INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Université d'Ottawa · University of Ottawa

Spread Spectrum Watermarking for Images and Video

by

Mercy George, M. Tech

A thesis submitted to the
School of Graduate Studies and Research
In partial fulfillment of the requirements for the degree of

Master of Applied Science
In Electrical Engineering

Ottawa-Carleton Institute for Electrical Engineering
School of Information Technology and Engineering
Faculty of Engineering
University of Ottawa

September 1999

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. N. D. Georganas for his support and encouragement throughout the course and this work. I am extremely thankful to him for developing my interest in the area of 'Digital Watermarking'.

I would also like to thank my co-supervisor, Dr. J.-Y. Chouinard for his great availability and consistent advice whenever I approached him. His efforts in presenting papers at the Canadian Conference on Electrical and Computer Engineering (CCECE'99) and Canadian Workshop on Information Theory (CWIT'99) are invaluable. Above all, he conveyed my attraction for secure communication.

I am thankful to all MCR Lab (Multimedia Communications Research Laboratory, University of Ottawa) members for their encouragement and suggestions throughout this work.

Finally, I would like to thank George Varughese, my beloved husband who supported me continuously and patiently while I completed this work.

ABSTRACT

This thesis describes a digital watermarking method using the principle of spread spectrum communication. This method could be used for owner authentication or for fingerprinting. Though this work concentrates on image and video watermarking, the same method could be used for audio and text (represented as an image) as well.

The major challenge in watermarking is that the watermarks should withstand all types of intentional and unintentional attacks, at the same time being imperceptible. Previously proposed digital watermarking techniques were not very robust, and the watermark was easy to remove. Using Direct Sequence Spread Spectrum, it is possible to embed invisible watermarks, which are highly robust to attacks and unintentional transformations. The robustness of this scheme to different signal manipulations and transformations is studied.

This thesis also covers the different aspects of watermarking and includes an information theoretic analysis of the spread spectrum watermarking for fingerprinting. Even though spread spectrum signals are very resistant to non-linear amplitude distortion and additive noise, they are not resistant to timing errors. Using a sliding correlator for instance, it is possible to detect and compensate for the loss in synchronization. Our simulation results demonstrate the robustness of the spread spectrum watermark to lossy compression, filtering, cropping, rotation, scaling, printing and scanning, multiple watermarking attacks, collusion attacks, etc.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AWGN	Additive White Gaussian Noise
BFSK	Binary Frequency Shift Keying
BMP	Bit-map
BPSK	Binary Phase Shift Keying
CBC	Cipher Block Chaining
CFB	Cipher Feedback
CPTWG	Copy Protection Technical Working Group
CSS	Content Scrambling System
DCT	Discrete Cosine Transform
DES	Data Encryption Standard
3 DES	Triple DES
DFT	Discrete Fourier Transform
DHSG	Data Hiding Sub Group
DOS	Disk Operating System
DSSS	Direct Sequence Spread Spectrum
DVD	Digital Video (or Versatile) Disk
DWT	Discrete Wavelet Transform
ECB	Electronic Code Book
GIF	Graphics Interchange Format
GOP	Group of Pictures
GTC	Transform Coding Gain
HD	High Density
HDTV	High Definition Television
HPF	High Pass filter
HVS	Human Visual System
IA	Image Adaptive
ICPAC	Interim Copy Protection Advisory Council
IDEA	International Data Encryption Algorithm
IDCT	Inverse Discrete Cosine Transform
IWT	Inverse Wavelet Transform
JDK	Java Development Kit
JND	Just Noticeable Difference
JPEG	Joint Photographic Expert Group
JSTEG	JPEG Steganography
LPF	Low Pass Filter
LPI	Low Probability of Intercept
LSB	Least Significant Bit
MPEG	Motion Picture Experts Group
MP3	MPEG Layer 3
NSEA	Non patented Simple Encryption Algorithm
OFB	Output Feedback
PN	Pseudo-Noise
PSNR	Peak Signal to Noise Ratio
RLC	Run Length Coding
RST	Rotation Scale and Translation
SNR	Signal to Noise Ratio

SS	Spread Spectrum
SWICO	Single Watermarked Image Counterfeit Original
WaRP	Watermark Review Panel
WAV	Waveform
WT	Wavelet Transform

LIST OF SYMBOLS

v_i:	Image pixel values
w_i:	Spread watermark values
m_i:	Watermarked image pixel values
p_i:	Pseudo-noise sequence
V_i:	Image DCT coefficients
W_i:	Watermark DCT coefficients
M_i:	Watermarked image DCT coefficients
α:	Watermark scale factor
r_c:	Chip-rate
N_{info}:	Number of information bits
P_e:	Probability of decoding error
P_F:	Probability of false alarm
P_D:	Probability of detection
T_{JND}:	Just noticeable difference threshold
$\rho(w_e, w_i)$:	Normalized correlation coefficient
T_{det}:	Detection threshold

1. INTRODUCTION

1.1 Background

The enormous progress in digital technologies during the last decades has contributed to the popularity of the use of electronic media for transmission and storage of multimedia contents. Information stored in digital format can be copied without quality loss and distributed efficiently at fairly low costs. These developments have also increased the potential for interception, manipulation, misuse and unauthorized distribution of information. With the rapid growth of multimedia applications on the network, security and legal issues of copyright protection have become more important. Conventional cryptographic systems permit only valid key holders access to encrypted data, but once such data is decrypted there is no way to track its reproduction or retransmission. Therefore cryptography provides little protection against data piracy, in which a publisher is confronted with unauthorized reproduction of information.

A ***digital watermark*** is intended to complement cryptographic processes. It is a perceptible, or preferably imperceptible, identification code that is permanently embedded in the data and remains present within the data after any decryption process. It can be used to identify the source or owner of the data, to verify if the data is tampered and to trace illegal copying of the data. Digital watermarking is a new research area, combining aspects of digital signal processing, cryptography, statistical communication theory and human perception. The watermarked data can be audio

(speech and music), images (photographs and graphics), video, VRML models or text. Watermarking is possible because of the limitations of the human auditory and visual systems.

Watermarking differs from ***authentication*** or ***digital signature*** that proves to a receiver that the message could only have come from one particular transmitter. Mostly, authentication messages in the form of conventional hash functions can easily be deleted by a pirate who wishes to use copyrighted material for illegal purposes. The goal is to give the copyright owner of a digital document, the possibility to attest technically the origin of the image. Watermarking does not address authentication explicitly, although certain forms of watermarks can be used for authentication as well. ***Fingerprinting*** is the process of attaching the identity of the receiver to a signal in a way that is difficult to erase. This allows the copyright owner to trace pirates if illegal copies are made.

The requirements for authentication and fingerprinting are contradictory. On one hand the broadcaster or copyright holder may want to easily recognize the fingerprint, preferably visually. This allows easy tracing of a decoder that is used for illegal purposes. This approach is very similar to the commonly used watermarking by means of the logo of a TV station that is continuously visible in one of the corners of the screen. On the other hand, the fingerprint should be hidden, in order not to disturb paying viewers with program-unrelated messages on their screen, or to avoid that any pirate can detect and erase the fingerprint electronically. In the latter case, it may involve specific equipment to detect and decode a fingerprint.

1.2 History

The concept of digital watermarking is derived from *Steganography*, an ancient art of hiding information. Steganography, derived from Greek, literally means 'covered writing'. Throughout history, people have hidden information by a multitude of methods and variations [3, 26]. For example, ancient Greeks wrote text *on wax-covered tablets*. To pass a hidden message, the wax is scraped off a tablet, the message is written on the underlying wood and the tablet is again covered with wax to make it appear blank and unused. *Invisible inks* also offered a common form of invisible writing. With invisible ink, a seemingly innocent letter could contain a very different message written between the lines. *Open coded messages*, which are plain text passages, which "sound" innocent can hide information. For example, extracting the second letter in each word in an innocent passage might reveal a very important message. Documents can also be marked and identified by modulating the position of lines and words.

Paper watermarks are designs or patterns put into paper during its production, by making the layer of pulp thinner or thicker when it is still wet, and hence, the name watermark. Paper watermarks can be seen holding the paper against the light or, in some cases, over a black surface. Usually, they show the manufacturer's name, and geometric designs, or images of animals, etc. The object of watermarks in paper is, essentially, identifying the paper, as a signature of the manufacturer, or as a security measure to avoid forgery of important documents as bank notes, passports, etc. Today, good quality writing paper, as well as art drawing paper or paper for bibliophile publishing, usually carry an identifying watermark.

In the last century, logarithm tables were protected by introducing tiny errors in the insignificant digits of a few random values. A different set of values was chosen for each copy of the table. If an owner of a logarithm table ever sold illegal copies of it. the tiny errors in the table would enable the police to trace an illegal copy back to its owner.

Microdots, developed by Germans are photographs the size of a printed period having the clarity of standard-sized typewritten pages. They permit the transmission of large amounts of data, including drawings and photographs. Like their paper ancestors, digital watermarks are added to presentation media as a guarantee of authenticity, quality, ownership and source. A digital watermark is a digital signal or pattern inserted into a digital document.

Even though the idea of watermarking is very old, the application of watermarks for copyright protection of digital documents was revived in 1994 by Brassil *et al.* [6]. The first paper on watermarking presented in a major conference (ICIP '94) is by Van Schyndel [52]. In the following years, there has been a large number of publications resulting from work in this area by researchers all over the world. These publications deal with embedding watermarks into text [6, 31, 32], audio [4, 7], digital images [37, 38, 44, 47], digital video [22, 50] and 3D polygonal models [36, 56].

1.3 Contributions

This thesis describes a performance analysis of the spread spectrum watermarking method for images and video. The software for watermarking images and video has been implemented in Java.

The method suggested by Hartung *et al.* [22] has been used in our work. The basic method given in [22], however, is not very robust to synchronization attacks like rotation, scaling, cropping, etc. Some additional features can improve the robustness of this scheme to a large extent. The use of an 'interleaver' helps to spread the information bits spatially, which makes the method highly robust to cropping attacks. We have implemented interleaving in our work. Synchronization loss was detected using a '2 D sliding correlator' and the loss was compensated to make the watermarking scheme robust to geometric transformations.

We have used the same method for watermarking images and video, both in the spatial domain and the spectral domain. A major advantage of this method over the earlier watermarking methods is that the watermark detection could be performed without using the original unwatermarked document.

An information theoretic analysis of the spread spectrum watermarking method is carried out to derive the probability of bit error for watermark extraction without using the original, and verified with experimental results. The effect of the watermark signal energy or the watermark scale factor, α and the information spreading rate called the chip-rate, r_c, on the perceptibility and the robustness of the embedded watermark has also been studied.

Publications resulting from this research:

1. "Digital Watermarking of Images and Video using Direct Sequence Spread Spectrum Techniques" *in the Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (IEEE CCECE '99), May 9-12, 1999,* pp. 116-121.
2. "Spread Spectrum Spatial and Spectral Watermarking for Images and Video" *in the Proceedings of the Canadian Workshop on Information Theory (CWIT '99), June 15-18, 1999,* pp.119-122.

1.4 Thesis Organization

Chapter 2 of this thesis explains the concept of watermarking by giving a general framework for watermark insertion, extraction and detection. This chapter lists some of the watermarking software available on the Internet and the state-of-the-art. It also discusses the requirements on digital watermarks, the different types of digital watermarks, various applications and some earlier works in digital watermarking. *Chapter 3* presents the challenge in watermarking by discussing the various possible attacks on watermarks. *Chapter 4* introduces the principle of spread spectrum communication and its applicability to watermarking. Some methods for defending against the attacks given in *chapter 3*, are also covered here. An information theoretic model of watermarking is given in *chapter 5*, both for copyright protection and for fingerprinting applications. This chapter also covers an analysis of the watermark channel capacity for spatial and spectral domain watermarking.

Chapter 6 is devoted to the experimental results on image and video watermarking. It gives the optimum values for the watermarking parameters based on the perceptibility and robustness of the watermark and presents the basic set-up used for image and video watermarking. The properties of robustness and imperceptibility are verified for all possible attacks and manipulations introduced in chapter 3. *Chapter 7* is a brief conclusion to the former chapters and also provides some suggestions for future work.

2. DIGITAL WATERMARKING – AN OVERVIEW

2.1 Digital Watermarking Framework

The watermark can be inserted in the original signal without deteriorating it. The insertion method must be invertible so that the watermark can be extracted. Watermark detection is usually done using a correlation receiver. The watermark insertion, extraction and detection procedures are shown in Fig. 1. The watermark can be a function of a number of parameters such as a user identity, a user key, original signal to be watermarked and its parameters, etc. The watermark insertion techniques vary with different applications. The watermark could be extracted from the watermarked signal using the user key. This extracted watermark is compared with all possible watermarks to detect the true owner of the watermarked signal.

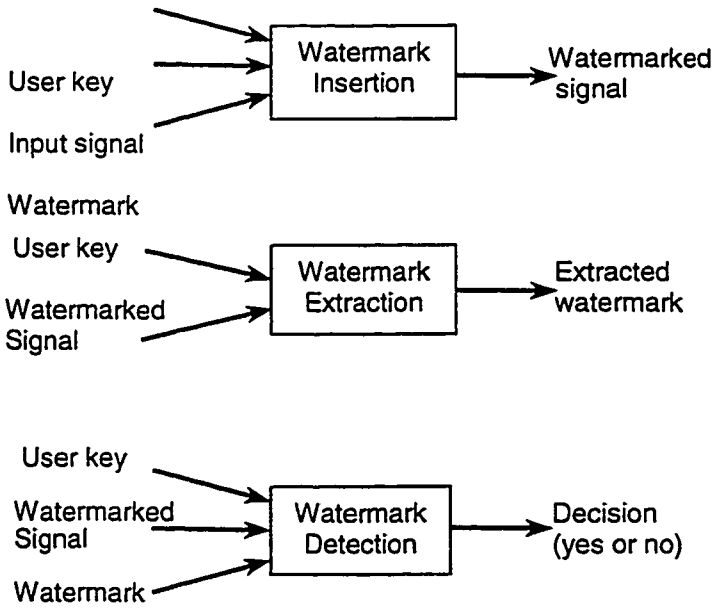


Figure 1: General framework for watermark insertion, extraction and detection.

2.2 State-Of-The-Art

Current read-only DVD (Digital Video Disk) systems prevent unauthorized copying from pre-recorded DVD disks by using licensed technologies, such as the digital "content-scrambling system" (CSS) to encrypt and play back movies. But with the advent of digital recording equipment such as recordable DVDs, digital tape recorders and personal computers with large storage capacity, additional copy-protection features are needed to prevent unauthorized copies of the copyrighted digital content.

The Data Hiding Sub Group (DHSG) of the Copy Protection Technical Working Group (CPTWG), is an ad hoc group of experts from computer, consumer-electronics and movie-studio industries that was formed to assess the technical merits of competing copy-protection proposals for movies, video and other content on digital media.

On February 18th, 1999, five leading Japanese and U.S computer and consumer electronics companies - Hitachi Ltd., IBM Corporation, NEC Corporation, Pioneer Electronic Corporation and Sony Corporation - announced that they have agreed to unify their approaches for creating invisible electronic watermarks for digital movies and video [59]. The five companies, which call themselves the Galaxy group, aim to provide such features via electronic watermarking technology that permits the copyright owner to specify whether content can be copied freely, can only be copied once or cannot be copied at all, thereby encouraging the owners to provide consumers with a variety of content.

Although invisible to the user, electronic watermarks contain information that can be recognized for example, by a detector chip in consumer digital recorders, or special detection software running on compliant PC systems as instructions for enabling or disabling its ability to make a copy. In addition to being invisible to users, electronic watermarks must survive through normal processes such as digital-to-analog conversion and repeated digital compression/decompression cycles while still remaining detectable by the digital recorder system.

By combining their expertise across both information technology and consumer electronic environments, the five companies now expect their watermarking method to provide superior performance for today's needs as well as for future high-definition television (HDTV) standards, digital-cinema distribution and other computer-based high-resolution digital-image applications. The five companies expect the first significant application of the new technology to be in future DVD (digital versatile disk) systems. Digital watermarks are also expected to be used in the copy protection of content distributed electronically via digital broadcasts and networks.

The Watermark Review Panel (WaRP) was formed at the December 1998 meeting of the CSS licensees to define and execute the final evaluations of the remaining proposals for video watermarking. Following the formation of the Galaxy group, only two proposals remain for evaluation by WaRP: the unified Galaxy proposal and one submitted by Philips, Macrovision and Digimarc. After completing its evaluations, WaRP will report its findings to the Interim Copy Protection Advisory Council (ICPAC) of the CSS licensing entity, which is expected to select by this

summer (1999), the watermark proposal that best meets the requirements of the CSS licensees.

2.3 Properties of Digital Watermarks

There is a number of desirable characteristics that a watermark should exhibit [10, 26], but different applications will have different requirements. Therefore there is no unique set of requirements that all watermarking techniques must satisfy. Some of the desirable properties are:

Unobtrusive The watermark should not be noticeable to the viewer nor should the watermark degrade the quality of the content. However, if a signal is truly imperceptible, then perceptually based lossy compression algorithms should, in principle, remove such a signal. Thus, to survive these compression algorithms, it will probably be necessary for a watermark to be noticeable to a trained observer.

Robust The watermark must be difficult (hopefully impossible) to remove. Attempts to remove or destroy a watermark by say, adding noise, should result in severe degradation in data fidelity before the watermark is lost. In particular, the watermark should be robust to:

- *Common Signal Processing*: These include, digital-to-analog and analog-to-digital conversion, resampling, requantization (including dithering and recompression), and common signal enhancements to image contrast and color, or audio bass and treble, etc.

- *Common Geometric Distortions on Image and Video Data*: Watermarks in image and video data should also be immune to geometric image operations such as rotation, translation, cropping and scaling.

- *Subterfuge Attacks - Collusion and Forgery*: The watermark should be robust to collusion by multiple individuals who each possess a watermarked copy of the data. That is, the watermark should be robust to combining copies of the same data set to destroy the watermarks. Further, if a digital watermark is to be used as evidence in a court of law, it must not be possible for colluders to combine their images to generate a different valid watermark with the intention of framing a third party.

Universal The same digital watermarking algorithm should apply to all three media under consideration. This is potentially helpful in the watermarking of multimedia products. Also, this feature is conducive to implementation of audio and image/video watermarking algorithms on common hardware.

Complexity The requirements on complexity or speed vary with the application. The speed of a watermark embedding algorithm is important for applications where documents are marked when they are distributed. The large bandwidth necessary for video also requires fast embedding methods. If watermarking is used for content/owner authentication (fragile watermark), each receiver has to retrieve the watermark and therefore the retrieval should be easy. If it is used for tracking illegal copies (robust watermark), watermark retrieval is needed only when copyright violations have to be investigated. Here the watermark insertion should be of low

complexity, but retrieval operation can be more complex in order to account for all possible kinds of attacks on the watermark.

Watermark retrieval without reference to original For some applications, it is necessary to recover the watermark without requiring the original, unmarked document.

Unambiguous Retrieval of the watermark should unambiguously identify the content owner or recipient. Further for fragile watermarks, the accuracy of owner identification should degrade gracefully upon attacks, like illegal changes made in the watermarked content.

2.4 Classifications Of Digital Watermarks

Digital watermarks can be classified based on factors like their nature, applications, insertion techniques etc.

- ***Visible and Invisible Watermarks***

Visible Watermarks: A watermark can be a visible "seal" placed over an image to identify the copyright owner. It is like the corporate logos used by the cable television industry to identify the source of the program, typically along the lower periphery of the television screen. Fig. 2 [from 55] is an example of an image, which is visibly watermarked. In this image the logo of '*hp*' could be seen in the background. But visible watermarks cause user distraction and are more prone to attacks. A good visible watermark should also resist falsification.

Figure 2: Example of an image with a visible watermark [55].

Invisible Watermarks: Invisible watermarks do not bring annoyance to the users, at the same time serving the purpose of copyright protection.

- ***Spatial and Spectral Watermarks (Based on insertion technique)***

Spatial Watermarks: The spatial watermarking techniques provide effective schemes for embedding an invisible watermark into the original image in the spatial domain. Most of the earlier schemes used this technique which produces images of high quality. However, these spatial watermarks are not very robust to attacks or transformations.

Spectral Watermarks: In spectral watermarking, the watermark is incorporated into the transform coefficients (e.g. Discrete Cosine Transform (DCT) coefficients) of an image. Spectral watermarking commonly uses the frequency sensitivity of the Human Visual System (HVS) to ensure that the watermark is invisible. Today a number of spectral watermarking algorithms are available which are robust to geometric

transformations, noise, filtering, compression, etc. As opposed to spatial watermarks that have relatively low bit capacity, spectral watermarks can embed a large number of bits without incurring noticeable visual artefacts.

- ***Private and Public Watermarks***

Private Watermarks: Private watermarks require the original image or a private key (for example, a pseudo-noise signal, when using the Spread Spectrum (SS) watermarking technique) to verify the mark. So only the owner can verify if his watermark is present in the document.

Public Watermarks: In practical systems, however, watermarking techniques are required that enable public decoding of the watermark. The idea is to make only parts of the pseudo-noise key public to verify the watermark, at the same time avoiding the possibility of removing the watermark by the public.

- ***Robust and Fragile Watermarks (Based on application)***

Robust Watermarks: Robust watermarks cannot be removed by an attacker without damaging the document. These watermarks are used to detect unauthorized copies of documents or images and to prove ownership. So the watermark should remain in the document after any type of attack like collusion, compression to low data rates, filtering, printing and rescanning, and geometric attacks such as cropping, resampling and rotation. Here, an attacker wants to remove the watermark at a minimal loss in quality, so that the owner cannot verify the presence of the watermark.

Fragile Watermarks: Fragile watermarks are used for content and/or author authentication [27, 56]. These are used for legal purposes, medical applications, news

reporting etc., where the content creator has to be verified while ensuring that the content has not been changed. Here the unrestricted distribution of copies of the image is much less a concern than verifying an image's origin and content. An attacker wants to make changes in the image without damaging the watermark, so that the image will pass as 'authentic'.

For authentication, it is important that even slight changes to the document be detected and localized. Embedding a false mark must be practically impossible and the watermark should not remain in the document after attacks like filtering, although it should survive cropping. These types of watermarks are known as fragile watermarks. The requirements and properties for robust and fragile watermarks are different.

- ***Adaptive and Non-Adaptive Watermarks***

Non-Adaptive Watermarks: In non-adaptive watermarking, a constant scale factor is used to insert the watermark. The watermarked coefficient, $M_i = V_i + \alpha \cdot W_i$ where V_i is the original coefficient, W_i is the watermark coefficient and α is the scale factor. The scale factor α determines the extent to which the watermark alters the original document.

Adaptive Watermarks: In image-adaptive watermarking, the watermark sequence is adapted to the local properties of the image using visual models. The Just Noticeable Difference (JND) thresholds are used to determine the maximum amount of watermark signal that can be tolerated at every pixel location without affecting the visual quality of the image. The JND thresholds are derived from the visual model

based on three different properties of the Human Visual System: frequency sensitivity, luminance sensitivity and contrast masking.

In audio–adaptive watermarking, the watermark is kept extremely low during silent segments and high during the noisy segments. This is because loud sounds tend to mask out quiet sounds.

2.5 Applications Of Watermarking

The watermark signal can serve various purposes [34], including:

Ownership assertion: In order to establish ownership over some content, user A can use a private key to generate a watermark and embed it into the original document. He then makes the watermarked document publicly available. Later, when user B claims he owns the image derived from user A's public image, user A can produce the unmarked original and demonstrate the presence of his watermark in user B's document. But since user B does not have the original unwatermarked document, he cannot do the same. For such a scheme to work, the watermark has to survive common image-processing operations. It also needs to be a function of the original image to avoid counterfeiting attacks.

Fingerprinting: To avoid unauthorised duplication and distribution of publicly available multimedia content, an author can embed a distinct watermark (or fingerprint) into each copy of the document. Upon finding unauthorised copies later, the origin of the copy can be determined by retrieving the fingerprint. In this application, the watermark needs to be invisible and invulnerable to deliberate attempts of forgery, removal or invalidation. Fig. 3 shows a scenario of video

transmission with watermarking for ownership assertion and fingerprinting. This is done by embedding a common copyright label (owner ID) and unique receiver IDs [19]. The video is also encrypted to make sure that only authorised users can view it. If the authorised user makes illegal copies, he can be caught by verifying the individual receiver ID contained in the copy.

