2002

Spread spectrum-based video watermarking algorithms for copyright protection

Serdean, Cristian Vasile

http://hdl.handle.net/10026.1/563

http://dx.doi.org/10.24382/3920
University of Plymouth

SPREAD SPECTRUM-BASED VIDEO WATERMARKING ALGORITHMS FOR COPYRIGHT PROTECTION

by

CRISTIAN VASILE SERDEAN

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

DOCTOR OF PHILOSOPHY

Satellite Research Centre
Department of Communication and Electronic Engineering
Faculty of Technology, University of Plymouth

October, 2002

Abstract

SPREAD SPECTRUM-BASED VIDEO WATERMARKING ALGORITHMS FOR COPYRIGHT PROTECTION

by

Cristian Vasile Serdean

Digital technologies know an unprecedented expansion in the last years. The consumer can now benefit from hardware and software which was considered state-of-the-art several years ago. The advantages offered by the digital technologies are major but the same digital technology opens the door for unlimited piracy. Copying an analogue VCR tape was certainly possible and relatively easy, in spite of various forms of protection, but due to the analogue environment, the subsequent copies had an inherent loss in quality. This was a natural way of limiting the multiple copying of a video material. With digital technology, this barrier disappears, being possible to make as many copies as desired, without any loss in quality whatsoever. Digital watermarking is one of the best available tools for fighting this threat.

The aim of the present work was to develop a digital watermarking system compliant with the recommendations drawn by the EBU, for video broadcast monitoring. Since the watermark can be inserted in either spatial domain or transform domain, this aspect was investigated and led to the conclusion that wavelet transform is one of the best solutions available. Since watermarking is not an easy task, especially considering the robustness under various attacks several techniques were employed in order to increase the capacity/robustness of the system: spread-spectrum and modulation techniques to cast the watermark, powerful error correction to protect the mark, human visual models to insert a robust mark and to ensure its invisibility. The combination of these methods led to a major improvement, but yet the system wasn't robust to several important geometrical attacks. In order to achieve this last milestone, the system uses two distinct watermarks: a spatial domain reference watermark and the main watermark embedded in the wavelet domain. By using this reference watermark and techniques specific to image registration, the system is able to determine the parameters of the attack and revert it. Once the attack was reverted, the main watermark is recovered. The final result is a high capacity, blind DWT-based video watermarking system, robust to a wide range of attacks.

Contents

List of figures

List of tables

Glossary

A/D. Analogue/Digital.

ACC. Accumulator.

AWGN. Additive White Gaussian Noise.

BER. Bit Error Rate.

BCH. Bose - Chaudhuri - Hocquenghem.

bpf. Bits Per Frame.

bps. Bits Per Second.

BPSK. Binary Phase Shift Keying.

CPTWG. Copy Protection Technical Working Group.

CRC. Cyclic Redundancy Code.

CWT. Complex Wavelet Transform.

D/A. Digital/Analogue.

DCT. Discrete Cosine Transform.

DFT. Discrete Fourier Transform.

DINT. De-interleaver.

DSSS. Direct Sequence Spread Spectrum.

DV. Digital Video.

DWT. Discrete Wavelet Transform.

EBU. European Broadcasting Union.

ECG. Electrocardiography.

EEG. Electroencephalography.

EKG. Electrocardiography.

ENC. Encoder.

EZW. Embedded Zero-tree Wavelet.

FEC. Forward Error Correction.

FFT. Fast Fourier Transform.

FMT. Fourier Mellin Transform.

HH, HL. High High, High Low.

HPF, hpf. High Pass Filter.

HVS. Human Visual System.

IDWT. Inverse DWT.

IDCT. Inverse DCT.

IFFT. Inverse FFT.

i.i.d. Independent Identically Distributed.

INT, INTER. Interleaver.

INT^{-1}. De-interleaver.

ITU. International Telecommunication Union.

JND. Just Noticeable Difference.

JPEG. Joint Photographic Experts Group.

LH, LL. Low High, Low Low.

LLT. Log-Log Transform.

lpf. Low Pass Filter.

LPT. Log-Polar Transform.

LSB. Least Significant Bit.

MJPEG. Motion JPEG.

MPEG. Moving Picture Experts Group.

MRS. Magnetic Resonance Spectra.

MTF. Modulation Transfer Function.

NC. Number of Cross-correlations.

PAL. Phase Alternate Line.

