can be used to monitor the content of television or radio broadcasting [32], or to control the access or the copying rights of digital content [8, 54, 81, 83]. The *Millennium Group* proposed to control the reproduction of DVD's by the addition of invisible fingerprints in the video streams. Associated with decoders in copying devices this can either grant or forbid the duplication of the digital content. Similar work is also done under the supervision of Philips and Sony to create a new audio disk encoding standard, called Direct Stream Digital (DSD), that includes watermarking for copy protection. In addition, simple *covert communications* have been investigated by many [9, 29], and some propose to use watermarking as a feature enhancement method [74]. A number of researchers

have also created multipurpose watermarking systems in order to add several layers of information to the host [40, 46].

Content authentication is the last application of watermarking that needs to be highlighted. This specific watermarking approach involves the embedding of a fragile key in the host. This key is destroyed if tampering occurs with the original content, hence the commonly accepted name *fragile watermarking.* Fragile watermarking can be used to assess the validity of the work at hand, thus increasing the value of its content [5, 39, 87, 88]. For this reason, content authentication is a very promising application of DW technologies, and the next chapter investigates it with more details.

From the above, it is clear that the use of watermarking on different media, and for different applications makes the number of publications quite overwhelming. Fortunately, extensive surveys have already been published [4, 19, 59, 81] and provide overviews of the state-of-the-art. Nevertheless, generic classifications can be obtained by the examination of previously proposed techniques.

3.3.3 Requirements of Watermarking Systems

In order to be truly efficient, digital watermarking systems have to consider the specific application and host data intended. In a fact, in DW, *one size does not fit all* [19]. Nevertheless, some general requirements of the technology can be extracted. The first, and most important one, is invisibility; that is, the watermarked work should not be perceptually different from the original

(*unmarked*) one. Security is another universal consideration of watermarking. Its basis must lie on Kerckhoff's assumption that *ones should assume that the method used to encrypt the data is known to the unauthorized party* [73]. From this, watermarking security can be viewed as encryption security, thus leading directly to the principle that it must lie mainly in the choice of the key and embedding protocol.

From the applications mentioned in Subsection 3.3.2, one can divide watermarks into two distinct types: *robust* and *fragile.* Used mainly for content authentication, fragile watermarks are meant to disappear if the image is corrupted. Robust watermarking, on the other hand, allows for the mark to still be detectable after the content has undergone tampering, and therefore, grants its permanent identification. Therefore, robustness *versus* fragility requirements are contradictory, and this is one of the most important reason why each kind of watermarking is required.

In summary, there are three main requirements of watermarking systems: invisibility, security and robustness or fragility.

3.3.4 Embedding Domains and Decoding Procedures

Although it is quite difficult to classify all the existing methods into families, the domain of embedding, that is, the kind of operation needed to insert the mark, can be used as a common trait for different techniques. A lot of effort has been put into the development of straightforward embedding in the spatial domain [5, 72, 75, 87] because of its ease of implementation and com-

puting efficiency. To add control to the frequency of embedding, some first transform the content using the discrete Fourier transform (DFT) to create the embedding domain [42]. Others want to match the current image compression standard and employ the discrete cosine transform (DCT) [17, 78]. Few methods are based on the Walsh-Hadamard Transform [7, 28] because of the frequency spreading it grants. In the context of copyright protection, techniques are also developed in order to increase the robustness of the watermarks. Fourier-Mellin approaches allow good resistance to the geometric alteration of the host [64]. On the other hand, the use of WT [39, 46, 86, 88] or WP [27, 77] have shown to increase the embedded marks' resistance to image processing operations. Moreover, these techniques allow for both the spatial and frequency localization of embedded marks. To sum up, one could classify watermarking schemes in two broad categories in terms of domain of embedding: spatial domain *versus* transform domains.

In the same way, the decoding procedures can be used to create an informative classification protocol. The simplest way to extract a mark from its host is when the unmarked data is available. Such techniques, referred to as *informed*[2], were the first to be investigated because of their relative simplicity [17, 73, 86]. Informed detection is still considered in copyright protection or fingerprinting since the owner of the work is likely to have access to the original *unmarked* data. In contrast, access to the original *unmarked* data is forbidden in content authentication or copy control applications. Techniques recovering

[2]Although this term is sometimes used to characterize embedding procedures that tailor the mark according to the host media [18], we consider our utilization as more accurate since all insertion systems are, in fact, *aware* or the original content.

the embedded mark without the use of the *unmarked* data are defined as *blind* decoding. Within these blind decoding techniques, some require access to a reference key to extract the mark [39, 42, 75, 76, 77, 87]. These are called *private*. Others, on the other hand, do not require the unmarked data either, nor do they need a key for decoding purposes [46, 64, 78, 88]. These methods are called *public* because everyone is allowed to access the watermarked information. They are mainly considered for copy control or feature enhancement purposes, while private approaches are mainly investigated for content authentication.

From this discussion, summarized in Figure 3.2, the basic categorization of watermarking techniques should become clear. From now on, the term *DWT-based blind private* watermarking technique[3] should be meaningful.

3.4 Watermarking for Copyright Protection

Many of the first papers published on DW are about its use for copyright and ownership protection related functions. Thereby, most of the bases and theories associated with the technology are laid out in relation to this particular application. For this reason, it is logical to introduce cornerstone papers on the subject.

The paper published by Cox, Kilian, Leighton and Shamoon in 1995 constitutes an important step towards the installation of watermarking as a

[3]In this case, the mark is embedded in the discrete wavelet domain and is detected with a secrete key but without access to the original content.

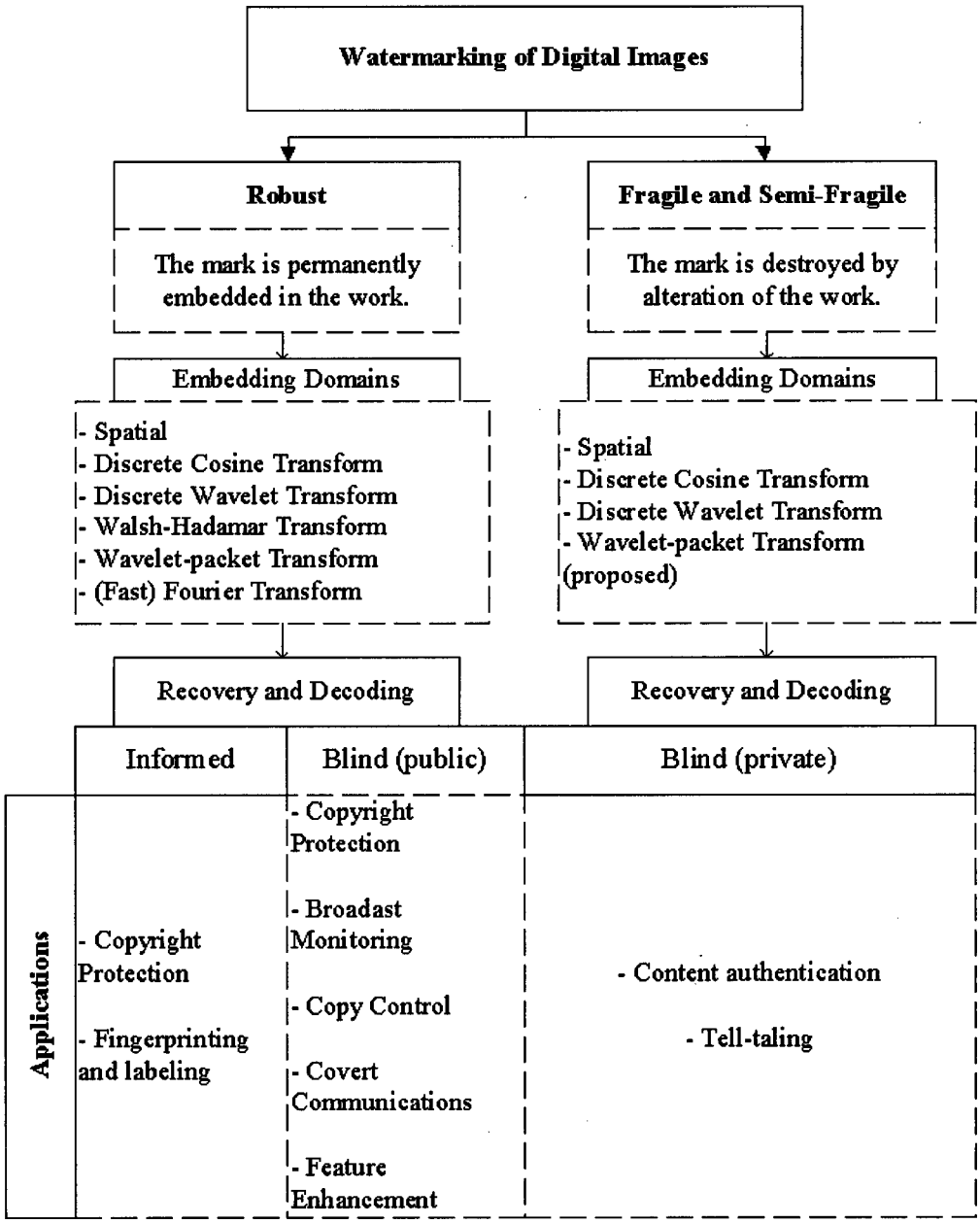


Figure 3.2: Our Generic Classification of Digital Watermarking Systems

technology in its own right [15]. Presenting a watermarking approach for the copyright protection of digital content, Cox et al. capture the most important concepts of robust watermarking. The necessary characteristics of robust watermarking schemes are outlined: fidelity preservation, robustness to attacks, unambiguous identification[4] and universality[5]. The authors are the first to argue that a watermark should be placed in perceptually significant components of a signal if it is to be robust to common signal distortions and malicious attacks. To compensate for the fact that this is contrary to the fidelity requirement, they adopt an embedding method similar to spread spectrum communications: that is, hiding a narrow band signal in a wideband channel that is the data to mark. In this case, the watermark, an independent identically distributed (*i.i.d.*) Gaussian random vector x, known to the authors, is added to the largest DCT coefficients v_i-which represent the most visually significant components-of the image. To assure perceptual transparency, a weighting parameter α is used to produce the watermarked DCT coefficients v_i' using the following three different strategies reflecting the degree of independence desired[6]:

$$v_i' = v_i + \alpha x_i \quad or \quad v_i' = v_i(1 + \alpha x_i) \quad or \quad v_i' = v_i(e^{\alpha x_i}) \tag{3.1}$$

As it distributes the mark over the entire content, the spread spectrum approach prevents attackers from *jamming* or detecting the embedded signal. At

[4]The retrieval of the watermark should unambiguously identify the owner of a work.

[5]Once again, universality has been found to be a questionable requirement, but it was part of the original paper.

[6]The second and third method are more robust against difference in scale between the DCT coefficients of the host and the mark to embed.

the same time, authorized parties can use their knowledge of the *i.i.d.* vector to form the originally embedded message from the comparison of watermarked image with its original version. Cox et al. demonstrate that their technique is robust to common signal processing procedures and geometric transformations, and is able to deal with simple collusive attacks, thus ensuring good copyright protection of images. They conclude by stating, without implementing, that watermarking systems should take explicit advantages of the characteristics of the human visual system, HVS.

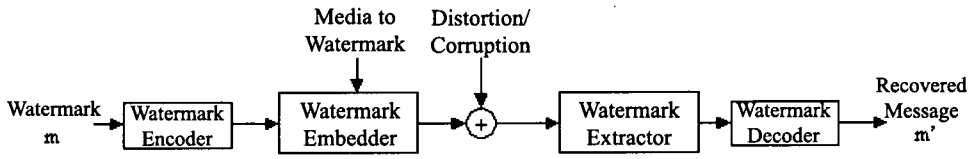


Figure 3.3: Watermarking as Communications

This paper directly led to another cornerstone paper. In [21], Cox, Miller and McKellips examine the similarities and differences between watermarking and traditional communications. Figure 3.3 describes a simple model of a watermarking system as a standard communication scheme and includes: a watermark encoder, a watermark embedded (as a modulator), a destructive channel, a watermark extractor (as a demodulator) and a watermark decoder. Using models based on Shannon's communication theories, the authors highlight the fact that the knowledge of the cover data as side information at the transmitter allows the design of more powerful watermark embedding algorithms. Cox et al. stress the importance of the use of characteristics of the HVS in the embedding process; both for maximizing the robustness, and for

minimizing the perceptual distortion introduced. They argue that an appropriate distortion model for watermarking applications includes a significant correlation between the distortion vectors (watermarks) and content vectors they are applied to. Therefore, their embedder makes use of the knowledge of the image rather than treating it as unknown noise. In addition, the inclusion of a perceptual model (Figure 3.4) assures the invisibility of the watermark based on the characteristics of the human eye [37]. The complete system of Figure 3.5 represents the underlying principles of watermarking as it takes into consideration the perception of the marked content by a potential user in parallel with the decoding procedure. The authors conclude by explaining the design of a blind optimal threshold-based detector. This discards the need to access the original image in the detection process, thus opening the field of watermarking to a wider range of applications.

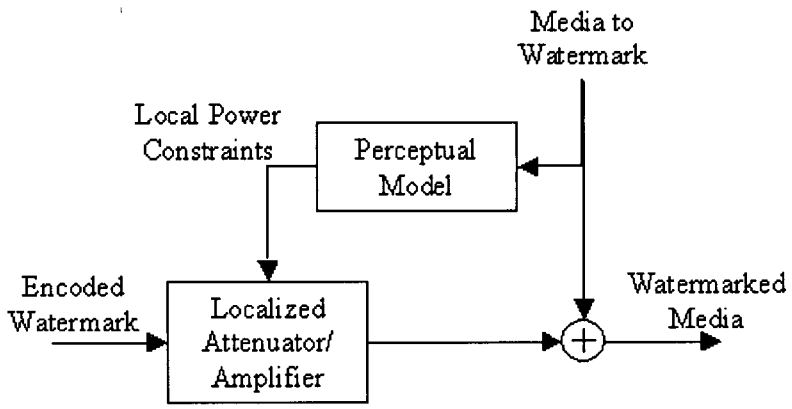


Figure 3.4: Watermark Embedder with Perceptual Model

To summarize, [15] is a great step towards the development of *modern* robust watermarking schemes. The authors list the requirements of robust wa-

termarking schemes, and acknowledge the challenges involved in the creation of an algorithm capable of producing watermarks that fulfill all the contradicting specifications. Then, [21] extends this work by explaining that the knowledge of the cover data as side information at the transmitter allows for the optimization of the watermark embedding algorithms, while enabling the blind recovery of the mark. Moreover, it shows that, to be fully efficient, a watermark embedding system has to take advantage of the knowledge of the host data and of the characteristics of the HVS. In conclusion, these two papers are important to highlight because they laid the basis for watermark embedding in images out. Other papers have greatly contributed to our understanding of the problems linked with watermarking systems, but the ones mentioned stand out both in their conceptualization and in their impact on our own work.

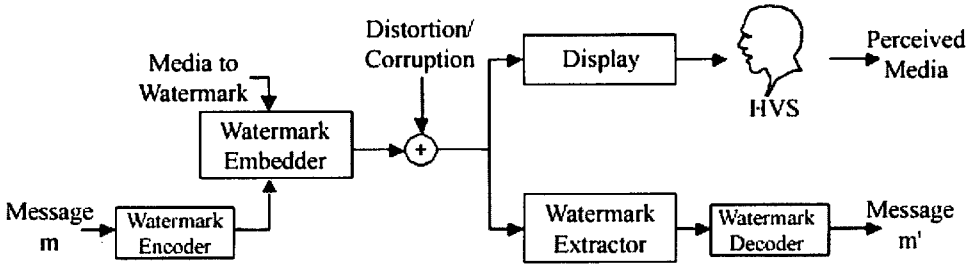


Figure 3.5: Complete Watermarking Scheme

3.5 Summary

In this chapter, we examined the basic concepts associated with DW. In the first place, we differentiated watermarking from other concealing technologies.

Afterwards, we positioned the development of the technology by giving an overview of its history, and pinpointing the main stages of its development. Then, we explained some of the most common applications of DW. We also gave generic classifications of proposed systems. In the last section, we reviewed cornerstone papers-dedicated to the use of watermarking for copyright protection-that have clearly exposed the challenges and requirements of efficient watermarking schemes. This chapter was meant to provide a general background on digital watermarking, and to give a better understanding of the issues under consideration. The next chapter focuses on the application of digital watermarking that we are the most interested in: *content authentication.*

Chapter 4

Image Authentication

"When I am finishing a picture, I hold some God-made object up to it (a rock, a flower, the branch of a tree or my hand) as a final test. If the painting stands up beside a thing man cannot make, the painting is authentic."

-Marc Chagall, painter (1887-1985)

4.1 Introduction

The rapid expansion of the Internet, and the overall development of digital technologies in the past years, have sharply increased the availability of digital media. One of the great advantages of digital contents is that they can be reproduced without loss of quality. On the other hand, digital contents may be easily modified, and sometimes, imperceptibly. In cases such as courtroom evidence and video security systems, any alterations of image, video or audio data have to be detected. Therefore, some work needs to be done in order to

develop security systems to protect the information contained in digital data.

As stated in the previous chapter, watermarking has come to be a widely accepted approach for copyright protection and ownership identification. Much effort has therefore, been dedicated to the development of robust watermarking schemes. Another possible, but much less investigated, application of digital watermarking is content authentication. Since the use of a fragile watermarking system, in which the integrity of the mark is affected by tampering with its host, allows for the detection of any unauthorized modifications of an image, a video or an audio sequence, it makes the validation of multimedia data possible, thus giving it legal value.

In this chapter, we review the bases of authentication schemes, and propose a new method based on digital watermarking. In the first part, we create a classification of authentication approaches considered. From this, we draw the requirements that such systems should fulfill in order to be efficient. Then, we detail specific methods introduced that serve as the basis in the development of our own system, which is finally introduced in the last section (4.5). It must be kept in mind that, although we are concerned specifically with the authentication of images, many principles introduced in this chapter can generalize to other media. This is why the terms *content, data, multimedia* or *host* are used, as well as *image*, to refer to the digital information to protect.

4.2 Approaches to authentication

Before we can explicitly state the requirements of image authentication schemes, we need to classify (and clarify) the different approaches considered. The first kind of system that has been proposed is named *complete authentication.* This family regroups techniques that treat multimedia data as integral information such that no further manipulation is allowed [41]. In that sense, it includes purely fragile watermark embedding techniques, as well as image signature methods. In the first case, a totally fragile watermark is embedded in the host data such that the key is broken if anything happens to the work [87]. In the later signature case, data features are extracted. These are later used to verify the authenticity of the data whenever it is needed. In [57], for example, we propose a wavelet packets-based image retrieval system that can also be used in the context of image authentication. Our method uses the correlation of wavelet packet coefficients to extract features inherent to individual images and then create their unique signatures. Due to inter-frequency cascading characteristics of wavelets coefficients [62], our system captures the salient points of the image. The signature can then be encrypted and saved in a database for future certification. Most image signature extraction procedures require the storage of additional information. To eliminate this constraint, the signature may be embedded in the host data in a robust way that achieves complete data authentication [84].

For some applications, it is of utmost importance that not a single bit of information is changed; for example, the meaning of a text message

can be completely changed by the modification of only few letters (a handful of bits). For these cases, complete authentication is needed. For other cases, however, such as audio or video contents, the modification of a few bits does not change the information carried. For that reason, schemes with more flexibility need to be developed. *Semi-fragile authentication*[1] schemes therefore, form the last group of systems introduced and the one we are more interested in. In this kind of authentication method, the altered multimedia is treated as authentic if the manipulations are imperceptible. It means that as long as the visual, audible or readable content of the data is kept unchanged, the later is considered authentic. The concept of image signature can be used to fulfil these requirements. For example, [6] extracts salient feature points of an image that are invariant to legitimate distortions, but not to malicious ones, and uses them to construct its semi-fragile image signature. In order to prevent the addition of side information, it is preferable to directly embed a mark in the host data in a semi-fragile way. Once inserted, the key stays unaffected by justifiable distortion. The key may also be designed so that any alterations--for example, malicious *versus* acceptable-in the data are made differentiable. Semi-fragile watermarking has been shown to be an effective method to grant multimedia data with legal value [81]. This is why this content authentication approach has attracted the most attention since the rise of interest for digital watermarking techniques [39, 46, 72, 88].

[1] Also referred to as *robust authentication.*

4.3 Requirements of Authentication Schemes

From our willingness to protect digital data against forgery and tampering, and also based on semi-fragile techniques already proposed, we can extract several requirements that authentication systems must fulfill. Here are the main points to keep in mind in the development and evaluation of certification systems. In the context of image protection, an effective authentication scheme should be able to do the following:

1. Determine whether an image has been altered or not;
2. Find the location in the image where the alterations, if any, are made;
3. Integrate authentication data within the host image rather than storing the data in a separate file[2];
4. Be robust to acceptable manipulations such as lossy compression or to other content-preserving manipulations defined by the original owner of the work to protect; and
5. Include security features preventing the forgery or manipulation of the reference mark. In essence, this means that the reference key used for authentication must be securely stored. In addition, the embedding protocol must depend on the secret key in order to enhance the security of the authentication scheme.

[2]Hence ruling out image signature procedures.

Some authors have also added the *recovery* capability as a prerequisite of image authentication systems [41]. This means that, after the detection process, it should be possible to find out the original content of the tampered areas, and also, that the recovered data shown, be of the same quality as the original. This concept is interesting, and it has also lead to the development of *erasable watermarking systems.* An erasable watermark can be removed from its associated cover work to obtain an exact copy of the original unwatermarked work. It is however, impossible to design an erasable watermarking scheme that can be uniquely applied on all the work of a specific family of digital contents[3] [20]. Erasable watermarking schemes are still highly prototypic and this is why, in the present thesis, we have strictly been concerned by the detection and localization of alterations, and have not attempted the subjects of reconstructing tampered regions or deleting embedded marks. Nevertheless, the use of digital watermarking for image authentication clearly presents some advantages.

The advantages of watermarking approaches for content authentication are twofold. First, the direct embedding of a mark in the host data removes the need to store a separate authentication signature (point 3). Second, as the watermark undergoes the same alterations as the host, the mark is modified by the host's corruption [20]. Using a reference pattern in the embedding and decoding procedures allows for the identification and delimitation of tampered regions. This satisfies points 1 and 2 of the list of requirements at once.

[3]For example, it is impossible to use the same erasable watermarking scheme on all digital images (§A.6).

In addition, some basic requirements of digital watermarking (see Subsection 3.3.3) are helpful in the authentication context. The fact that the embedded mark must stay invisible allows the watermarked data to be as close as possible to the original data, therefore, preserving the original content. Furthermore, security concerns (point 5) are already considered by watermarking systems' requirements, which state that such systems must be planned under Kerckhoff's security assumption.

From this, it is easy to recognize why watermarking is seen as a plausible candidate for image authentication systems, and to understand the growing interest in the subject.

4.4 Previous Work

The raising of interest for content authentication has accelerated the development of fragile watermarking systems. As for other watermarking types, the fragile watermarking techniques proposed can be divided in two general categories in terms of the embedding process: the ones acting directly in the spatial domain and the others, working in different transform domains. Each has pros and cons that we highlight here.