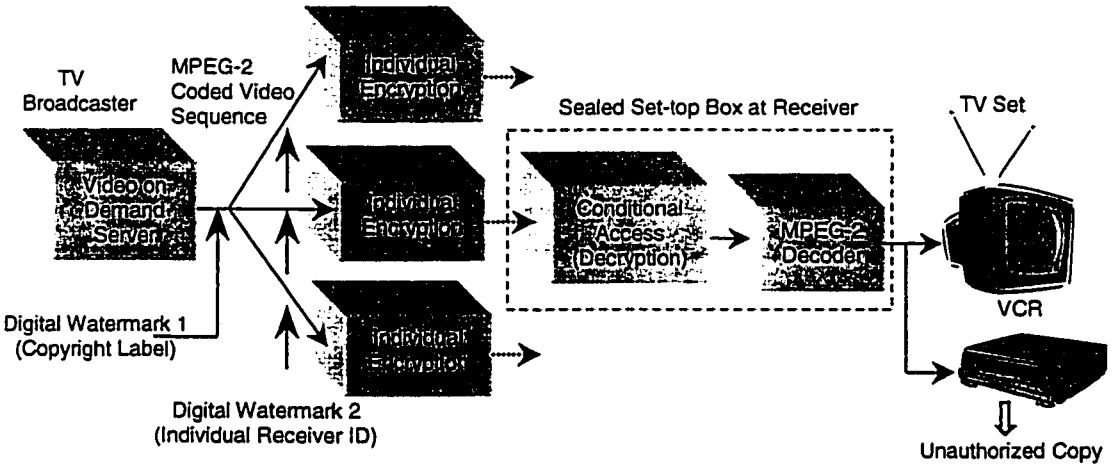


Figure 3: Video transmission scheme with watermarking .

Authentication and integrity verification: When multimedia content is used for legal purposes, medical applications, news reporting or commercial transactions, the originator of the content has to be verified while ensuring the content has not been changed, manipulated or falsified. When the watermarked data are checked, the watermark is extracted using a unique key associated with the source, and the integrity of the data is verified through the integrity of the extracted watermark.

Content labelling: The watermark bits embedded into the data can contain an annotation, giving further information about the contents. For example, a photographic image could be annotated to describe the time and place the photograph

was taken, a procedure that could be done automatically by the processor in the camera.

Usage Control: In a closed system in which the multimedia content needs special hardware for copying and viewing, a digital watermark can be inserted to indicate the number of copies permitted. Every time a copy is made, the watermark can be modified by the hardware, and at some point the hardware would not create any more copies of the data.

An example is the Digital Video Disc, also known as Digital Versatile Disc (DVD). Recently the DVD consortium called for proposals for a copyright marking scheme to enforce serial copy management. The idea is that the DVD players sold to consumers should allow unlimited copying of home videos and time-shifted viewing of TV programmes, but it should not be abused for commercial piracy [41]. The proposed implementation is that videos will be either unmarked, or marked 'never copy' or 'copy once only'. Compliant players would not record a video marked 'never copy' and when recording one marked 'copy once only' would change its mark to 'never copy'. Commercially sold videos would be marked 'never copy', while TV broadcasts and similar material would be marked 'copy once only' and home videos would be unmarked.

Content Protection: In certain applications, a content owner may want to publicly and freely provide a preview of the multimedia content being sold. To make the preview commercially worthless, the content could be stamped with a visible watermark very difficult to remove in an automated way.

The specific requirements of each watermarking technique vary with the application. There is no watermarking technique that could satisfy all the requirements of all the applications. Consequently, each watermarking technique has to be designed within the context of the entire system in which it is to be used.

2.6 Previous Work In Image Watermarking

Several data embedding methods are proposed for grayscale, dithered binary, facsimile, color still images and video. All the earlier works used spatial watermarking methods [49]. But these spatial watermarks are not very robust to image compression, noise and common geometric distortions, especially image cropping and rotation. The transform (spectral) domain methods emphasize the importance of varying the watermark strength in different DCT coefficients to ensure imperceptibility and robustness.

2.6.1 Spatial Watermarking Methods

Method 1: The digitized grayscale image data $\{x_i\}$; $i \in N$ are converted to a sequence in which the first element is x_i and subsequent elements are the differences between successive points, i.e., $\Delta_i = x_i - x_{i-1}$. Next the person(s) embedding and extracting the message agree on the use of a particular cipher key table, which assigns a value c_i, either zero or one, to each Δ_i. To embed a binary sequence $B = \{ b_i: b_i = 0 \text{ or } 1 \}$; $i \in N$, look up the value of c_i corresponding to Δ_i. If $c_i = b_i$, then keep Δ_i as is. If $c_i \neq b_i$, go to the nearest Δ_j such that $c_j = b_i$ and substitute Δ_j in place of Δ_i. The hidden message can be retrieved by looking up the value for c_i corresponding to Δ_i [35].

Method 2: Facsimile document signals are digitized by scanning in the horizontal direction [35]. The message-embedding scheme is based on the fact that the data will be compressed using Run-Length Coding (RLC) and modified Huffman coding schemes. A binary message is embedded by shortening or lengthening runs by one pixel at the boundaries of the runs.

Method 3: Bender *et al.* proposed a method [4] called patchwork, where a subset of pixels in an image is divided into two distinct sets. The brightness of one set of pixels is shifted by a positive number, and those from the other set are shifted by the corresponding negative number. Only a limited amount of information can be embedded in an image using this approach.

Method 4: In a method [14] by Dauzenberg and Boland, images are partitioned into blocks, and the mean is shifted by one or zero to embed a binary code. But here the amount of information that can be embedded depends on the number of blocks in the image.

Method 5: Kutter *et al.* [28] use amplitude modulation and a secret key to embed a signature in color images. The signature bits are repeatedly embedded in the blue channel to ensure robustness. The blue channel is used because the HVS (Human Visual System) is less sensitive in this color domain. A single bit is embedded in a randomly selected pixel in an image by modifying the blue channel by a fraction of the luminance. The embedded message is retrieved using prediction of the original value of the pixel based on a linear combination of its neighboring pixel values.

Method 6: Puate and Jordan [45] use Fractal Compression analysis to embed a signature in an image. A binary signature is embedded in an image by varying the regions in which pattern-matching searches are performed. Although the method is robust to JPEG compression, only a limited amount of data could be embedded using this method. Also since fractal analysis is computationally expensive and some images do not have many large, self-similar patterns, the technique may not be suitable for general use.

2.6.2 Spectral Watermarking Methods

Method 1: In a method suggested by ÓRuanaidh [37], the image blocks are transformed using DCT, Walsh transform, or Wavelet transform (WT). The data is embedded by incrementing a selected coefficient to encode a "1" and decrementing it to encode a "0". Coefficients are selected according to a criterion based on energy content. In a second approach [38], the watermark is embedded in the discrete Fourier transform phase.

Method 2: In the algorithm by Cox *et al.* [11], the watermark is embedded in the largest magnitude DCT coefficients to provide greater robustness to compression algorithms. In order to place a length n watermark into an $N \times N$ image, the $N \times N$ DCT of the image is computed and the watermark is inserted into the n highest magnitude coefficients of the transform matrix, excluding the DC component.

Watermark insertion:

The *watermark insertion* algorithm consists of 4 steps.

1. Compute the DCT/WT and identify perceptually significant regions of the image suitable for watermark embedding.

2. Construct the watermark $w = w_1, w_2,\ldots,w_n$, where each w_i is chosen according to a normal distribution, $N(0,1)$.
3. Insert the watermark in the frequency domain of the image by setting the frequency component $V_{u,v}$ in the original image to

$$M_{u,v} = V_{u,v}(1+ \alpha \cdot W_{u,v}), \tag{2.1}$$

$$M_{u,v} = (V_{u,v} + \alpha \cdot W_{u,v}) \text{ or} \tag{2.2}$$

$$M_{u,v} = V_{u,v}\,(e^{\alpha \cdot W_{u,v}}) \tag{2.3}$$

where α is a scale factor,

$V_{u,v}$ is the original DCT/WT coefficient

$M_{u,v}$ is the watermarked DCT/WT coefficient and

$W_{u,v}$ is the DCT/WT watermark coefficient.

4. Compute the inverse DCT/WT of the sum to recover a transparently marked image.

Watermark extraction:

The *watermark extraction* consists of 4 steps as given below:

1. Compute the DCT/WT of a (possibly) watermarked image.
2. Compute the DCT/WT of the original image.
3. Compute the difference in the results from the previous two steps to a watermark W_e.
4. Compare the extracted mark W_e with the original watermark $W_{u,v}$.

This method is robust to compression, dithering, clipping, printing and rescanning etc., but the strength of the technique can be diminished when multiple

watermarked copies of the same image are available to an 'attacker'. Also the original image and the watermarked images are needed for watermark extraction.

The DCT-based technique offers the advantage of direct watermarking of JPEG bit streams by partially decompressing the bit stream. Specifically the bit stream is passed through the entropy decoder and inverse quantizer and then watermarked.

The WT-based method gives better results as it best matches the properties of the Human Visual System (HVS). Wavelet based methods produce watermark with both a spatially local and a spatially global support. The watermarked component with local spatial support is robust to signal processing such as cropping. The watermark component with global spatial support is robust to operations like low pass filtering (LPF). The DCT-based approach provides watermark with only global spatial support.

Method 3: A block-based version of Cox's method that employs visual models has been developed by Podilchuk and Zeng [44]. Their algorithm uses the Just Noticeable Difference (JND) paradigm employed by perceptual coders. Here the value of the watermark is larger at the edges and highly textured areas in the image, where it is least visible. The results are slightly better than the original method of Cox *et al.*

For *Image-Adaptive watermarking*, the JND threshold is calculated at each pixel location and it is used as the scale factor. Fig. 4 illustrates the method for ***image-adaptive watermark insertion***. The JND threshold is a measure of the capacity of a pixel to hold a watermark bit without it being noticeable.

The watermark insertion process assuming that the watermark is added only to the perceptually significant coefficients, could be represented as:

$$M_{u,v} = \begin{cases} V_{u,v} + J_{u,v} \cdot W_{u,v} & , \ if\ V_{u,v} > J_{u,v}\ and\ J_{u,v}/V_{u,v} < T_{JND} \\ V_{u,v} & , \ otherwise \end{cases} \tag{2.4}$$

where T_{JND} is a threshold value which determines the cut-off for perceptually significant frequency components as determined by the JND threshold values.

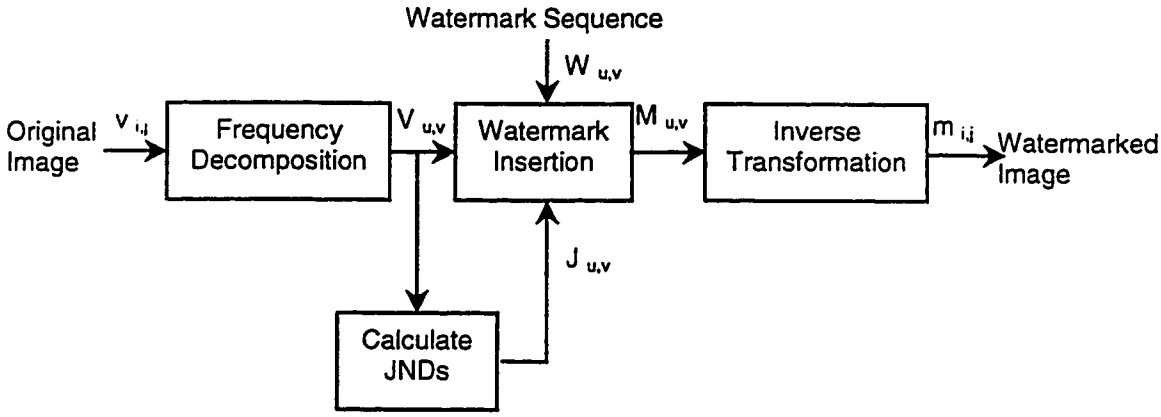


Figure 4: Image adaptive watermark insertion in the frequency domain.

Fig. 5 shows the ***watermark detection*** method using correlation detection. Detection is done by obtaining the correlation between a specific watermark sequence and the difference between the original and the received images. The correlation value is compared to a threshold, T_{det} to determine whether the received image contains the watermark in question. The detection scheme can be expressed as:

$$W_e = M_{u,v} - V_{u,v} \tag{2.5}$$

$$\rho(w_e, w) = \frac{w_e \cdot w}{\sqrt{E_{w_e} \cdot E_w}} \tag{2.6}$$

where $V_{u,v}$ and $M_{u,v}$ are the transform coefficients of the original and the received watermarked images. w and w_e are the original and extracted watermark sequences in the spatial domain. If $\rho(w_e,w)>T_{det}$, watermark w is detected and if $\rho(w_e,w)\leq T_{det}$, watermark w is not detected. The threshold T_{det} is determined based on the tradeoff between the probability of detection and the probability of false alarm.

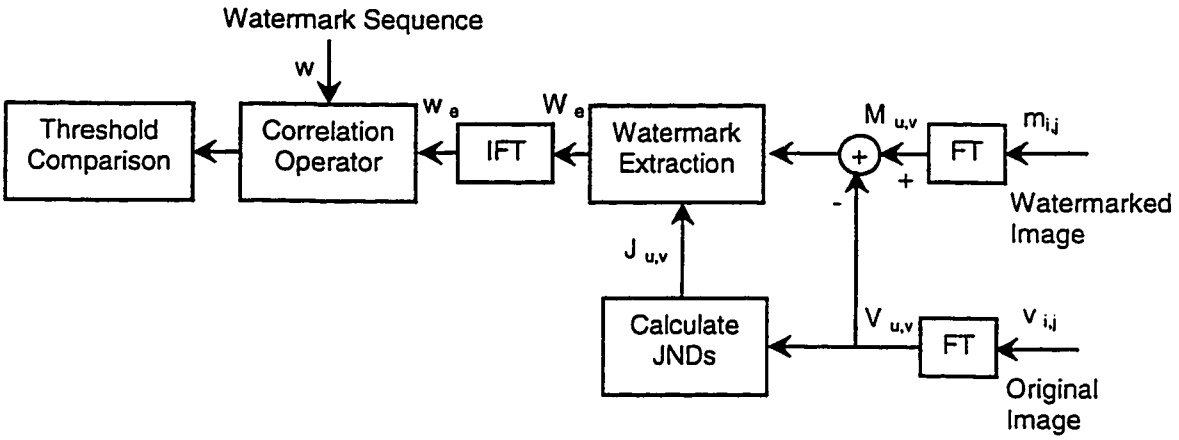


Figure 5: Image adaptive watermark detection using correlation detector.

3. CHALLENGE IN WATERMARKING

3.1 Introduction

All of the existing watermarking schemes have focused on the means to label a document *invisibly* and to ensure *robustness* of the inserted labels against malicious attacks. As a result, the concerns regarding what watermarks can achieve or fail to achieve have not been properly addressed. That is, the question as to whether or not these watermarking techniques can really be used to prove beyond doubt that a document in dispute has actually been derived from a user's original has not been considered. The answer to this question is 'no', with most invisible watermarking schemes which cannot resolve rightful ownership of a document watermarked with multiple ownership labels. Above all, watermark removal software like Stirmark [42] and Unzign [43] which are available on the Internet can remove the watermark inserted by almost all of the existing methods by geometric transformations and filtering.

3.2 Classification Of Attacks On Digital Watermarks

The possible attacks [23] on watermarks can be broadly put in four categories as shown in Fig. 6. These are attacks that do not impair the quality of the host data (data in which the watermark is embedded) significantly.

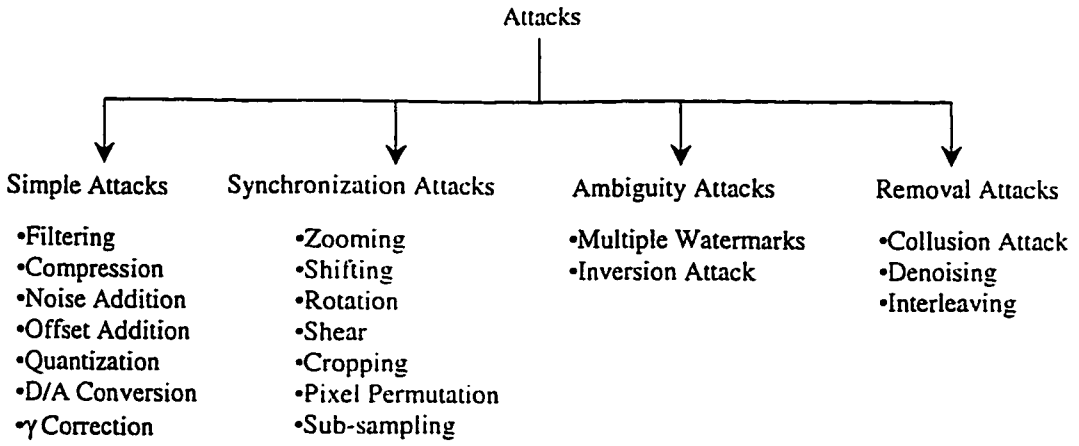


Figure 6: Classification of attacks on digital watermarks.

3.2.1 Simple Attacks

These are conceptually *simple attacks* that attempt to impair the embedded watermark by manipulations of the watermarked data (host data plus watermark), without attempting to identify and isolate the watermark. Examples include linear and non-linear filtering, compression (JPEG, MPEG), addition of noise, addition of an offset, quantization in the pixel domain, digital-to-analog conversion and gamma correction.

3.2.2 Synchronization Attacks

Synchronization attacks are detection disabling attacks that attempt to break the correlation between the watermarked image and the pseudo-noise signal and to make the recovery of the watermark impossible for a watermark detector. Most of the existing watermarking schemes require proper synchronization of the watermarked data and the watermark, for effective extraction of the watermark. Geometric distortions like zooming, shift in spatial or temporal (for video) direction, rotation,

shear, cropping, pixel permutations, sub-sampling, removal or insertion of pixels, etc., result in loss of synchronization which can make watermark detection infeasible. A typical property of this type of attack is that the watermark remains in the attacked data and can be recovered with increased complexity of the watermark decoder.

3.2.3 Ambiguity Attacks

Ambiguity attacks attempt to confuse people by producing fake original data or fake watermarked data.

Multiple Watermarking: When watermarking is used for authentication, the extracted watermarks should identify the rightful owner.

Suppose Alice watermarks an image with her ownership label, S_A, using some watermarking technique and makes the watermarked image publicly available. Now Bob cannot extract Alice's label but he can insert his label, S_B, into the already watermarked image. Since now the image contains both the labels S_A and S_B, how can anyone determine who is the true owner? Thus, the *multiple watermarking attack* embeds one or several additional watermarks such that it is unclear which was the first watermark of the true owner.

Of course, somewhere out in the dark, there are the so-called original images (or, unwatermarked images). Alice should obtain Bob's original image and check if it contains her watermark. Similarly, Bob can ask for her original image and check for his mark. Since Alice's original image does not contain Bob's label and Bob's original image (which is actually the watermarked image of Alice) contain Alice's label, it could be proved that Alice is the true owner of the image in dispute.

Fig. 6 (from Craver *et al.* [12]) illustrates such a scenario. Suppose there is a watermarked image $\hat{I}$ in which the watermarks of both Alice and Bob have been detected, and both claim rightful ownership. I^0_A is Alice's claimed original image, and I^0_B is the claimed original of Bob. D_A, C_A, D_B and C_B represents the decoding and watermark comparator functions used by Alice and Bob respectively. To check the presence of Alice's and Bob's watermarks, the watermarks from the watermarked image are first extracted by Test I, as in Fig. 7(a), and compared with their respective watermarks, and similarly the decoding tests can be achieved using the two 'original' images by Test II as in Fig. 7(b). The results of the tests, together with the logical determination of ownership by the watermark decoding tests, are tabulated in Table1. '1' indicates the presence of watermark, '0' indicates the absence, and '×' represents "don't cares".

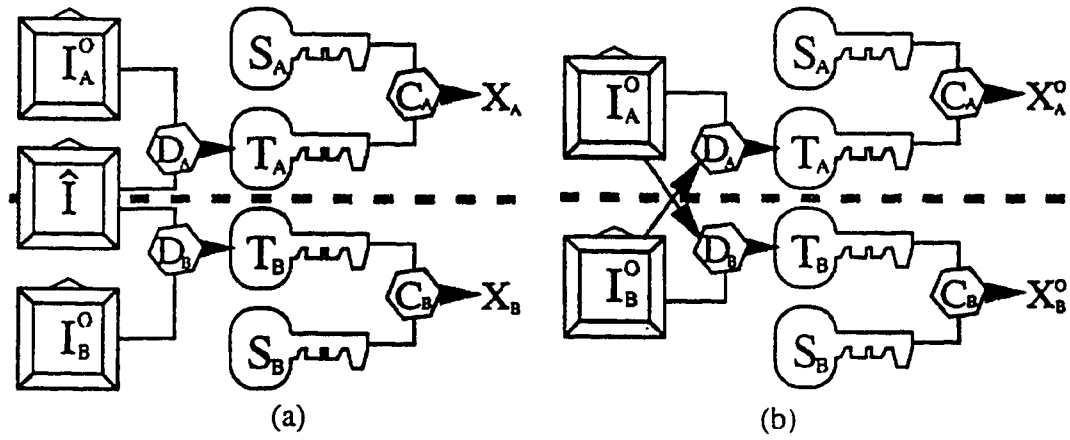


Figure 7: Decoding tests to extract watermarks for ownership determination [12].
(a) Test I: Alice and Bob test an image for the presence of a watermark.
(b) Test II: Alice and Bob test each other's original images for the presence of a watermark.

Scenario	Test I		Test II		Derived Ownership
	x_A	x_B	x_A^0	x_B^0	
Case 1	1	0	X	X	Alice
Case 2	0	1	X	X	Bob
Case 3	1	1	1	0	Alice
Case 4	1	1	0	1	Bob
Case 5	1	1	1	1	???

Table 1: Determination of ownership from watermark decoding tests.

Inversion Attack: A more sophisticated attack is suggested by Craver *et al.* in [12] called *the inversion attack* or the *SWICO* (Single Watermarked Image Counterfeit Original) *attack*. The above problem of multiple watermarking was solved because Alice's original image did not contain Bob's label. But what if it contains Bob's label? By proper selection of his ownership label, it is possible that Bob can claim ownership and Alice's original image can be made to contain his label.

Let S be Alice's label embedded in the original image. If the watermarking scheme is invertible, Bob can *remove* a randomly selected watermark S′ instead of embedding one. This means that, Bob can identify some features in a watermarked image and claim them to be his watermark, which he removes from his fake original (which is Alice's watermarked image). So these features and thereby S′ will be present in Alice's original image (which is the true original).

To avoid such a situation it should be ensured that the watermarking scheme is non-invertible. Another solution might be to use time stamping, so that the watermarked image with an earlier timestamp would be the original one.

3.2.4 Removal Attacks

Removal attacks attempt to analyze the watermarked data, estimate the watermark or the host data, separate the watermarked data into host data and watermark, and discard only the watermark.

Collusion attacks are a threat when differently watermarked versions of the same data are distributed, i.e. for fingerprinting applications. Averaging of the differently watermarked copies can remove the watermark. Boneh and Shaw have shown that it is possible to construct collusion-secure watermark signals [5]. The basic idea is to compose the watermarks out of static components (that do not vanish by averaging) and dynamic components. The watermark should be designed in such a way that for every possible combination of colluding parties there are parts of the codes that do not average to zero. Thereby it is possible to determine all colluding parties from the colluded (averaged) version of the watermarked data. But a drawback is that the length of the proposed collusion-secure codes increases exponentially with the number of different codes.

Another possible attack when several differently watermarked copies are available is by *interleaving* the watermarked data from different copies. An attacker could switch between different watermarked copies of the data, and thus easily scramble the embedded watermark information. To avoid such attacks, the spatial position of the embedded bits should be chosen randomly.

An attack possible on video is *inter-frame collusion attack.* If a different watermark is used for each frame, an attacker can compare frames that change very

little within a scene. Simple collusion between these frames can remove a large part of the watermark. Also, the computational effort to test each frame for watermark extraction will be large. Using the same watermark for all the frames also poses problems, since an attacker could then collude with frames from completely different scenes.

Scene-adaptive watermarking [50] using temporal wavelet transform could solve this problem. This places the elements of the video that change little over the course of the scene (background, etc.) in the low time-resolution bands of the transform. The parts with fast motion would be decomposed into higher resolution bands. This effectively isolates the two types of motion in the scene. A different watermark is then placed in each resolution band. This results in the same sequence being embedded in the slow-changing parts of a scene to avoid collusion. The sequences added into the sections with large amounts of motion are different and change frequently throughout the scene. The inverse transform of the watermarked wavelet coefficients forms the marked video sequence.

3.3 Stirmark Attack

Stirmark [42] is a generic tool developed for robustness testing of image marking algorithms. In its simplest version, Stirmark simulates a resampling process, i.e., it introduces the same kind of errors into an image as printing on a high quality printer and then scanning it again with a high quality scanner. It applies a minor geometric distortion: the image is slightly stretched, sheared, shifted and/or rotated by an unnoticeable random amount and then resampled using either bi-linear or Nyquist

interpolation. In addition, a transfer function that introduces a small and smoothly distributed error into all sample values is applied. This emulates the small non-linear analog/digital converter imperfection typically found in scanners and display devices. Stirmark introduces a practically unnoticeable quality loss in the image if it is applied only once. However after a few iterated applications, the image degradation becomes noticeable.

4. SPREAD SPECTRUM WATERMARKING

4.1 Introduction

Most of the watermarking methods proposed for images and video are based on concepts from spread spectrum radio communications, namely additive embedding of a pseudo-noise watermark pattern and watermark recovery by correlation. Even methods that are not presented as spread spectrum methods often build on these principles.

Cox et al. noted the applicability of spread spectrum techniques to digital image watermarking [9]. Spread spectrum has several advantageous features. It offers cryptographic security and is capable of achieving error free transmission of the watermark at the limits given by the maximum channel capacity.

4.2 Spread Spectrum Digital Communication System

4.2.1 Spread Spectrum Communication Model

In spread spectrum communication, a message may be hidden in the background noise by spreading its bandwidth with coding and transmitting the resultant signal at a low average power [40]. Because of its low power level, it has a low probability of being intercepted (detected) by a casual listener and hence is also called a *low-probability-of-intercept* (LPI) signal.

The block diagram in Fig. 8 illustrates the basic elements of a spread spectrum digital communication system. The information sequence, at the input of the

transmitting side and the output of the receiving side is binary. The pseudo-random pattern generators on both sides have to generate the same pseudo-noise (PN) binary valued sequence, $\{p_i\}$. This pseudo-random sequence is impressed on the transmitted signal at the modulator and removed from the received signal at the demodulator. The resulting modulated signal is called a direct sequence spread spectrum (DSSS) signal. For demodulation, the two PN sequences should be perfectly synchronized.

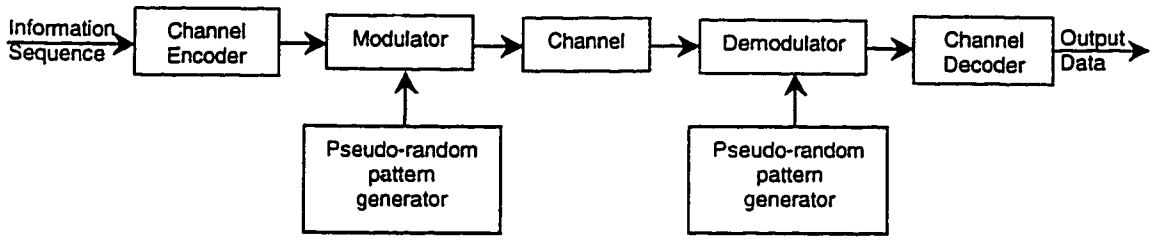


Figure 8: Model of spread spectrum digital communication system.

4.2.2 Direct Sequence Spread Spectrum Signals

The modulator in Fig. 8 could be a Binary Phase Shift Keying (BPSK) modulator. Let the information rate at the input of the encoder be R bits/sec and the available channel bandwidth be W Hz. In order to utilize the entire available channel bandwidth, the phase of the carrier is shifted pseudo-randomly according to the pattern from the PN generator at a rate W times/sec. The reciprocal of W, denoted by T_c, defines the duration of a rectangular pulse, which is called a *chip* while Tc is called the *chip interval.*

Let $T_b = 1/R$ be the duration of a rectangular pulse corresponding to the transmission time of an information bit. Then the bandwidth expansion factor may be expressed as:

$$B_e = \frac{W}{R} = \frac{T_b}{T_c} \tag{4.1}$$

In practical systems, this ratio is an integer, r_c called the chip-rate which is the number of chips per information bit.

$$r_c = \frac{T_b}{T_c} \tag{4.2}$$

r_c is the number of phase shifts that occur in the transmitted signal during the bit duration $T_b = 1/R$. Fig 9 illustrates the relationship between the PN signal and the data signal or the information sequence.

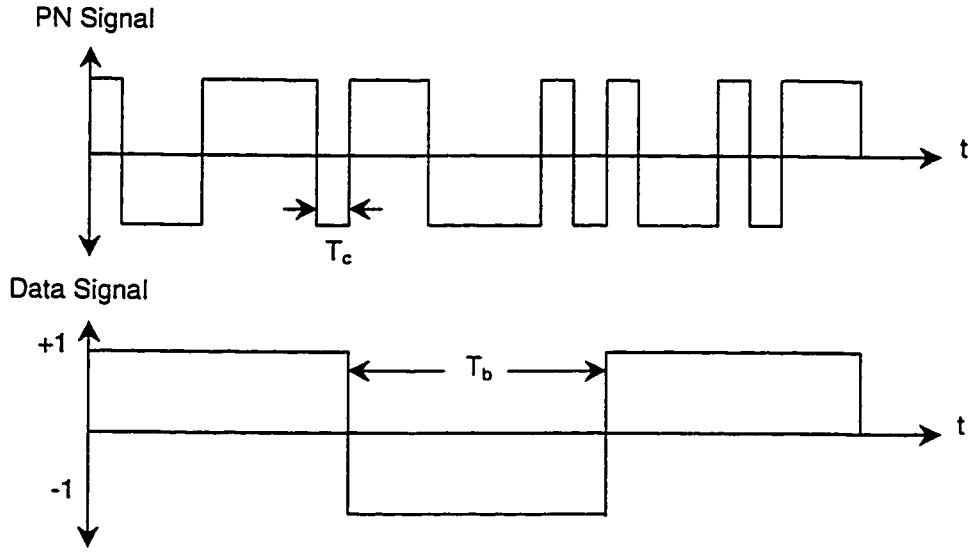


Figure 9: The PN and data signals for a DS spread spectrum system.

The spreading sequence: The pseudo-noise sequence, $\{p_i\}$ where $p_i \in \{-1,1\}$ is called the *spreading sequence*, and it has several special properties [16]. Assume that the spreading sequence, $\{p_i\}$ is defined for all i and has period N, i.e., for any k,

$$\sum_{i=0}^{N-1} p_i p_{i+kN} = N \tag{4.3}$$

Two of the key properties of $\{p_i\}$are:

(i) $\{p_i\}$ should have a mean value of approximately zero, i.e.,

$$\frac{1}{N}\sum_{i=0}^{N-1} p_i \approx 0 \tag{4.4}$$

(ii) The discrete-time periodic autocorrelation function is given by,

$$\frac{1}{N}\sum_{i=0}^{N-1} p_i p_{i+k} = \begin{cases} 1, & i = 0 \\ -1/N, & 0 < |i| < N. \end{cases} \tag{4.5}$$

4.3 Image Watermarking Using Spread Spectrum Principle

Cox *et al.* [9] introduced the concept of "spread spectrum watermarking". They observed that for a watermark to be robust, it must be somehow inserted into the most perceptually significant image components. Otherwise, it could be erased by suppressing the least perceptually significant components from a watermarked image, without altering the perceived image quality. Clearly, in order for the changes introduced to the perceptually significant components of an image to remain invisible, these changes must be small. However small perturbations are very sensitive to noise.

In spread spectrum communication, a narrow band signal is converted to a wide band signal by modulation so that the energy in any narrow spectral band is

small. However by appropriately combining these weak signals at the demodulator, the original information signal is recovered. The same idea is used to insert a watermark. Many small changes are introduced into the most significant image components. Since, during the extraction process, the location and value of these changes are known, it is possible to concentrate the information of all these small changes to come up with a robust decision on the presence or absence of a watermark. Furthermore, to destroy such a watermark, noise of high amplitude would be required in all these perceptually significant components, thus drastically reducing the perceived image quality too.

Spread spectrum is an example of a symmetric key crypto-system. System security is based on proprietary knowledge of the keys (or the seeds for pseudo-random noise generators) which are required to embed, extract or detect an image watermark. Given the keys or seeds, the sequences themselves can be generated with ease. A good spread spectrum sequence is one which combines desirable statistical properties such as uniformly low cross correlation with cryptographic security. Examples of sequences used in spread spectrum systems used in digital watermarking include m-sequences, Gold codes, Kasami codes and Legendre sequences [40].

Most of the effective watermarking methods developed so far are based on the spread spectrum principle. The method proposed by Hartung *et al.* in [22] for watermarking compressed and uncompressed video also stems from this idea. But in their method, the watermark is inserted in all the DCT coefficients. The strength of the method comes from the redundancy caused by embedding the same information

bit in a number of pixels. Their method of watermark generation, insertion and extraction is outlined below and it has been used in our work.

4.3.1 Spread Spectrum Watermark Generation

Let N be the total number of pixels in an image. Let r_c be the chip-rate used to spread the watermark bits. Then a total of $N_{info} = N/r_c$ information bits could be embedded in the image. Let $\{a_j\}$ where $a_j \in \{-1,1\}$ be the sequence of information bits that has to be embedded into the image. This discrete signal is spread by a large factor, that is the chip-rate r_c, to obtain the spread sequence $\{b_i\}$.

$$b_i = a_j, \quad \text{where } j \cdot r_c \leq i < (j+1) \cdot r_c \tag{4.6}$$

The purpose of spreading is to add redundancy by embedding one bit of information into r_c pixels of the image. The spread sequence $\{b_i\}$ is then modulated by a pseudo-noise sequence $\{p_i\}$ where $p_i \in \{-1,1\}$. p_i serves for frequency spreading.

$$w_i = b_i \cdot p_i \tag{4.7}$$

where w_i is the spread spectrum watermark, which is arranged into a matrix with size equal to the image size.

4.3.2 Spread Spectrum Watermark Insertion

The watermark could be inserted in the spatial domain or the spectral domain.

Spatial Watermarking: In spatial watermarking, the spread spectrum watermark w_i, is scaled with a factor, α (called the watermark embedding depth or scale factor) and is directly added to the image pixel values v_i, to give the watermarked image m_i.

$$m_i = v_i + \alpha \cdot w_i \tag{4.8}$$

Fig. 10 shows the spread spectrum technique for watermark generation and embedding. Due to the noisy nature of the pseudo-noise signal, p_i, the watermark signal, w_i is also a noise like signal and thus difficult to detect, locate and manipulate.

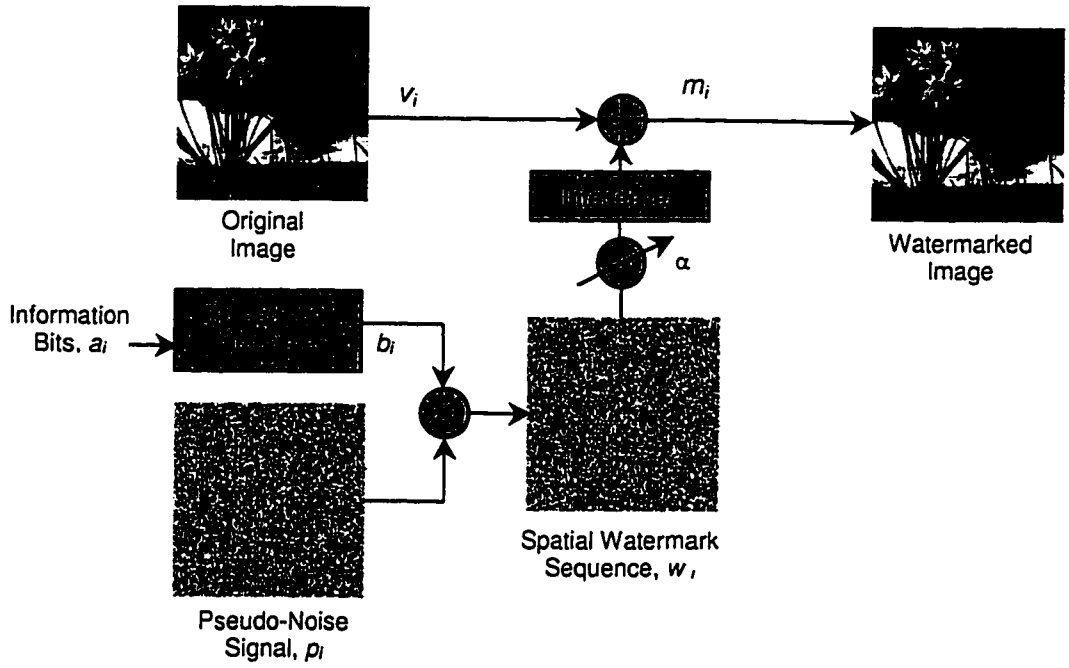


Figure 10: Spread spectrum watermark generation and insertion in the spatial domain.

Interleaving is used in spread spectrum communication systems to get independent chip contributions separated in time. These contributions provide time diversity that can help to improve the resilience against channel variability. In this case channel variability is related to the non-stationarity of common images. If pixels in the tiling are reordered in such a way that pixels modulated by each information bit are scattered over the whole image, then the resulting spatial diversity can be used to improve the performance of the watermark detector and decoder.

Spectral Watermarking: In spectral watermarking, the watermark bits w_i, are frequency transformed using the Discrete Cosine Transform, scaled with a scale

factor, α and added to the image DCT coefficients. The watermark DCT coefficients W_i, are added to the corresponding image DCT coefficients V_i, excluding the *dc coefficient*, to give the watermarked DCT coefficients M_i :

$$M_i = V_i + \alpha \cdot Wi \tag{4.9}$$

Performing the inverse DCT on M_i will give the watermarked image.

4.3.3 Spread Spectrum Watermark Extraction

4.3.3.1 Watermark extraction without using the original image

For fingerprinting, we do not want the users to extract the watermark. Only the owner who has the original document should be able to detect or extract the watermark. But in all other applications like authentication, captioning, access control etc., the embedded information needs to be accessed by the users. An example is when a user's web browser should extract and display a caption or the identity of the owner of a downloaded document. Another application is when it is required that a web crawler or search engine should automatically find all illegal copies of an image that belong to a particular photo archive or all images with a certain embedded caption.

Using the spread spectrum method, the watermark could be extracted without using the original, unwatermarked image by means of a correlation receiver as shown in Fig. 11. But the pseudo noise sequence $\{p_i\}$ is needed for watermark extraction. The input watermarked image is first high pass filtered to separate and remove major components of the image itself. The second step is demodulation, which is the multiplication of the filtered watermarked image with the same pseudo-noise signal

$\{p_i\}$ that was used for embedding. This is followed by summation over a window of length equal to the chip-rate, yielding the correlation sum s_j for the j'th information bit.

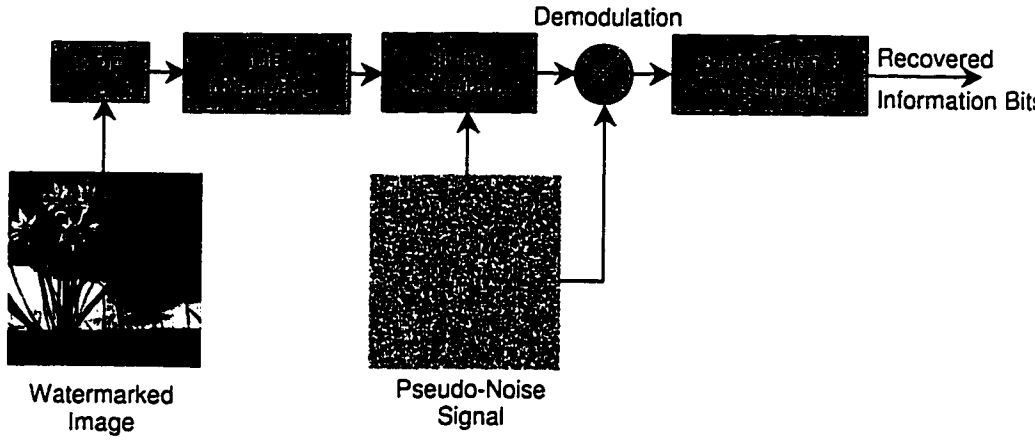


Figure 11: Spread spectrum watermark extraction without the original.

The watermarked image, $m_i = v_i + \alpha \cdot w_i$ where $w_i = b_i \cdot p_i$. The high pass filter removes major components of the image, v_i. Therefore,

$$s_j \approx \sum_{i=j \cdot r_c}^{(j+1) \cdot r_c - 1} p_i \cdot \alpha \cdot w_i \tag{4.10}$$

$$\approx \sum_{i=j \cdot r_c}^{(j+1) \cdot r_c - 1} p_i^2 \cdot \alpha \cdot b_i \approx a_j \cdot r_c \cdot \alpha \tag{4.11}$$

$$sign(s_j) = sign(a_j \cdot r_c \cdot \alpha) = sign(a_j) = a_j \tag{4.12}$$

This is because $r_c > 0$, $\alpha > 0$, $p_i^2 = 1$ and $a_j = \pm 1$. Thus the embedded bit can be retrieved without any loss. This means that the embedded information bit is 1 if the correlation is positive and –1 if it is negative. But since the image is not completely

removed by the high pass filter, it can result in some error in the extracted watermark bits.

If the wrong pseudo-noise sequence is used or if it is not in synchronization with the pseudo-noise sequence used for watermark embedding, the scheme does not work and the recovered bits are random. If synchronization is lost, it is possible to re-synchronize the sequence using a sliding correlator.

Sliding correlator to detect synchronization loss: Detection-disabling attacks like Stirmark exploit the fact that the human visual system is not sensitive to small shifts and global modifications, as long as there are no severe local modifications. The Stirmark attack introduces small shifts, rotation or zooming. In radio spread spectrum, the sliding correlator is a popular method to re-synchronize one-dimensional spread spectrum signals in the case of synchronization loss. This concept can be extended to images and video for counter attacking these attacks (Fig. 12).

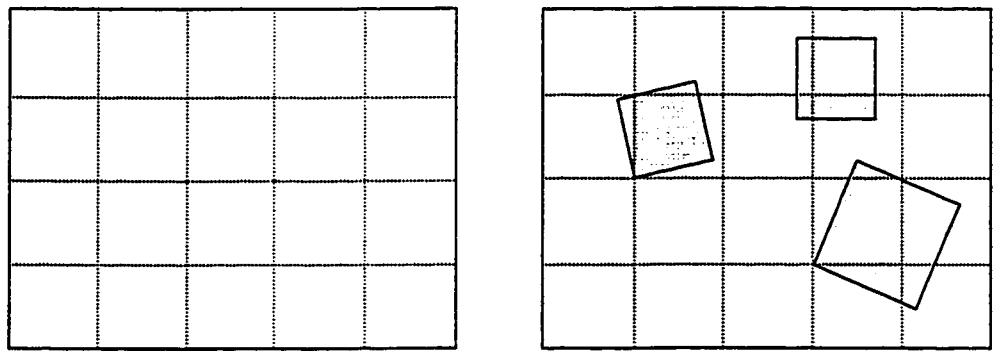


Figure 12: Block based sliding correlation.

(a) Subdivision of the image into blocks. (b) Example combinations of shift, zoom and rotation for the sliding correlator.

The attacked image is divided into blocks of arbitrary size, e.g. 16 × 16 as shown in. 12(a). For each of the blocks, all possible combinations of shift, rotation, zoom etc. are applied as illustrated in Fig. 12(b), and the correlation between the modified block and the original pseudo-noise signal is calculated. The modification with the highest correlation for each block is assumed to be the one resulting from the attack, and is used to undo the effect of the attack for the block.

4.3.3.2 Watermark extraction using the original image

The recovery of the watermark bits is more robust, if the original, unwatermarked image is available. The image can be completely eliminated by subtraction before the demodulation process.

4.3.4 Spread Spectrum Watermark Detection

After extraction of the watermark (information bits $\{a_j\}$) using the pseudo-noise sequence $\{p_i\}$, the normalized correlation coefficient $\rho(w_e,w_i)$ is evaluated for the extracted watermark w_e ,and all the other original watermarks $\{w_i\}$, to identify the particular user.

$$\rho(w_e, w_i) = \frac{\sum_{n=1}^{N_{info}} w_e(n) \cdot w_i(n)}{\sqrt{\sum_{n=1}^{N_{info}} w_e^2(n)} \cdot \sqrt{\sum_{n=1}^{N_{info}} w_i^2(n)}} \tag{4.13}$$

$$\rho(w_e,w_i) = \begin{cases} 1; & \text{if the watermark is extracted without any error.} \\ 0; & \text{if 50 \% of the watermark bits are in error, the worst case.} \end{cases}$$

where $N_{info} = N/r_c$ is the number of information bits embedded in the image. N is the total number of pixels in the image and r_c is the chip-rate. Each user has a unique watermark w_i. $i = 1, 2, 3, \ldots N_u$ where N_u is the number of users.

4.4 Rotation, Scale and Translation Invariant Watermarking

A major drawback of spread spectrum watermarking (commonly used along with DCT or DWT) is that it is not tolerant to timing errors. For this reason, a sliding correlator has to be used to detect the changes and re-synchronize the watermarked image. Therefore it would be advantageous to use a transform which is invariant to loss of synchronization. Ó Ruanaidh *et al.* [39] describes how Fourier-Mellin transform-based invariants can be used for digital image watermarking.

Let the image be a real valued continuous function $f(x_1,x_2)$ defined on an integer valued Cartesian grid $0 \le x_1 < N_1$, $0 \le x_2 < N_2$. The 2-D Discrete Fourier Transform (DFT) is defined as:

$$F(k_1,k_2) = \sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} f(x_1,x_2)\cdot e^{-\left(\frac{j2\pi x_1 k_1}{N_1}\right)}\cdot e^{-\left(\frac{j2\pi x_2 k_2}{N_2}\right)} \tag{4.14}$$

The inverse transform is,

$$f(x_1,x_2) = \frac{1}{N_1 N_2}\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1} F(k_1,k_2)\cdot e^{\left(\frac{j2\pi x_1 k_1}{N_1}\right)}\cdot e^{\left(\frac{j2\pi x_2 k_2}{N_2}\right)} \tag{4.15}$$

The DFT of a real image is generally complex valued, which leads to a magnitude and phase representation for the image given by:

$$A(k_1,k_2) = |F(k_1,k_2)| \tag{4.16}$$

$$\phi(k_1,k_2) = \angle F(k_1,k_2) \tag{4.17}$$

Translation Invariance: From the property of Fourier transform, shifts in the spatial domain cause a linear shift in the phase component. i.e.,

$$F(k_1,k_2)\cdot e^{[-j(ak_1+bk_2)]} \leftrightarrow f(x_1+a, x_2+b) \tag{4.18}$$

Both $F(k_1,k_2)$ and its dual $f(x_1,x_2)$ are periodic functions. So it is assumed that translations cause the image to be "wrapped around", which could be referred to as circular translation. Since spatial shifts affect only the phase representation of an image, the DFT magnitude is circular translation invariant.

Rotation and Scale Invariance: Translation invariants may be converted to rotation and scale invariants by means of a *log-polar mapping*.

Consider a point $(x,y) \in \Re^2$ and define:

$$x = e^{\mu}\cos\theta \text{ and } y = e^{\mu}\sin\theta \tag{4.19}$$

where $\mu \in \Re$ and $0 \le \theta < 2\pi$. For every point (x,y), there is a point (μ,θ) that uniquely corresponds to it. The new coordinate system has the following properties:

(i) Scaling is converted to a translation.

$$(\rho x, \rho y) \leftrightarrow (\mu + \log\rho, \theta) \tag{4.20}$$

(ii) Rotation is converted to a translation.

$$(x\cos(\theta+\delta) - y\sin(\theta+\delta), x\sin(\theta+\delta) + y\cos(\theta+\delta)) \leftrightarrow (\mu, \theta+\delta) \tag{4.21}$$

Therefore it is possible to implement a rotation and scale invariant transform by applying a translation invariant transform in the log-polar co-ordinate. Taking the

Fourier transform of a log-polar map is equivalent to computing the Fourier-Mellin transform:

$$F_M(k_1,k_2) = \int_{-\infty}^{\infty}\int_{0}^{2\pi} f(e^{\mu}\cos\theta, e^{\mu}\sin\theta)e^{[i(k_1\mu+k_2\theta)]}\,d\mu d\theta \tag{4.22}$$

The modulus of the Fourier-Mellin transform is rotation and scale invariant.

4.5 Video Watermarking Using Spread Spectrum Principle

Video watermarking uses the same method as image watermarking. In most cases, the video is stored as compressed MPEG stream. The watermark could be inserted after partial decoding of the MPEG stream. The watermark needs to be added to only the I pictures. During MPEG decoding it gets replicated in the P and B pictures as well.

MPEG-1 Compression Standard: The MPEG video sequence is divided into units of Group–of-Pictures (GOP's), consisting of an Intra (I) picture, and an arrangement of Predictive (P) pictures, and Bidirectional (B) pictures. The I picture is coded without reference to other pictures, the P picture is coded with reference to previous I or P pictures and the B picture is coded with reference to an immediate previous I or P picture as well as an immediate future P or I picture. The I picture serves as the natural entry point to facilitate random seek or channel switching. A GOP is defined by its length, N, the distance between I pictures and M, the distance between P pictures. Fig. 13 shows a GOP structure with $M = 3$ and $N = 9$.