4.4.1 Fragile Watermarking in the Spatial Domain

Fragile watermarking techniques that embed hidden information in the spatial domain, such as [5, 60, 72, 74, 87], are definitely more straightforward, and therefore, less computationally expensive, than the ones using transforms.

Therefore, this kind of embedding is probably more suitable for real-time implementation. In [87], Yeung and Mintzer propose one of the first watermarking methods for high-quality color and grey-scale image verification and authentication. A watermark image is embedded into the source image in the spatial domain by the modification of the pixel values. The stamped image produced is visibly identical to the original one. A verification *key*, stored and known only to authorized parties, is also produced and is used in the verification process in order to extract the image inserted in the host. The extraction procedure can detect and localize spatial alterations done on previously watermarked images. The technique therefore provides a way of ensuring data integrity, adds to the security of the digital content, and allows the recipients of an image to verify the image as well as to display the ownership information of that image. The embedding process is however, fragile to unintentional image distortions introduced by basic image processing operations (e.g. compression) done for storage purposes.

Another spatial embedding watermarking method is that proposed by Tefas and Pitas [72]. In addition to allowing the identification of modified regions in tampered images, it is able to reject small distortions introduced by high quality image compression (for which [87] is fragile). A pseudo-random watermark is embedded on randomly selected pixels using a neighbour-dependant function. In the detection process, the pixels surrounding the marked ones are used to create a mapping of false detections. The identification of changes in small details of the image is based on mathematical

morphology; altered pixels are linked together in order to indicate tampered areas. Finally, the decision about the image's authenticity is made by comparing the ratio of correctly detected watermark with a predefined threshold. A similar technique to further enhance the resistance of the watermarked images to medium quality JPEG compression has also been proposed in [55]. However, both techniques (i.e. [87] and [72]) suffer from the following major drawback of spatial domain watermarking: the difficulty of the frequency localization of modifications. In fact, because the marks are inserted in certain particular pixels, it is often impossible to localize frequency alterations applied to the entire image. The reason why the localization of frequency alterations is important is twofold: one, it is a step towards *telltaling*, the characterization of the specific process used for the alteration of the content; and two, it provides a measure of the relative degree of image distortion.

In addition to the impossibility of identifying frequency tampering, image authentication systems based on the embedding of watermarks in the spatial domain have the drawback of being more susceptible to malicious attacks. In fact, search and collage attacks (defined below) are a threat to spatial-based- and particularly block-based-watermarking authentication approaches [20]. In a search attack, the aggressor, who has access to the watermark decoder, creates altered versions of the work and processes them through the decoder by brute force, until one is declared authentic. Since the mark is embedded directly in the pixel intensity values, it is possible, although lengthy, to extract a pattern from the multiple watermarked images and then use it to create

authentic images. On the other hand, collage attacks are much more possible and easy to realize. In that case, the attacker has access to several similarly marked works. He can therefore, use parts of different genuine images and assemble them to form a new authentic image. This is really easy to do and it allows the unregistered modifications of (*supposedly*) tamper-proof images. For examples, see Figures 5.21 and 5.22 in the next chapter.

In summary, spatial-based authentication watermarking methods show speed advantage that can be favored for real-time implementations. This is why such techniques have often been extended to the authentication of video data [5]. However, for all the reasons mentioned above, more compliant techniques must be developed for still image authentication.

4.4.2 Fragile Watermarking in Transform Domains

The techniques using transform domain are, of course, a little bit more complex and computationally expensive than the spatial domain ones. Yet, they offer a higher degree of robustness against common image processing operations [16]. One could wonder why robustness is important for fragile watermarking systems. This is simply because it is highly preferable that basic image processing operations-ones that are typically used for storage of watermarked images-do not alter the embedded marks.

Some authors have proposed taking advantage of the knowledge of current image compression standards to develop semi-fragile watermarking techniques in the discrete cosine transform (DCT) domain [41, 85]. In [41], Lin

and Chang introduce an authentication scheme that accepts JPEG lossy compression performed on the watermarked image up to a pre-determined lowest quality factor while rejecting crop and replacement processes. Their authentication procedure is based on JPEG invariant properties of DCT coefficients. Their technique also allows for the recovery of original visual information after tampering. To achieve these goals, two binary sequences are created. The authentication bits (Φ), used to determine if any tampering has occurred, are computed from the relationship between two DCT coefficients of the same position in two separate (8 by 8) image blocks. This value is used since it is invariant to JPEG compression at a given quality factor. On the other hand, the recovery bits (Ψ), used to reconstruct the approximation of the original blocks of pixels after tampering, are obtained by the reduction, compression and encoding of the original (*unmarked* and *uncompressed*) image. The two are then embedded independently by the quantization of DCT coefficients using secret *block-seclection* functions in relation with JPEG quantization tables. Selecting quantization levels greater than JPEG ones guarantees that the embedded marks stay unaltered up to a lowest compression quality threshold. In the decoding step, the private authentication process first reconstructs the authentication bits, and then, reconstructs altered regions, if needed. Finally, the capacity of the system to endure JPEG compression with QF $>$ 50, and to reconstructed altered regions, is explicitly demonstrated.

Although DCT approaches show some potential, it is the wavelet domain that attracts the most attention among all the transform domains used

as it has been shown to yield the highest degree of robustness to simple image processing operations [86]. Furthermore, as DCT techniques (like [41]) are mainly block-based, they are also highly susceptible to collage attacks. In terms of decomposition the main advantage of wavelets over Fourier and DCT analysis is that they allow for combined spatial and frequency resolutions. Wavelet transform allows for the decomposition of the signal in narrow levels of detail, while keeping the basis signal space limited [24]. This is certainly of great importance when dealing with real signals, especially when spatial localization is to be considered. Moreover, as stated earlier, the availability of numerous *mother* wavelets gives flexibility to the analysis and allows it to be truly adaptive to a particular application. It is also possible to develop new basis functions to fulfill specific requirements. Finally, the use of the wavelet domain-as opposed to spatial or DCT domains-to embed the watermark provides simultaneous spatial localization and a frequency spread of the watermark in the host image [39]. All these gains certainly explain why WT attracts so much attention for a wide range of image processing applications, including digital watermarking for image authentication [39, 46, 47, 88] and the upcoming image compression standard, that is, *JPEG-2000.*

In [39], Kundur and Hatzinakos present a semi-fragile watermarking technique for the tamper proofing of still images. They propose to embed a mark in the discrete wavelet domain by the quantization of the image's corresponding wavelet coefficients. The first operation is the decomposition of the image by the computation of its discrete wavelet transform (DWT). The

authors make use of the *Haar* wavelet exclusively, and propose an algorithm in which the changes in the wavelet coefficients guarantee integer changes in the spatial domain. Once the image is decomposed in L levels of detail, a watermark can be inserted. First, an author identification key is produced by the generation of a pseudo-random binary sequence (zeros and ones) of length N_w. This sequence is kept secret and known only by the original owner of the work. Then, a quantization map is created based on a user-defined quantization step Δ. The rounding of specific DWT coefficients to even or odd quantization step values embeds the zeros and ones of the watermark[4] (see Figure 4.1). The selection of embedding locations is pseudo-random and well spread spatially and throughout each resolution level to be able to assess changes to all image components. The location information is stored in the coefficient selection key (*ckey*). In addition, an image-dependant quantization key (*qkey*) is introduced to improve security against forgery, and monitor specific changes to the image. The last step of the embedding process is the construction of the *tamper-proofed* image by the computation of the inverse discrete wavelet transform, IDWT.

In the decoding process, the DWT is performed on the *possibly* tampered image and locations of original watermark embedding are selected using *ckey*. Then, the embedded mark is blindly extracted by the computation of quantization levels' parity using Δ and *qkey*. This allows for the comparison of the mark extracted with the originally embedded one. The approach permits tamper detection in localized spatial and frequency regions, therefore making

[4]Referred to even and odd quantizations.

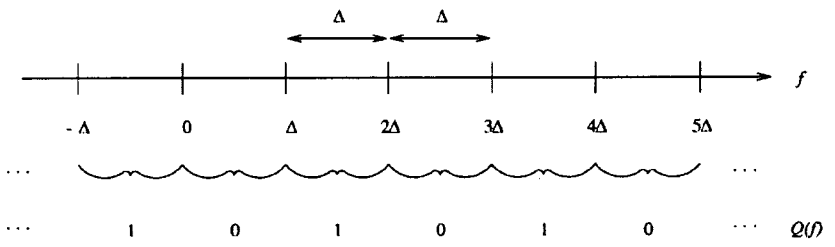


Figure 4.1: Quantization Scheme used in [39]

possible the identification of specific modified frequencies in an image. To assess the extent of tampering (the difference between the embedded mark w and the extracted one $\tilde{w}$), a tamper assessment function, TAF, is computed with the following:

$$TAF(w, \tilde{w}) = \frac{1}{N_w} \sum_{i=1}^{N_w} w(i) \oplus \tilde{w}(i) \tag{4.1}$$

Comparing the TAF with a predefined threshold $\mathcal{T}$, allows the user to make application-dependant decisions concerning the credibility of the received data. Examining how a known embedded watermark has been changed gives the possibility to investigate how a work has been corrupted. This type of watermarking is referred to as a *telltale* watermarking. Thus, the users are allowed to make context-dependant decisions on the validity of the images received. However, the total capacity[5] of the system, given by the mark's length N_w, is not specified. In addition, no strategy is propose to deal with a combination of malicious tampering and incidental distortion for the choice of Δ or $\mathcal{T}$.

In the same line of thought, Yu et al., developed, in [88], a digital images

[5]The capacity of a watermarking scheme is defined as the amount of information that is to be embedded in the host.

authentication procedure that allows for the detection of malicious tampering while staying robust to incidental distortion introduced by compression. As in [39], they embed a binary watermark in the wavelet transform domain. Once again, the insertion is done by the even or odd quantization of selected wavelet coefficients. Quantization-based watermarking is the simplest protocol because it requires the least storage of information. It is however, very sensitive to image modification. For this reason, the authors propose to make the embedded watermark more robust by quantizing the mean value of weighted magnitudes of wavelet coefficients. The quantization of regions of wavelet coefficients is performed using a predetermined function Q. The same function is used in the blind detection process as well, to privately retrieve the mark by *reversed quantization*, that is, determining the parity (in terms of quantization level) of the mean value of the WP coefficients. In order to distinguish malicious tampering from incidental distortion, the amount of modification on wavelet coefficients introduced by incidental *versus* malicious tampering is modeled as *Gaussian* distributions with small *versus* large variance. The probability of watermark error due to incidental alterations is smaller than malicious tampering because they produce a comparatively smaller variance difference with the embedded marks. To state the validity of possibly tampered images, a tamper response function (TRF) is defined for each decomposition level. It compares original quantization values $x_l(i,j)$ with wavelet coefficients $x_l(i^*,j^*)$ of the possibly tampered image, as shown below:

$$TRF(x_l(i^*,j^*),x_l(i,j)) = \frac{\max\{|i^*-i|,|j^*-j|\}}{\sum_{k=1}^{(Density(x_l(i^*,j^*)+1)} k^2} \tag{4.2}$$

The TRF allows for the estimation of tampering depth. Furthermore, the computation of the *Chess-Board* distance among altered coefficients permits the mapping of the tamper response. This serves as the basis for the decision rules to measure the malevolence of attacks. The integration of the tamper response at each scale of the wavelet decomposition allows for the discrimination of malicious tampering from incidental ones. This grants a certain degree of robustness to the system as the method is able to blindly authenticate JPEG compressed images. In spite of this, the authors do not explicate the degree to which the image can be compressed, and never explain how the quantization parameters are chosen.

The main flaw with the two techniques described above is that they both involve post-processing operations to determine the validity of the content. In [39], the user has to set a threshold below which a mark can be considered authentic, while in [88], the tampering distribution has to be examined. Furthermore, in [88], the users might have to determine the tampering manually at each scale if the tampered area is too small, or if there are many small unconnected tampered regions. In fact, both systems in themselves are not robust to JPEG compression, and only the detection processes allow this specific operation to go unnoticed. In conclusion, truly robust automated image authentication techniques in the wavelet-domain have yet to be developed.

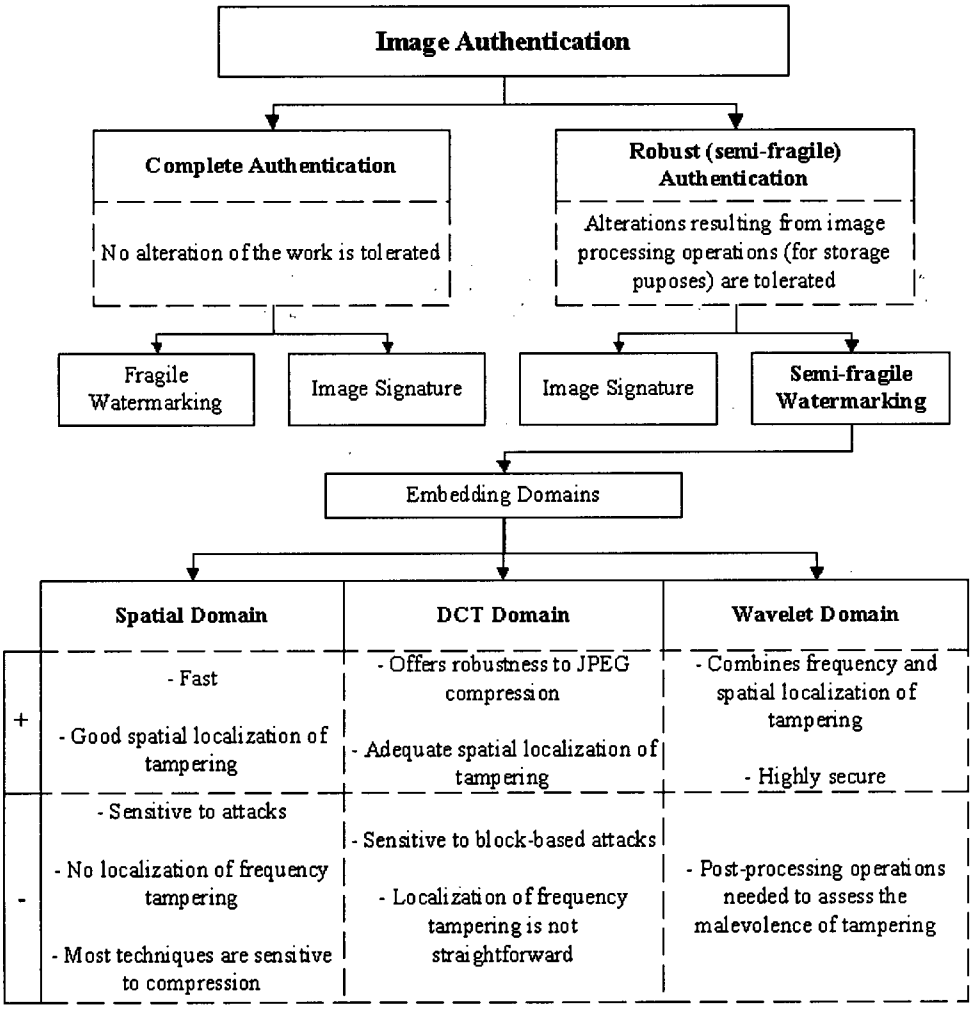


Figure 4.2: Our Classification of Image Authentication Techniques

4.5 Our WP-Based Image Authentication

The last two techniques presented in Subsection 4.4.2 protect digital images from malicious tampering and unauthorized processing, while allowing the compression of images with small compression ratios. However, these techniques require the user to determine the malevolence of an attack. In addition, they necessitate a certain degree of interaction in both the embedding and decoding procedures. On the other hand, our technique overcomes both these drawbacks as the *accepted attacks* are predetermined prior to the embedding process. Furthermore, our embedding procedure is adaptive since it is designed to automatically take maximum advantage of the characteristics of the human visual system (HVS). In addition, we propose to further improve the frequency resolution of standard discrete wavelet transform by the use of wavelet packets. The use of WP leads to narrower frequency bands at higher frequencies, and assures the capture of images' salient points [57]. Moreover, it allows for much higher precision and flexibility in the selection of the bands to be used in the embedding process.

Our motivation for developing a semi-fragile authentication system that does not require the post-processing of the tamper detection in the recovery process comes from the fact that, in our mind, the original owner or creator of the work-not the end user-should decide the extent to which an alteration is judged acceptable. From this, we believe that techniques including user interaction in the detection process have a serious security flaw that could prevent their commercial use. This is why we introduce a novel technique for

the content authentication of digital images, which is able to detect and localize malicious image alterations while offering a certain degree of robustness to image compression. In this section, we go over the details of the embedding and decoding procedures, highlighting principally the quantization and evaluation steps of each process, respectively. The experimental results obtained, as well as the comparison with a commercially available (and spatial-based) technique, are presented in the next chapter.

4.5.1 Embedding Process

The starting assumption of our approach is that any modification to an image leads to changes in the corresponding wavelet coefficients and embedded watermark [88]. As explained, small modifications in the wavelet coefficients do not change the image significantly, while minor changes in the image alter the coefficients locally, but noticeably. This characteristic is a good premise for watermark invisibility and fragility. In fact, this is the first reason why we have chosen the wavelet domain for our embedding procedure. The main steps of the technique developed are presented here, along with the specific advantages of wavelet packets. Our embedding scheme is summarized in Figure 4.7.

1. An author's identification key of 64 bits is randomly generated. Although this is an arbitrary length, 64 bits are enough to grant uniqueness of the key and protect individual creators and/or owners. The key is stored and kept secret. Only the owner-or other authorized parties-has access to the binary information contained in it.

2. The first four bits in the author's key are used to select the wavelet decomposition (wavelet function and number of decomposition levels) applied to the image. As previously stated, one of the advantages of WT is the great flexibility offered by the multitude of basis functions available. In the present application, it increases the security of our scheme since it is impossible for a would-be pirate to know which specific wavelet domain has been used for the embedding [52]. The selected number of decomposition levels had to be chosen so as to allow good frequency resolution, and to yield an important enough number of coefficients per band for embedding. Moreover, the computational cost for analysis had to be kept reasonable. We also had to select *mother wavelets* in order to have finite (compact) support and improved frequency resolution.

Generally speaking, the use wavelet packets adds another degree of freedom because it allows selecting frequency independently of the scale [77]. A specific advantage for our application is that it leads to decomposition bands of same size. This is important because it simplifies the frequency-space relation. Another important point is to choose the wavelets to achieve optimal spatial and frequency resolutions. Much work has yet to be done in the domain of wavelet function design or selection for particular applications. This is a field of research in itself, and the investigation of the effects of wavelet function selection on the final image would be an interesting subject for another research project.

As far as we are concerned, we use two families of compactly supported

wavelets: *Daubechies* and *Coiflets*. Both are formed with asymmetrical functions, which allow for perfect reconstruction and show good compressibility because of their smoothness[6]. From this, and although the use of 4 bits from the authentication key leads to 16 different possibilities for the selection of decomposition parameters, we have limited our investigation to 4 levels of decomposition using Daubechies 12 or 16 and Coifflets 12, 18, 24 or 30. The choice of 4 levels in the decomposition yields the minimal number of frequency bands necessary for the embedding, (see steps 4 and 5) thus minimizing the computational cost. The selected wavelets have less space localization than the simple *Haar* wavelet, and hence, yield better frequency resolution. Although they yield worst spatial localization, we have experimentally found them to be more suitable to the intended purpose since the embedded mark is truly spread over the image's content. As an indicator, the discrete filter functions used are presented in Figures B.5 and B.6 in the appendix.

3. WP decomposition of the image is performed based on the specification extracted from the key in **2**. In our case, we use a special implementation of the 2-D wavelet decomposition in *Matlab*: *Wavekit*[7]. This toolbox decomposes the image using multidimensional wavelet packets using successive filter banks and allows for the visualization of its level of detail (Figure 4.3).

4. The specific wavelet packet coefficients where the mark will be embedded are obtained. Our system identifies 64 groups of $K + 1$ bands. These groups are formed by one principal band surrounded by K secondary bands.

[6]The smoothness is associated with the number of vanishing moments, which increases with the order of the function used [24].

[7]http://www.math.rutgers.edu/~ojanen/wavekit/

Figure 4.3: Two Levels of Daubechies 12 and Coiflets 30 Wavelet packet decomposition and Associated Original Images

The principal band is always located in the top-left corner of the group, that is, it corresponds to the lowest frequencies of the group.

The first principal band is selected to be the LL or approximation band, which corresponds to the lowest frequencies of the entire decomposition. Then, the other 63 principal bands are evenly distributed in the wavelet packet decomposition to cover the entire frequency content[8]. In each principal band, our system picks N *subbands* (regions of M wavelet packet coefficients). These subbands are spread along each principal band to cover the spatial content globally at all scales. The position of these regions is fixed in each band for any given decomposition. However, the subbands are shifted (by one subband size) from one main band to the next (see Figure 4.5) to cover the entire image content (see Figure 4.6(a)).

Each group of bands (a combination of one principal band and K secondary bands) is used in Step **5.** for the embedding of one authentication bit from the author's key. For this reason, N regions of wavelet packet coefficients -or subbands-are established in the secondary bands as well. The regions belonging to a secondary band correspond exactly to the same spatial areas as the WPC regions of the matching principal band. A coefficients selection example is shown in Figure 4.4 for four groups of four bands, each having four subbands of four coefficients (that is, $K = 3$, $N = 4$ and $M = 4$).

The selection of 64 principal bands with their secondary bands allows for the embedding of the 64-bit author's identification key. This identification

[8]This is shown by the computation of the FFT of the inverse WP transform of the selected coefficients (Figure 4.6(b)).

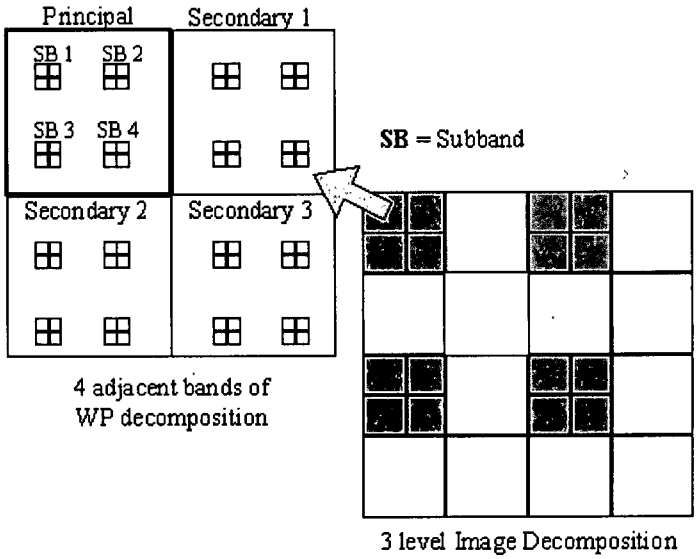


Figure 4.4: Coefficients Selection Approach (steps 4.)

step is important because we have to make sure the entire frequency spectrum is covered, as well as the entire image (in the spatial domain), in order to be able to assure authentication of the entire content. Furthermore, since we need to make sure that the images' characteristics are globally protected at each scale, the selected regions have to be spread in space, and their number should be sufficient to cover the image completely. In this context, the application of a four-level WP decomposition and the use of three secondary bands ($K = 3$) for each of the sixty-four principal bands with eight regions ($N = 8$) of four coefficients ($M = 4$) per band, is adequate for the protection of (256 by 256) images. A wavelet packets' domain mapping of coefficients selected for embedding is presented in Figure 4.5. In order to demonstrate that our selection process fulfils the above mentioned requirements (global space and

frequency covering), Figure 4.6 presents the mapping of the selected embedding coefficients in the spatial and frequency domains.