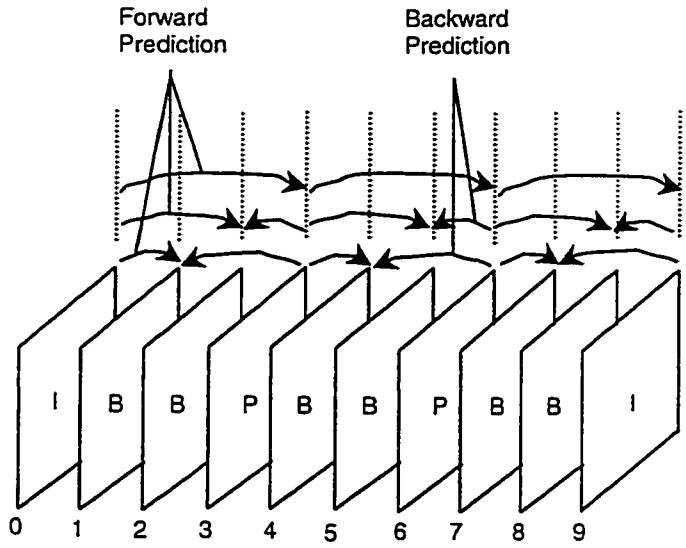


Figure 13: An MPEG Group of Picture (GOP) structure.

Each picture within a GOP is composed of one or more 'slices', the basic unit at which resynchronization information is associated. Since loss of synchronization in I or P pictures can have a propagative effect throughout the GOP, it is common to have several slices in these pictures but only one slice in B pictures, in order to minimize resynchronization overhead. Typically in I or P pictures, a slice consists of a row of 'macroblocks'. A macroblock, in turn, consists of a 16×16 block (or alternately 4, 8×8 blocks) of luminance along with each of 8×8 blocks of Cb and Cr chrominance blocks.

In terms of compression, I pictures are the least efficient as they exploit only spatial redundancies. P pictures are more efficient as they exploit causal-temporal as well as spatial redundancies. B pictures are the most efficient as they also exploit non-causal temporal redundancies in addition to those that P pictures exploit.

5. INFORMATION THEORETIC MODEL OF WATERMARKING

5.1 Analysis Of Watermarking For Copyright Protection

A general model of a watermarking system used for copyright protection, and its analysis is given in [24]. Fig. 14 shows this model. The purpose of this watermarking system is to provide a means to enforce ownership of multimedia information. This is done by introducing imperceptible alterations into a source output in a secret fashion in such a way that those alterations encode copyright information and using a properly designed verification test, the ownership can be verified. It is assumed that the original source output is not available for the verification test.

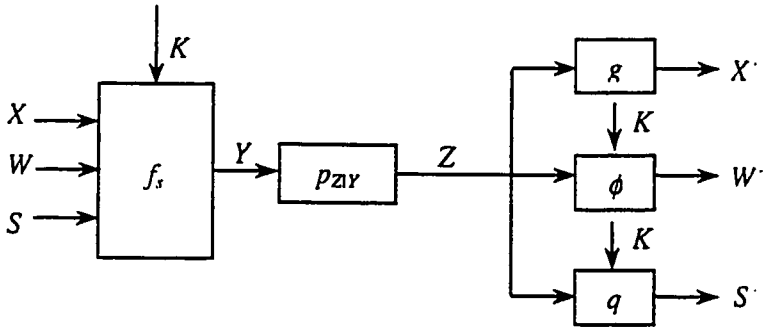


Figure 14: General model of a watermarking system.

Let X be the host data with discrete alphabet $\mathcal{X}$ and W be the watermark in a discrete alphabet $\mathcal{W}$. Let $S \in \{0,1\}$ be a random variable which indicates whether X will be watermarked. Let K be a secret key on a discrete alphabet $\mathcal{K}$, known only to the owner of the host data. When $S=1$, the host data X is transformed into a

watermarked version Y with alphabet $\mathcal{Y}$, using a watermarking function f_1: $\mathcal{X}\times\mathcal{W}\times\mathcal{K} \rightarrow \mathcal{Y}$. When $S = 0$, the host data X is transformed using a function $f_0 : \mathcal{X} \rightarrow \mathcal{Y}$.

The watermarked data Y is transmitted through a noisy channel and gets transformed into $Z \in \mathcal{Y}$. The channel models the unintentional distortions and attacks suffered by Y. This noisy channel can be defined by the distribution $p_{Z|Y}(z \mid y)$ which is independent of K.

Finally, Z is processed to obtain a point $X' \in \mathcal{X}$ which is used by the recipient instead of X. It is assumed that the original host data is not available to the recipient. The function q: $\mathcal{Y}\times\mathcal{K} \rightarrow \{0,1\}$ models the *watermark detection test* used to obtain an estimate S' of S i.e., to decide whether Z has been watermarked using K. The function ϕ: $\mathcal{Y}\times\mathcal{K} \rightarrow \mathcal{W}$, models the *watermark decoding test* and obtains an estimate W' of the hidden watermark W.

5.1.1 Imperceptibility

The watermarking system must guarantee that the functions f_0, f_1 and g introduce imperceptible alterations to X. This is expressed as the mean distortion constraints,

$$E[d(X, g(f_0(X)))] \leq D_0 \tag{5.1}$$

$$E[d(X, g(f_1(X, W, K)))] \leq D_1 \tag{5.2}$$

5.1.2 Decoding Error Probability

The probability of decoding error, which is due to the uncertainty about the host data X is defined [24] as,

$$P_e \triangleq \Pr\{W' \neq W\} = \sum_K p_k(K) \Pr\{\phi(Z,K) \neq W \mid K\} \tag{5.3}$$

5.1.3 False Alarm and Detection Probabilities

Let the critical region for the watermark detection test performed with K, i.e. the set of points in Y where $S' = 1$ is decided for that key be:

$$\varepsilon_K \triangleq \{Z \in Y \mid q(Z,K) = 1\} \tag{5.4}$$

The watermark detection test is completely defined by the sets $\{\varepsilon_K, K \in \mathcal{K}\}$. The probability of false alarm, P_F and the probability of detection, P_D are defined [24] as,

$$P_F \triangleq \Pr\{S' = 1 \mid S = 0\} = \sum_K p_k(K) \Pr\{Z \in \varepsilon_K \mid S = 0\} \tag{5.5}$$

$$P_D \triangleq \Pr\{S' = 1 \mid S = 1\} = \sum_K p_k(K) \Pr\{Z \in \varepsilon_K \mid S = 1, K\} \tag{5.6}$$

The performance of the watermark detection test could be optimized by increasing the distance between distributions $p(Y \mid S = 1, K)$ and $p(Y \mid S = 0)$. But the maximum achievable distance is limited by the perceptual distortion constraint and by the entropy of the host data.

5.1.4 Attacks

Assuming that the attacker has unlimited computational power and that the algorithms for watermarking, detection and decoding are public, the security of the watermarking system relies exclusively on the secret key K of the copyright owner.

Impersonation attack: In authentication systems, an attacker tries to fake an owner by generating the point Z in Y which maximizes his chances of deciding $S' = 1$ for the secret key of the true owner. The probability of success in this attack is,

$$P_I \triangleq \max_Z \sum_K p(K) \cdot q(Z,K) \tag{5.7}$$

Elimination attack: In fingerprinting, an attacker wants to delete the watermark information. That is, an elimination attack is to alter a watermarked data output Y to obtain a negative result ($S' = 0$) in the watermark detection test for the secret key used by the true owner. But the alteration made by the attacker should be imperceptible, which imposes the constraint $E[d(Z,Y)] \leq D_E$, where D_E is a distortion threshold. Therefore the maximum probability of success for a given Y can be expressed as,

$$P_E(Y) \triangleq \max_{Z:d(Z,Y)\leq D_E} \sum_K p(K \mid Y)(1-q(Z,K)) \tag{5.8}$$

After averaging out over Y, the average probability of success in the elimination attack is,

$$P_E = \sum_Y p(Y) \max_{Z:d(Z,Y)\leq D_E} \sum_K p(K \mid Y)(1-q(Z,K)) \tag{5.9}$$

The security level of the watermarking system can be measured by the uncertainty about the key, given a watermarked data Y, which is the conditional entropy $H(K \mid Y)$, called the key equivocation. A large equivocation results in greater uncertainty to the attacker about the key and if the watermarking function and the detection test are properly designed, this uncertainty can be used to reduce the probability of success in an elimination attack.

5.2 Analysis Of Watermarking For Fingerprinting

Fig.14 is modified and redrawn as Fig.15 to adapt it for fingerprinting application. The notations used in Section 4.3 for the spread spectrum watermarking method is adopted here. Let V be the host data with discrete alphabet $\{\mathcal{V}\}$ and W_i be the watermark (information sequence) in a discrete alphabet $\{\mathcal{W}\}$ where $i=1,2,\ldots,N_u$, when there are N_u recipients of the data. In fingerprinting, a different watermark sequence is used for each recipient. Let K be a secret key on a discrete alphabet $\mathcal{K}$, (say, the pseudo-noise sequence $\{p_i\}$ in our method). The host data V is transformed into a watermarked version M with alphabet $\mathcal{M}$, using a watermarking function $f(\cdot): \mathcal{V}+\mathcal{W}\times\mathcal{K}\rightarrow\mathcal{M}$.

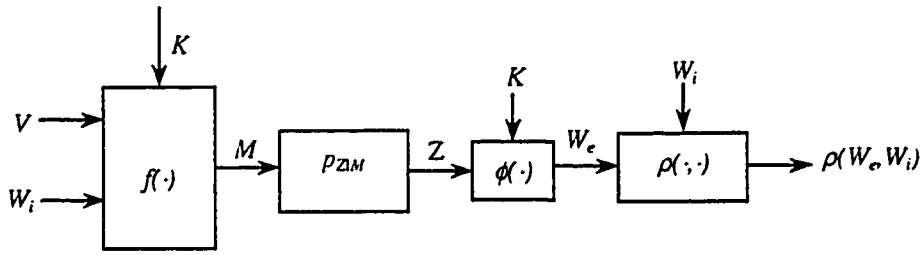


Figure 15: Model of a watermarking system for fingerprinting.

The watermarked data M is transmitted through a noisy channel and gets transformed into $Z\in\mathcal{M}$. The channel models the unintentional distortions and attacks suffered by M. This noisy channel can be defined by the distribution $p_{Z|M}(z \mid m)$ which is independent of K.

Finally, the distorted data Z, which may be an illegal copy is checked to trace the recipient whose copy has been used to obtain Z. This could be done without using the original host data. The function $\phi(\cdot)$: $\mathcal{M} \times \mathcal{K} \rightarrow \mathcal{W}$, models the *watermark extraction* and obtains an estimate W_e of the hidden watermark W. The function $\rho(\cdot,\cdot)$ models the correlation detection which obtains the correlation coefficient, $\rho(W_e, W_i)$ by comparing W_e with W_i, which are the possible information sequences used.

5.2.1 Imperceptibility

The watermarking system must guarantee that the function $f(\cdot)$ introduces imperceptible alterations to V. This is expressed as the mean distortion constraint,

$$E[d(V, (f(V, W_i, K))] \leq T_{JND} \tag{5.10}$$

where T_{JND} is the just noticeable difference threshold, which is a measure of the capacity of a pixel to hold a watermark bit without it being noticeable.

5.2.2 Robustness and Unambiguousness

The watermark is robust if an attacker cannot eliminate it. In fingerprinting, an attacker in most cases is a recipient who wants to make illegal copies.

In the spread spectrum watermarking method, the watermarked image is represented as $m_i = v_i + \alpha \cdot b_i \cdot p_i$ as per equation. (4.8). The embedded watermark W_i stands for the information sequence $\{a_j\}$ in section 4.3.1 or $W_i = \{a_1, a_2, a_3, \ldots\ldots, a_{N_{info}}\}$. The extracted watermark, W_e could be represented as a_{je} where

$a_{je} = sign\left\{\sum_{i=j\cdot r_c}^{(j+1)\cdot r_c} m_i \cdot p_i\right\}$, where $j = 1, 2, 3, \ldots, N_{info}$ and $i = 1, 2, 3, \ldots, N$. N_{info} is the number of information bits and N is the total number of pixels in the image. i. e.,

$$a_{je} = sign\left\{\sum_{i=j\cdot r_c}^{(j+1)\cdot r_c} (v_i + \alpha \cdot b_i \cdot p_i) p_i\right\} \tag{5.11}$$

The normalized correlation coefficient given in equation. (4.13) could be rewritten as:

$$\rho = \frac{\sum_{j=1}^{N_{info}} a_j \cdot a_{je}}{N_{info}} \tag{5.12}$$

If $\rho > T_{det}$, the information sequence $\{a_j\}$ is detected and if $\rho \le T_{det}$, the information sequence $\{a_j\}$ is missed. Here T_{det} is called the detection threshold. Therefore the probability of missed detection is given by,

$$P_{md} = \Pr\{\rho \le T_{\det}\} \tag{5.13}$$

$$= \Pr\left\{\sum_{j=1}^{N_{info}} a_j \cdot sign\left\{\sum_{i=j\cdot r_c}^{(j+1)\cdot r_c} v_i \cdot p_i + \alpha \cdot b_i \cdot p_i^2\right\} \le T_{\det} \cdot N_{info}\right\} \tag{5.14}$$

$$= \Pr\left\{\sum_{j=1}^{N_{info}} a_j \cdot sign\left\{\sum_{i=j\cdot r_c}^{(j+1)\cdot r_c} v_i \cdot p_i + \alpha \cdot a_j\right\} \le T_{\det} \cdot N_{info}\right\} \tag{5.15}$$

For a watermark to be unambiguous, an attacker should not be able to frame another recipient. In other words, the probability of false alarm is the probability that a sequence $\{a_j\}$ is detected falsely in a document in which it was not embedded and is given by,

$$P_{fa} = \Pr\left\{\sum_{j=1}^{N_{info}} a_j \cdot sign\left\{\sum_{i=j\cdot r_c}^{(j+1)\cdot r_c} v_i \cdot p_i\right\} > T_{\det} \cdot N_{info}\right\} \tag{5.16}$$

5.3 Information Theoretic Bounds For The Capacity Of The Data Hiding Channel

A block diagram of the data-hiding channel is shown in Fig. 16. S is the message to be transmitted through the channel, which has two sources of noise: I, the noise due to the original host image, and P, the noise due to processing. $\check{S}$ is the "corrupted" message. Note that the receiver does not have access to the cover image [46].

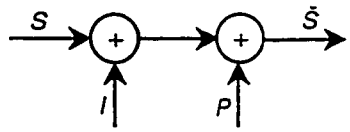


Figure 16: The data hiding channel.

5.3.1 Capacity of the Spatial Domain Data Hiding Channel

Fig. 17 gives a simple model for an additive noise channel. Here, $X \sim [f_X(x), \sigma_x^2]$ is the message to be transmitted, $Z \sim [f_Z(z), \sigma_z^2]$ is the additive noise in the channel, and $Y \sim [f_Y(y), \sigma_y^2]$ is the received signal at the output of the channel. Assuming X and Z are independent, implies that $\sigma_y^2 = \sigma_x^2 + \sigma_z^2$.

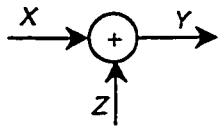


Figure 17: A simple additive noise channel.

The channel capacity is given by,

(5.13)

$$C = \max_{f_X(x)} I(\mathbf{X};\mathbf{Y}) = \max_{f_X(x)} \left(h(\mathbf{Y}) - h(\mathbf{Y} \mid \mathbf{X})\right) = \max_{f_X(x)} \left(h(\mathbf{Y}) - h(\mathbf{Z})\right)$$

where *I(X;Y)*, is the *mutual information* between *X* and *Y*. For a given variance σ_y^2, the maximum differential entropy value of $h(Y)=\frac{1}{2}\log_2(2\pi e\sigma_y^2)$ bits is achieved when *Y* has a normal distribution. Therefore, the maximum value of the mutual information between *X* and *Y* is achievable if both pdfs (probability density functions) $f_Z(z)$ and $f_X(x)$ are normally distributed.

To calculate the maximum achievable entropy, the noise *Z* is passed through an ideal *information processor*, (see Fig. 18) which does not alter the amount of information in *Z*, but changes its statistics to a Gaussian distribution for its output Z_g. Since the output of the information processor has the same entropy as the input, the variance of the output, σ_{zg}^2, can be obtained by solving the equation,

$$h(\mathbf{Z}_g) = h(\mathbf{Z}) = \frac{1}{2}\log_2(2\pi e\sigma_{zg}^2) \text{ bits} \tag{5.14}$$

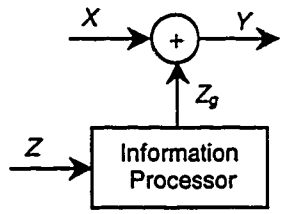


Figure 18: Additive noise channel modified to obtain equivalent gaussian noise.

It is known that the Gaussian distribution has the highest entropy for a given variance. Alternately, the Gaussian distribution has the least variance for a given

entropy. Thus it is always true that $\sigma_{zg}^2 \leq \sigma_z^2$. σ_{zg}^2 could be called the *entropy equivalent Gaussian variance*. The maximum value of $h(Y)$ is therefore obtained as

$$\max_{f_X(x)} h(\mathbf{Y}) = \max_{f_X(x)} h(\mathbf{X}+\mathbf{Z}_g) = \frac{1}{2}\log_2(2\pi e(\sigma_{zg}^2 + \sigma_x^2)) \text{ bits.} \tag{5.15}$$

In order to calculate the channel capacity, $f_Z(z)$ can be replaced by $N[0, \sigma_{zg}^2]$.

$$C = \max_{f_X(x)} h(\mathbf{Y}) - h(\mathbf{Z}_g) = \frac{1}{2}\log_2\left(1+\frac{\sigma_x^2}{\sigma_{zg}^2}\right) \text{bits.} \tag{5.16}$$

The two channel noises given in Fig. 16 can be replaced by a single Gaussian noise source with the combined variance of $\sigma_{ig}^2 + \sigma_p^2$, where σ_{ig}^2 is the equivalent Gaussian variance for the image noise I, and σ_p^2 is the variance of the processing noise. If σ_s^2 is the message signal energy, the capacity of the data-hiding channel can be expressed as,

$$C_h = \frac{1}{2}\log_2\left(1+\frac{\sigma_s^2}{\sigma_{ig}^2+\sigma_p^2}\right) \text{bits.} \tag{5.17}$$

The image noise I is due to the original image pixels, which are assumed to be uniformly distributed random variables u taking values between 0 and 255 with variance σ_i^2. If MN is the number of pixels in an image, then the energy (or variance) of the zero-mean message signal is calculated as,

$$\sigma_s^2 = \frac{\sum_{i=1}^{MN} S_i^2}{MN} \tag{5.18}$$

where, S_i is the message signal added to the i^{th} pixel. The entropies, $h(g)$ of a Gaussian random variable g, with variance σ_g^2 and that of a uniform random variable u with variance σ_g^2 are expressed as ,

$$h(g) = \frac{1}{2}\log_2(2\pi e\sigma_g^2) \tag{5.19}$$

$$h(u) = \frac{1}{2}\log_2(12\sigma_u^2) \tag{5.20}$$

From equation. (5.20), *the entropy equivalent Gaussian noise* (or the Gaussian random variable that has the same entropy as the uniform random variable u of variance σ_i^2) has a variance given by,

$$\sigma_{ig}^2 = \frac{12}{2\pi e}\sigma_i^2 \tag{5.21}$$

Statistics from many test images show that $\sigma_i = 55$. Using equation. (5.21) σ_{ig} = 46. Assuming the processing noise is due to JPEG compression at 50% quality, then it is measured for test images that the processing noise has a standard deviation $\sigma_p \approx 6.7$. This would yield a capacity value, C_h of 0.0022 bits/pixel (140 bits for a 256 × 256 image). Fig. 19 shows a plot of the channel capacity, obtained using equation. (5.17), for spatial domain message embedding for different values of PSNR assuming there is no processing noise (i.e. when $\sigma_p = 0$). The Signal-to-Noise Ratio (SNR) is given by:

$$SNR = 10\log_{10}\left(\frac{\sigma_i^2}{\sigma_s^2}\right) \tag{5.22}$$

The Peak Signal-to-Noise Ratio (PSNR) is defined for the maximum value of σ_i = 255. Therefore,

$$PSNR = 10\log_{10}\left(\frac{255^2}{\sigma_s^2}\right) \tag{5.23}$$

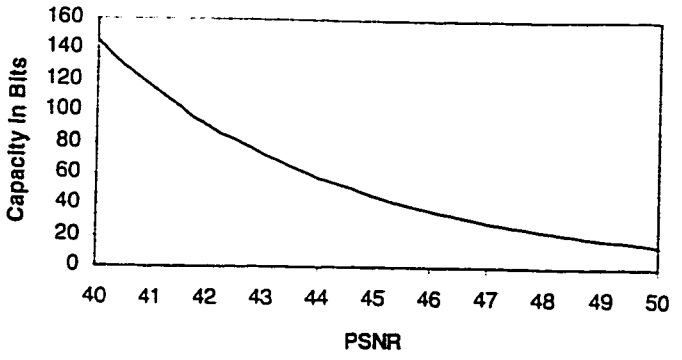


Figure 19: Channel capacity for spatial domain message embedding.

5.3.2 Capacity of the Spectral Domain Data Hiding Channel

In spectral domain data hiding (see Fig. 20) the image is decomposed into L sub-bands using a suitable transform, say DCT. This is equivalent to having L parallel sub-channels with two noise sources in each sub-channel. Let σ_{ij}^2, $j = 1,2,.....,L$, be the variances of the image noise and σ_{igj}^2 be the equivalent Gaussian variances for each sub-band. If σ_{pj}^2 is the Gaussian variance of the processing noise in the j^{th} sub-channel, then the total capacity of the L parallel sub-channels is given by,

$$C_h = \frac{MN}{2L} \sum_{j=1}^{L} \log_2 \left(1 + \frac{v_j^2}{\sigma_{igj}^2 + \sigma_{pj}^2} \right) \text{bits} \qquad (5.24)$$

for an image of MN pixels. In equation (5.24), v_j is the *visual threshold* of band j. In other words, v_j^2 is the maximum message signal energy permitted in band j based on its perceptual quality effects.

The values of σ_{igj} for a test image are as given in Table 2 giving an average value of 19. This is low compared to the variance in the spatial domain, which has a value of 55. In other words. a transform with good energy compaction or higher

Transform Coding Gain (GTC) would result in more imbalance of the coefficient variances and therefore increase the capacity. Fig. 21 shows the channel bit capacity for spectral domain watermarking using equation. (5. 24) assuming there is no processing noise.

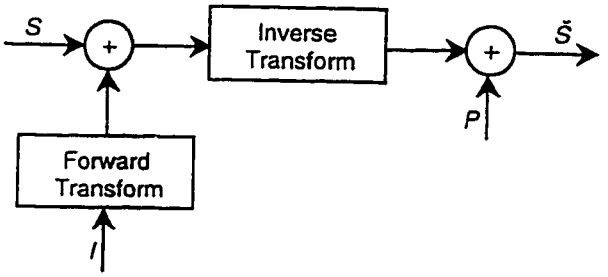


Figure 20: The spectral domain data hiding channel.

300	71	26	14	8	6	3	2
101	51	27	14	8	5	3	2
72	39	27	14	8	5	3	2
44	32	22	14	9	5	3	2
30	23	16	12	8	5	3	2
21	17	13	8	7	4	3	2
13	10	9	6	4	3	3	1
9	7	6	4	3	2	2	1

Table 2: Image variances, σ_{ij} for the 64 DCT sub-bands.

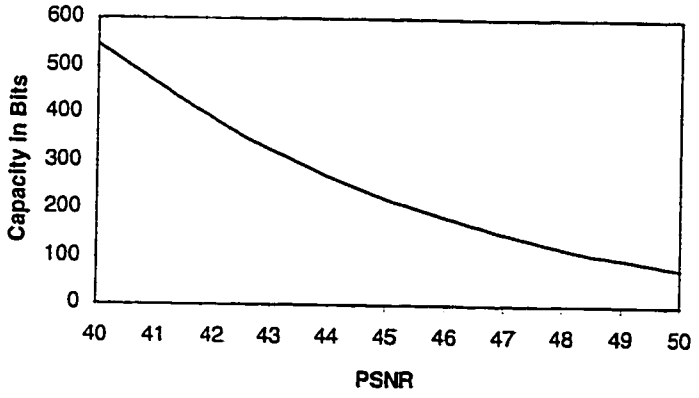


Figure 21: Channel capacity for spectral domain message embedding.