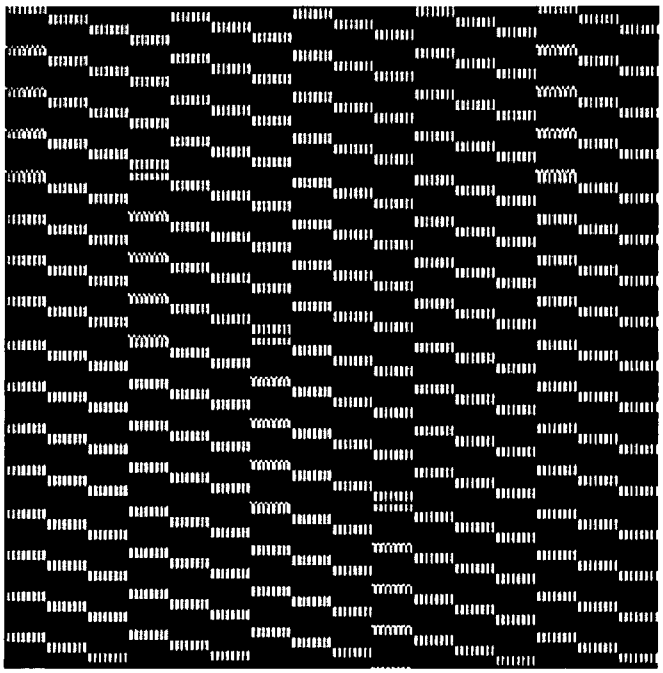
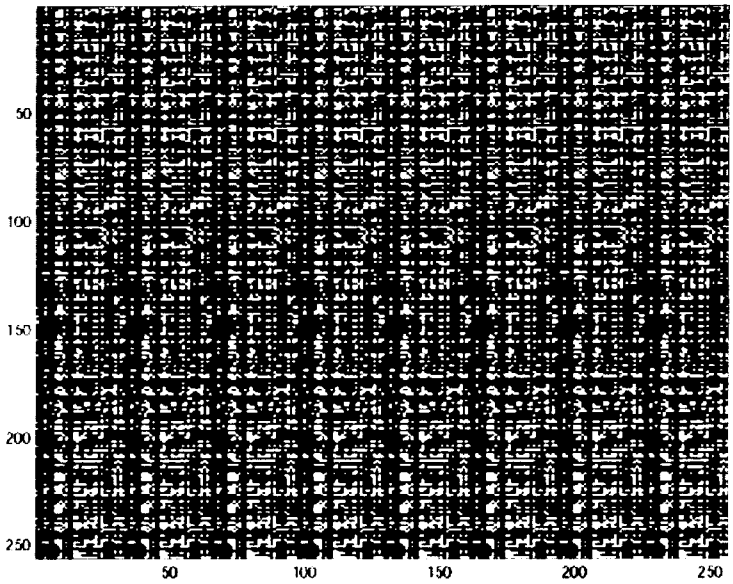
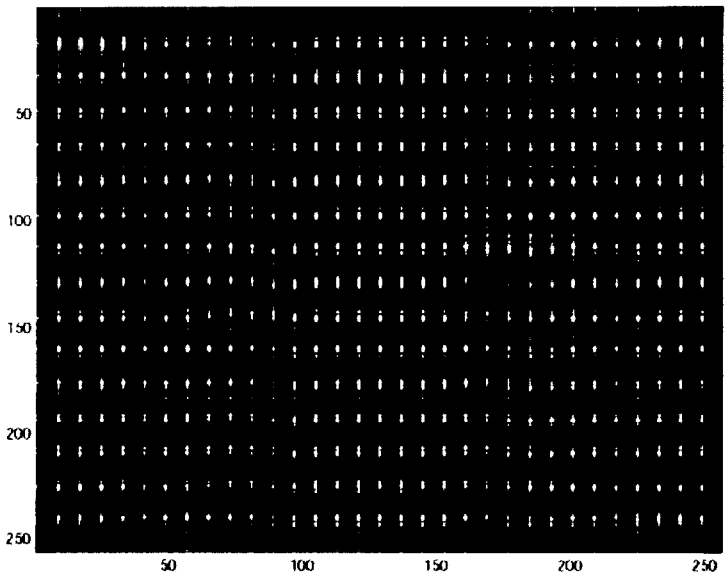


Figure 4.5: Selected Coefficients in the WP (Coifflets 12) domain

5. The author's secret key is used one last time for its embedding. As mentioned, each of the 64 bits is uniquely embedded in a group of one principal and K secondary bands. First, the means $Mean(i, j)$ of all the selected regions of WPC (in the principal and secondary bands) are individually computed. Then, the original quantization levels $q(i, j)$ are obtained (Equation 4.3) based on an optimal quantization step Δ (see Subsection 4.5.2). Afterwards, each bit of the author's identification key is inserted in the WP domain by the modification of the (individual) mean of the WPC regions belonging to each selected group of bands (one principal and K secondary). Rounding



(a)



(b)

Figure 4.6: Spatial (a) and Frequency (b) Mapping of Selected Coefficients

the mean to an even quantization level embeds a zero, while rounding the mean to an odd quantization level embeds a one. This is done by rounding the obtained quantization levels $q(i,j)$ to the nearest even/odd quantization levels (to form $q'(i,j)$) and then adjusting the mean of the WPC regions to the computed values (as demonstrated in Equations 4.4 and 4.5). The advantage of using this quantization technique is that the embedded information is modified as a function of the host. By opposition, an adversary can forge a fragile watermark more easily if the embedded pattern is not dependant of the cover work [20]. Therefore, quantization-based approaches increase the security of authentication systems by assuring the uniqueness of the resulting embedded mark. In addition, as we do not use integer-to-integer wavelet decomposition, the intensity values of the watermarked image that will be produced in **6.** are not guaranteed to be integers. Therefore, the use of quantization assures that the level of modifications introduced in the image is important enough to be kept by the discretization of the pixel values done in the image saving process.

$$q(i,j) = \lfloor \frac{Mean(i,j)}{\Delta} \rfloor \tag{4.3}$$

$$\begin{aligned} key[n] = 0 \quad q'(i,j) &= \begin{cases} q(i,j) & \text{if } q(i,j) \text{ even} \\ q(i,j)+1 & \text{if } q(i,j) \text{ odd} \end{cases} \\ key[n] = 1 \quad q'(i,j) &= \begin{cases} q(i,j)+1 & \text{if } q(i,j) \text{ even} \\ q(i,j) & \text{if } q(i,j) \text{ odd} \end{cases} \end{aligned} \tag{4.4}$$

$$Mean'(i,j) = q'(i,j) \cdot \Delta \tag{4.5}$$

In our embedding process, an optimal step Δ is used for the rounding of the mean of the defined regions belonging to the 64 principal bands. On the other hand, $\Delta/2$ is used for the regions belonging to the $64 \cdot K$ secondary bands. This is done to minimize the introduced distortion. Section 4.5.2 presents the procedure to obtain the optimal Δ for Laplacian distribution of unitary variance ($\sigma^2 = 1$). As the WPC of the selected bands do not have unitary variance, we have to weight each quantization step as a function of the distribution of the particular embedding band in order for it to truly represent the optimal quantization step. To achieve this, we compute the power (σ^2) contained in each of the selected bands, and use it as the modulating factor. In this way, our system takes advantage of the human visual system's characteristics by giving more weight to marks embedded in regions with more details for which the HVS is less sensitive [37] (see Figure 5.9). In fact, our system takes only implicit advantage of the characteristics of the visual system as it does not-contrary to what others have done [6]-weight the embedded marks as a function of the sensitivity of the human eye at each frequency. However, this is shortly shown to be more than sufficient to assure the invisibility of the marks.

6. Finally, we apply the appropriate wavelet packet synthesis bank on the available coefficients-some modified and some not-to reconstruct the visual data and form the watermarked image. As shown in Figure 4.7, the image produced is visually identical to the original *unmarked* image[9].

[9]More details are perceptible by comparing Figure 5.8 with Figure 5.7.

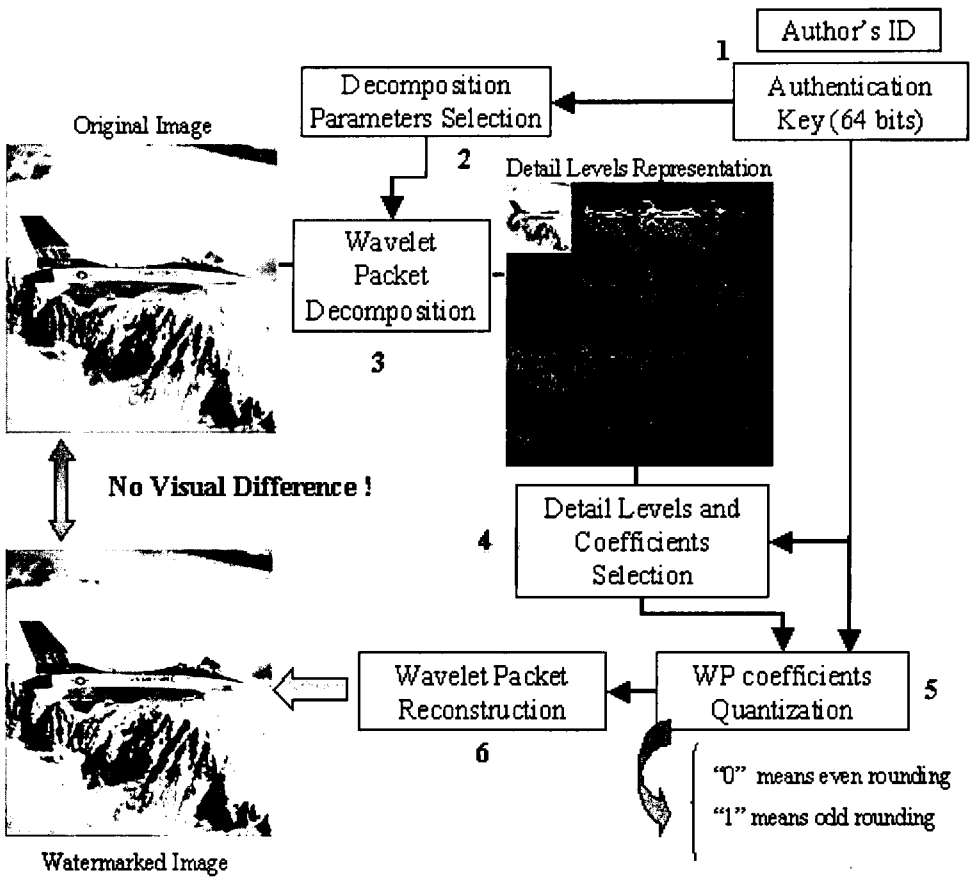


Figure 4.7: Embedding Scheme Developed

4.5.2 Optimal Quantization Step

As we have just explained, our watermarking algorithm uses the rounding of WP coefficients. The quantization step Δ should be chosen so as to maximize the embedding weight, while minimizing the distortion introduced. Thus, the choice of the optimal quantization step is of high importance. As we have not made explicit use of the HVS, we need to assure that minimal error is produced by quantization. In terms of watermarking, it means that we want to minimize the mean squared quantization error, which is as follows:

$$MSQE = \sigma_q^2 = 2 \sum_{i=1}^{M/2-1} \int_{(i-1)\Delta}^{i\Delta} (x - (\frac{2i-1}{2})\Delta)^2 f_x(x)dx + 2\int_{(M/2-1)\Delta}^{\infty} (x - (\frac{M-1}{2})\Delta)^2 f_x(x)dx \quad (4.6)$$

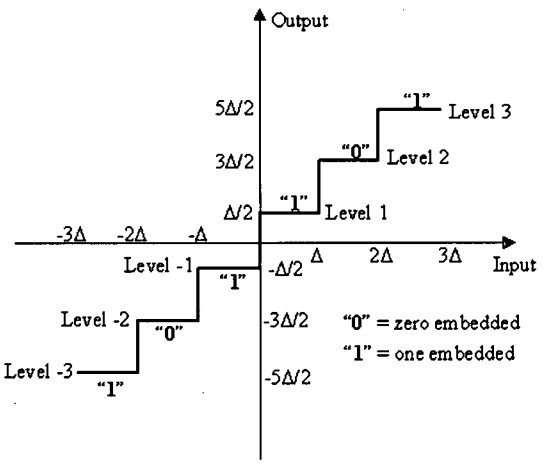


Figure 4.8: Input/Output Relation in the Quantization Process

In Equation 4.6, $f_x(x)$ is the probability distribution function (pdf) of the variable x, and M is the number of quantization levels to be used. Based

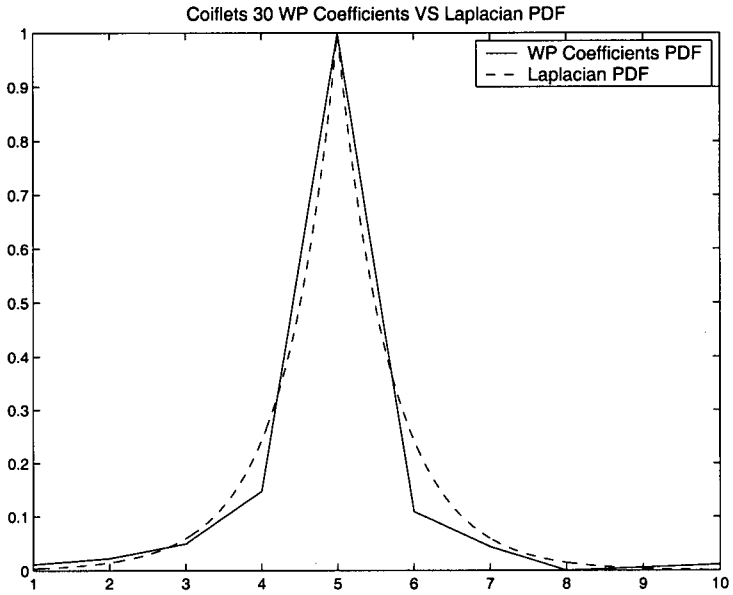
on the generic quantization scheme of Figure 4.8 [65], the minimization to achieve as a function of Δ is the following:

$$\frac{\delta\sigma_q^2}{\delta\Delta} = -\sum_{i=1}^{M/2-1}(2i-1)\int_{(i-1)\Delta}^{i\Delta}(x-(\frac{2i-1}{2})\Delta)f_x(x)dx$$
$$-(M-1)\int_{(M/2-1)\Delta}^{\infty}(x-(\frac{M-1}{2})\Delta)f_x(x)dx = 0 \qquad (4.7)$$

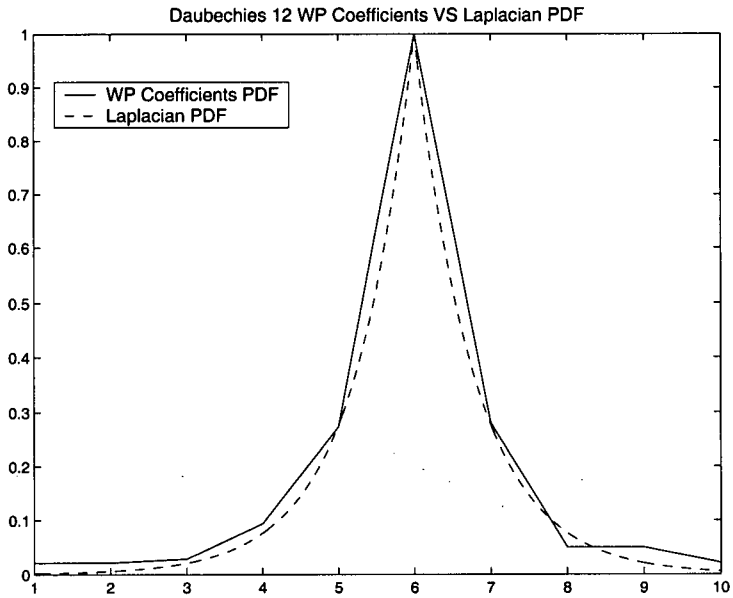
It has been shown that wavelet coefficients have *Laplacian* probability distribution functions [49]. In order to emphasize that this still holds for WP coefficients, we present, in Figure 4.9, the distribution associated with a high frequency band of a four level decomposition of an image using Coiflets 30, and the one associated with a low frequency band of a 4 level decomposition using the Daubechies 12 basis[10].

Fortunately, the problem of minimization has already been solved for this type of distribution [2]. It means that the optimal quantization steps are already available as a function of the number of quantization levels. In compression applications, the size of the alphabet-that is, the number of quantization levels-is selected to reflect the compression ratio desired. Here, it is the main parameter that controls the fragility of the embedded mark. Therefore, the steps used can be selected to reflect the degree of protection that is necessary to achieve. Of course, the visibility of the watermark also has to be kept in mind. As far as fragile watermarking is concerned, the visibility concern is similar to the fragility concern: the greater the number of quantization levels used, the less distortion introduced, and the more fragile the mark is. Typically, we use 32 quantization levels, as we have found that this value

[10]Five different images are used for the creation of each figure.



(a)



(b)

Figure 4.9: Probability Density Function of WP Coefficients: (a) Coiflets 30 and (b) Daubechies 12

Table 4.1: Optimum Step Sizes for Laplacian Distribution with $\sigma^2 = 1$ (from [65])

Alphabet size	**Laplacian Distribution**	
	Step Size	**SNR** (dB)
2	4.40	3.00
4	1.0873	7.05
6	0.8707	9.56
8	0.7309	11.39
10	0.6334	12.81
12	0.5613	13.98
14	0.5055	14.98
16	0.4609	15.84
32	0.2799	20.46

yields good perceptual transparency and authentication capabilities[11]. Therefore, the optimal normalized quantization step Δ used in our experiments is 0.2799.

4.5.3 Watermark Decoding Process

At the other end of the communication channel or after the image has been stored, the watermarked content needs to be authenticated. We developed a decoding procedure in order to extract the embedded mark, decide on the authenticity of the received image and localize the tampering if needed. The first four steps (step **1.** to step **4.**) of our decoding procedure are identical to the embedding ones. The author's unique key is used in order to decompose

[11]The values presented in Table 4.1 always have to be modulated by the variance of the band used, as explained in the previous section.

the image in levels of detail according to specific parameters. The regions of WP coefficients used in the embedding are selected. This allows us to recover a verification key without any use of the original *unmarked* image (see step **5.**). Then, the extracted mark is examined (steps **6.** and **7.**), and the areas with watermarked errors are labelled as tampered areas (**8.**).

5. The $64 \cdot N$ regions of wavelet packet coefficients belonging to the principal bands, and the $64 \cdot K \cdot N$ regions belonging to the secondary bands are examined. First, their mean $\hat{Mean}(i,j)$ is computed. Then, using the knowledge of the optimal quantization steps, we extract a verification sequence by rounding to the nearest quantization level: an even quantization level meaning zero and an odd quantization level meaning one (see Equations 4.8 and 4.9). We call this step *inverse quantization* even if it does not involve the removal of the effects of quantization. The principal bands' quantization subbands are first scanned. In the same way, the regions from the secondary bands are also examined to verify the authenticity of each of them. As the total number of regions of embedding is $64 \cdot (K+1) \cdot N$, this extraction procedure yields a verification sequence ($key_{verification}$) of $64 \cdot (K+1) \cdot N$ bits. This binary sequence is however, not used directly for tampering assessment as we chose to evaluate the embedding regions in the WP domain by intraband and interband comparisons.

$$\hat{q}(i,j) = round[\frac{\hat{Mean}(i,j)}{\Delta}] \tag{4.8}$$

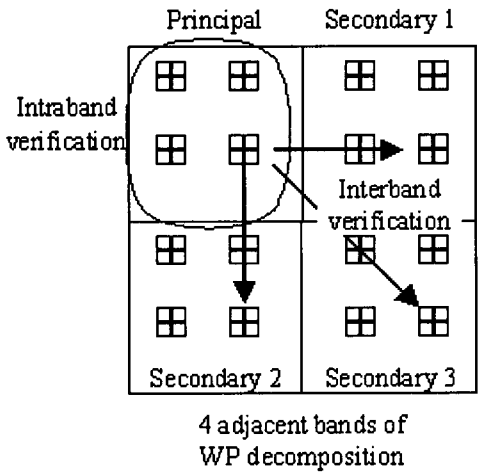


Figure 4.10: Intraband/Interband Verification Scheme (steps **6.** and **7.**)

$$\begin{aligned} key_{verification}[n] &= 0 \quad \text{if } \hat{q}(i,j) \text{ even} \\ key_{verification}[n] &= 1 \quad \text{if } \hat{q}(i,j) \text{ odd} \end{aligned} \tag{4.9}$$

6. Intraband comparison: associations are made between WPC regions belonging to the same principal band to decide if the image has suffered from any frequency tampering. Basically, we verify if the bits extracted from a given principal embedding band are of the same parity and if they conform to the corresponding originally embedded bit.

7. Interband comparison: the verification is now performed throughout the WPC regions associated with the same spatial area to decide whether the image has been spatially altered or not. In this step we verify if all the bits obtained from the $K+1$ WPC regions (from one principal band and K secondary bands) belonging to a given spatial region are of the same parity as the initially inserted bit.

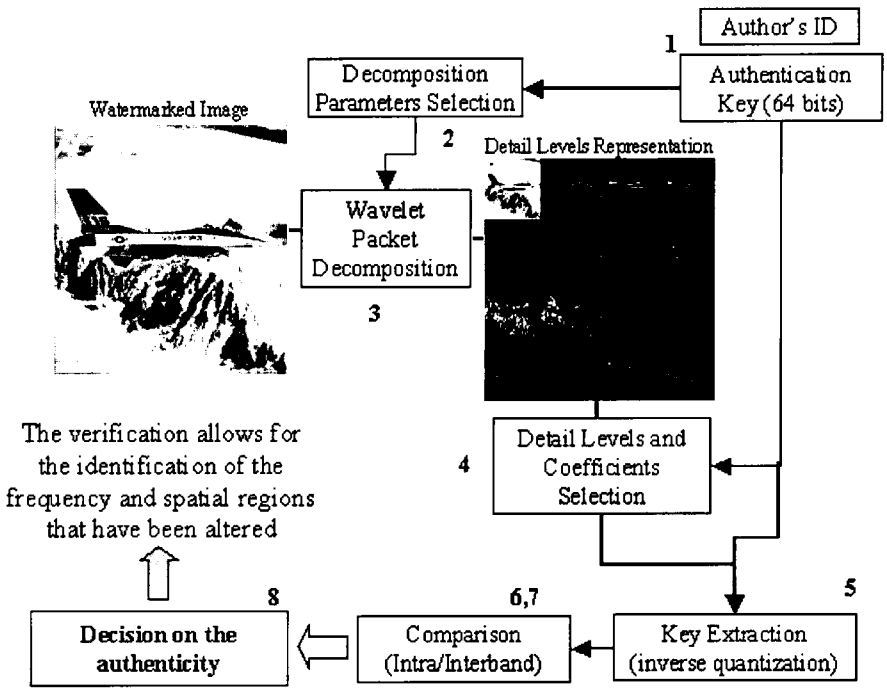


Figure 4.11: Decoding Scheme Developed

Intraband and interband verifications remove the need for post-processing operations by verifying the conformity of the embedded mark within a level of detail and along different bands for the same spatial region. In that way, it permits the quick decision of the spot or region of tampering in the image, and on the type of tampering, which can be either frequency alterations or modifications of pixel values.

The rules of detection are quite straightforward and based on experimentations with the embedding of 64 bits in 64 main bands, having 8 subbands of 4 coefficients, and 3 corresponding secondary bands each. First, we state the authenticity of a given spatial region by ensuring that a minimum number

of embedded marks (here three regions out of four) corresponding to it are untouched. For example, if the WPC region belonging to the principal band is authentic, then only two of the secondary regions have to be authentic too. However, if the principal region is not authentic, then all the secondary regions have to be genuine in order for the mark to be considered valuable. Obviously, a similar approach is adopted in the intraband verification; the comparisons are now performed within the same principal decomposition band. The entire FB is declared authentic if at least seven out of eight regions are still marked with the original key. Otherwise, the band is marked as frequency tampered. In addition, we include an overall detection rule that requires 504 regions of embedding (out of a total of 512) extracted from the principal bands to have the same quantization level parity as originally embedded for the image to be considered genuine[12]. As each bit is embedded in 8 regions of a unique principal band, requesting 504 authentic ones assures the detection of one single different bit. Of course, other sets of rules can be defined in order to fully reflect the degree of protection desired. For that reason, they can be considered as the detection parameters to be chosen for each application.

8. Finally, based on the results of steps **5.**, **6.** and **7.**, a decision is made on the authenticity. If it is decided that the image has been tampered with, the altered frequency or spatial region(s) is(are) identified. Localization is useful because the knowledge of when or where a work has been altered can be used to understand the motive of tampering, the possible candidate adversaries, and whether the alteration is legitimate. Localization of the tampering

[12]This is considering the use of 64 principal bands with 8 regions of embedding §4.5.1.

can be considered as a special case of *telltale* watermarking. To determine whether or not an attack is malicious or acceptable, a basic rule of thumb is to consider the conclusions that might be drawn from the images' use. If the distortions introduced do not change these conclusions, then the attack is considered not malicious [20]. In our case, high-quality JPEG and JPEG-2000 compression operations are considered as the only legitimate image processing operations as they are used for storage purposes, but do not alter the visual content. Techniques destined to extend the use of our system to medium quality JPEG compression are presented in Section 5.5. Nevertheless, the system, as proposed, shows enough robustness to be invariant to commonly used high quality compression standards.

4.6 Summary

This chapter was exclusively dedicated to digital data authentication procedures. First, we laid the basis of such systems, and grouped previously proposed approaches. Then, we used this classification to highlight the most important requirements of content authentication schemes. In addition to being able to determine if digital content has been tampered with, authentication schemes must allow for the evaluation of where and how an image has been altered. In addition, they must include security features preventing the reference mark to be unnoticeably forged or manipulated. Then, we have presented previous approaches working in the spatial, DCT and WT domains. We also highlighted problems with each technique, and explained how we overcame

them in our digital watermarking image authentication method. We proposed to embed a secret author's identification key by rounding the mean of wavelet packet coefficients' regions to even or odd quantization levels. We developed an adaptive quantization procedure that makes use of the knowledge of the (*Laplacian*) distribution of WPC to maximize the mark embedding weight, while guaranteeing the invisibility of the watermark. Finally, we introduced an interband/intraband verification procedure, and explained how it is used in our decoding procedure to extract the watermark present in the image and decide if any tampering has occurred either in space or frequency. These procedures make obsolete the use of any post detection operations employed to judge the overall tampering. Experimental results obtained with the proposed technique are presented in the following chapter, along with its comparison with a commercially available software and strategies to augment its robustness to JPEG compression.

Chapter 5

Experimental Results

"A rock pile ceases to be a rock pile the moment a single man contemplates it, bearing within him the image of a cathedral."

–Antoine De Saint-Exupery, writer and poet (1900-1944)

5.1 Introduction

This chapter explains the experimental results, which verify the capabilities of our watermarking scheme. In order to do this, we use real and computer generated square-size images. We create a set of seven hundred watermarked images[1] and test their visual quality and content authenticity. Since we used the *Wavekit* software for *Matlab* to perform the wavelet packet decomposition, our system is currently limited to squared size images. In reality, input images must have the dimensions 2^n by 2^n to allow for the computation of the two

[1]In fact, we create 720 watermarked versions of 16 original images, all shown in Figures 5.1 and 5.2, using 45 different embedding signatures.

dimensional WT. This is due to the factor 2 subsampling operation performed after each filtering step, and the fact that there is no interpolation executed between the decomposition's scales. In future developments, however, we can design our own implementation of WP in order to avoid this limitation. As stated earlier, the ultimate goal of our work is to lay the basis of a watermarking scheme for image authentication, and not to develop a new wavelet analysis tool for *Matlab*. Moreover, optimization certainly has to be made before the *commercial* exploitation of our system is considered. This definitely involves the use of an instrument other than *Matlab* to implement the functions for wavelet packet decomposition. At any rate, we use *Wavekit* here, and the images presented should make the capacities of our wavelet packets-based watermarking system clear.

In the first section, we confirm the invisibility of the embedded marks, as well as the authentication capabilities of our system. Then, we prove its ability to detect and localize tampering, even in the presence of compression. Several wavelet functions are considered and their impact on the authentication procedure is commented upon. In Section 5.4, we compare our system with a commercially available-spatial-based-watermarking tool in terms of the visibility of the marks, tampering localization aptitude and resistance to attacks. Finally, different strategies meant to increase the robustness of our system are examined in the last section.

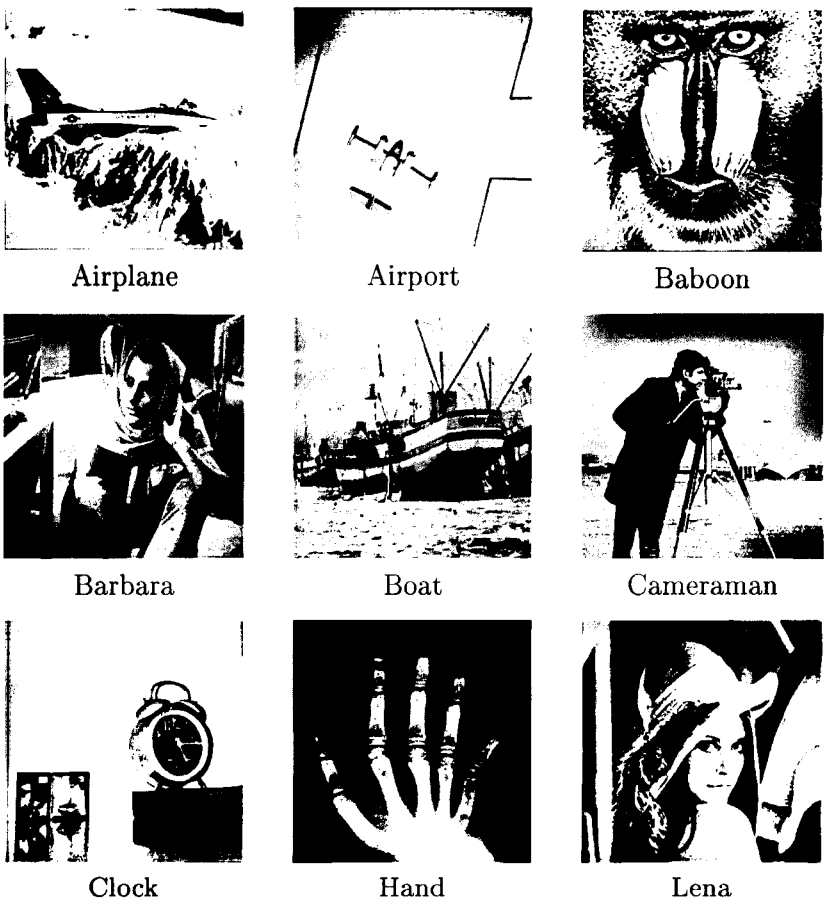


Figure 5.1: Test Images

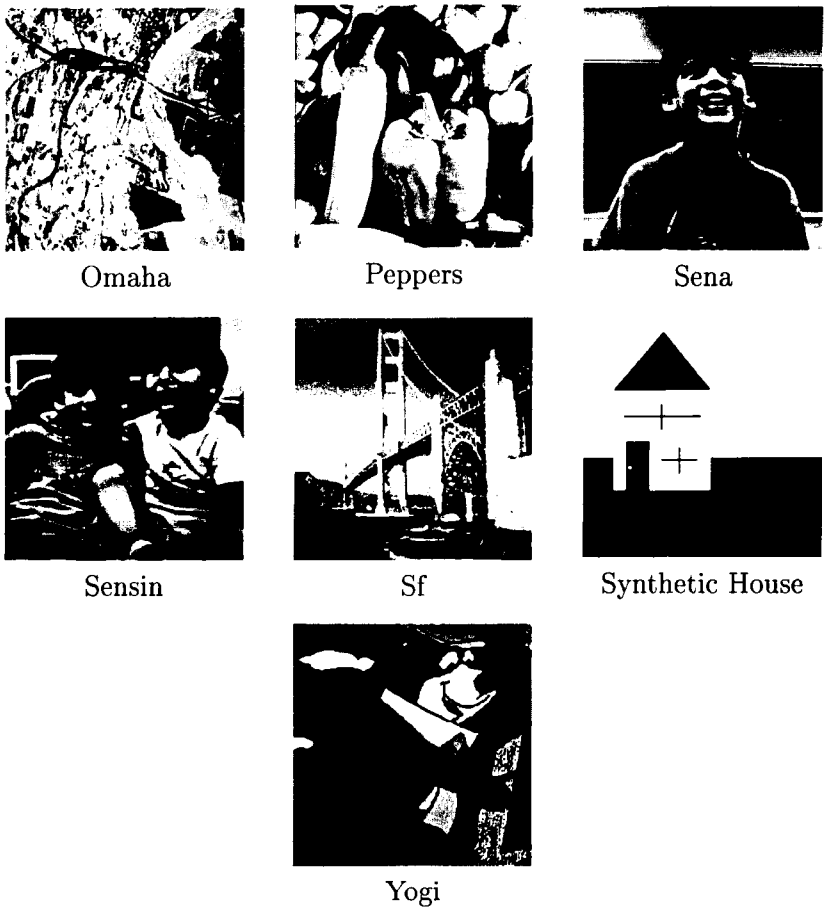


Figure 5.2: Test Images

5.2 Embedding, Decoding and Visibility

First we have made sure that our embedding system does not introduce visual artefacts in images. We have first measured the visual quality of the marked images by qualitative observations. However, in order to produce more objective results we have also used the *Peak Signal-to-Noise Ratio* (PSNR) defined in Equation 5.1, which measures the difference between an original image $I(i,j)$ and its modified version $I'(i,j)$. With the set of images produced, we have obtained an average PSNR of 42.47 *dB*. This is above the usually tolerated degradation level of 40 *dB*. Figures 5.3 to 5.8 clearly show that the vast majority of the watermarked images are not perceptually different from the unprotected ones[2] and that the differences are, in fact, more important in the regions of detail (see Figure 5.9). Figure 5.10 summarizes the results obtained with four different images. For clarity, not all images tested are included. Figure B.7 is shown in the appendix to complete the one presented here.

$$PSNR(dB) = 10 \cdot \log[\frac{max(I(i,j))^2}{\sum_{N,M}(I'(i,j) - I(i,j))^2}] \tag{5.1}$$

Although *Daubechies* functions in general, seem to yield slightly larger *PSNR*, the choice of the wavelet function does not clearly influence the visual quality of watermarked images (Table 5.1). In fact, Coiflets 12 yields the best results of all, in terms of *PSNR*. The length of the mother wavelet seem to have a greater influence on the visual quality of marked images-a shorter function means smaller distorted spatial regions-but the results are not yet absolute

[2]Visual differences between the watermarked image and its original result from saving and/or printing operations.

Figure 5.3: Original Barbara Image

Figure 5.4: Watermarked Barbara Image (Coiflets 12 with $PSNR = 41.76dB$)

Figure 5.5: Watermarked Barbara Image (Coiflets 24 with $PSNR = 41.88dB$)

Figure 5.6: Watermarked Barbara Image (Daubechies 12 with $PSNR = 42.72dB$)

Figure 5.7: Original Airplane Image

Figure 5.8: Watermarked Airplane Image (Coiflets 24, $PSNR = 43.15dB$)

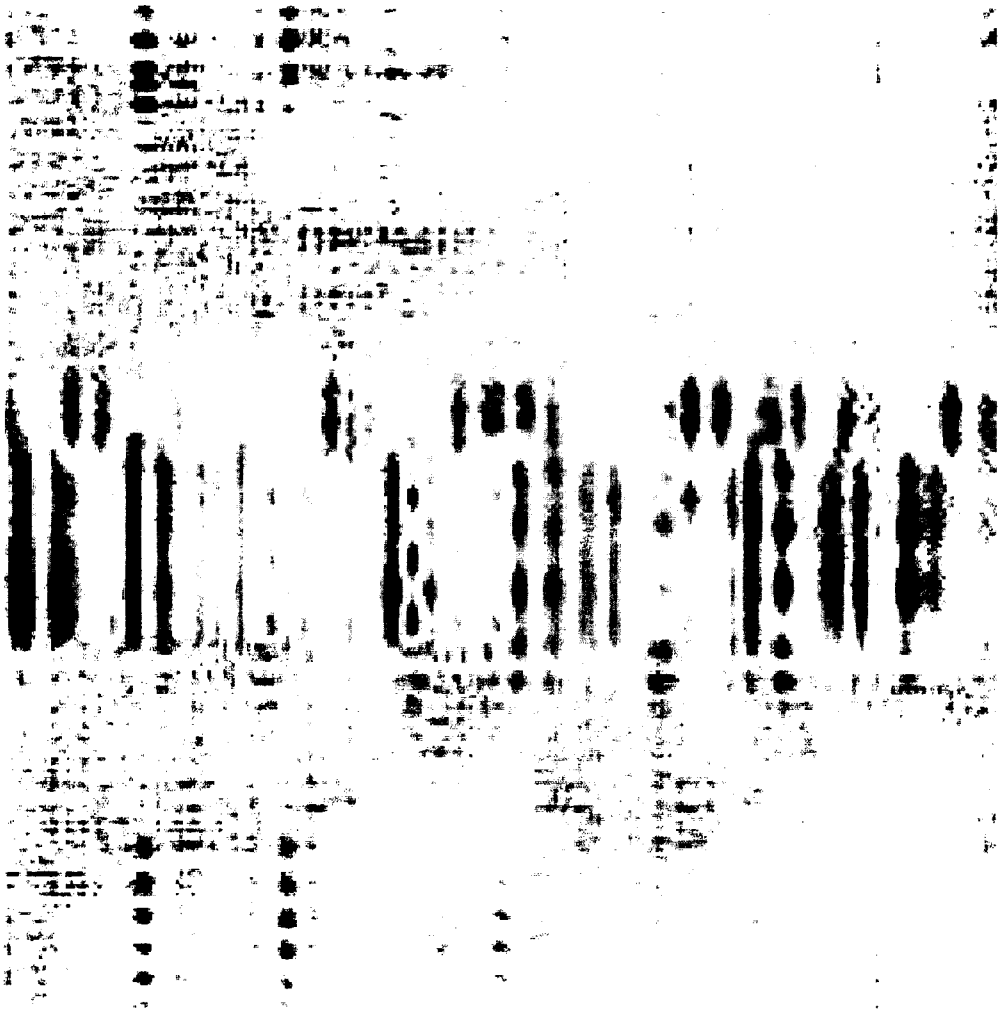


Figure 5.9: Difference between the Original and the Watermarked Airplane Images (the grayscale has been magnified for visualization, black regions referring to large differences)

Table 5.1: Average PSNR for Different Wavelet Functions

Wavelet Function	**Average PSNR** *dB*
Coiflets 12	43.40
Coiflets 18	42.05
Coiflets 24	41.78
Coiflets 30	41.98
Daubechies 12	43.25
Daubechies 16	43.25
Overall	42.47

enough. The characteristics of the original images, however, greatly affect the ability of the embedded mark to be unnoticed. For example, it is difficult for our algorithm to find adequate regions of embedding in the *Synthetic House* or *Yogi* images, as they do not include much detail. In fact, as they are computer generated, they contain several regions of uniform intensity. Therefore, small perturbations are quickly noticeable by the human eye. On the other hand, the *Omaha* image is easy to mark imperceptibly since it contains a lot of detail[3]. However, as two images of different perceptual quality can have the same *PSNR*, the *PSNR* does not seem to measure the absolute visual quality of images. Nevertheless, the *peak signal-to-noise ratio* is the current standard of comparison amongst watermarking techniques. Moreover, a commonly accepted distortion quantification technique that fully takes into consideration the characteristics of the HVS has yet to be developed. For these reasons, we have adopted the *PSNR* in our measurements. From these, a general rule

[3]To compare the characteristics of *Synthetic House* and *Omaha*, see Figure 5.2.

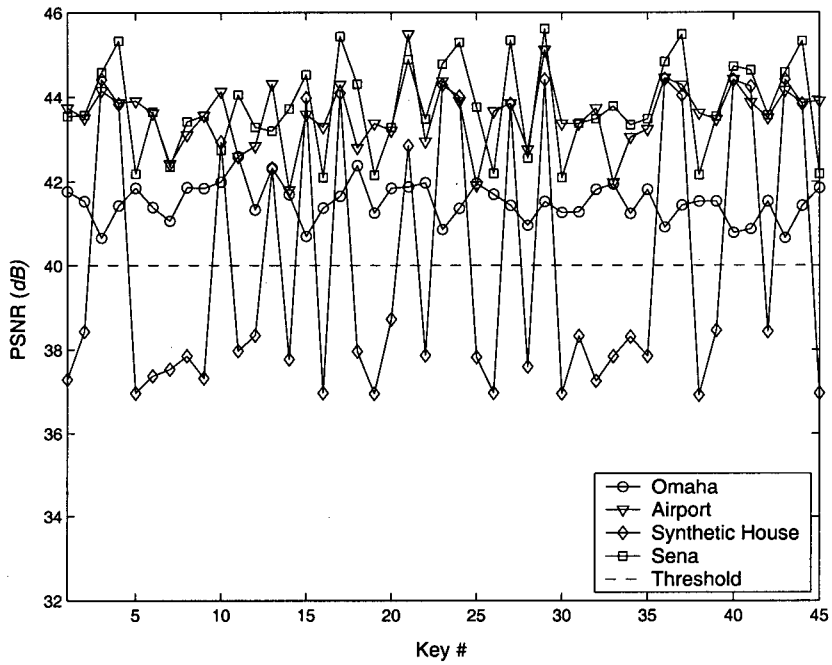


Figure 5.10: PSNR Values for Different Embedding Keys

can be extracted: our system is more suitable for *photographic-like* gray-scale images since they have more detail in which to hide a watermark.

Another important factor is to test the ability of our system to authenticate genuine (not tampered) images. Therefore, while testing the invisibility of the marks, we also tested the capacity of the embedded marks to de detected by our intra/inter-band verification approach. We had to ensure the percentage of *false negative*-the number of times where authentic images are declared tampered-and the number of *false positive*-the number of times where unmarked, wrongly marked or tampered images are declared genuine-are both kept to a minimum.

The later is the easiest to prove since our system is designed to avoid this kind of situation. The embedding of a binary mark-here, a sequence of 64 zeros and ones-limits the average percentage of random positive matches to 50 % for images with no watermark.

As our detection process requires 98.4%[4] of the extracted mark to be identical to the original one for the image to be considered genuine, the number of images falsely declared valid is inherently low. Theoretically, false positives are possible. However, in practice, we have not encountered any in the seven hundred images tested. From this, we can comment on the low false positive rate and the capacity of our system to reject unmarked or wrongly marked images.

On the other hand, more experiments are needed to confirm the authentication ability of our scheme. Using the same set of 720 images, we applied the authentication procedure. We found that, even if some low-level marks are sometimes destroyed by the simple discretization of the pixel value (i.e. saving the image), our intra/interband verification approach is able to declare the image authentic 99.87% of the time (see Figure 5.11[5]). Once again, results show that the choice of the wavelet function in the image decomposition does not appreciably influence the detection rates. These range from 99.70% of marked regions unchanged by the discretization of pixel values[6] with Coiflets 18 wavelets to 99.97% (of marked regions unchanged by the discretization of pixel values) with Daubechies 16 wavelets (see Table 5.2). Then again, the

[4] 504 regions of embedding out of a total of 512, see §4.5.3.

[5] Again, not all the images are represented but more results are shown in B.8.

[6] This is done naturally to store the images in the bitmap format.

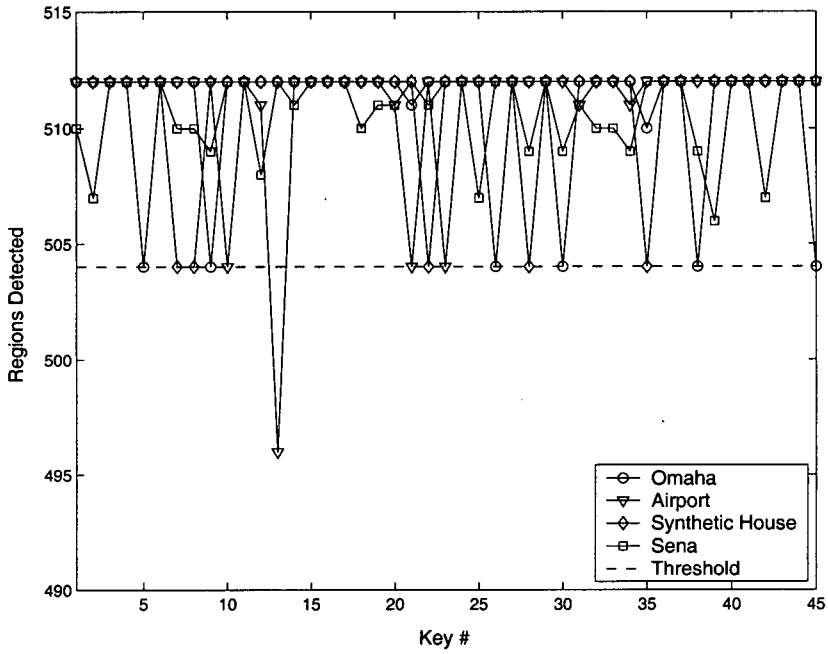


Figure 5.11: Detection Rates Achieved for Authentic Images

characteristics of the original images influence the detection capacities. For example, the well known-and highly detailed-*Barbara*, *Peppers* and *Airplane* images, which are easy to watermark, all yielded 100% detection rate, while the less textured *Clock* image is harder to authenticate. Nevertheless, the results unmistakably prove that the decoding method proposed in Subsection 4.5.3 permits the authentication of untampered images.