5.4 Conclusion

An information theoretic analysis of digital watermarking has been given in this chapter. Section 5.1 covers the analysis of watermarking for copyright protection given by Hernandez *et al.* In section 5.2, we have extended this analysis to spread spectrum watermarking for fingerprinting applications based on our model given in section 4.3. For copyright protection, the watermark has to be always extracted at the receiver side, without the knowledge of the original host data. The extracted watermark gives information about the owner of the data as well as it could prove the authenticity of the data. Whereas in fingerprinting, the watermark contains information about the recipient and needs to be extracted only in case of a dispute. Here the extracted watermark is compared with the watermarks of all recipients to find the true recipient of the data. Finally in section 5.3, the channel capacity of spatial and spectral domain watermarking has been plotted for different values of PSNR based on the analysis given by Ramkumar *et al.*

6. EXPERIMENTAL RESULTS ON IMAGE AND VIDEO WATERMARKING

6.1 Introduction

There has been a rapid growth in the use of digital imaging in the last five years. Digital images and video are now widely distributed on the Internet and via CD-ROM or DVD. Digital imaging allows to make unlimited number of copies of an 'original' without any loss in quality and they can be easily distributed. Duplication of digital video signals does not result in the inherent decrease in quality suffered by analog video. This has resulted in an ever increasing demand for copyright protection. Digital watermarking can be used for the authentication of the content or owner and also to identify illegal copies.

This chapter gives the results obtained for spread spectrum watermarking of images and video using *robust* (destination-based) watermarks, used for finger-printing. In fingerprinting applications, a different watermark (receiver ID) is inserted in each copy and it should survive all possible attacks which try to destroy the watermark.

6.2 Watermarking Parameters

Watermark embedding depth or scale factor, α: The scale factor, α scales the watermark amplitude as given in equations (4.8) and (4.9). The visibility of the watermark is controlled by varying α. Increasing α ensures that the watermark is

more robust to attacks and image manipulations. But the value of α should also guarantee that the perceptual constraints are fulfilled.

Chip-rate, r_c: The chip-rate controls the number of bits that could be embedded in a given image. As seen from equation (4.6), each message bit is replicated in a number of pixels equal to the chip-rate. This replication introduces the necessary redundancy to compensate for the low energy that the watermark will have, compared to the energy of the image in which it is embedded. The larger the value of chip-rate, the smaller the number of embedded message bits. But a large value of chip-rate ensures watermark extraction with less error.

Table 3 gives the normalized correlation coefficient values obtained using equation (4.13), for different scale factors and chip-rates for watermark extraction without using the original image. Fig. 22 shows the corresponding plot for scale factors 1, 2 and 3 and chip-rates varying from 2 to 400. It can be noted that the watermark is extracted without any error ($\rho = 1$) when chip-rate, $r_c \geq 400$.

r_c	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
2	0.2727	0.3254	0.399
5	0.3858	0.4847	0.519
10	0.516	0.6335	0.678
20	0.646	0.7638	0.8114
30	0.7235	0.8508	0.881
40	0.7816	0.8926	0.925
50	0.8169	0.9191	0.945
60	0.8334	0.9349	0.9551
70	0.889	0.9637	0.976
80	0.9129	0.9805	0.985
90	0.9202	0.9781	0.991
100	0.9282	0.9787	0.9908
110	0.9396	0.9865	0.9899
200	0.9573	0.9866	0.9878
300	0.9817	0.9909	0.9909
350	0.9893	1	1
400	1	1	1

Table 3: Correlation coefficients for different values of the scale factor and chip-rate.

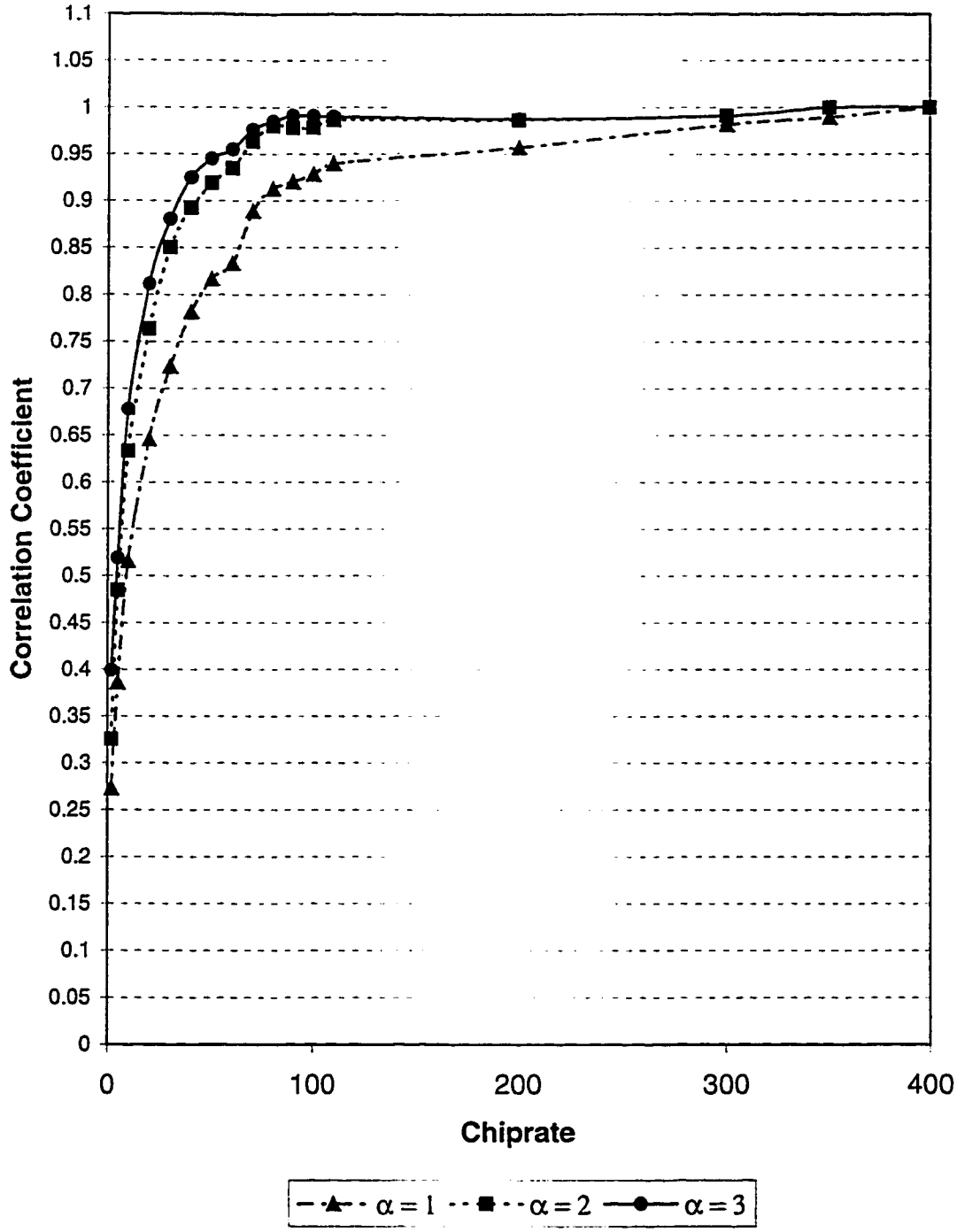


Figure 22: Detector response for watermark extraction without using the original.

6.3 Watermarking Grayscale, YUV And RGB Images

The watermark pattern for images in different format can be expressed in terms of the watermark embedding depth, α. For *gray scale* images, the watermark pattern is of the form:

$$W_{gray}(v_{ij}) = \pm\alpha \tag{6.1}$$

for all pixels v_{ij}. For *YUV images*, the watermark pattern is defined in such a way that it effects the luminance value, y only:

$$W_{yuv}(v_{ij}) = (\pm\alpha,0,0) \tag{6.2}$$

RGB images can be obtained from *YUV* images using the following rule:

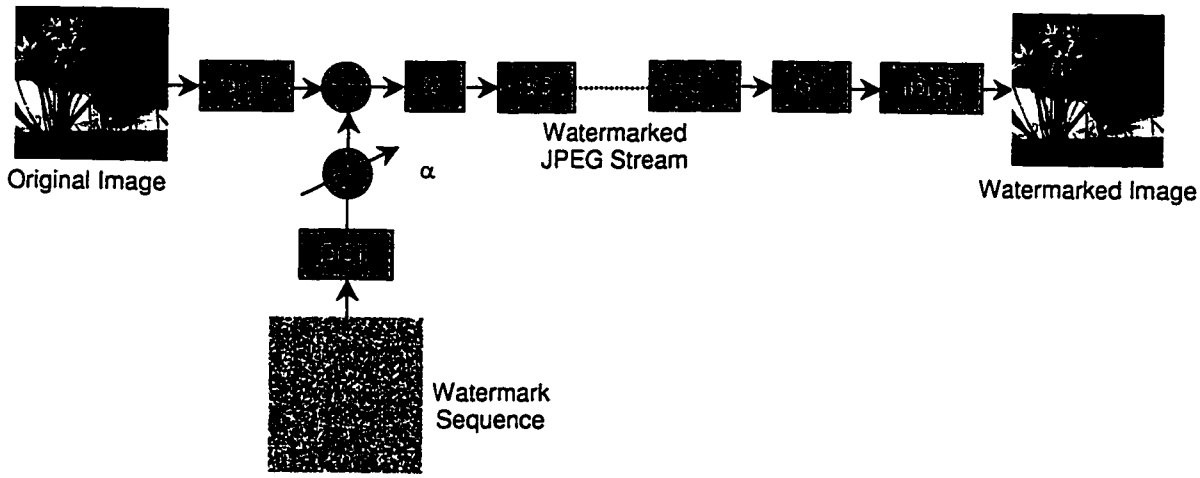
$$(r,g,b) = (y+v, y-\frac{1}{6}u-\frac{1}{2}v, y+u) \tag{6.3}$$

Therefore for *RGB images*, the watermark pattern could be defined as adding the watermark to all the three components, *r*, *g* and *b*:

$$W_{rgb}(v_{ij}) = (\pm\alpha,\pm\alpha,\pm\alpha) \tag{6.4}$$

6.4 Image Watermarking Set-Up

Watermark insertion could be incorporated with the standard JPEG compression scheme. If the image is already compressed, the watermark could be added after partial decoding of the available JPEG stream. Fig. 23 shows watermark insertion incorporated with the standard JPEG coding scheme. This set-up has been used in our work.



DCT - Discrete Cosine Transform
Q - Quantization
EC - Entropy Coding

IDCT - Inverse DCT
Q^{-1} - Inverse Quantization
EC^{-1} - Inverse Entropy Coding

Figure 23: Image watermarking in the spectral domain.

6.5 Video Watermarking Set-Up

In practical systems, the video sequences are stored and transmitted in compressed format. In this case, if different copies have to be watermarked with individual watermarks, decoding, watermarking in the spatial domain and re-encoding is not feasible. A partial decoding of the compressed video i.e., entropy decoding and inverse quantization is enough to watermark the video in spectral domain. Fig. 24 shows the basic set-up used for spectral domain video watermarking, when the original video is available as compressed MPEG stream. This set-up has been used in this thesis.

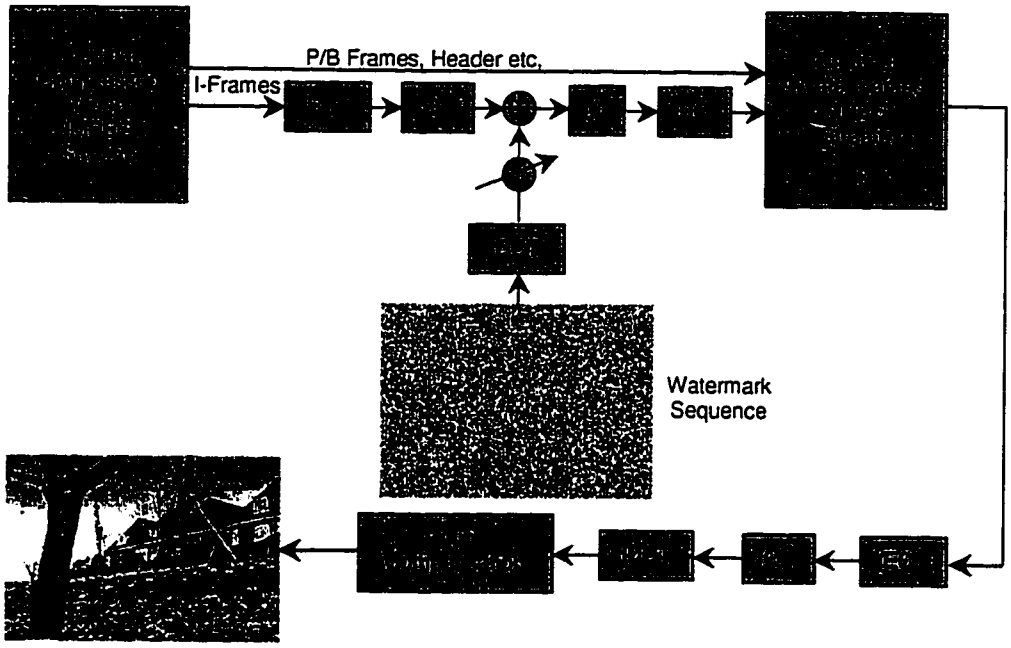


Figure 24: Compressed video watermark insertion in the spectral domain.

The basic steps in video watermarking are:

I. Generate a watermark sequence for each frame of the video sequence in the same manner as for images.

II. Arrange the watermark sequence into a two-dimensional signal having the same dimensions as the video frames.

III. For each block of 8 × 8 pixels coded in the bit stream, the corresponding 8 × 8 block of the watermark sequence is transformed using DCT.

IV. The 8 × 8 block of the I-picture in the video sequence and the 8 × 8 block of the watermark sequence are added in the DCT domain, and the resulting watermarked 8 × 8 block is encoded and written to the new bit stream,

V. All other parts of the compressed bit stream are simply copied to the new bit stream.

6.6 Simulation Results

Simulation results [17, 18] demonstrate the robustness of the technique to all the four types of attacks mentioned in Chapter 3, which are the simple attacks, synchronization attacks, ambiguity attacks and removal attacks. The robustness and invisibility of this watermarking scheme was analyzed for a number of JPEG and MPEG streams.

6.6.1 Imperceptibility

Image Watermarking: Fig. 25(a) shows the original image, which is of size 256 × 256 pixels. The watermarked images with $\alpha = 1$, 3 and 6 and chip-rate $r_c = 500$ are shown in Fig. 25(b), (c) and (d) respectively.

For imperceptibility of the watermark, the Signal-to-Noise Ratio (SNR) as given in equation (6.5) should be below –25 dB [48].

$$SNR = 10\log_{10}\left[\frac{\sigma_W^2}{\sigma_V^2}\right] \text{ dB} \tag{6.5}$$

where σ_W^2 and σ_V^2 are the variances of the watermark and the image sequences respectively. It has been noted that σ_V^2=2827 for the image sequence entitled "summer". The values for σ_W^2, for different values of the scale factor, α and the corresponding SNR values are given in Table 4. It can be seen from Fig. 25, that the watermark is not perceptible when α<3, for which the SNR< -25dB. Though the

watermark is not visible when $\alpha<3$, the difference in the pixel values in the watermarked images can be noted from the histogram as shown in Fig. 26.

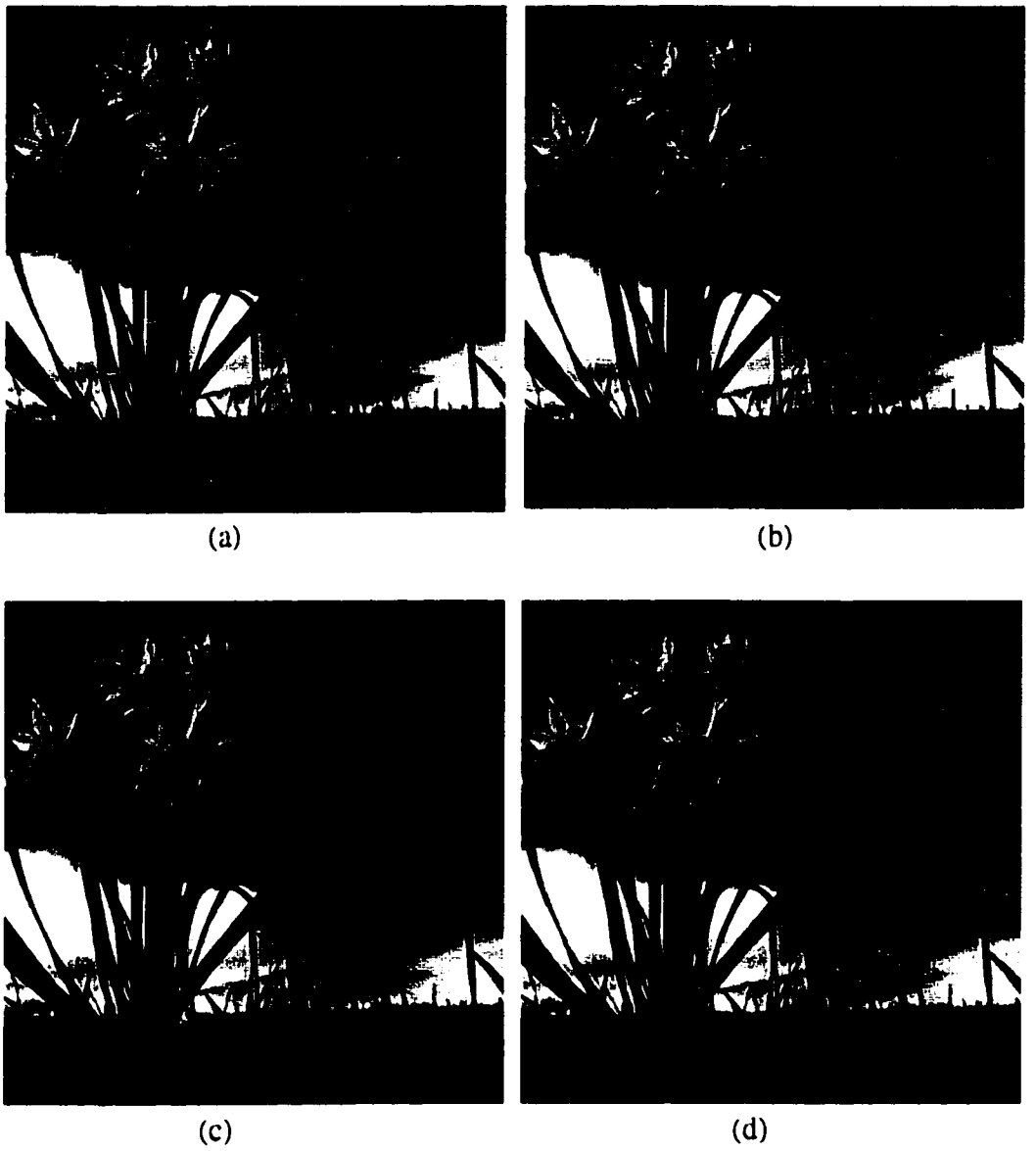
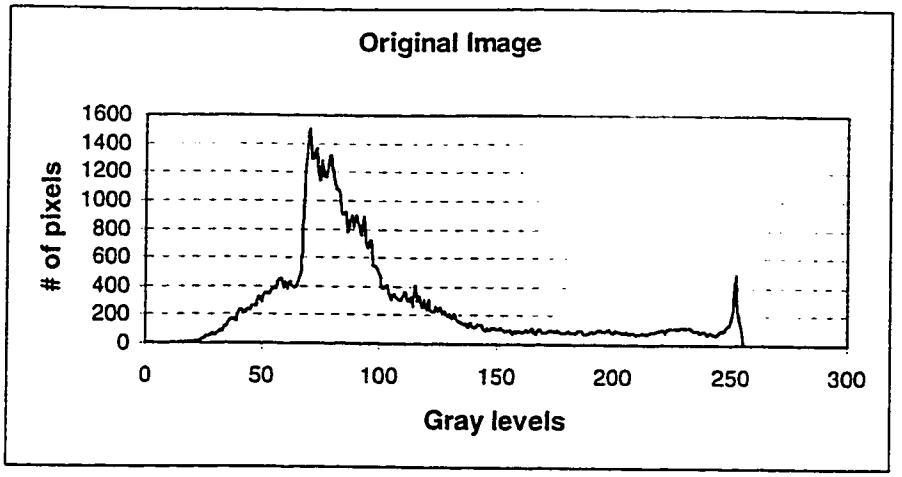
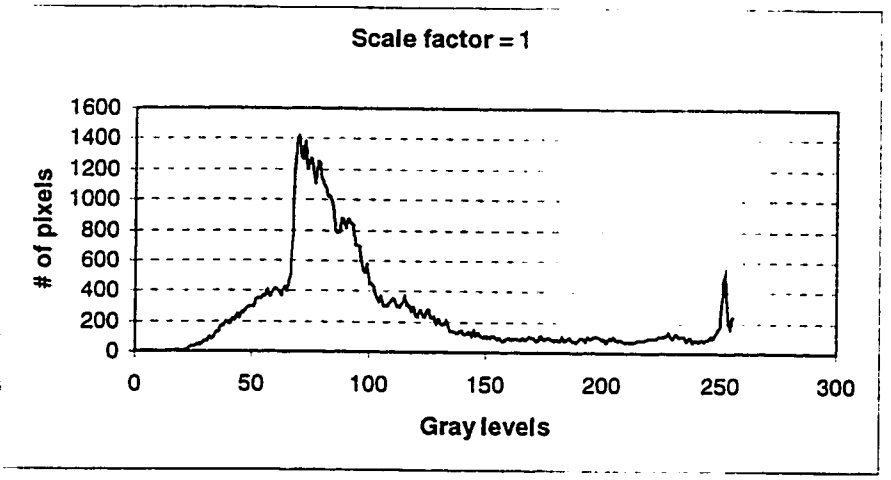


Figure 25: Original and watermarked images.

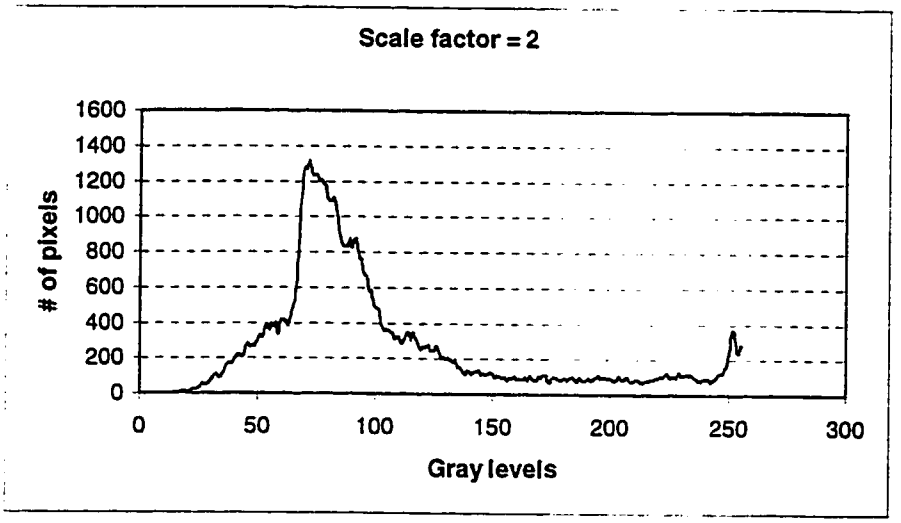
(a) Original image entitled "summer". (b) Watermarked image with $\alpha = 1$.
(c) Watermarked image with $\alpha = 3$. (d) Watermarked image with $\alpha = 6$.



(a)



(b)



(c)

Figure 26: Histogram of the original and watermarked images.

(a): Histogram of the original image entitled "summer". (b): Histogram of the watermarked image with $\alpha = 1$. (c): Histogram of the watermarked image with $\alpha = 2$.

σ_w^2	α	SNR(dB)
1	1	-34.51
2	1.414	-31.51
3	1.732	-29.74
4	2	-28.49
5	2.236	-27.52
6	2.449	-26.73
7	2.645	-26.06
8	2.828	-25.48
9	3	-24.97

Table 4: SNR in dB for different values of the scale factor, α.

For $\alpha = 1$, the watermark is not perceptible, at the same time it could be extracted without any error ($\rho = 1$) using the original. Extraction without using the original gave $\rho = 0.985$ which is well above the detection threshold $T_{det} = 0.3$. Thus the watermark is still detectable. In Fig. 25 (d) the watermark is more visible as the embedding depth, α is increased to 6.

Video Watermarking: The watermark is added to the luminance value of I-frames. Watermark detection for video gave similar results as for still images. Fig. 27(a) shows the original MPEG frames, which are of size 352 × 288 pixels. The watermarked frames with $\alpha = 1$, 2 and 10 with chip-rate, $r_c = 500$ are shown in Fig. 27(b), (c) and (d) respectively. The histograms of the MPEG frames coded as I picture and B picture are given in Figs. 28 and 29 respectively. Fig. 29 shows that although the watermark is added to only the I pictures, it get replicated in the pictures coded as P or B during decoding.