In this section, we have made explicit that our WP-based system is able to embed a mark imperceptibly in an image. Furthermore, we have demonstrated that the marks are detectable with the proposed extraction procedure, and that genuine images can be authenticated easily. Finally, these results have been shown to be independent of the wavelet function chosen for the

Table 5.2: Average Detection Rate for Different Wavelet Functions

Wavelet Function	**Average Detection** % of valid regions
Coiflets 12	99.90
Coiflets 18	99.70
Coiflets 24	99.85
Coiflets 30	99.82
Daubechies 12	99.87
Daubechies 16	99.97
Overall	99.87

decomposition, at least amongst the functions tested. Now, we have to make sure altered images can also be identified, and their tampering highlighted.

5.3 Tampering Detection

An important aspect of our system is its ability to localize image tampering. For that reason, we have tampered with previously watermarked images and ran the verification process. This was done to make sure our system is able to detect and highlight the doctoring, which is the malicious modification of an image with the intention of adding or removing information. Figure 5.12 shows the tampering done on the previously watermarked *Barbara* image by the addition of a second bookshelf to the right of the existing one, as well as the tampering detection obtained with our WP-based system. In addition, as we wanted our system to be robust to high quality JPEG compression, we have compressed the tampered images and ran the verification process again. Figure 5.13 shows the results obtained for the *Barbara* image. From these, we

found that the ability of our system to detect and localize tampering is good, even in the presence of JPEG compression. To complete and to evaluate the ability of the authentication scheme to detect frequency tampering, we have marked the *Baboon* image and tampered with it by low pass filtering[7]. This operation, although not perceptible, gets rid of small detail in the image. The results obtained for the frequency localization are lastly, presented in Figure 5.14.

As shown, the spatial tampering recognition does not pinpoint specific pixels, but defines regions of highly probable tampering. This is due to the following: first, the spatial correspondence between WP coefficients and pixels is not one-to-one since the wavelet and scaling functions are not of unitary length (the functions actually span over more than one pixel), and second, we use clusters of WP coefficients. Nevertheless, we have found that 5 by 5 pixels doctoring[8] can be detected with our authentication method, but that it is more reliable in detecting and accurately localizing modifications larger than 10 by 10 pixels. Furthermore, as shown in Figure 5.14, frequency tampering is identified in terms of altered bands, with the white ones being the tampered bands. Figure 5.14 (f) clearly shows that the low frequency bands (top-left corner) are untouched, while tampered bands become more plentiful in the higher frequency bands, which corresponds to the modification of the frequency spectrum of the watermarked image (Figures 5.14 (d) and (e)). It is also possible to highlight the corresponding most severely altered spatial re-

[7]A 3 by 3 *Gaussian* filter was applied to the entire image.
[8]In 256 by 256 images.

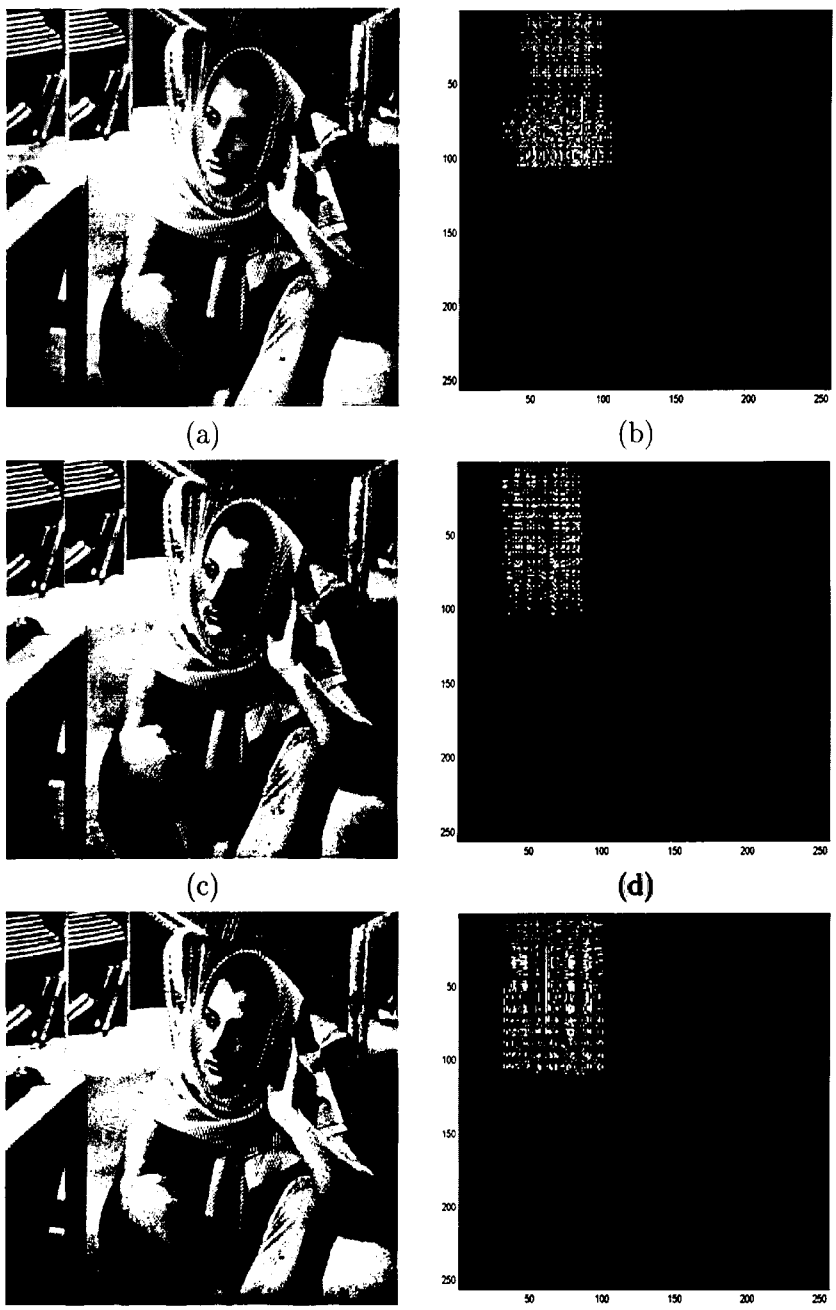


Figure 5.12: Tampered Watermarked Barbara Images (bookshelf added to the right of the existing one) with the Detection Results using Coiflets 12 (a,b), Coiflets 24 (c,d) and Daubechies 12 (e,f)

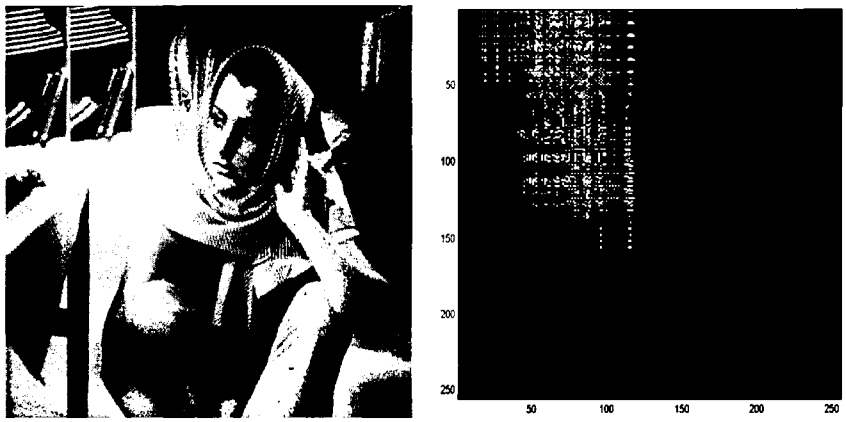


Figure 5.13: Compressed (3:1) Tampered Watermarked Image (Coiflets 12) and Detection of Spatial Tampering

gions. Accordingly, these results guarantee the ability of our scheme to detect malicious tampering meant to destroy spatial detail or spectral information in high quality images. This is an important step towards *telltale* watermarking, as it gives a better understanding of how an image has been corrupted by examining how a known embedded watermark has been modified. In addition, our system is able to withstand a certain degree of JPEG and JPEG-2000 compression, up to a ratio 3 : 1 for the former, and 2.23 : 1 for the later. This, as will be discussed, forms an advantage compared to available schemes. In summary, as previously shown in [56], our system is truly able to detect malicious tampering, both in the spatial and frequency domains, even in the presence of high quality compression.

To finish this section, we highlight general observations. First, we have found that, for 256 by 256 grey images, the use of 4 levels' decomposition increases the robustness to JPEG compression-hence, the use of 4 levels in

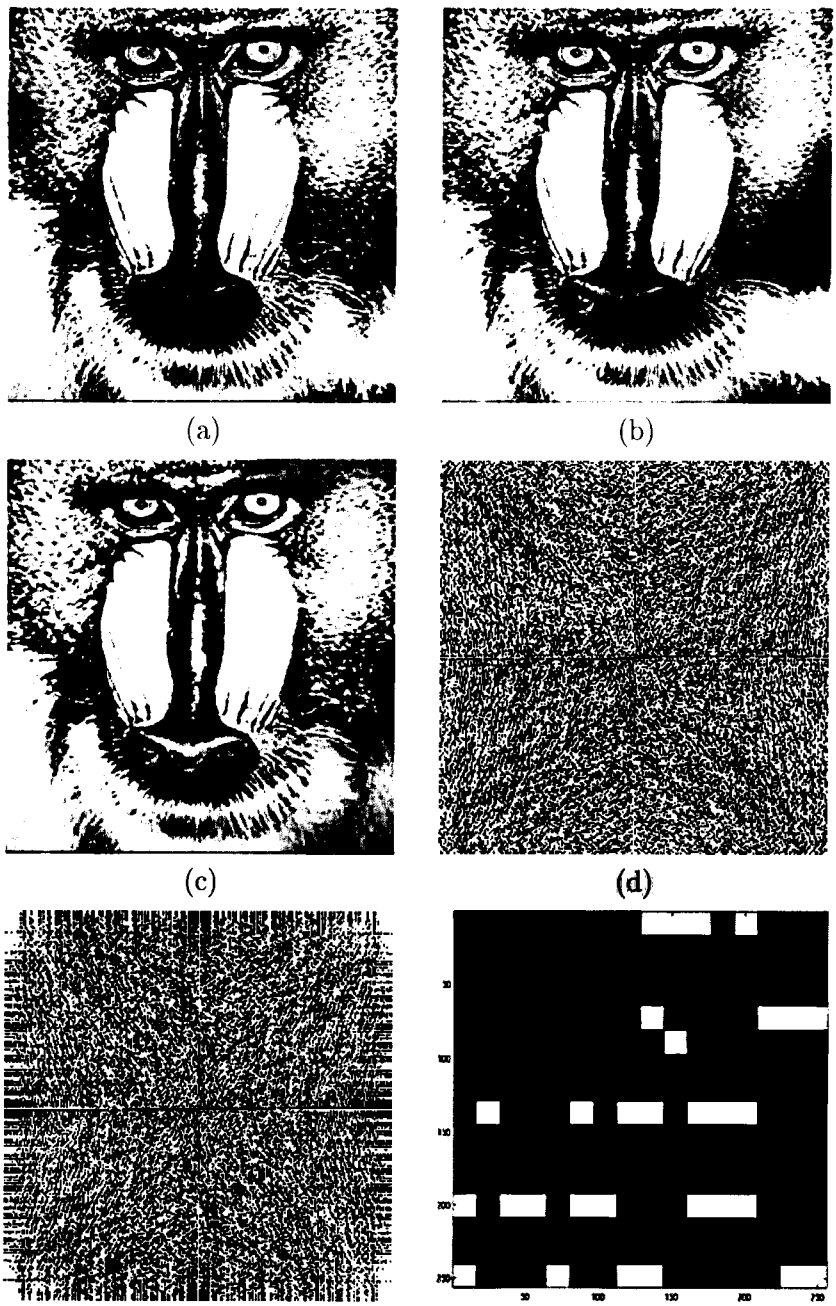


Figure 5.14: Original (a), Watermarked (b) and LP (watermarked) Baboon images (c). Frequency spectrums of the Watermarked (d) and LPF images (e). Frequency Detection of Tampering with our WP-based Approach (f)

all the results presented. This was anticipated since the use of only 4 levels allows deeper mark embedding in the most significant parts of images. As the presence of fewer decomposition levels generates fewer decomposition bands, it also worsens the frequency resolution. This is however, compensated by better spatial resolution resulting from the wavelets' fundamental principles (Figures 2.8 and 2.12). In the same line of thought, decompositions based on Coiflets functions yield better authentication and more accurate frequency and spatial localization of tampering. Intrinsic characteristics of Coiflets wavelets, such as the highest number of vanishing moments (both for $\psi(n)$ and $\phi(n)$) and the shorter effective length of the functions, may allow for better control of the effects of WPC quantization on the resulting images and, therefore, gives more assurance that embedded marks will stay after the *discretization* of the images' intensity. All the same, in order to complete our evaluation, we need to see how our authentication approach compares with other available tools.

5.4 Comparison with Eïkonamark

As *copyright protection* is the most common application of digital watermarking, some efforts have been put forth for the development of testing benchmarks for *robust* watermarking schemes. *Stirmark*[9], *Certimark*[10] and *Optimark*[11] are examples of systems that try to solve problems associated with the evaluation and comparison of watermarking systems. Stirmark is propably the most wide-

[9]http://www.cl.cam.ac.uk/ fapp2/watermarking/stirmark/
[10]http://www.igd.fraunhofer.de/igd-a8/projects/certimark/
[11]http://poseidon.csd.auth.gr/optimark/

ley accepted, as it was the first to be developed, as early as 1997. However, Optimark solves the complexity constraint by allowing for the performance assessment by a unique quality factor. By the use of a unique *score*, it grades the quality of a watermarking technique. It is therefore, a promising candidate for future reference. Nevertheless, all those benchmarks are meant to compare robust watermarking schemes only.

In the absence of evaluation standards for fragile watermarking systems, we need to compare our scheme with previsously proposed ones on a one on one basis. For that reason, we have chosen a commercialy available watermarking software called *Eïkonamark*. It is a crude implementation of [5], and comes from the *Alpha-tec* corporation [82]. Using only the image authentication capabilities of *Eïkonamark* (which also offers robust embedding for copyright protection), we compared the quality of the produced images, the ability to detect tampering, and finally, the resistance to collage attacks achieved by this software with the results obtained with our system.

5.4.1 Image Quality and Tampering Detection

As the first comparison step, we produced Eïkonamark watermarked (Eïkonamarked) images and compared them with the ones obtained with our novel approach. Figures 5.15 to 5.18 show original images in comparison with their Eïkonamarked version. From the comparison of the average *PSNR* value obtained with Eïkonamarked (38.42 *dB*), with the one obtained with our system (42.62 *dB*), it is clear that the commercial software degrades the images more

than our method does.

Figure 5.15: Original Barbara Image

Another important characteristic is the ability of the systems to correctly identify modified regions in marked images. To assess this for Eïkonamark, we have performed the same doctoring tests as the ones presented in Section 5.3.

Figure 5.16: Eïkonamarked Barbara Image ($PSNR$= 38.51 dB)

Figure 5.17: Original Cameraman Image

Figure 5.18: Eïkonamarked Cameraman Image ($PSNR$= 38.37 dB)

Figure 5.19: Tampered Eïkonamarked Barbara Image and Detection with Eïkonamark

The tampering detection quality achieved with Eïkonamark (Figure 5.19) is comparable with the results obtained with our system (shown in Figure 5.12). However, as Eïkonamark is spatially based, it yields a slightly better delimitation of the modified areas. On the other hand, we found that it is possible for doctored regions to go unnoticed in several cases. First, if the tampering of a 256 by 256 gray scale image is smaller than 10 by 10 pixels, we have found that Eïkonamark is not able to detect the tampering. At a size of 10 by 10 pixels, the doctoring is detected in the sense that the authentication key is not 100% found, but the tampering cannot be localized. In fact, if the size of the region of tampering is 10 by X (or X by 10), where $X \leq 30$, the watermark is not noticeably broken, and the modified area cannot be spatially pin pointed. Moreover, Eïkonamark might generate a somewhat better detection of spatial tampering for *medium-size* corrupted areas, but it does not allow for the localization of frequency tampering. In addition, we have

found the system to be highly susceptible to high quality JPEG compression, as shown in Figure 5.20.

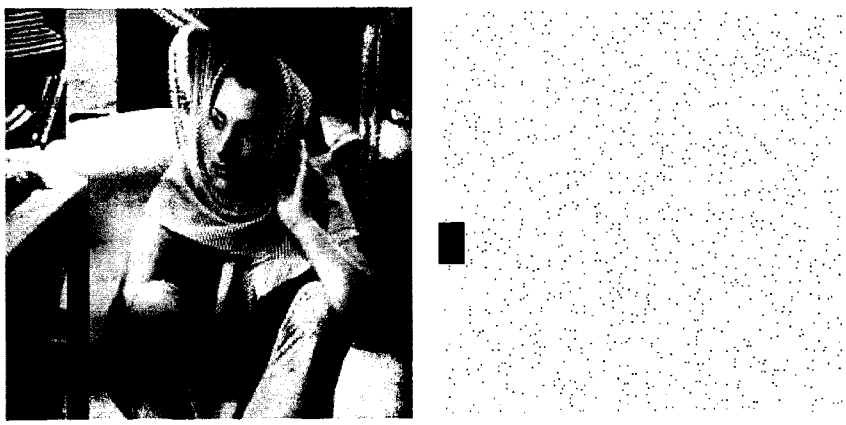
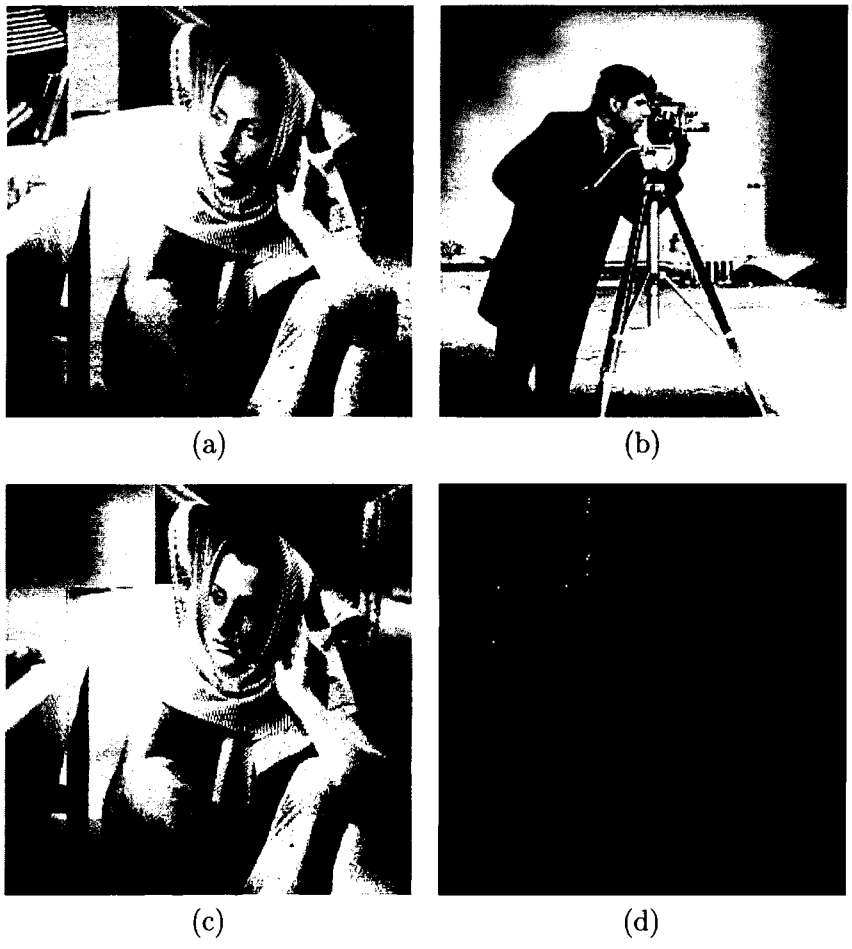


Figure 5.20: Compressed (3:1) Eïkonamarked Barbara Image and Authenticity Detection

5.4.2 Resistance to Collage Attacks

Since the goal of authentication methods is to detect any unlawful modification or tampering, it is of utmost importance to be able to uncover any kind of alterations. Forged attacks are often problematical since they are designed to bypass implemented protections. In that sense, collage attacks (see Subsection 4.4.1) are easy to realize and have been proven effective at defeating authentication procedures. For that reason, the last aspect of our comparison considers the capacity of the marked images to resist collage attacks, or more precisely, the ability of the authentication processes to detect them. For our system and Eïkonamark, we first obtained two watermarked images. Then, we produced a tampered image by using parts of the two *authentic* images.

Finally, tamper detection results are presented, both for Eïkonamark and our WP-based approach.



(a) (b)

(c) (d)

Figure 5.21: Eïkonamarked Barbara (a) and Cameraman (b) Images with the Mixed Version (c) , notice the disappearance of books from top left corner, and the Tampering Detection Result with Eïkonamark (d)

From Figure 5.21, it is obvious that Eïkonamark is easy to defeat by the use of collage attack. Even if the attack is made quite apparent for visualization reasons, the spatial-based approach is not able to detect the combination of the

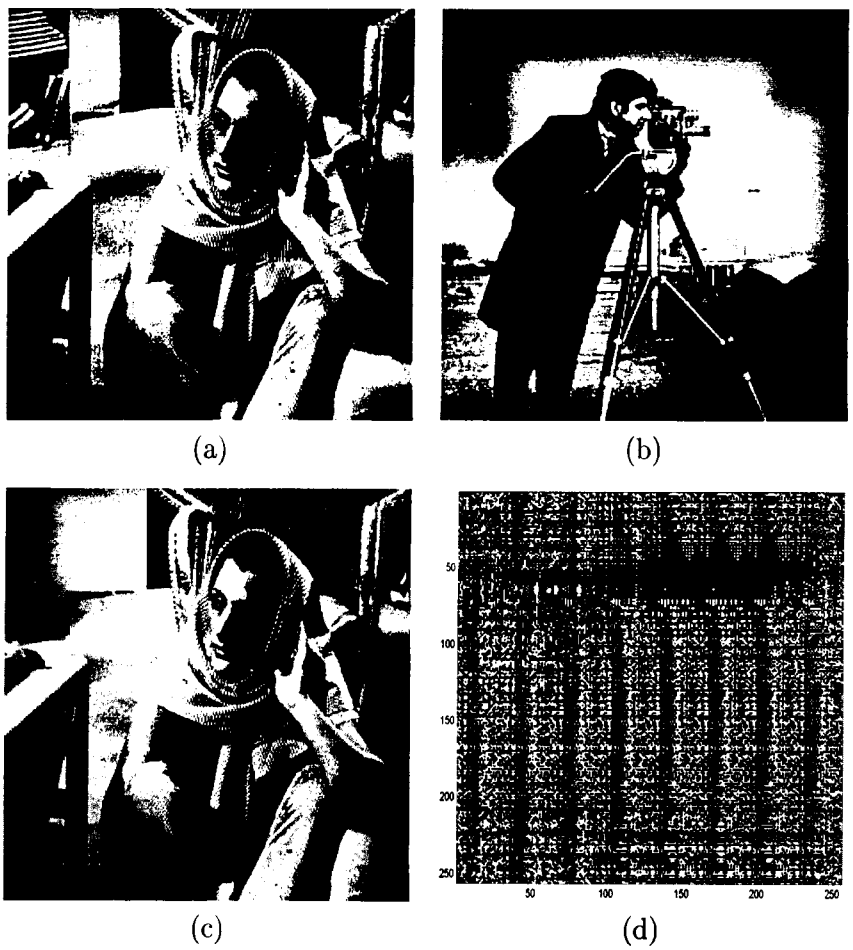


Figure 5.22: Watermarked Barbara (a) and Cameraman (b) Images with the Mixed Version (c), and the Tampering Detection Result with our WP-based Approach (d)

original images. Of course, this is a major security flaw as collage attacks are easily implemented and are quite effective at removing (or adding) important visual information. By comparison, our system is able to declare the image *tampered.* As shown, our WP-based approach is not able to locate the changes due to the high level of modifications detected. In fact, in this case, the localization of the doctoring is not of utmost importance as the image inspected is formed by the combination of two genuine images. Therefore, no region can be considered more or less tampered than the others. It is the image, as a whole, that needs to be considered unauthentic. From this, the results obtained are obviously satisfactory, as we have shown that the proposed method is more secure than the commercially available Eïkonamark software. In addition, as stated earlier, straightforward spatial embedding eases search attacks, while the embedding in a wavelet domain-unknown to potential attackers-prevents it or, at least, makes it substantially more lengthy and difficult.

5.4.3 Summary of Comparisons

To sum up, we have demonstrated that our system introduces less visual distortion when embedding marks in gray-scale images. We have shown that the quality of spatial tampering detection is equivalent for both systems, with a slight advantage, as far as localization of large tampered regions, for Eïkonamark. However, our system includes the recognition of frequency tampering in addition to spatial ones. Moreover, it surpasses Eïkonamark in its capacity to tolerate image processing operations since it already includes ro-

bustness to high quality JPEG compression. Furthermore, our system is able to detect collage attacks. The only advantage of Eïkonamark is that it is less computationally expensive. This is mainly due to the nature of the two systems-one is spatial-based while the other works in a transform domain-but also to the tool used in our implementation, which is not fully optimized. In conclusion, our system outperforms Eïkonamark in most of the aspects investigated. With minor changes and optimization, it can certainly be developed into commercial software.

5.5 Robustness to JPEG Compression

The first goal of our project is, indeed, to develop a watermarking scheme for image authentication that can withstand a certain degree of image compression. As already shown, the system proposed in the previous chapter can tolerate a reasonable degree of image compression. Generally speaking, JPEG recommends a quality factor between 75 and 95[12] for a compressed image to be visually indistinguishable from the original one, and between 50 and 75 to be merely acceptable[13]. In applications where small image detail have to be kept intact, such as identification photos, medical images or video security systems, we need to make sure that images are visually unchanged by compression. For that reason, a quality factor of 95 is the lowest tolerable in these

[12]The quality factor controls the compression ratio by deciding the quantization table used in the DCT coefficients quantization. If $QF < 50$ then $k = \frac{50}{QF}$, otherwise $k = \frac{200-2*QF}{100}$ where k is the multiplication factor for the quantization table. As an indication, a QF of 92 results in a 3:1 compression ratio.

[13]http://www.faqs.org/faqs/jpeg-faq

contexts. In other cases for which the visual content of images is important but not vital, more substantial compression gains might be valuable in order to permit efficient storage of the data.

In this work, we want to add an additional module to our authentication scheme in order to allow for a higher compression ratio when needed. Since we always keep in mind that the image visual content must not be changed by the storage operation, our goal is to allow compression up to a quality factor of 85. According to [88], in quantization based watermarking techniques, the robustness of an embedded mark can be improved by either enlarging the quantization step (Δ), or reducing the amount of modification caused by image processing. The amount of distortion produced by compression cannot be changed as the operations involved are standardized. On the other hand, augmenting the quantization interval involves increasing the distortion introduced, thus, violating the watermarking invisibility requirement. Moreover, we have found the increase of the quantization step not to be very effective as far as enhancing the overall robustness to JPEG (or JPEG-2000) compression. To abridge, this means that additional techniques have to be developed if one wants to consider medium quality JPEG compression as an acceptable alteration of the work which our authentication scheme should also protect. In order to do so, we have experimented with different procedures that we describe in the following subsections.

5.5.1 Predistortion in the Spatial and Wavelet Domains

In [67], Smith and Comiskey introduce a spread spectrum information hiding technique. Their system embeds information in digital images by making small modifications to a large number of pixels. They use a pseudo-random carrier to code and distribute the information over the entire image. In order to increase the resistance of the mark to JPEG compression, they propose a *predistortion* technique. The original direct sequence carrier is compressed and uncompressed prior to the modulation and demodulation operations in order to compensate for the distortion from JPEG. Their idea is to use the compression routine to filter out, in advance, all the energy that would otherwise be lost later by the compression of the marked image. In light of this, the authors argue that:

> Tricks analogous to this are probably possible whenever the information hider has a model of the type of distortion that will be applied.

From this, it is reasonable to assume that predistortion techniques can be used to augment the robustness of our authentication scheme to JPEG compression. Consequently, we have experimented with different predistortion related approaches that we explain below. All are derived from experimentations and observations rather than fundamental compression theories.

The first and most simple tactic we consider is to correct the distortion introduced by compression using post-watermarking operations in the spatial domain. By inspection, we found that the difference between a watermarked

image and its JPEG compressed version (QF = 85) is, perceptually speaking, the same as the difference between the original (*unmarked*) image and its compressed version. Therefore, this should mean that the compression affects the watermarked image in the same way as the unmarked one. However, it also indicates that the watermark does not alter the image significantly enough to be kept by medium quality JPEG compression (QF smaller than 90). From these observations, we have tried several approaches. First, we have simply tried to compress the image prior to the watermark embedding. Since medium quality JPEG removes low energy components in images, pre-compression assures that the marks must be embedded in higher energy WP coefficients. It does not, however, guarantee that the mark values have enough energy to sustain post-embedding compression. For this reason, the first approach is not successful. Therefore, we have tried a second pre-distortion method in the spatial domain. Our idea is to magnify the spatial differences (between the original and *marked* images) introduced by watermarking to *over-compensate* for the compression to be performed. We wanted to make the perturbation introduced significant enough to be kept by the JPEG quantization procedure, and consequently, kept in the compressed version of the watermarked image. This tactic did not, however, produce a significant improvement of the robustness of our system without introducing visual distortion to test images. From this, we conclude that we need to experiment with other pre-distortion techniques to see if any can serve the present purpose.

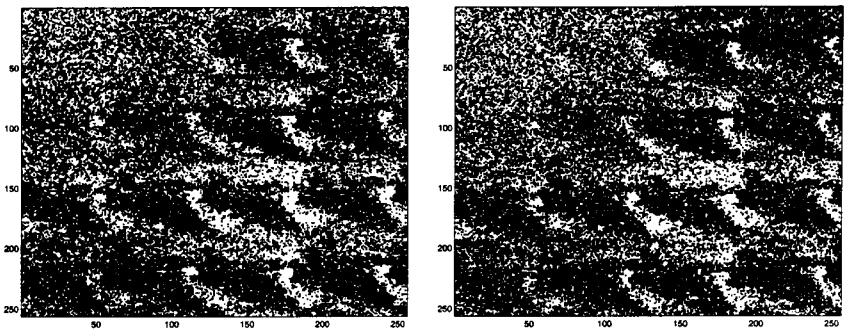


Figure 5.23: WP Regions of 2 level Lena Image Decomposition with Coiflets 24 and Daubechies 16 that are Unaltered by JPEG Compression (QF=85)

The second approach considered uses the wavelet packets' domain. First, we found that the quantization step Δ computed for each embedding band stays unchanged between a marked image saved in *bitmap* and the same one saved in JPEG (QF = 85). This confirms that the overall energy of each band is unchanged by compression, and explains the degree of robustness already achieved by our authentication technique. Then, we found that some coefficients of WP decomposition stay unaltered by compression (with a quality factor of 85). Furthermore, it seems that these regions are similar for the same level of different wavelet decomposition (see Figure 5.23[14]). Accordingly, it means that the watermark perturbation introduced by compression comes from the modification of separate individual coefficients and not from overall alteration[15]. We first tried to compensate for the distortion by diminishing the quantization alphabet size, thus enlarging the quantization steps. As stated earlier, this did not allow for the improvement of the resistance of our authen-

[14]**Note:** Here, the unaltered regions are represented in white.

[15]In fact, this was our main motivation in the use of mean of regions of WP coefficients for our embedding.

tication watermark to JPEG compression without affecting the visual quality of the work at hand. Another tactic is to make the embedding regions larger--that is, take the mean of more WPC at once. Despite the fact that it may augment the system's robustness to some extent, it would also decrease our ability to localize image tampering, a compromise which is unacceptable.

As a result, we considered the strategy previously evaluated in the spatial domain instead: we have to overcompensate the perturbation in the embedding step in order for a post-compressed image to contain the appropriate mark. Compressing the watermarked image, and then extracting the available key accomplished this. From comparison of the mark obtained with the embedded one, it is possible to see where the compression causes modification of the mean of WP coefficients regions. Then, for these regions only, we tried to correct the alterations prior to compression by introducing as much distortion in the other direction (see Figure 5.24) in order to counteract its effects. This did not, however, produce the intended results because the compression is highly dependant on the frequency content of the images, which is modified by the introduction of *overly quantized* coefficients, but led to another technique to increase the robustness of our scheme.

In the same line of thought, we implemented a sequential watermarking technique. Initially, the original image is marked using our previously presented scheme (Figure 4.7). After, the image is compressed and the watermark is extracted and compared with the original. Then, modifications by shifting in the WP domain are performed on the *unauthentic* marks. This is

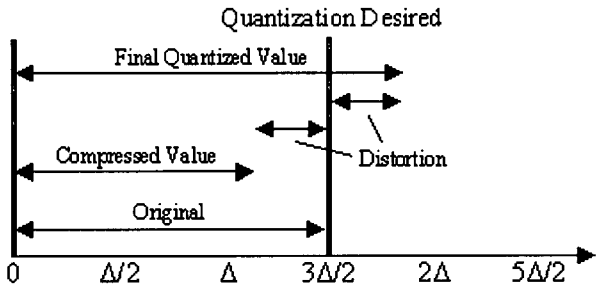


Figure 5.24: Overcompensation in the WP domain

done to obtain regions in each band for which the mean is not affected by compression. Subsequently, the same embedding system is applied recursively in order to insert the modified marks (Figure 5.25).

Theoretically, the *recursive embedding* system shows great promise. In practice, however, problems arise. First, the implementation of the recursive technique is computationally too costly. Even if each embedding is not lengthy (around 8 to 12 seconds), its repetition becomes intolerable. Secondly, it is extremely difficult to keep track of the modified *versus* authentic coefficients, as well as of the areas of embedding. Even though it is possible to do so in the embedding process, we have not been able to come up with an efficient manner to perform the recovery in the decoding process. Furthermore, by the shifting of WP regions, it is never certain that the image to protect is fully spatially covered. Finally, as the lossy part of the JPEG process is its quantization performed in the DCT domain, it is never guaranteed that unmodified regions are found, or that areas that are unmodified in one iteration stay untouched in order.

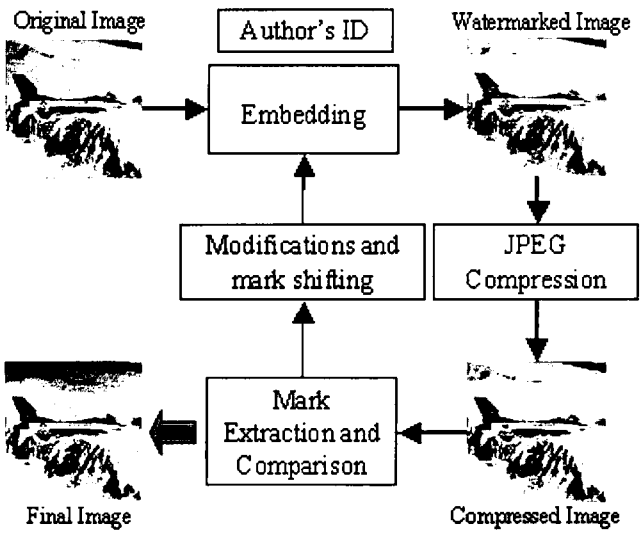


Figure 5.25: Recursive Embedding Scheme

All the techniques presented lead us to believe that predistortion techniques cannot be used in the proposed scheme. It seems that the tactic proven to improve the robustness of a spread-spectrum spatial domain watermarking method [67] cannot enhance the resistance of wavelet domain watermarking schemes. From this, we have to conclude that our system cannot be made more resistant to JPEG compression than intrinsically possible, unless a completely different authentication approach is developed.

5.6 Summary

In this chapter, we demonstrated the image authentication capabilities of our system. First, we showed that minimal visual perturbation was introduced by the insertion of a secret author identification mark in the WP domain.

Furthermore, our intra/interband verification technique was proven to allow good spatial and frequency localization of tampering without requiring any reference images. In the same way, we have found the use of the wavelet packets-based embedding domain to maximize the robustness of the marks, which allows our system to work in the presence of high quality JPEG compression. This was also shown to be a major advantage of our method as compared to *Eikonamark*, a commercially available watermarking tool, and so were its ability to localize frequency tampering and its robustness to collage attacks. Finally, multiple strategies intended to augment the robustness of our system and extend its use to medium quality JPEG compressed images were presented. However, none were found to improve the overall performance of our watermarking system.

To sum up, our WP-based digital watermarking system was shown to yield excellent results in terms of its authentication capabilities, the quality of watermarked images produced, its robustness to high quality JPEG compression and its resistance to collage attacks.

Chapter 6

Conclusions and Future Research

"The only true wisdom is in knowing you know nothing."

-Socrates, philosopher

6.1 Overview

The protection of visual content is becoming an important issue as the use of digital images increases. In this context, this thesis has studied the utilization of watermarking for content protection of digital images. In summary, the first part of the thesis gives an overview of wavelet transforms, while the most important part is dedicated to watermarking technologies. The bases of digital watermarking and content authentication methods are laid down. The emphasis is, however, mainly put on the implementation of a wavelet

packets-based digital watermarking system.

The main contribution of the thesis is the development of a novel, semi-fragile watermarking-technique for image authentication. We elaborate a secure watermarking technique for which the specific domain of embedding is known only by the author. For the embedding process, we formulate an optimal wavelet packet coefficients quantization protocol that takes maximum advantage of the host image's characteristics. We also create an authentication procedure to detect unauthorized tampering with images performed either in the frequency or spatial domains. This interband/intraband procedure verifies that certain regions in certain bands are not tampered with. In addition to our contribution to watermarking for image authentication, we introduce an easy description of wavelets and their implications in signal and image processing. We also, for the first time in the field, summarize and categorize the different watermarking techniques, and use this classification to summarize previously proposed techniques.

The main concepts we review and develop are summarized one last time in the following.

6.2 Digital Watermarking and Content Authentication

Digital watermarking allows for the imperceptible insertion of information into multimedia data. The supplementary information, called a *watermark*, is em-

bedded into the cover work through its slight modification. Watermarking has attracted a lot of attention in recent years, and several applications have been found for it (e.g. copyright protection, fingerprinting or copy protection), and reviewed in this thesis. From this, we divide watermarking techniques into two general approaches in terms of the capacity of the embedded mark to resist alteration of the host: *robust versus fragile* embedding. Although more attention is given to the first category in the literature, this thesis is particularly focused on the protection of information contained in digital media, and hence, in the development of a fragile watermarking scheme. To be precise, we proposed to protect a digital image's visual content by the embedding of a semi-fragile watermark in the wavelet packets' domain.

Different conceptual approaches are considered to protect the authenticity of digital media. This yields the introduction of several semi-fragile watermarking methods. However, each of the previous techniques has flaws that must be circumvented. In particular, spatial-based authentication techniques lack the ability to identify tampering performed on the spectral content of images. Furthermore, they are sensitive to search and collage attacks. On the other hand, techniques working in the wavelet domain are robust to attacks, and allow for the identification of frequency tampering as WT allows both spatial and frequency localization of alterations. However, standard wavelet-based techniques previously introduced ask for some user interaction in the decoding process to decide on the severity of tampering. This represents a security flaw that endangers their commercialization.

6.2.1 Our Wavelet Packets-Based Authentication Scheme

This thesis introduces a novel technique for digital image authentication based on semi-fragile watermarking. Using the wavelet packets' domain for the watermark embedding, our technique overcomes the above mentioned problems arising from the use of previously proposed methods. This novel approach embeds an author's unique (secret) binary identification key (of 64 bits) in an image to allow the image's authentication. First, the secret key is used as input parameter for the selection of the specific wavelet packet decomposition. This characteristic improves the overall security of our system as a randomly generated key, known only by the author, controls the information about the embedding domain. Then, our authentication scheme embeds the author's key in some regions in the WP domain. The coefficients in the regions are rounded so as not to affect the visual quality of the picture. To determine the optimal quantization achievable (i.e. find the quantization step to grant *detectability* to the mark while minimizing the mean square quantization error in the WP distribution), an adaptive quantization procedure that makes use of the knowledge of the (*Laplacian*) distribution of WP coefficients is introduced. Then, the 64-bit author's key is embedded by the rounding of selected wavelet packet coefficients to even/odd quantization levels. This novel procedure permits us to maximize the embedding weight, while minimizing the distortion introduced. Finally, the watermarked image is obtained by the computation of the inverse wavelet packet transform on the quantized coefficients.

The authentication of the image content is performed blindly using the author's secret key. First, the WP decomposition is computed on the image to be verified and the mark is extracted. Then, interband/intraband verification procedures are developed to complete the watermark decoding. Basically, the authenticity of the mark is verified across frequency and space to decide if any tampering has occurred in either frequency or space. These verification measures make obsolete the use of any post detection operations for judging the overall incidence of tampering. Furthermore, this allows for the differentiation of malicious tampering *versus* small image alterations introduced by JPEG compression on a genuine image. In that sense, our method includes robustness in the context of high quality image compression, while allowing image tampering detection and localization.

6.2.2 Review of Results

Experimental results demonstrate the capability of our WP-based watermarking approach to authenticate digital images. First, the visual quality of watermarked images produced shows the ability of our system to embed a secret mark in an image, while keeping the level of distortion introduced to a minimum. We also confirm that the mark is detectable with the proposed extraction procedure and that genuine images can be authenticated easily. In addition, the false negative detection rate is kept at a minimum. As well, the watermarking technique is proven to allow for accurate detection and good localization of image tampering performed either in the spatial domain (e.g.

addition or removal of objects) or in the frequency domain (e.g. filtering), even in the presence of high quality JPEG or JPEG-2000 compressions. Furthermore, these results are shown to be independent of the wavelet decomposition chosen.

Afterwards, the WP-based approach is compared with *Eïkonamark*, a commercially available spatial-based watermarking tool, in terms of watermarked image quality, tampering detection ability and resistance to collage attacks. Our system is shown to yield better watermarked images' quality than *Eïkonamark*, while spatial tampering detection is found to be comparable for both systems. *Eïkonamark*, however, did not allow the localization of frequency tampering, nor did it tolerate high quality JPEG compression. Furthermore, our system clearly outperforms Eïkonamark in terms of its resistance to collage attacks. These are obviously satisfactory results as they show that the proposed method is more robust and more secure than this commercially available software.

To complete our investigation, we experimented with many different *predistortion* strategies in order to extend the use of our authentication system to medium quality JPEG compressed images. Unfortunately, neither the techniques working in the spatial domain, nor those performing predistortion of the watermark in the WP domain can assure the resistance of the mark to medium quality JPEG compression without introducing unacceptable visual distortion. This leads us to conclude that other strategies would have to be considered in order to allow image authentication schemes to be robust enough

for medium quality JPEG compression.

In summary, our WP-based digital watermarking system gives terrific results. First, the interband/intraband decoding technique we developed yields outstanding authentication and tampering detection and localization capabilities. Second, the optimal quantization approach favored produces watermarked images of excellent visual quality. Third, the control of the embedding domain by an author's secret (and unique) key enhances the global security of our scheme. Last, the overall system has been shown to perform better than a commercial software in its robustness to JPEG compression and resistance to collage attacks.

6.3 Future Research

This thesis clearly improves existing image authentication techniques and introduces new concepts for the use of watermarking. Nonetheless, there are still many aspects that can be further investigated, and regarding which the overall image authentication technology can be improved.

First, the optimization of the embedding process for particular types of images may augment the embedding capacity-the amount of information carried by a host. As certain kinds of images possess particular traits, our system might be able to take advantage of those special characteristics. In the same line of thought, the overall performance of our system may be further enhanced by the use of color images, since their capacity to accept an invisible mark is greater that one of gray-scale images, due to the presence

of chrominance information, in addition to the luminance. Coding techniques can also be used to increase the capacity of embedding. The addition of an error control coding module [51] to augment the reliability of the information carried, may achieve this goal.

As we have seen, wavelet packet decomposition is not limited to two-dimensional signals. Therefore, the concepts developed in the context of image authentication may be adapted for the use of our WP-based authentication scheme on other media, such as text, audio or video.

Finally, future work may also include the augmentation of the mark's resistance to JPEG and/or JPEG-2000 compressions by the extraction of compression-invariant images' characteristics in the wavelets' domain and their use in the embedding process. In [55], the resistance of the watermark to JPEG compression is granted by assuring that the distortion caused on an image's DCT coefficient by compression is equal to the corresponding DCT coefficient of the spatially embedded watermark signal. As JPEG is based on the discrete cosine transform while our system works in the wavelet packets' domain, the augmentation of our scheme's resistance to JPEG compression using a similar technique definitely requires deep modification of the proposed scheme. On the other hand, this tactic can probably be used much more simply in the wavelet domain in order to grant our system more resistance to JPEG-2000, which is wavelet-based.

6.4 Closing Remarks

Semi-fragile watermarking techniques provide an effective means of protecting the content of digital media. Furthermore, their use on digital images allows for the detection and localization of unauthorized tampering, while permitting the efficient storage of visual information. These combined characteristics are of primary importance in applications, such as courtroom evidence or medical imaging for which the information contained in images is of utmost importance, while the number of images available requires proficient storage techniques. In this context, the use of semi-fragile watermarking techniques clearly augments the value of digital images. For all these reasons, the development of certification systems for digital data will become an increasingly important issue in the future. Thus, the introduction of our wavelet packets-based digital watermarking technique for image authentication is a small step in the advancement of overall digital security.

Bibliography

[1] M.D. Adams. Reversible wavelet transforms and their application to embedded image compression. Master's thesis, University of Victoria, Victoria, B.C., 1999.