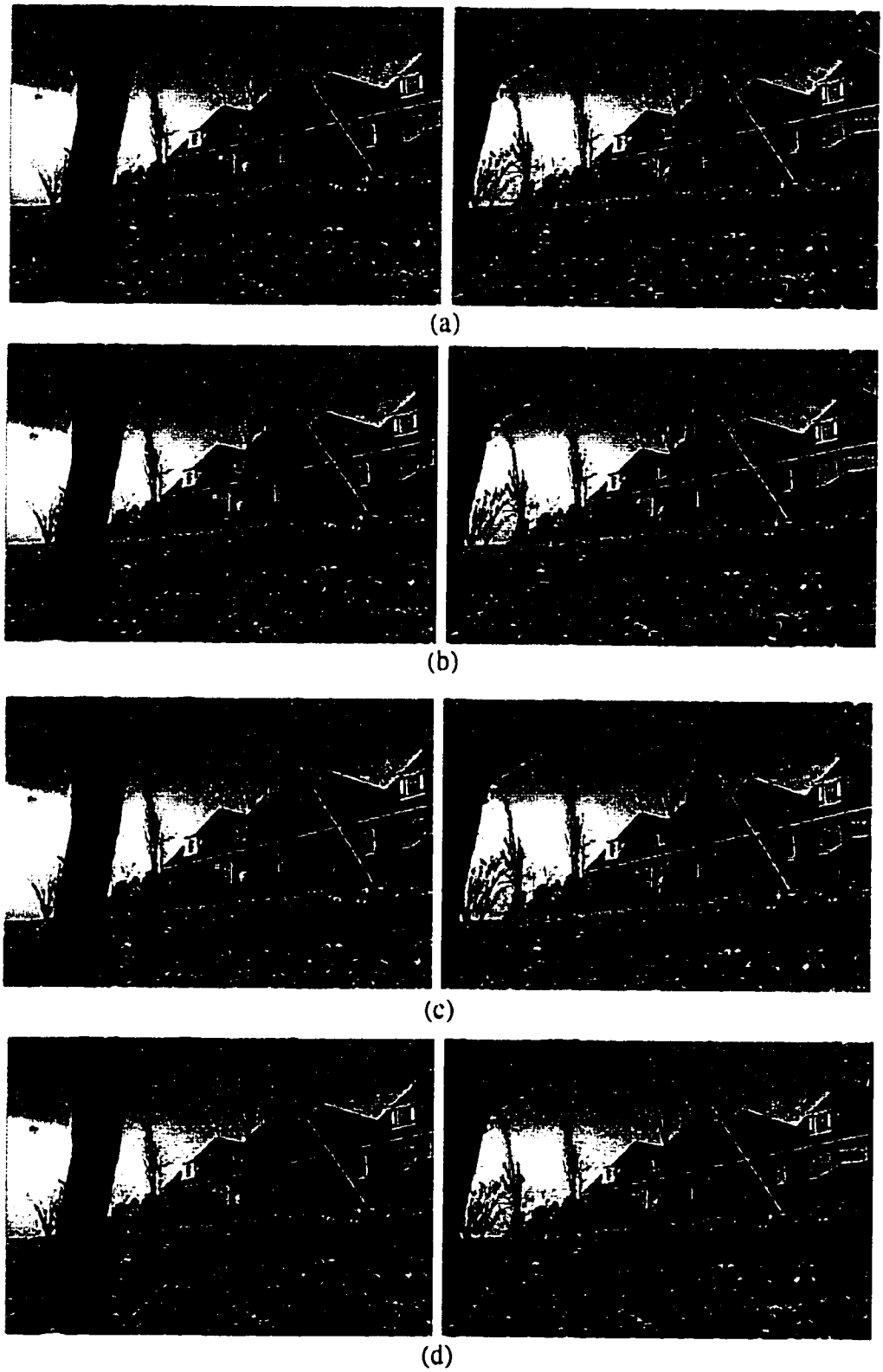
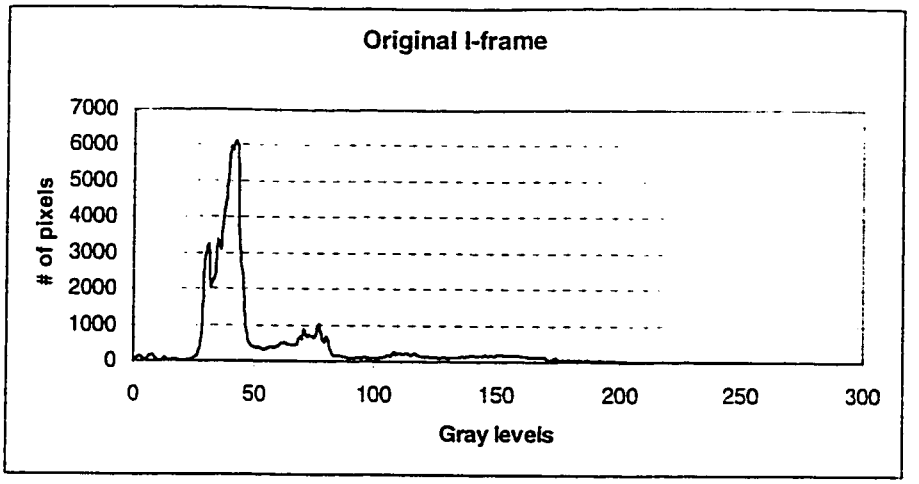
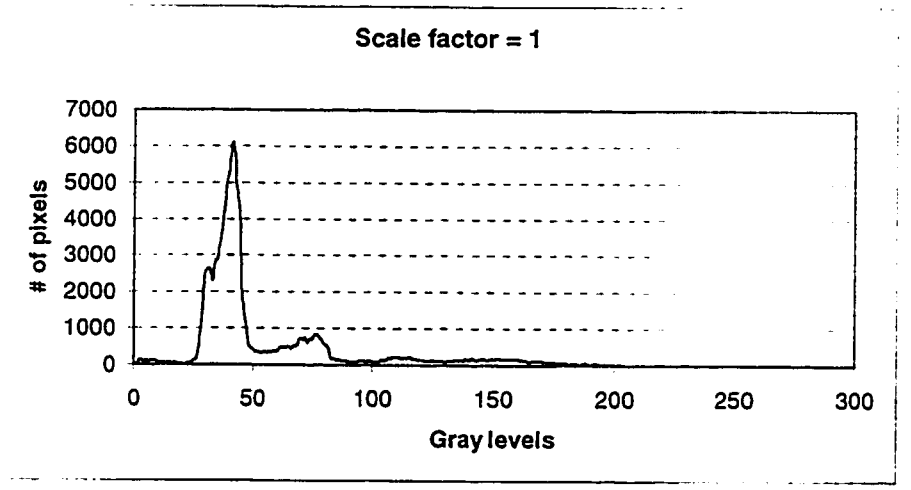


Figure 27: Original and watermarked MPEG frames.

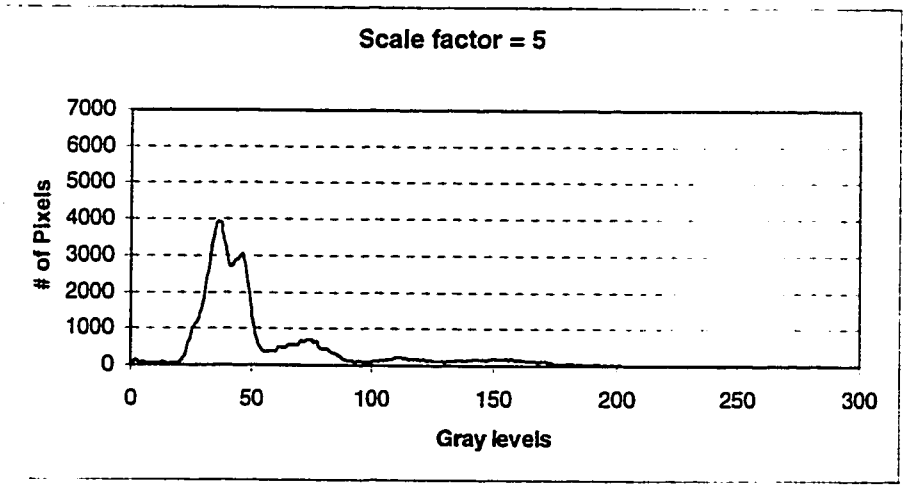
(a): Original MPEG frames entitled "flower". (b): Watermarked frames with $\alpha = 1$. (c): Watermarked frames with $\alpha = 2$. (d): Watermarked frames with $\alpha = 10$.



(a)



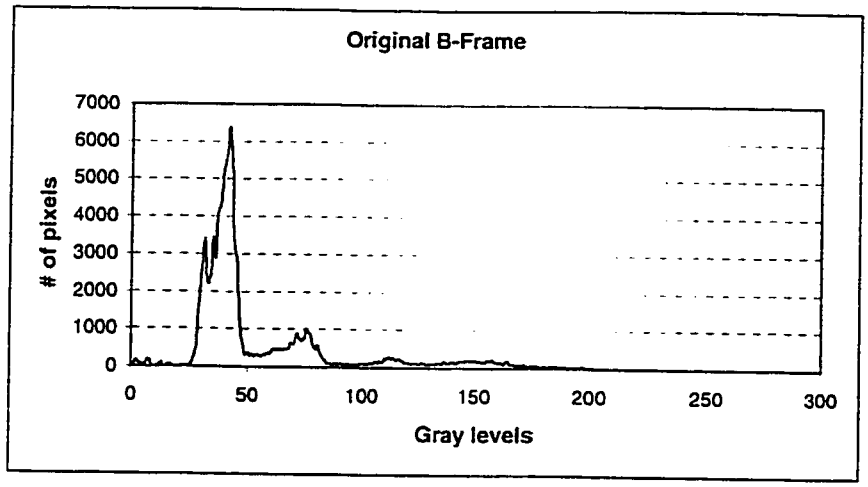
(b)



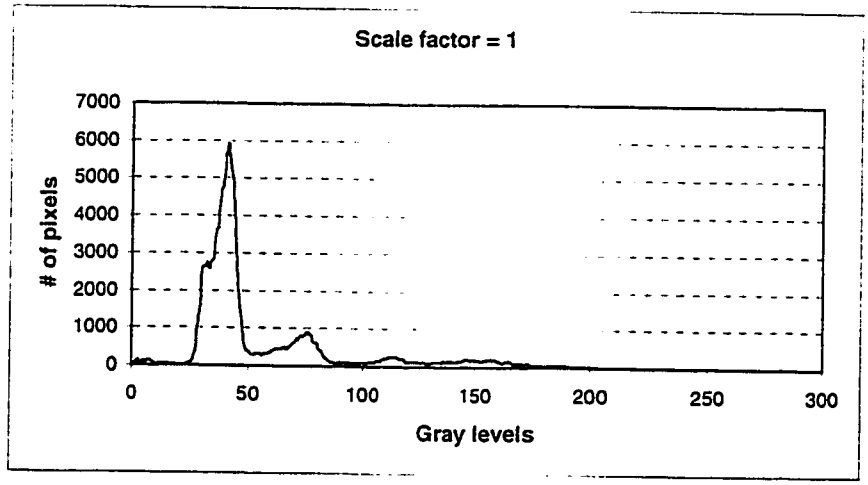
(c)

Figure 28: Histogram of the original and watermarked MPEG I-frames.

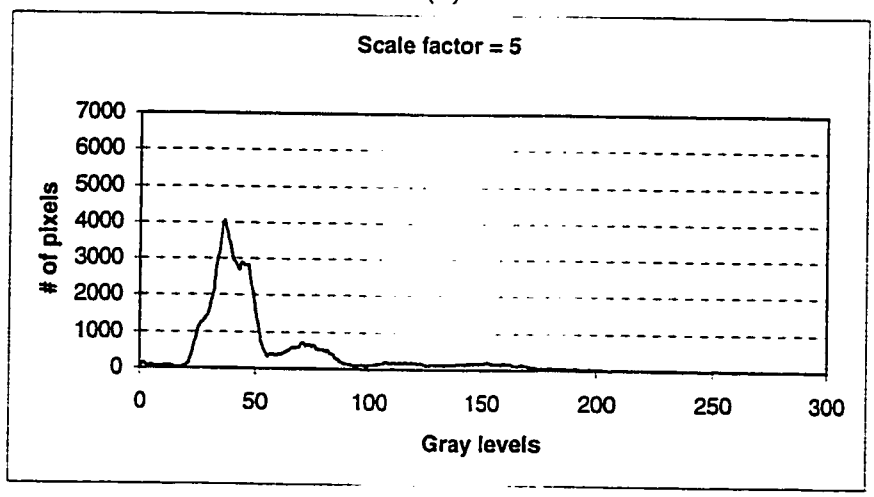
Histogram of (a): original I-frame entitled "flower". (b): watermarked I-frame with $\alpha = 1$. (c): watermarked I-frame with $\alpha = 5$.



(a)



(b)



(c)

Figure 29: Histogram of the original and watermarked MPEG B-frames. Histogram of (a): original B-frame entitled "flower". (b): watermarked B-frame with $\alpha = 1$. (c): watermarked B-frame with $\alpha = 5$.

6.6.2 Robustness and Unambiguousness

6.6.2.1 Uniqueness Of Watermark

Fig. 30 shows the uniqueness of the watermark extracted from an image watermarked with a key (seed) equal to 200 and for chip-rate, r_c = 500. The normalized correlation coefficient, ρ is equal to unity for only one watermark which was generated with the key, 200. The value of ρ is in the range (-0.26,0.26) for all the random watermarks. Therefore the threshold T_{det}, for detection could be chosen as 0.26 for chip rate, r_c = 500.

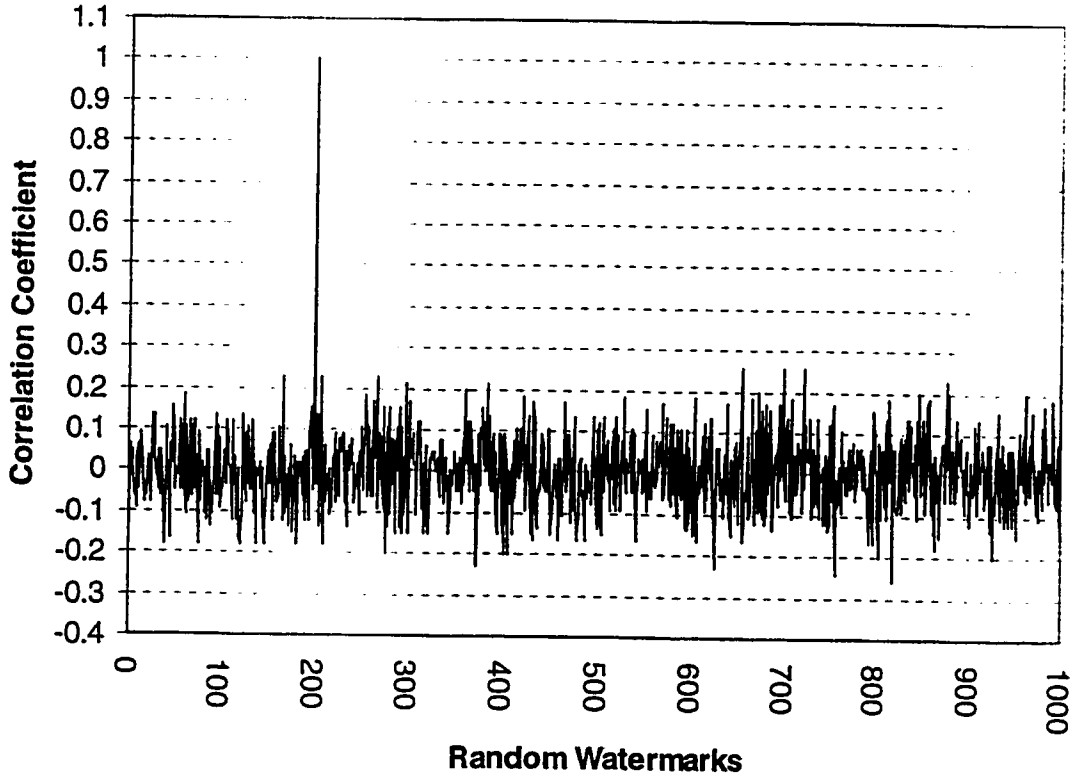


Figure 30: Detector response for 1000 random watermarks for chip rate = 500.

The correlation coefficient, ρ, could be considered as a measure of robustness. If the watermark is robust, $T_{det} < \rho \leq 1$, after all possible attacks on the watermarked image. Fig. 31 shows a similar plot, but for which the chip rate, $r_c = 100$. T_{det} could be chosen as 0.1 in this case.

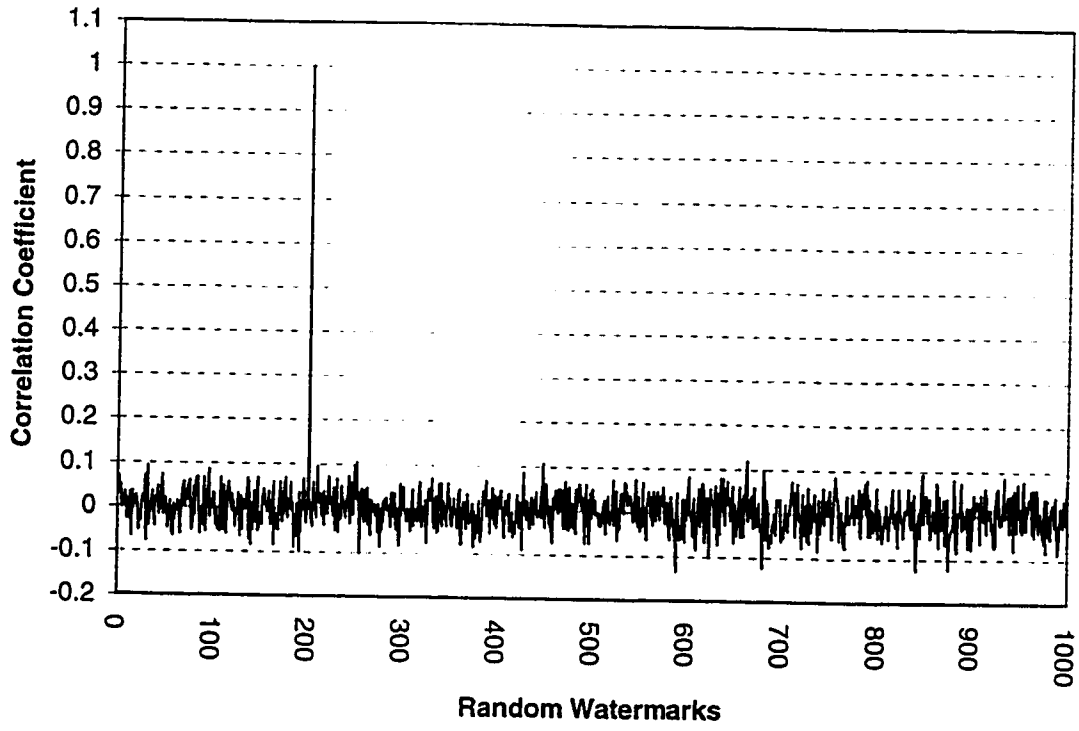


Figure 31: Detector response for 1000 random watermarks for chip rate = 100.

For lower values of the chip-rate, the number of information bits, N_{info} would be larger. As per the weak law of large numbers (equation 6.6), the correlation coefficient obtained using equation 4.13, now approaches zero the mean value more closely. The weak law of large numbers states that the sample average of a random sequence, X approaches the statistical mean η_X with high probability when the number of samples N tends towards infinity. That is,

$$\lim_{N\to\infty} \Pr\left\{\left|\left[\frac{1}{N}\sum_{n=1}^{N} X_n\right] - \eta_X\right| < \varepsilon\right\} = 1 \tag{6.6}$$

where ε is arbitrarily small.

6.6.2.2 Probability Of Bit Error For Watermark Extraction Without Using The Original Image

When watermark extraction is carried out without the original, we use a high pass filter before correlation detection to eliminate major (i.e., low frequency) components of the image. But since the image is not completely eliminated by the high pass filter, some of the extracted bits could be in error. The probability of bit error, P_b due to the interference from the original image, could be derived using the additive noise channel model. Let the watermarked image be, $M_i = V_i + W_i$, where V_i is the original image and W_i is the added watermark. Let E_W be the watermark energy and σ_V^2 be the image variance. The Signal-to-Noise Ratio (SNR) is thus E_W/σ_V^2. However, the spread spectrum receiver enjoys a theoretical SNR advantage of G_p. E_W/σ_V^2, where G_p denotes the processing gain and equals the chip rate, r_c, in this scenario. Therefore, under the AWGN model assumption for the image pixels and under the assumption that the message bits are equiprobable, the probability of bit error, P_b is given by:

$$P_{b(BPSK)} = Q\left[\sqrt{G_p \cdot \frac{E_W}{\sigma_V^2}}\right] \tag{6.7}$$

where,

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-y^2/2}\, dy \tag{6.8}$$

Table 5 gives the theoretical and simulation results. The theoretical values for non-coherent Binary Frequency Shift Keying (BFSK) are also given for comparison.

$$P_{b(BFSK)} = \frac{1}{2} e\left[-G_p \cdot \frac{E_W}{4\sigma_V^2} \right] \tag{6.9}$$

Simulation results are given for the probability of bit error without using the high pass filter (0 HPF), high pass filtering once (1 HPF) and high pass filtering twice (2 HPF) using the 3×3 filter given in equation 6.9. These results are plotted in Fig. 32. The SNR = E_w/σ_i^2 in dB, is plotted along the x-axis, where $E_w = \frac{1}{N}\sum_{n=0}^{N-1} w^2(n) = \alpha^2$. It is noted that high pass filtering twice would help to extract the watermark with almost no error. The simulation results are seen not to properly match the theoretical results, mainly because of the AWGN assumption, which is not true in the case of images.

α	SNR (dB)	Theoretical (BPSK)	Theoretical (BFSK)	0 HPF	1 HPF	2 HPF
0.5	-40.53	4.17E-01	4.95E-01	4.70E-01	3.40E-01	1.35E-02
1	-34.51	3.37E-01	4.78E-01	4.30E-01	2.02E-01	1.81E-04
1.5	-30.99	2.64E-01	4.53E-01	3.98E-01	1.17E-01	2.39E-05
2	-28.49	2.00E-01	4.19E-01	3.59E-01	5.54E-02	0.00E+00
2.5	-26.55	1.47E-01	3.79E-01	3.28E-01	2.57E-02	0.00E+00
3	-24.97	1.04E-01	3.36E-01	2.95E-01	1.45E-02	0.00E+00
3.5	-23.632	7.05E-02	2.91E-01	2.69E-01	4.82E-03	0.00E+00
4	-22.472	4.63E-02	2.46E-01	2.46E-01	2.21E-03	0.00E+00
4.5	-21.449	2.92E-02	2.04E-01	2.20E-01	5.02E-04	0.00E+00
5	-20.5338	1.78E-02	1.66E-01	2.03E-01	1.00E-04	0.00E+00
5.5	-19.706	1.04E-02	1.31E-01	1.76E-01	3.82E-05	0.00E+00
6	-18.95	5.80E-03	1.02E-01	1.54E-01	1.91E-05	0.00E+00
6.5	-18.255	3.10E-03	7.72E-02	1.35E-01	0.00E+00	0.00E+00
7	-17.6113	1.60E-03	5.73E-02	1.21E-01	0.00E+00	0.00E+00
7.5	-17.012	8.09E-04	4.16E-02	1.05E-01	0.00E+00	0.00E+00
8	-16.4514	3.86E-04	2.95E-02	9.06E-02	0.00E+00	0.00E+00
8.5	-15.925	1.77E-04	2.05E-02	7.82E-02	0.00E+00	0.00E+00
9	-15.4284	7.75E-05	1.39E-02	6.68E-02	0.00E+00	0.00E+00
9.5	-14.9587	3.26E-05	9.24E-03	5.59E-02	0.00E+00	0.00E+00

Table 5: Probability of bit error for watermark extraction without the original.