[2] W.C. Adams and C.E. Giesler. Quantizing characteristics for signals having laplacian amplitude probability density function. *IEEE Transactions on Communications*, 26(8):1295–1297, 1978.

[3] Ashok Ambardar. *Analog and Digital Signal Processing.* PWS-Kent Publishing Company, Division of Wadsworth, Inc., 20 Park Plaza, Boston, MA 02116, USA, 1995.

[4] M. Barni, C.I. Podilchuk, F. Bartolini, and E.J. Delp. Watermark embedding: Hiding a signal within a cover image. *Special Issue of IEEE Communication Magazine on Digital Watermarking for Copyright Protection: A Communication Perspective*, 39(8):102–108, August 2001.

[5] F. Bartollini, A. Tefas, M. Barni, and I. Pitas. Image authentication techniques for surveillance applications. *Proceedings of the IEEE*, 89(10):1403–1418, October 2001.

[6] S. Bhattacharjee and M. Kutter. Compression tolerant image authentication. In *IEEE International Conference on Image Processing (ICIP'1998)*, volume I, pages 435–439, October 1998.

[7] F.M. Boland, J.J.K. O Ruanaidh, and C. Dautzenberg. Watermarking digital images for copyright protection. In *IEE Conference on Image Processing and its Applications*, volume 410, pages 326–330, jul 1995.

[8] J. Brassil, S. Low, N. Maxemchuk, and L. O'Gorman. Electronic marking and identification techniques to discourage document copying. In *Proceedings of Infocom '94*, pages 1278–1287, June 1994.

[9] J. Brassil, S. Low, N. F. Maxemchuk, and L. O'Gorman. Hiding information in document images. CISS '95, 1995.

[10] C. Busch, W. Funk, and S. Wolthusen. Digital watermarking: From concepts to real-time video applications. *Computer Graphics and Applications*, 16:25–35, January 1999.

[11] Q. Cheng and J. Sorensen. Spread spectrum signaling for speech watermarking. In *IEEE International Conference on Acoustics Speech and Signal Processing*, volume 3, pages 1337–1340, May 2001.

[12] Ronald R. Coifman, Yves Meyer, Stephen R. Quake, and Mladen Victor Wickerhauser. Signal processing and compression with wavelet packets. Preprint, Yale University, New Haven, April 1990.

[13] Ronald R. Coifman, Yves Meyer, and Mladen Victor Wickerhauser. Wavelet analysis and signal processing. In Mary Beth Ruskai:1992:WTA, Gregory Beylkin, Ronald Coifman, Ingrid Daubechies, Stéphane Mallat, Yves Meyer, and Louise Raphael, editors, *Wavelets and Their Applications*, pages 153–178. Jones and Bartlett, Boston, 1992.

[14] R.R. Coifman and Y. Meyer. Orthonormal wave packet bases. Technical report, Dept. of Mathematics, Yale University, 1990. Preprint.

[15] I.J. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. Technical Report 95-10, NEC Research Institute, 1995.

[16] I.J. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for images, audio, and video. In *IEEE International Conference on Image Processing (ICIP'96)*, volume III, pages 243–246, 1996.

[17] I.J. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for images, audio, and video. *Transactions on Image Processing*, 6(12):1673–1687, December 1997.

[18] I.J. Cox and M.L Miller. Electronic watermarking: the first 50 years. In *IEEE Fourth Workshop on Multimedia Signal Processing*, pages 225–230, 2001.

[19] I.J. Cox, M.L. Miller, and J.A. Bloom. Watermarking applications and their properties. In *IEEE International Conference on Information Technology*, pages 6–10, 2000.

[20] I.J Cox, M.L. Miller, and J.A. Bloom. *Digital Watermarking.* Morgan Kaufmann Publishers, 2929 Campus Drive, Suite 260, San Mateo, CA 94403, USA, 2002.

[21] I.J. Cox, M.L. Miller, and A.L McKellips. Watermarking as communications with side information. *Proceedings of the IEEE*, 87(7):1127–1141, July 1999.

[22] A. Croisier, D. Esteban, and C. Galand. Perfect channel splitting by use of interpolation/decimation/tree decomposition techniques. In *International Conference Information Science and Systems*, pages 443–446, 1976.

[23] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure Applied Mathematics*, XLI(41):909–996, November 1988.

[24] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA, 1992. Notes from the 1990 CBMS-NSF Conference on Wavelets and Applications at Lowell, MA.

[25] S. Decker. Engineering considerations in commercial watermarking. *Special Issue of IEEE Communication Magazine on Digital Watermarking for Copyright Protection: A Communication Perspective*, 39(8):128–133, August 2001.

[26] O. D'Souza. Digital and analogue watermarking of video recordings. Technical Report 71-76, Chubb Vision Australia, 1997.

[27] M. Ejima and A. Miyazaki. A wavelet-based watermarking for digital images and video. In *IEEE International Conference on Image Processing (ICIP'2000)*, volume III, pages 678–681, Vancouver, BC, Canada, September 2000.

[28] C. Fei, D. Kundur, and R. Kwong. The choice of watermark domain in the presence of compression. In *IEEE International Conference on Information Technology: Coding and Computing*, pages 79–84, 2001.

[29] David J. Fleet and David J. Heeger. Embedding invisible information in color images. In *IEEE Signal Processing Society 1997 International Conference on Image Processing (ICIP'97)*, volume III, pages 532–535, Santa Barbara, California, October 1997.

[30] A. Grossman and J. Morlet. Decomposition of functions into wavelets of constant shape. *SIAM Journal on Mathematic Analysis*, 15(4):723–736, 1984.

[31] Alfred Haar. Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910. In German.

[32] F. Hartung and M. Kutter. Multimedia watermarking techniques. *Proceedings of the IEEE*, 87(7):1179–1107, July 1999.

[33] Frank Hartung and Bernd Girod. Fast public-key watermarking of compressed video. In *IEEE Signal Processing Society 1997 International Conference on Image Processing (ICIP'97)*, volume III, pages 528–531, Santa Barbara, California, October 1997.

[34] E. F. Hembrooke. Identification of sound and like signals, 1961. United States Patent number 3,004,104.

[35] C.T. Hsu and J.L. Wu. Dct-based watermarking for video. *Transactions on Consumer Electronics*, 4(1):206–216, February 1998.

[36] D. Hunter. *Handmade Paper and its Watermarks: A Bibliography.* B. Franklin, New York, 1962.

[37] N. Jayant, J. Johnston, and R. Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10):1385–1422, October 1993.

[38] N. Komatsu and H. Tominaga. Authentication system using concealed images in telematics, 1988.

[39] D. Kundur and D. Hatzinakos. Digital watermarking for telltale tamper proofing and authentication. *Proceedings of the IEEE*, 87(7):1167–1180, July 1999.

[40] G.C. Langelaar and R.L. Lagendijk. Optimal differential energy watermarking of dct encoded images and video. *IEEE Transactions on Image Processing*, 10(1):148–158, January 2001.

[41] C. Lin and S. Chang. Semi-fragile watermarking for authenticating jpeg visual content. In *SPIE Security and Watermarking of Multimedia Content II*, pages 140–151, San Jose, CA, January 2000.

[42] C.-Y. Lin, M. Wu, J.A. Bloom, I.J. Cox, M.L Miller, and Y.M. Lui. Rotation, scale, and translation resilient watermarking for images. *IEEE Transactions on Image Processing*, 10(5):767–782, October 2001.

[43] J. Liu, K. Najarian, and E. El-Kwae. Comparative evaluation of wavelet-based digital image watermarking. In *International Conference on Acoustics Speech and Signal Processing (ICASSP'2001)*, volume III, pages 211–214, 2001.

[44] S. H. Low, N. F. Maxemchuk, J. T. Brassil, and L. O'Gorman. Document marking and identification using both line and word shifting. In *Infocom '95*, Boston, MA, April 1995.

[45] C.-S. Lu, S.-K. Huang, C.-J. Sze, and H.-Y Mark Liao. Cocktail watermarking for digital image protection. *IEEE Transactions on Multimedia*, 2(4):209–224, December 2000.

[46] C.-S. Lu and H.-Y.M. Liao. Multipurpose watermarking for image authentication and protection. *IEEE Transactions on Image Processing*, 10(10):1579–1592, October 2001.

[47] A. Lumini and D. Maio. Blind watermarking system for digital images in the wavelet domain. In *SPIE International Symposiom Electronic Imaging Security and Watermarking of Multimedia Contents II*, pages 524–535, January 2000.

[48] S. Mallat. Multiresolution approximation and wavelet orthonormal bases of L^2. *Trans. Amer. Math. Soc.*, 315:69–87, September 1989.

[49] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[50] J.R. Hernandez Martin and M. Kutter. Information retrieval in digital watermarking. *Special Issue of IEEE Communication Magazine on Digital Watermarking for Copyright Protection: A Communication Perspective*, 39(8):110–116, August 2001.

[51] L. M. Marvel and C. T. Retter. The use of side information in image steganography. In *IEEE International Symposium on Information Theory and Its Applications (ISITA'2000)*, November 2000.

[52] P. Meenvald and A. Uhl. Watermark security via wavelet filter parametrization. In *IEEE International Conference on Image Processing (ICIP'2001)*, volume II, pages 1027–1030, October 2001.

[53] Yves Meyer. *Ondelettes et Opérateurs.* Hermann, Paris, France, 1990. In two French volumes.

[54] F. Mintzer, G.W. Braudaway, and A.E. Bell. Opportunities for watermarking standards. *Communications of the ACM*, 41(7):57–64, July 1998.

[55] N. Nikolaidis and I. Pitas. Robust image watermarking in the spatial domain. *Signal Processing*, 766(3):385–403, 1998.

[56] A.H. Paquet and R.K. Ward. Wavelet-based digital watermarking for image authentication. In *IEEE Canadian Conference on Electrical and Computer Engineering*, volume I, pages 879–884, Winnipeg, Manitoba, May 2002.

[57] A.H. Paquet, S. Zahir, and R.K. Ward. Wavelet packets-based image retrieval. In *IEEE International Conference on Acoustics Speech and Signal Processing*, volume IV, pages 3640–3643, May 2002.

[58] F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn. Information hiding - a survey. *Proceedings of the IEEE*, 87(7):1062–1078, July 1999.

[59] C.I. Podilchuk and E.J. Delp. Digital watermarking: Algorithms and applications. *IEEE Signal Processing Magazine*, 18(4):33–46, July 2001.

[60] M.P. Queluz and P. Lamy. Spatial watermark for image verification. In *SPIE Conference on Security and Watermarking of Multimedia Contents II*, volume 3971, pages 120–130, January 2000.

[61] E.A. Robinson. *Random Wavelets and Cybernetic Systems.* Griffin and Co., London, 1962.

[62] J. Romberg, H. Choi, R. Baraniuk, and N. Kingsbury. Multiscale classification using complex wavelets and hidden markov tree models. In *IEEE International Conference on Image Processing (ICIP'2000)*, volume II, pages 371–374, Vancouver, BC, Canada, September 2000.

[63] J.J.K. O Ruanaidh, W.J Dowling, and F.M. Boland. Watermarking digital images for copyright protection. *IEE Proceedings on Visual Image Signal Processing -Special Section*, 143(4):250–256, August 1996.

[64] Joseph J. K. Ó Ruanaidh and Thierry Pun. Rotation, translation and scale invariant digital image watermarking. In *IEEE Signal Processing Society 1997 International Conference on Image Processing (ICIP'97)*, volume III, pages 536–539, Santa Barbara, California, October 1997.

[65] Khalid Sayood. *Introduction to Data Compression.* Morgan Kaufmann Publishers, 2929 Campus Drive, Suite 260, San Mateo, CA 94403, USA, 1996.

[66] Eero P. Simoncelli and Edward H. Adelson. Non-separable extensions of quadrature mirror filters to multiple dimensions. *Proceedings of the IEEE: Special Issue on Multi-dimensional Signal Processing*, 78(4):652–664, April 1990.

[67] J.R. Smith and B.O. Comiskey. Modulation and information hiding in images. In *Presented at the Workshop on Information Hiding - Springer-Verlag Lecture Notes in Computer Science*, volume 1174, May 1996.

[68] Eric J. Stollnitz, Tony D. DeRose, and David H. Salesin. *Wavelets for Computer Graphics: Theory and Applications.* Morgann Kaufmann, San Francisco, CA, 1996.

[69] Gilbert Strang and Truong Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, MA, 1996.

[70] M.D. Swanson, B. Chau B.Z Zhu, and A.H. Tewfik. Multiresolution video watermark using perceptual models and scene segmentation. In *IEEE International Conference on Image Processing (ICIP'97)*, volume III, pages 558–561, 1997.

[71] Mitchell D. Swanson, Bin Zhu, and Ahmed H. Tewfik. Transparent robust image watermarking. In *1996 SPIE Conf. on Visual Communications and Image*, volume III, pages 211–214, 1996.

[72] M.A. Tefas and I. Pitas. Image authentication based on chaotic mixing. In *IEEE International Symposium on Circuits and Systems (ISCAS'2000)*, volume I, pages 216–219, May 2000.

[73] Ahmed H. Tewfik. Digital watermarking; special issue on. *Signal Processing Magazine*, 17(9), September 2000.

[74] A.Z. Tirkel, C.F. Osborne, and R.G. van Schyndel. Image watermarking - a spread spectrum technique. In *IEEE 4th International Symposium on Spread Spectrum Techniques and Applications*, volume II, pages 785–789, 1996.

[75] M.-J. Tsai, K.-Y. Yu, and Y.-Z. Chen. Joint wavelet and spatial transformation for digital watermarking. *IEEE Transactions on Consumer Electronics*, 46(1):241–245, February 2000.

[76] M.-J. Tsai, K.-Y. Yu, and Y.-Z. Chen. Wavelet packet and adaptive spatial transformation of watermark for digital image authentication. In *IEEE International Conference on Image Processing (ICIP'2000)*, volume I, pages 450–453, Vancouver, BC, Canada, 2000.

[77] J. Véhel and A. Manoury. Wavelet packet based digital watermarking. In *IEEE International Conference on Image Processing (ICIP'2000)*, volume III, pages 417–420, Vancouver, BC, Canada, 2000.

[78] R. Venkatesan and M.H. Jakubowski. Image watermarking with better resilience. In *IEEE International Conference on Image Processing (ICIP'2000)*, volume I, pages 403–406, Vancouver, BC, Canada, September 2000.

[79] M. Vetterli. Multidimensional subband coding: Some theory and algorithms. *Signal Processing*, 6(2):97–112, April 1984.

[80] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding.* Prentice Hall, 1995.

[81] G. Voyatzis and I. Pitas. The use of watermarks in the protection of digital multimedia products. *Proceedings of the IEEE*, 87(7):1197–1207, July 1999.

[82] AlphaTec Watermarking. http://www.alphatecltd.com/watermarking/.

[83] K.K. Wong, C.H. Tse, K.S. Ng, T.H. Lee, and L.M. Cheng. Adaptive watermaking. *Transactions on Consumer Electronics*, 43(4):1003–1009, November 1997.

[84] Ping Wah Wong. A public key watermark for image verification and authentication. In *IEEE International Conference on Image Processing (ICIP'1998)*, volume I, pages 455–459, October 1998.

[85] M. Wu and B. Liu. Watermarking for image authentication. In *IEEE International Conference on Image Processing (ICIP'1998)*, volume II, pages 437–441, October 1998.

[86] X.-G. Xia, C. Boncelet, and G. Arce. A multiresolution watermark for digital images. In *IEEE Signal Processing Society 1997 International Conference on Image Processing (ICIP'97)*, volume III, pages 548–551, Santa Barbara, California, October 1997.

[87] M.M. Yeung and F. Mintzer. An invisible watermarking techniques for image verification. In *IEEE International Conference on Image Processing (ICIP'1997)*, volume II, pages 680–683, 1997.

[88] G.J. Yu, C.-S. Lu, H.-Y. M. Liao, and J.-P. Sheu. Mean quantization blind watermarking for image authentication. In *IEEE International Conference on Image Processing (ICIP'2000)*, volume III, pages 706–709, Vancouver, BC, Canada, 2000.

Appendix A

In this brief appendix, we present some mathematical equations, proofs and explanations to support statements and theories developed throughout the thesis.

A.1 Fourier Analysis

The Fourier series of a continuous signal of period T is given by the following:

$$x_p(t) = \sum_{k=1}^{\infty} a_k \cos(k\omega_0 t) + b_k \sin(k\omega_0 t) \tag{A.1}$$

where the expansion coefficients are given as follows:

$$a_k = \frac{2}{T}\int_T x(t)cos(k\omega_0 t)dt \tag{A.2}$$

$$b_k = \frac{2}{T}\int_T x(t)sin(k\omega_0 t)dt \tag{A.3}$$

The Fourier transform of a discrete signal $x[n]$ is computed with the following:

$$\mathcal{F}\{x[n]\} = X(\omega) = \sum_{-\infty}^{\infty} x[n]\exp(-j\omega n) \tag{A.4}$$

and the original signal can be reconstructed using the following:

$$x[n] = \mathcal{F}^{-1}\{X(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\exp(j\omega n)d\omega \tag{A.5}$$

A.2 Orthonormality of Haar Basis

To make sure that Haar basis functions given in Equation 2.4 form an orthonormal basis for signals from the space $l_2(\mathcal{Z})$, we need to prove the following:

1. $\{\varphi_k\}_{k\epsilon\mathcal{Z}}$ is an orthonormal family.

2. $\{\varphi_k\}_{k\epsilon\mathcal{Z}}$ is complete.

First, if we consider 1, we need to show that $\langle\varphi_k, \varphi_l\rangle = \delta[k-l]$. If we take k as even, that is, $k = 2i$, the basis functions overlap only if $l \geq 2i$ or if $l \leq 2i+1$. For those two cases, we have the following:

$$\langle\varphi_{2i}, \varphi_{2i}\rangle = \varphi_{2i}^2[2i] + \varphi_{2i}^2[2i+1] = \frac{1}{2} + \frac{1}{2} = 1$$

$$\langle\varphi_{2i}, \varphi_{2i+1}\rangle = \varphi_{2i}[2i]\cdot\varphi_{2i+1}[2i] + \varphi_{2i}[2i+1]\cdot\varphi_{2i+1}[2i+1] = 0$$

The same argument can be made for odd l's, and thus orthogonality of the basis is proven. Now, to prove 2, we need to show that any signal belonging to $l_2\mathcal{Z}$ can be expanded using Haar basis. This is equivalent to demonstrating that there exist no signal $x[n]$ with $\|x\| > 0$ that has a zero expansion or $\|\langle\varphi_k, x\rangle\| = 0$ for all k's. To prove 2, we demonstrate that the opposite is impossible, as follows.

$$\|\langle\varphi_k, x\rangle\| = 0 \Leftrightarrow \|\langle\varphi_k, x\rangle\|^2 = 0 \Leftrightarrow \sum_{k\epsilon\mathcal{Z}} |\varphi_k[n], x[n]\rangle|^2 = 0 \tag{A.6}$$

Since $\varphi_k[n]$ and $x[n]$ are non negative terms, Equation A.6 holds only if:

$$X[k] = \langle \varphi_k[n], x[n] \rangle = 0 \quad \text{for all } k$$

First, if we take k as even and consider $X[2k] = 0$, the inverse transform yields $x[2k] = -x[2k+1]$ for all k. Then, the inverse transform of $X[2k+1] = 0$ for odd k's, gives $x[2k] = x[2k+1]$ for all k. Therefore, the only way the two conditions are satisfied is when $x[2k] = x[2k+1] = 0$, which is in contradiction with our first premise. Subsequently, this shows that there is no sequence $x[n]$, $\|x\| > 0$, such that $\|X\| = 0$, and therefore, proves completeness of the Haar basis $\langle \varphi_k \rangle$. In conclusion, Haar basis functions form an orthonormal basis as requested.

A.3 Conditions of Filters $H_i(z)$ and $F_i(z)$

There are problems linked with the appropriate choice of filters that need to be solved in order for the signal reconstructed from analysis→synthesis using filter banks to be an equivalent (in fact a delayed version) of the original signal. Since the analysis filters H_0 and H_1 are not ideal brick wall filters, their responses overlap, therefore creating aliasing. The synthesis filters must be designed in consideration of the problem. Thus, the conditions that the filters must satisfy can be split in two as follows:

1. Alias Cancellation is as follows:

$$F_0(z) = H_1(z) \qquad \text{and} \qquad F_1(z) = -H_0(-z)$$
$$F_0(z)H_0(-z) + F_1(z)H_1(-z) = 0 \tag{A.7}$$

2. Perfect Reconstruction is as follows:

$$F_0(z)H_0(z) - F_0(-z)H_0(-z) = 2z^{-l} \tag{A.8}$$

These can be summarized using the following matrix notation:

$$\begin{bmatrix} F_0(z) & F_1(z) \end{bmatrix} \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} = \begin{bmatrix} 2z^{-l} & 0 \end{bmatrix} \tag{A.9}$$

A.4 Definition of Multiresolution

A multiresolution analysis consists of the following sequence of embedded closed subspace:

$$\ldots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \ldots \tag{A.10}$$

such that

1. *Upward Completeness* is as follows:

$$\overline{\bigcup_{m \epsilon \mathcal{Z}}} V_m = L_2(\mathcal{R}). \tag{A.11}$$

2. *Downward Completeness* is as follows:

$$\bigcap_{m \epsilon \mathcal{Z}} V_m = \{0\}. \tag{A.12}$$

3. *Scale Invariance* is as follows:

$$f(t) \in V_m \Longleftrightarrow f(2^m t) \in V_0. \tag{A.13}$$

4. *Shift Invariance* is as follows:

$$f(t) \epsilon V_0 \Longrightarrow f(t-n) \in V_0 \text{ for all } n \in \mathcal{Z}. \tag{A.14}$$

5. *Existence of a Basis:* there exists $\varphi \epsilon V_0$, such that the following is an orthonormal basis for V_0:

$$\{\varphi(t-n)|\ n\ \epsilon\ \mathcal{Z}\} \tag{A.15}$$

A.5 Steps towards Multiresolution

Here are the main steps leading to (and necessary for) a multiresolution application for wavelet decomposition. From [69], we have the following:

1. An increasing sequence of scaling subspace V_j
2. Wavelet subspace W_j that gives $V_j + W_j = V_{j-1}$
3. The dilation requirement $f(t)$ in Vj to $f(2t)$ in V_{j-1}
4. The basis $\varphi(t-k)$ for V_0 and $\psi(t-k)$ for W_0
5. The basis $\varphi(2^j t-k)$ for V_j and $\psi(2^j t-k)$ for W_j
6. The basis of all wavelets $\psi(2^j t-k)$ form the whole space L^2

A.6 Erasable Watermarking

From [20], the main steps towards erasable watermarking are :

1. All the information in the work is used to compute a signature.
2. The signature is embedded in the mark in an erasable (*invertible*) manner.

3. The recipient of the work extracts and records the embedded signature.

4. The watermarked signature is erased from the cover work. As this point, the work should be identical to the original one.