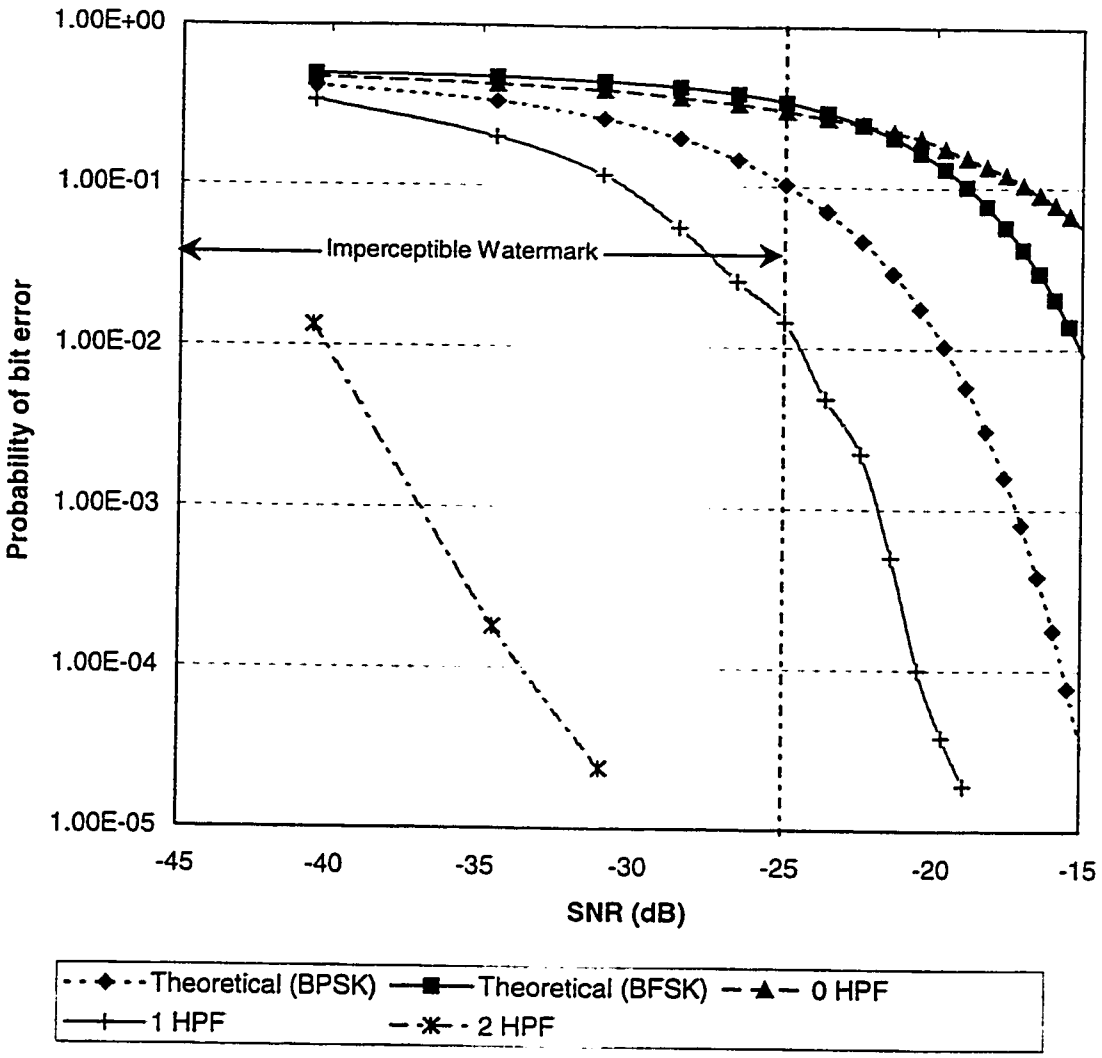


Figure 32: Probability of bit error for watermark extraction without the original.

6.6.2.3 JPEG Compression

Fig. 33 shows a plot of the correlation coefficient for different quality factors for JPEG compression of the watermarked image with chip-rate, $r_c = 500$ and scale factor, $\alpha = 2$. JPEG coding is a lossy compression scheme, the loss being caused by quantization. The lower the quality factor q, the higher is the compression. The memory size for the images for $q = 5$ and $q = 100$ are 3 kBytes and 62 kBytes

respectively. It can be noted from Fig. 33 that the correlation coefficient ρ, is above the threshold when the quality factor, q is greater than 15. But the image quality is also degraded, as can be seen from Fig 34 (b) so that it is useless when $q < 15$.

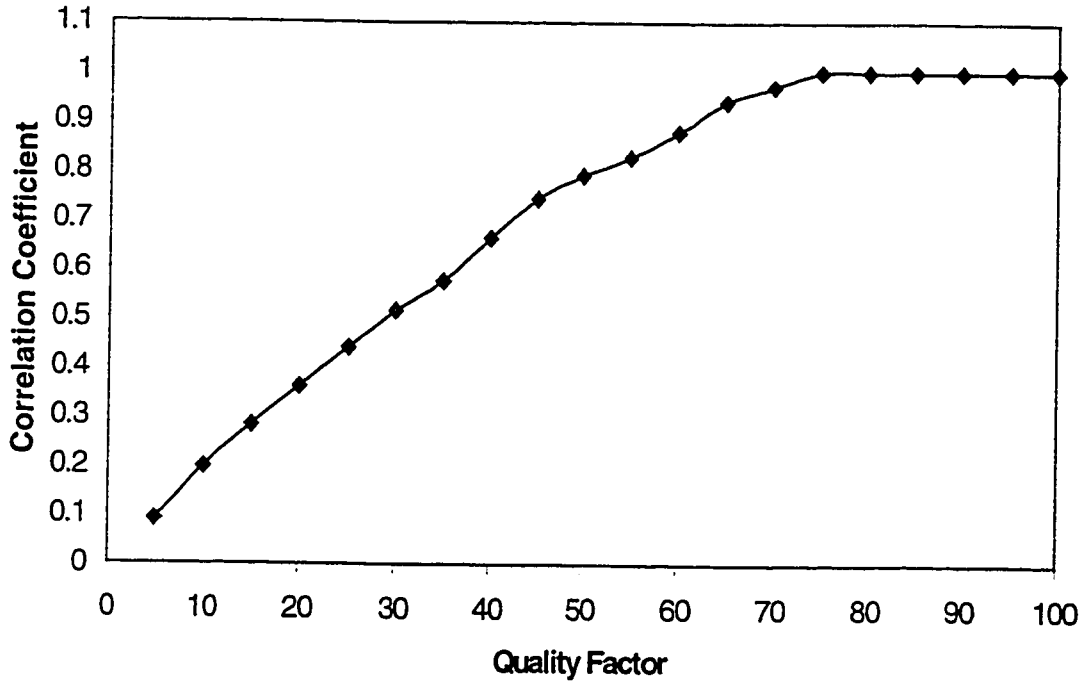


Figure 33: Detector response to JPEG compression.

(a) (b) (c)

Figure 34: Original and watermarked images after lossy JPEG compression.
(a) Original image. (b) Watermarked image with $q = 70$.
(c) Watermarked image with $q = 10$.

6.6.2.4 High Pass Filtering

High pass filtering (HPF) will make the watermark more visible as can be seen from Fig. 35 (b). In fact, for extraction without using the original image, high pass filtering before demodulation helps to extract the watermark with less error.

(a) (b)

Figure 35: Effect of high pass filtering on the watermarked image.
(a): Watermarked image with $\alpha = 4$. (b): Watermarked image after high pass filtering.

Table. 6 gives the correlation coefficient values for different scale factors, when the watermark is extracted without using the HPF and high pass filtering once or twice. The values are obtained for chip rate equal to 500.

α \ HPF	0 HPF	1 HPF	2 HPF
1	0.03	0.6363	1
2	0.2272	0.9242	1
3	0.4545	0.9848	1
4	0.5606	1	1

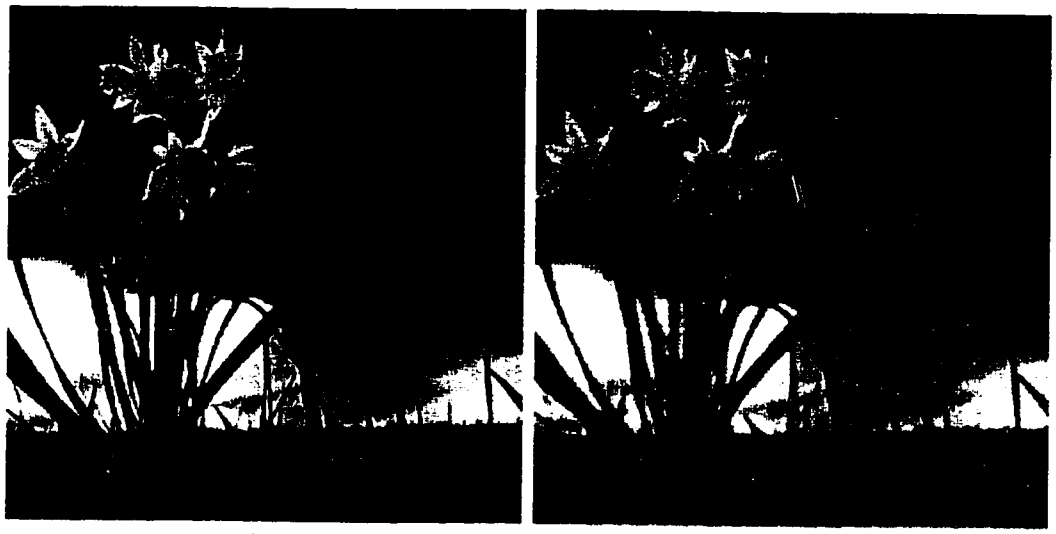
Table 6: Correlation coefficients for different values of the scale factor, α with and without HPF.

The following 3×3 filter has been used as the high pass filter:

$$HPF = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{bmatrix} \tag{6.10}$$

6.6.2.5 Low Pass Filtering

An attacker could low pass filter the watermarked image to eliminate the high frequency components of the watermark. However, a low pass filter will smooth out the image (Fig. 36(b)) and introduce severe distortion in areas with complex textures. It is not possible to filter out the watermark without damaging the image. Table 7 shows the correlation coefficient values, ρ, for different scale factors, α and a chip-rate, $r_c = 500$, with and without low pass filtering. It can be noted that the correlation coefficient increases, or the detection error decreases as the value of scale factor is increased.



(a) (b)

Figure 36: Effect of low pass filtering on the watermarked image.
(a): Watermarked image with α = 4. (b): Watermarked image after low pass filtering.

α \ LPF	0 LPF	1 LPF	2 LPF	3 LPF
0.5	1	0.1969	0.1772	0.1666
1	1	0.4545	0.2272	0.197
1.5	1	0.5606	0.2727	0.1969
2	1	0.6515	0.3636	0.2575
2.5	1	0.7272	0.3787	0.2727
3	1	0.7727	0.4091	0.2727
3.5	1	0.8181	0.4545	0.2727
4	1	0.8636	0.4697	0.333

Table 7: Correlation coefficients for different values of the scale factor, α with and without LPF.

Fig. 37 depicts the correlation coefficient for low pass filtering once with chip-rates 500 and 1000. The following 3×3 filter has been used as the LPF:

$$LPF = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{6.11}$$

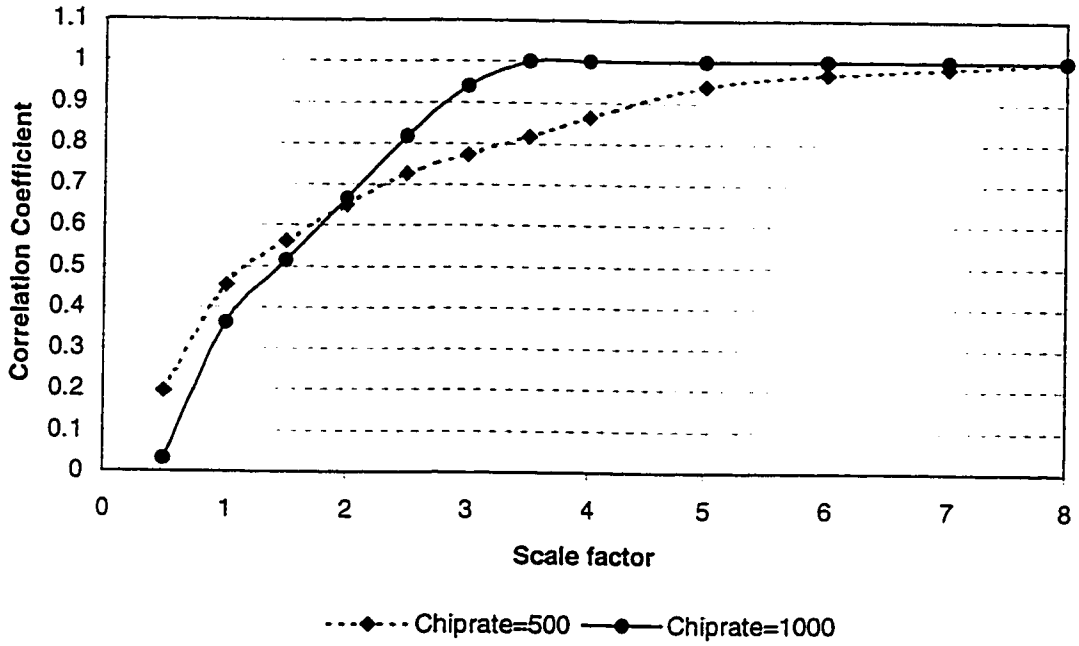


Figure 37: Detector response to low pass filtering for different scale factors.

6.6.2.6 Cropping

Fig. 38(a) shows the cropped image watermarked with $\alpha = 1$. The upper left quarter (128 x 128 portion) of the watermarked image is clipped off. To extract the watermark, the missing portion is replaced by the corresponding portion from the original unwatermarked image as in Fig. 38(b). Fig. 39 shows the detector response to cropping with and without interleaving. Without interleaving, the watermark could be detected without error when the chip-rate is large enough ($r_c > 128$, the crop length) to add enough redundant bits. Interleaving or spatially spreading the watermark bits ensures that cropping the edges will not remove the bits completely.

(a) (b)

Figure 38: Cropped and restored watermarked images.

(a): Watermarked image with a quarter clipped off. (b): Restored image after the lost portion is replaced by the corresponding portion from the original.

Interleaving drastically improves the resilience against cropping. The cropping attack can actually be seen as an extreme case of variable channel, which completely destroys certain pixels and keeps unchanged the remaining pixels. For this reason, the spatial diversity provided by the interleaving process helps to recover all the information bits from the unaltered portion of the image. All the information bits will lose on average the same number of pixels if the watermarked image is cropped, whereas without interleaving the effect of a cropping attack is greater for those bits located at the removed area of the image.

Furthermore, if the reordering performed by the interleaver is different for each key, an additional level of security is provided since the uncertainty about the spatial location of pixels modulated by each information bit prevents attacks concentrated on specific bits of the hidden message.

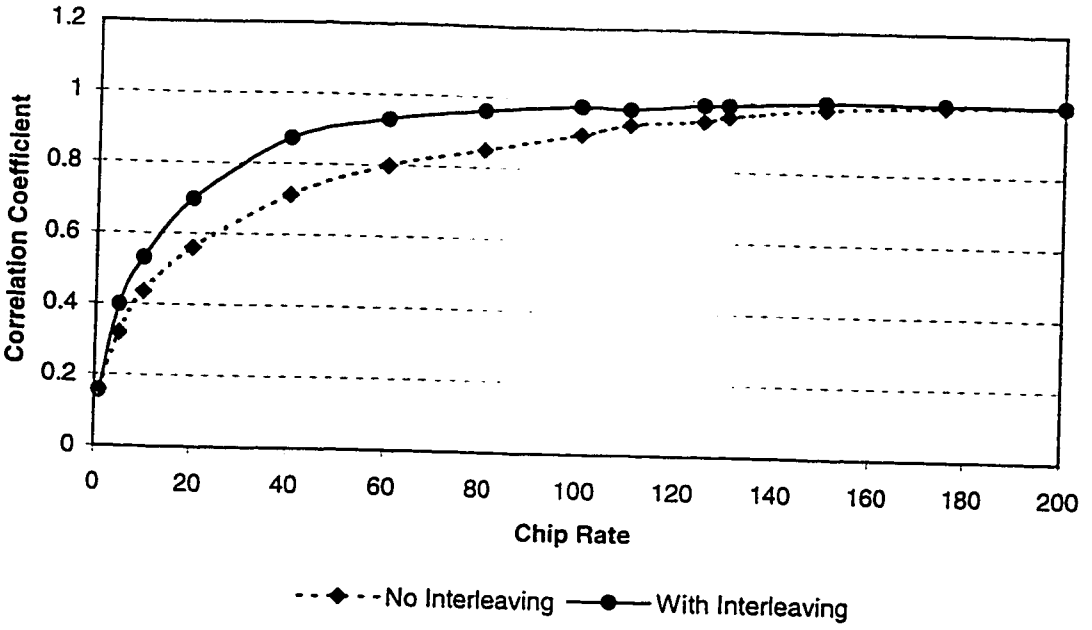


Figure 39: Detector response to cropped watermarked image with and without interleaving.

6.6.2.7 Rotation

In Stirmark attack, the watermarked image is rotated by small angles (up to 5°) to remove the correlation with the pseudo-noise sequence used for extraction. The edges of the rotated image are cropped to make the rotation not noticeable. Fig. 40(b) shows a watermarked image, which is rotated clockwise by 5°.

A block based *sliding correlator* used before extraction can detect the rotation and it can be compensated. The watermarked image is divided into smaller blocks and for each block, different combinations of shift, rotation, zoom, etc., are tried and the correlation between the modified block and the original pseudo noise signal is calculated. Table 8 shows that the watermark could still be detected, when the rotation angle is detected using the sliding correlator and compensated.

(a) (b)

Figure 40: Effect of rotation on the watermarked image.
(a): Watermarked image entitled "lena". (b): Watermarked image rotated by 5°.

Angle	1°	2°	3°	4°	5°
ρ	0.81	0.72	0.63	0.57	0.51

Table 8: Correlation coefficients for rotation by different angles.

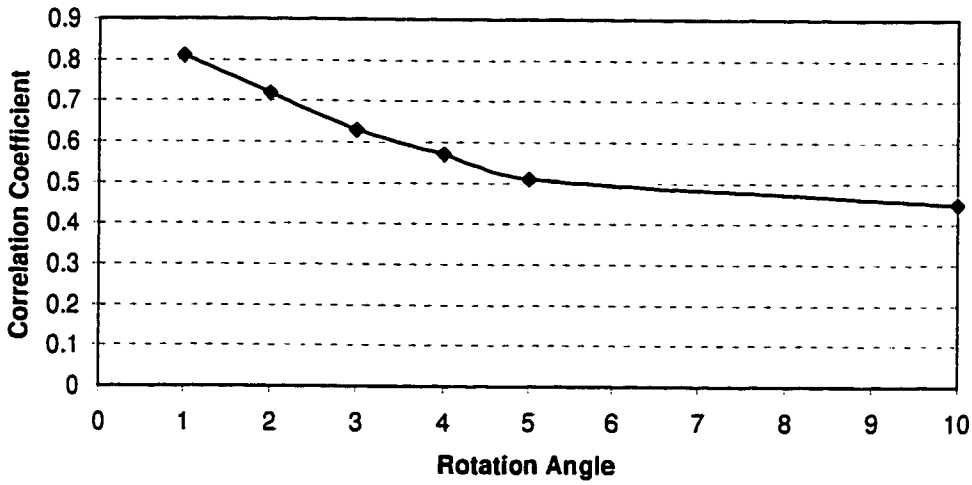


Figure 41: Detector response to rotation of the watermarked image.

6.6.2.8 Printing, scanning and re-scaling

Printing and re-scanning the watermarked image also causes scaling and cropping which has to be compensated. Table 9 shows that the watermark could be detected when the scale factor, $\alpha \geq 2$. Fig. 42 is the detector response for watermarked image after it has been printed, scanned and rescaled. Fig. 43 shows the original and the printed, scanned and rescaled watermarked images.

α	1	2	3	4	5
ρ	0.21	0.32	0.4	0.57	0.6

Table 9: Correlation coefficients for printing and rescanning for different values of scale factor.

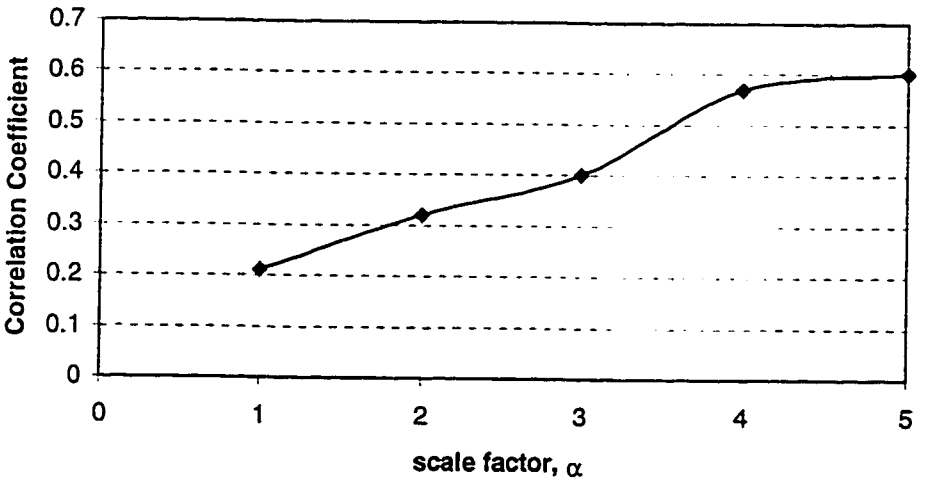


Figure 42: Detector response for printed, scanned and re-scaled watermarked image.

(a) (b)

Figure 43: Original and printed, scanned and re-scaled watermarked image. (a): Original image. (b): Watermarked image for $\alpha = 2$, after printing, scanning and re-scaling.

6.6.2.9 Multiple Watermarking

In authentication applications, even though a user cannot extract the owner's watermark, it is possible that he can add his own watermark and then claim ownership over the watermarked image. The owner's watermark should remain in the image even after many fake watermarks are added above it. Robustness to multiple watermarking was studied by successively embedding 20 different watermarks. This attack is equivalent to adding noise to the frequency bins containing the watermark. Fig. 44(b) shows an image which has 20 different watermarks embedded in it. Fig. 45 is the detector response to 1000 randomly generated watermarks, including the 20 specific watermarks with keys 50, 100, 150, 200, ..., 950, 1000. The 20 spikes clearly indicate that all the embedded watermarks could be detected.

Figure 44: Original and multiple watermarked images.
(a): Original image. (b): Image successively watermarked with 20 different watermarks.

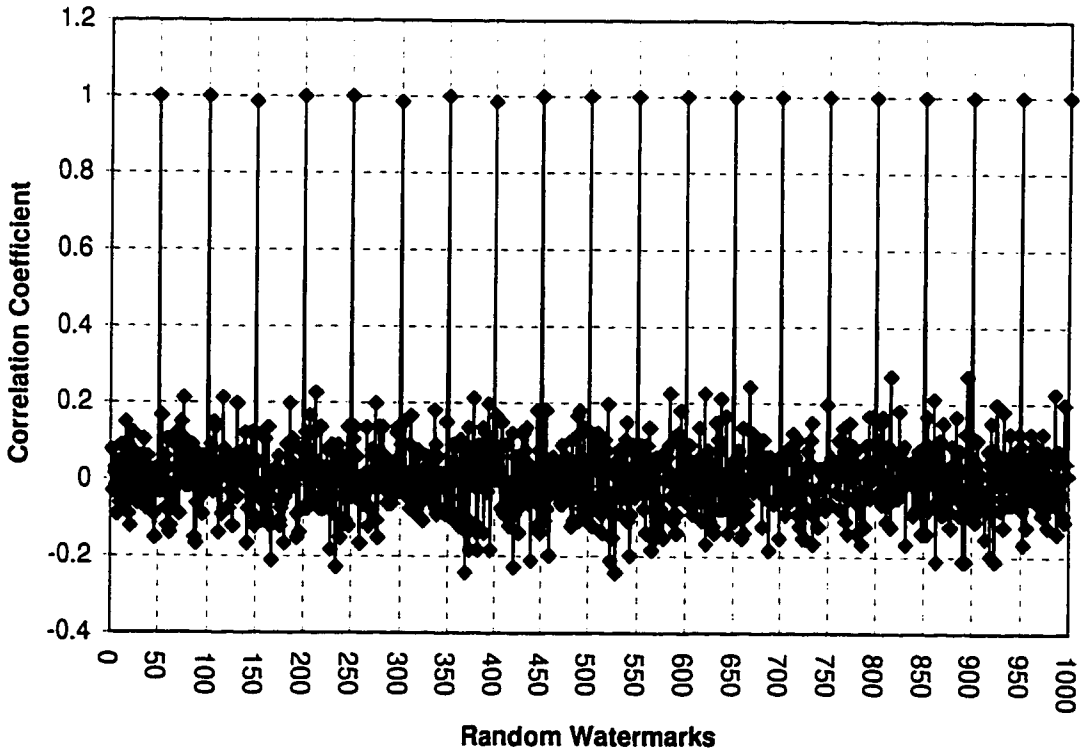


Figure 45: Detector response for multiple watermarking.

It can be noted that the image with multiple watermarks embedded in it is degraded, but not as much as in the case of simple attacks. But significant degradation occurs as the process is repeated by adding more and more watermarks.

The fake owner's original image (the one he claims to be the original) will contain the owner's watermark, whereas the true owner's original will not contain any watermarks. Thus in case of dispute, the original images can be checked which could identify the true owner.

6.6.2.10 Collusion Attack

In fingerprinting, where a different watermark is embedded in each copy, it is possible that a number of users could get together and compare their copies to detect some or all of the watermark bits. If a large number of differently watermarked copies are available, an averaging could eliminate the watermark. This is another clever attack where the watermark could be removed by not degrading the image.

In the experiment, separately watermarked images are taken and averaged. Fig. 46 shows the detector response to 1000 randomly generated watermarks which include the 5 watermarks present in the image. The 5 spikes show that the watermarks used for collusion could be detected. Fig. 47 and 48 are the responses when 20 and 40 watermarked images are colluded. It can be noted from Fig. 48, that not all of the watermarks involved in collusion are distinguishable: almost half the watermarks are undetected. Fig. 49(b) is the image formed by colluding 20 images. If a sufficient number of differently watermarked images are available, it is possible to remove the watermark by averaging, without degrading the image.

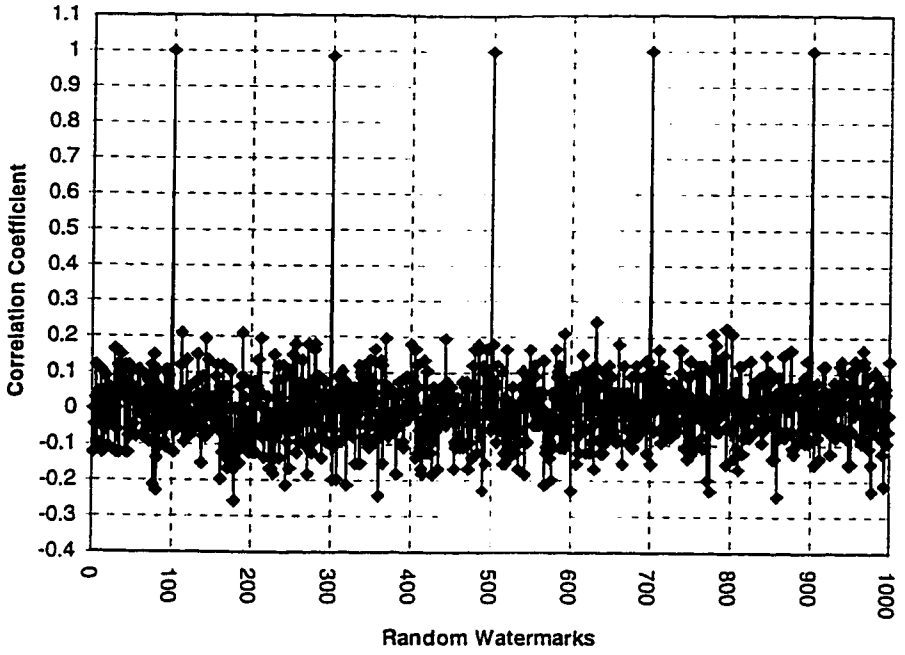


Figure 46: Detector response for collusion of 5 watermarked images.

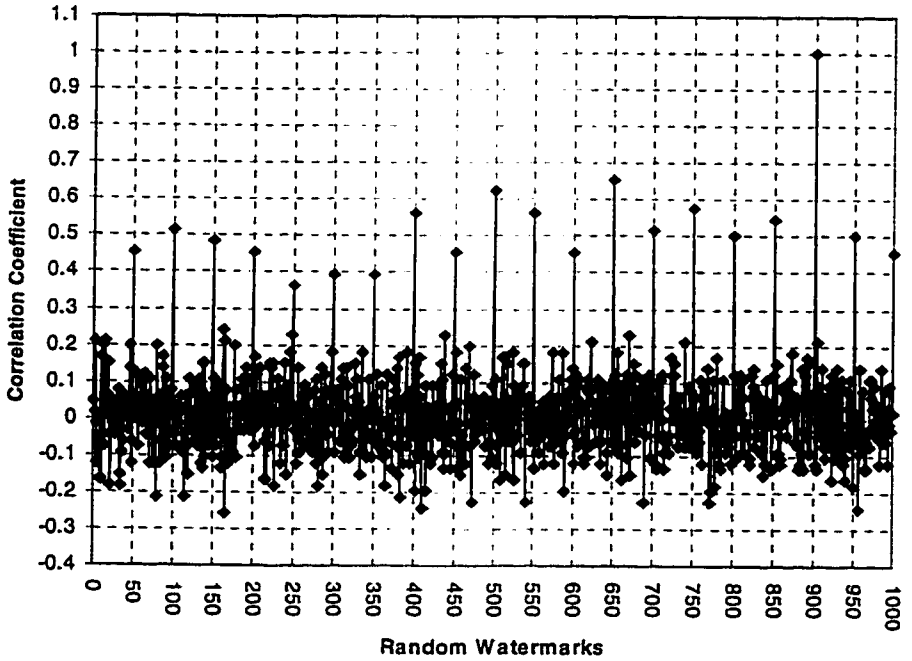


Figure 47: Detector response for collusion of 20 watermarked images.

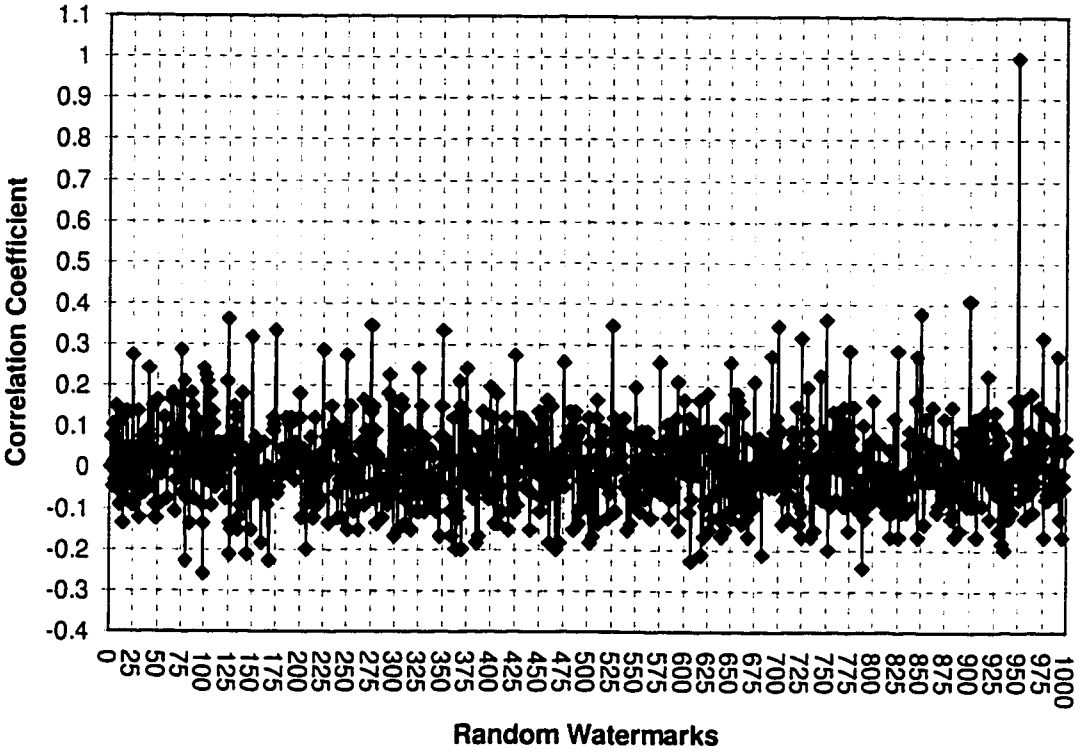


Figure 48: Detector response for collusion of 40 watermarked images.

(a) (b)

Figure 49: Original and colluded watermarked image.

(a): Original image. (b): Image obtained by averaging 20 differently watermarked images.

6.7 Conclusion

The major requirements for watermarks used for fingerprinting are imperceptibility and robustness against intentional and unintentional attacks. In spread spectrum watermarking, both these requirements can be satisfied. To ensure imperceptibility of the watermark, the scale factor, α should be kept less than or equal to 3. Robustness of the watermark is measured in terms of the normalized correlation coefficient, ρ. The value of ρ can be varied by varying the scale factor, α or the chip-rate, r_c. For $\alpha \leq 3$, r_c should be ≥ 400 for perfect reconstruction of the watermark when extraction is carried out without the original. Therefore, $\alpha = 2$ and $r_c = 500$ could be chosen as the optimum values, which can withstand all possible attacks and manipulations. Fig. 32 also shows that the watermark could be extracted without much error by high pass filtering the watermarked image twice before correlation detection.

7. CONCLUSIONS

7. 1 Thesis Summary

Spread spectrum watermarks are resistant to cropping (provided it is re-synchronized), non-linear distortions of amplitude and additive noise. Also, if it has good statistical properties, it should be mistaken for noise and go undetected by an eavesdropper. There are however some drawbacks to using direct sequence spread spectrum. These watermarks are not tolerant of timing errors. Synchronization is of utmost importance during watermark extraction. If watermark extraction is carried out in the presence of the host image, then synchronization is relatively trivial. The problem of synchronizing the watermark signal is much more difficult in the case where there is no host image. If the watermarked image is translated, rotated and scaled, then synchronization necessitates a search over a four dimensional parameter space (X-offset, Y-offset, angle of rotation and scaling factor). The search space grows even larger if one takes into account the possibility of shear and a change of aspect ratio.

In this work, we present a performance analysis of image and video watermarking using direct sequence spread spectrum principle. The robustness of the method has been studied by measuring the normalized correlation coefficient after the watermarked image is subject to various signal processing operations or attacks like compression, filtering, rotation, scaling, cropping, multiple watermarking, collusion attacks etc. A block based sliding correlator used before watermark extraction could

re-synchronize the watermarked image with the pseudo-noise signal used for watermark insertion. The watermark was readily detected with and without using the original image or video in all of the distortions and attacks. But collusion attack could remove the watermarks if a sufficient amount of differently watermarked copies is available to the attackers.

The theoretical probability of bit error has been derived for watermark extraction without using the original image, and compared with the experimental results. The effect of watermark signal energy and the chip-rate on the perceptibility and robustness of the embedded watermark has been studied. An information theoretic analysis of the spread spectrum watermarking has been carried out for fingerprinting applications.

7. 2 Suggestions for further research

Although video watermarking uses the same method as for image watermarking, video suffers an additional threat of *inter-frame collusion attack.* A possible solution to this attack is scene adaptive watermarking using temporal wavelet transform. Although this has been suggested in section 3.2.4, it has not been implemented. We have chosen DCT as the frequency transform for spectral domain watermarking so that the watermarking could be incorporated with the standard MPEG video compression standard.

Also choosing a different transform, which is insensitive to rotation, scale and translation, the watermark could be made robust to synchronization attacks. As

suggested in section 4.4, the Fourier-Mellin transform is a good choice, that gives rotation, scale and translation (RST) invariant watermarks.

The expression for the probability of bit error for watermark extraction without using the original, needs to be refined to closely match the experimental results.

The watermark bit capacity has to be experimentally verified for spread spectrum watermarking. The bit capacity should be measured by varying the chip-rate to achieve a correlation coefficient, $\rho > T_{det}$, the detection threshold, for watermark extraction without the original. But T_{det} varies with the chip-rate, r_c, as could be seen from Figs. 30 and 31, and therefore a relationship between r_c and T_{det} has to be derived.

8. APPENDIX

Steganographic Software

A wide variety of steganographic software are now available on the Internet, which hides information in text, image or audio [58]. The idea of steganography is to hide information for secret communication. The following is a list of this software.

1. **DiSi-Steganograph** (*by Nikola Injac*) is a very small, DOS-based steganographic program that embeds data in PCX images.
2. **EZStego** (*by Romana Machado*) is a Java based steganographic software which modifies the LSBs of still pictures (supports only GIF and PICT formats) and rearrange the color palette.
3. **Gif-It-Up v1.0** (*by Lee Nelson*) is a stego program for Windows 95 that hides data in GIF files. It replaces color indexes of the GIF color table with indexes of 'color friends' (a color friend is a color in the same table and as close as possible).
4. **Hide and Seek** (*by Colin Maroney*) is a stego program that hides any data into GIF images. It flips the LSBs of pseudo-randomly chosen pixels. The data is first encrypted using the blowfish algorithm.
5. **JPEG-JSTEG** (*by Derek Upham*) hides data inside a JPEG file.
6. **MandelSteg and GIFExtract** (*by Henry Hastur*) hide data in fractal GIF images. MandelSteg will create a Mandelbrot image, storing the data in the specified bit of the image pixels, after which GIFExtract can be used by the recipient to extract that bit-plane of the image.

7. **MP3Stego** (*by Fabien A. P. Petitcolas,* Computer Laboratory, University of Cambridge) hides data in popular MP3 sound files.
8. **Nicetext** (*by Mark Chapman and George David,* Department of EE & CS, University of Wisconsin Milwaukee) transforms ciphertext into innocous text which can be transformed back into the original ciphertext.
9. **Pretty Good Envelope** (*by Robert G. Durnal*) hides data in almost any file. In fact it embeds a binary message in a larger binary file by appending the message to the covert file as well as a 4-byte pointer to the start of the message. To retrieve the message, the last 4 bytes of the file are read, the file pointer is set to that value, and the file read from that point.
10. **Snow** (*by Mathew Knaw,* University of Melbourne, Australia) is used to conceal messages in ASCII text by appending white spaces to the end of lines.
11. **Steganography Tools 4** (*by Andy Brown*) encrypts the data with IDEA, MPJ2, DES, 3DES and NSEA in CBC, ECB, CFB and OFB modes and hides it inside graphics (by modifying the LSBs of BMP files), digital audio (WAV files) or unused sectors of HD floppies. The embedded message is usually very small.
12. **Stegodos** is a set of DOS programs that encodes messages into GIF or PCX images. The data embedded by modifying the LSBs of the picture is noticeable in most cases.
13. **Steganosaurus** (*by John Walker,* Switzerland) is a Unix program that will convert any binary file into nonsense text, but which statistically resembles text in the language of the dictionary supplied.

14. **SteganoWav** (*by Peter Heist*) is a Java (JDK 1.0) program that hides information in 16-bit wav files using a spread spectrum technique.

9. REFERENCES

[1] J. M. Acken, "How Watermarking Adds Value to Digital Content," in *Communications of the ACM*, vol. 41, Jul. 1998, pp. 75-77.

[2] D. Andrew, "MPEG Encoding of Full Screen Video for Compact Disc Interactive," in *Digital Video – Concepts and Applications Across Industries*, IEEE Press, pp.558-560.

[3] H. Berghel, "Watermarking Cyberspace," in *Communications of the ACM*, vol. 40, Nov. 1997, pp. 19-24.

[4] W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for Data Hiding," in *IBM Systems Journal*, vol. 35, Nos 3&4, 1996.

[5] D. Boneh and J. Shaw, "Collusion-Secure Fingerprinting for Digital Data," in *IEEE Transactions on Information Theory*, vol. 44, Sept. 1998, pp. 1897-1905.

[6] J. T. Brassil, S. H. Low, N. F. Maxemchuk and L. O. Gorman, "Electronic Marking and Identification Techniques to Discourage Document Copying," in *Proceedings IEEE Infocom* 1994, pp. 1278-1287.

[7] L. Boney, A. H. Tewfik and K. H. Hamdy, "Digital Watermarks for Audio Signals," in *Proceedings EUSIPCO 1996*, Sept. 1996.

[8] G. Caronni "Assuring Ownership Rights for Digital Images" in H. H. Brueggemann and W. Gerhardt-Haeckl, editors, Reliable IT Sytems VIS '95, Vieweg, Germany, 1995, pp. 251-264.

[9] I. J. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "Secure Spread Spectrum Communication for Multimedia" in *Technical Report, N. E. C Research Institute*, 1995.

[10] I. J. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "A Secure, Robust Watermark for Multimedia" in *Proceedings of the First International Workshop on Information Hiding, Univ. of Cambridge*, Lecture Notes in Computer Science, LNCS 1174, May 1996.

[11] I. J. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," in *IEEE Transactions on Image Processing*, vol. 6, Dec. 1997, pp. 1673-1686.

[12] S. Craver, N. Memon, B. L. Yeo and M. M. Yeung, "Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks and Implications," in *IEEE Journal on Selected Areas in Communications*, vol. 16, May 1998, pp. 573-586.

[13] S. Craver, B. L. Yeo and M. M. Yeung, "Technical Trials and Legal Tribulations" in *Communications of the ACM*, vol. 41, Jul. 1998, pp. 45-54.

[14] C. Dautzenberg and F. Boland, "Watermarking images," in *Technical Report*, Dept. of Electrical Engineering, Trinity College, Dublin, 1994.

[15] G. Depovere, T. Kalker and J. P. Linnartz, "Improved Watermark Detection Reliability Using Filtering Before Correlation," in *Proceedings IEEE ICIP*, Oct. 1998, pp. 430-434.

[16] P. G. Flikkema, "Spread-Spectrum Techniques for Wireless Communication," in *IEEE Signal Processing Magazine*, May 1997, pp. 26-36.

[17] M. George, J.-Y. Chouinard, N. D. Georganas, "Digital Watermarking of Images and Video using Direct Sequence Spread Spectrum Techniques" in *Proceedings IEEE Canadian Conference on Electrical and Computer Engineering (IEEE CCECE '99) May 9-12, 1999*, pp. 116-121.

[18] M. George, J.-Y. Chouinard, N. D. Georganas, "Spread Spectrum Spatial and Spectral Watermarking for Images and Video" in *Proceedings Canadian Workshop on Information Theory (CWIT '99), June 15-18, 1999*. pp. 119-122.

[19] F. Hartung and B. Girod, "Digital Watermarking of Raw and Compressed Video" in *Systems for Video Communication*, Oct 1996, pp. 205-213.

[20] F. Hartung and B. Girod, "Digital Watermarking of MPEG-2 Coded Video in the Bitsream Domain" in *Proceedings IEEE ICASSP*, vol. 4, Apr. 1997, pp. 2621-2624.

[21] F. Hartung and B. Girod, "Fast Public-Key Watermarking of Compressed Video" in *Proceedings IEEE ICIP*, Oct. 1997.

[22] F. Hartung and B. Girod, "Watermarking of Uncompressed and Compressed Video" in *Signal Processing*, vol. 66, May 1998, pp. 283-301.

[23] F. Hartung, J.K. Su and B. Girod, "Spread Spectrum Watermarking: Malicious Attacks and Counterattacks," in *Proceedings SPIE, Security and Watermarking of Multimedia Contents*, vol. 3657, Jan. 1999, pp 147-158.

[24] J. R. Hernandez and F. P. Gonzalez, "Shedding More Light on Image Watermarks," in *Proceedings of the Second International Workshop in Information Hiding*, Lecture Notes in Computer Science, LNCS 1525, April 1998, pp. 191-207.

[25] C. T. Hsu and J. L. Vu, "DCT-Based Watermarking for Video," *in IEEE Transactions on Consumer Electronics*, vol. 44, Feb 1998, pp. 206-216.

[26] N. F. Johnson and S. Jajodia, "Exploring Steganography: Seeing the Unseen," in *Computer*, Feb. 1998, pp 26-33.

[27] D. Kundur and D. Hatzinakos, "Towards a Telltale Watermarking Technique for Tamper-Proofing," in *IEEE Proceedings of the ICIP* 1998, pp. 409-413.

[28] M. Kutter, F. Jordan and F. Bossen, "Digital signature of color images using amplitude modulation," in *Proceedings. SPIE-EI97*, 1997, pp. 518-526.

[29] D. Le Gall, "MPEG: A Video Compression Standard for Multimedia Applications," in *Communications of the ACM*, vol. 34, Apr. 1991, pp. 47-58.

[30] J. P. Linnartz, T. Kalker and G. Depovere, "Modelling the False Alarm and Missed Detection Rate for Electronic Watermarks," in *Proceedings of the Second International Workshop in Information Hiding*, Lecture Notes in Computer Science, LNCS 1525, April 1998, pp. 329-343.

[31] S. H. Low, N. F. Maxemchuk and A. M. Lapone, "Document Identification for Copyright Protection Using Centroid Detection," in *IEEE Transactions on Communications*, vol. 46, Mar. 1998, pp. 372-382.

[32] S. H. Low, N. F. Maxemchuk and A. M. Lapone, "Performance Comparison of Two Text Marking Methods," in *IEEE Journal on Selected Areas in Communications*, 1998.

[33] F. Mintzer, G. W. Braudaway and A. E. Bell, "Opportunities for Watermarking Standards," in *Communications of the ACM*, vol. 41, Jul. 1998, pp. 57-64.

[34] N. Memon and P. W. Wong, "Protecting Digital media content," in *Communications of the ACM*," vol. 41, Jul. 1998, pp 35-43.

[35] K. Matsui and K. Tanaka, "Video Steganography: How to Secretly Embed a Signature in a Picture", in *IMA Intellectual Property Project Proceedings*, January 1994, pp. 187-206.

[36] R. Ohbuchi, H. Masuda and M. Aono, "Watermarking Three-Dimensional Polygonal Models Through Geometric and Topological Modifications," in *IEEE Journal on Selected Areas in Communications*," vol. 16, May 1998, pp 551-560.

[37] J. J. K. Ó Ruanaidh, W. J. Dowling and F. M. Boland, "Watermarking Digital Images for Copyright Protection," in *IEE Proceedings: Vision, Image and Signal Processing*, vol. 143, Aug. 1996, pp. 250-256.

[38] J. J. K. Ó Ruanaidh, W. J. Dowling and F. M. Boland, "Phase Watermarking of Digital Images," in *IEEE Proceedings of the ICIP* 1996, pp. 239-242.

[39] J. J. K. Ó Ruanaidh and T. Pun, "Rotation, Scale and Translation Invariant Digital Image Watermarking," in *Signal Processing* 66, May 1998, pp. 303-318.

[40] J. G. Proakis, "Digital Communications," Chapter 13 – Spread Spectrum Signals for Digital Communications, pp. 695-753, Edition 3.

[41] F. Petitcolas, R. J. Anderson and M. G. Kuhn, "Attacks on Copyright Marking Schemes," in *Proceedings of the Second International Workshop in Information Hiding*, Lecture Notes in Computer Science, LNCS 1525, April 1998, pp. 218-238.

[42] F. Petitcolas and M. G. Kuhn, "Stirmark 3.0 watermark robustness testing software," available at http://www.cl.cam.ac.uk/~fapp2/watermarking/image_watermarking/stirmark/

[43] F. Petitcolas, "Unzign watermark robustness testing software," http://www.cl.cam.ac.uk /users/fapp2/steganography/image_watermarking/ unzign/

[44] C. I. Podilchuk and W. Zeng, "Image-Adaptive Watermarking Using Visual Models," in *IEEE Journal on Selected Areas in Communications*, vol. 16, May 1998, pp. 525-539.

[45] J. Puate and F. Jordan. (1996). Using fractal compression scheme to embed a digital signature into an image. Available online: http://www.epfl.ch/~jordan /watermarking.html.

[46] M. Ramkumar and A. N. Akansu, "Information Theoretic Bounds for Data Hiding in Compressed Images," in *Workshop on Multimedia Signal Processing 1998.*

[47] J. Smith and B. Comiskey, "Modulation and Information Hiding in images," in *Proceedings of the First International Workshop in Information Hiding*, Lecture Notes in Computer Science, LNCS 1173, Springer, May/June 1996, pp. 207-226.

[48] J.K. Su, F. Hartung and B. Girod, "A Channel Model for a Watermark Attack," in *Proceedings SPIE, Security and Watermarking of Multimedia Contents*, vol. 3657, Jan. 1999, pp 159-170.

[49] M. D. Swanson, M. Kobayashi and A. H. Tewfik, "Multimedia Data-Embedding and Watermarking Technologies," in *Proceedings of the IEEE*, vol. 86, Jun. 1998, pp. 1064-1087.

[50] M. D. Swanson, B. Zhu and A. H. Tewfik, "Multiresolution Scene-Based Video Watermarking Using Perceptual Models," in *IEEE Journal on Selected Areas in Communications*, vol. 16, May 1998, pp. 540-550.

[51] A. Z. Tirkel, R. G. Van Schyndel and C. F. Osborne, "A Two-Dimensional Digital Watermark", in *Proceedings ACCV'95*, December 1995, pp. 378-383.

[52] R. G. Van Schyndel, A. Z. Tirkel and C. F. Osborne, "A Digital Watermark," in *Proceedings IEEE ICIP*, 1994, pp 86-90.

[53] R. B. Wolfgang, C.I. Podilchuk and E.J. Delp, "Perceptual Watermarks for Digital Images and Video," submitted to *Proceedings of IEEE*, 1998.

[54] K. K. Wong, C. H. Tse, K. S. Ng, T. H. Lee and L. M. Cheng, "Adaptive Watermarking," in *IEEE Transactions on Computer Electronics*, vol. 43, Nov. 1997, pp. 1003-1008.

[55] M. M. Yeung, "Digital Watermarking," in *Communications of the ACM*, vol. 41, Jul. 1998, pp. 31-33.

[56] M. M. Yeung and B. L. Yeo, "Fragile Watermarking of Three-Dimensional Objects," in *IEEE Proceedings of the ICIP*, vol. 2, Oct. 1998, pp. 442-446.

[57] J. Zhao, E. Koch and C. Luo, "In Business Today and Tomorrow," in *Communications of the ACM*, vol. 41, Jul. 1998, pp. 67-72.

[58] The information hiding home page, http://www.cl.cam.ac.uk/~fapp2/steganography/stego_soft.html

[59] Sony Corporation of America, Press release, http://www.sony.com/SCA/press/feb_19_99b.html