5. In the same way as in 1, a signature is computed from the work and compared with the extracted signature from 3.

6. Authenticity decision is based on the similarities/differences between the extracted (3) and computed (4) signatures.

An erasable watermark can be removed from its associated cover work to obtain an exact copy of the original unwatermarked work. It is however impossible to make an erasable mark that can be embedded in 100 % of digital content. This is to due to the fact that digital works are represented with a finite number of bits. Therefore, there is a fixed (although very large) number of possible works. For example there are 2^{524288} possible 256x256 8-bits images. Erasability requires that the original work can be recovered from the watermarked work. This asks for the unicity of the mapping between the original and watermarked works. This means that for the 2^{524288} possible original images, there are exactly 2^{524288} corresponding unwatermarked images. If the original and watermarked works are represented with the same number of bits, it means that all the original works must be considered watermarked. Therefore, the only way to achieve 100 % effectiveness is to allow for 100 % false positive, which goes against an important requirement of watermarking schemes. For more detail on the subject, see Chapter 10 of [20].

Appendix B

This second appendix presents figures that we thought were interesting, as well as highly informative, for the readers, but not absolutely essential to the core of our work.

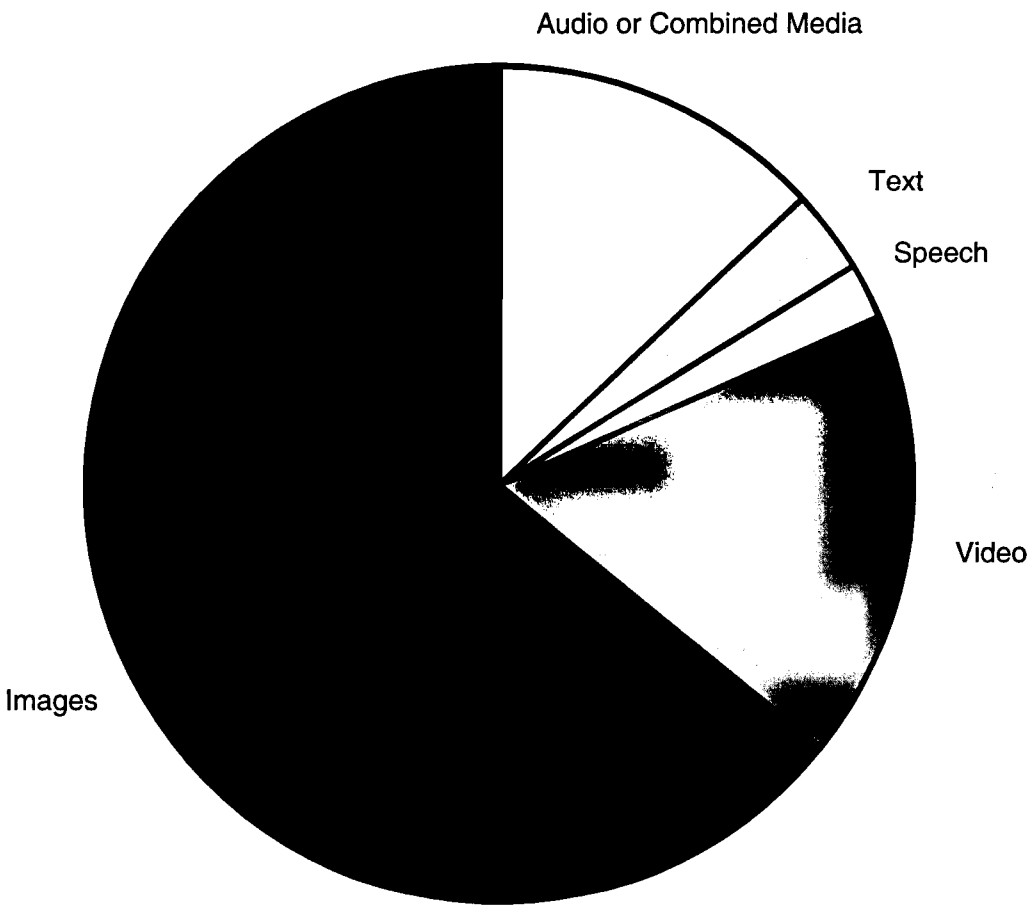


Figure B.1: Proportion of each Media used for Digital Watermarking

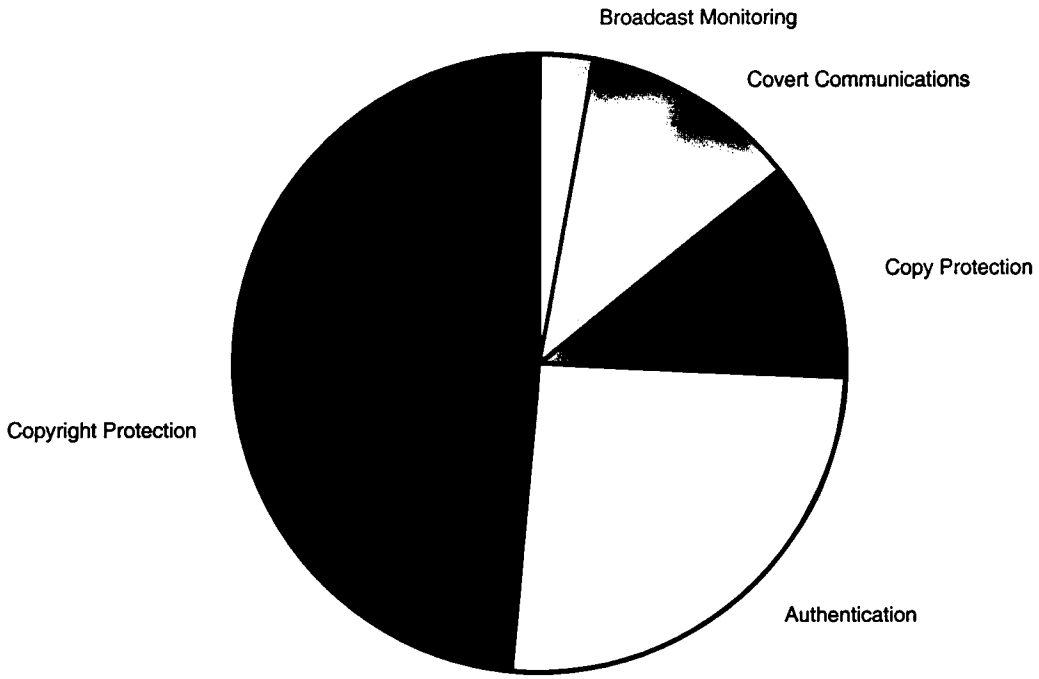


Figure B.2: Different Applications of Digital Watermarking

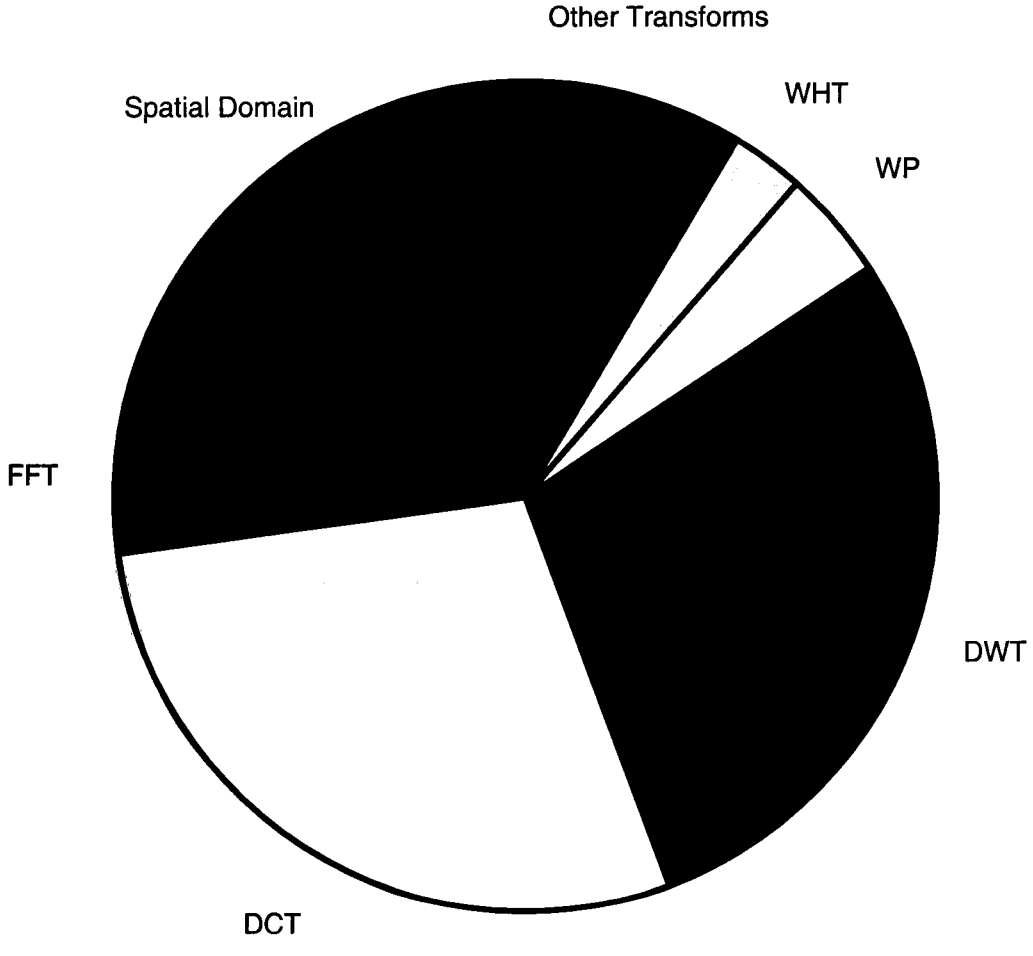


Figure B.3: Embedding Domain used for Digital Watermarking

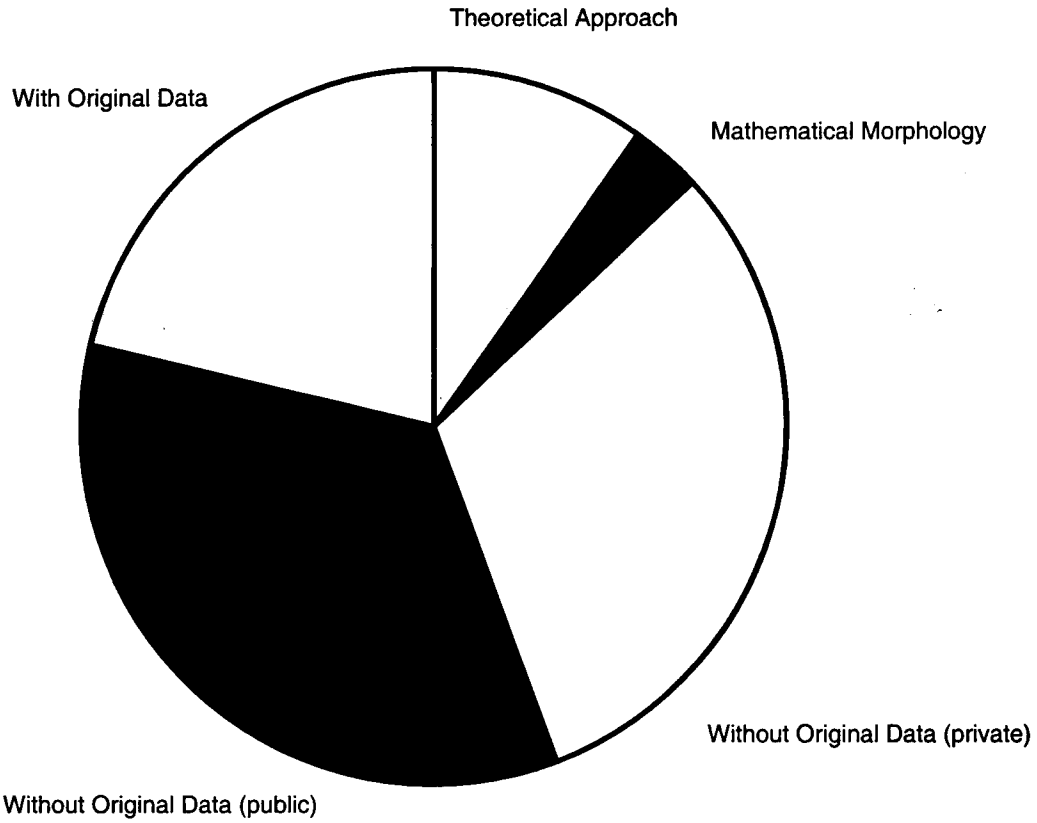


Figure B.4: Decoding/Detection Procedure used for Digital Watermarking

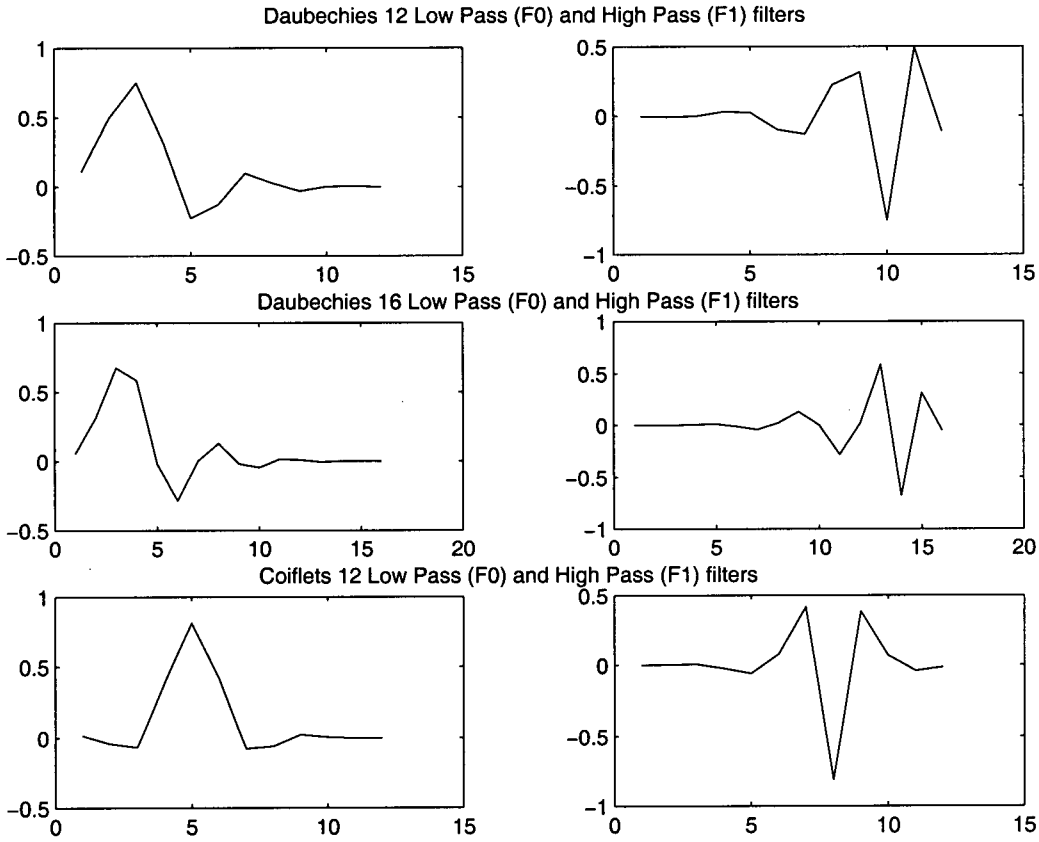


Figure B.5: Discrete Filters used in our Implementation

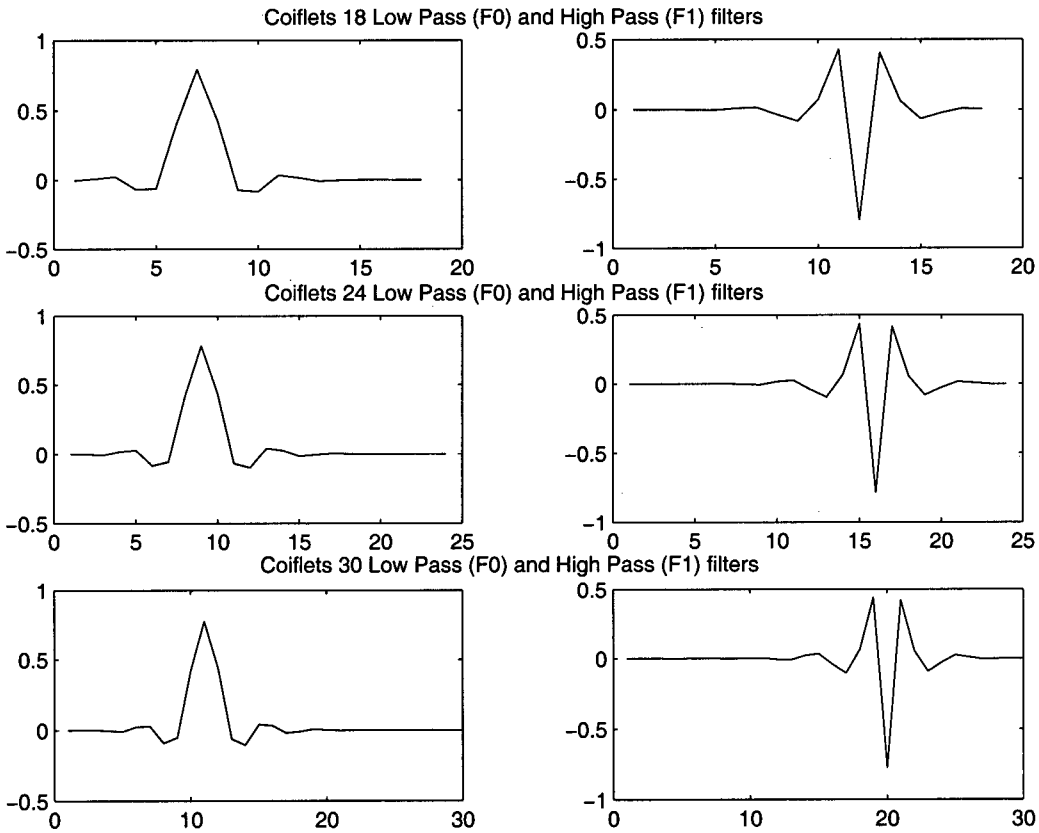


Figure B.6: Discrete Filters used in our Implementation

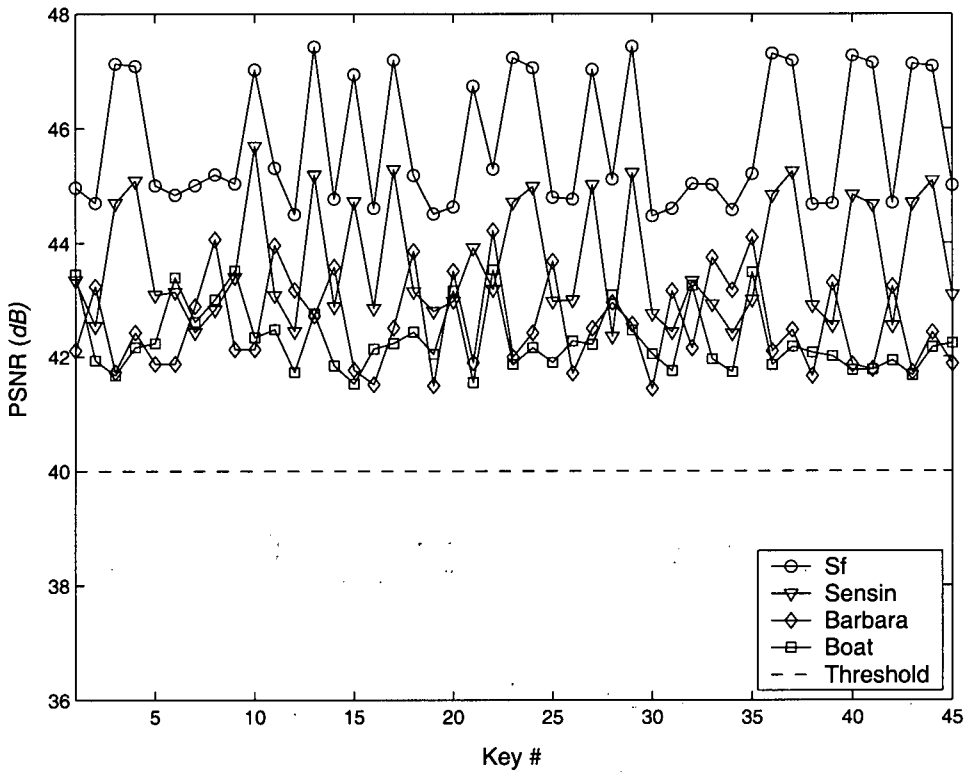


Figure B.7: PSNR Values for Different Embedding Keys

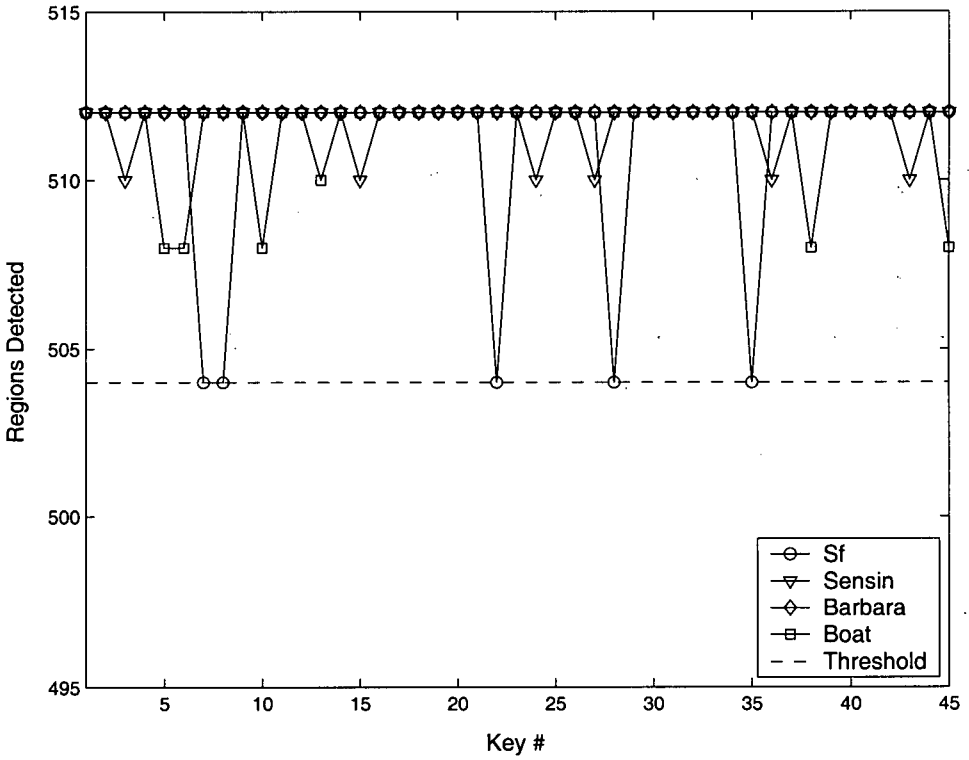


Figure B.8: Detection Rates achieved for Authentic Images