PCCC. Parallel Concatenated Convolutional Code.

pdf. Probability Density Function.

PN. Pseudo Noise.

POMF. Phase Only Matched Filter.

QPSK. Quadrature Phase Shift Keying.

RS. Reed-Solomon.

RSC. Recursive Systematic Code.

RST. Rotation, Scale and Translation.

SCCC. Serial Concatenated Convolutional Code.

SDMI. Secure Digital Music Initiative.

SISO. Soft Input Soft Output.

SNR. Signal to Noise Ratio.

SPIHT. Set Partitioning In Hierarchical Trees.

SPOMF. Symmetrical POMF.

STFT. Short Time Fourier Transform.

TC. Turbo Codes.

TEL. Tolerable Error Level.

VCR. Video Cassette Recorder.

VDP. Video Dependent, Perceptual.

VHS. Video Home System.

VIP. Video Independent, Perceptual.

VWG. Video Watermarking Group.

WM. Watermark.

X-corr. Cross-correlation.

ACKNOWLEDGMENTS

I would like to express my sincere thanks to my supervisors, Professor Martin Tomlinson and Dr. Graham Wade for their support and guidance during my PhD. Additionally, I am very grateful to Dr. Graham Wade for his feedback, and for providing me with his expertise in writing scientific publications and in particular for my PhD thesis.

I would like to address special thanks to my colleague and friend Dr. Adrian Ambroze for all his help during my PhD, and particularly for supplying me with some of his C code. I gratefully acknowledge all his programming help and advice.

I gratefully acknowledge the support of EPSRC, which through a research studentship made all this work possible. I would like to extend my gratitude to BBC R&D and TRL Technology Ltd. who contributed financially to my studentship.

Many thanks to my friends: Zaki Ahmed, Adrian Ambroze, Peter van Eetvelt, Keith Flack, Bogdan Ghita, Licha Mued, Phil Rodwell, James Slader and Levente Toth for their company, help and "considerable amount of good quality humour that compensated for the weather".

I would also like to extend my gratitude to all the administrative staff from University of Plymouth, particularly to Mrs. Brenda Garraghan, Mrs. Barbara Fuller and Ms. Sue Locke for all their help and efficiency in solving many bureaucratic issues.

Many thanks to all my Romanian Professors for providing me with good technical education, in spite of the difficult economic conditions. Special thanks to Prof. Monica Borda for offering me the opportunity of coming to Plymouth as an MSc exchange student, and for all her subsequent help.

Finally but not at least, I would like to express my warmest thanks to my family, for all their support and sacrifice during all my undergraduate and graduate studies. My endless gratitude goes to my mother who encourages and supports me, and yet misses me day by day; and to my late father, God rest his soul, for all his effort and invaluable teachings.

I would also like to thank to my relatives for their support, particularly to Fam. Raicu and Neag.

DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This work was financed by the EPSRC grant GR/M37691 and additional industrial contributions from BBC Research & Development, Kingswood Warren and TRL Technology Ltd., Tewkesbury.

This research programme included an extensive literature survey and attendance of relevant international conferences. The work has been regularly presented at the research seminars organised by the University of Plymouth, in several international conferences and in technical meetings with BBC, TRL Technology and other partners.

This work was concretised in publishing two journal papers (one as a co-author) and six international conference papers (one as a co-author). The full list of publications is provided in section 8.3.

Signed: …………………………
Date: …………..………………

This thesis is dedicated to my mother Victoria and to the loving memory of my father Vasile (1940-1990).

Chapter

1

"Do what you feel in your heart to be right – for you'll be criticized anyway. You'll be damned if you do, and damned if you don't."

Eleanor Roosevelt (1884-1962)

Introduction

Nowadays virtually all multimedia production and distribution is digital. The advantages of digital media, for creation, processing and distribution are all well known: superior quality, more quicker and easier to edit and modify, possibility of software processing rather than the more expensive hardware alternative (if the real time processing is not a requirement), and maybe the most important advantage is the unlimited copying of digital data without any loss of quality whatsoever. This latter advantage is not desired at all by the media producers and content providers, in fact is perceived like a major threat, because it may cause them considerable financial loss.

Once the digital technology is widely available to the public, the piracy suddenly becomes a major issue. This generates the need for protecting the copyrighted material against piracy. Some typical examples are the recent court battles between the music industry and Napster, Kazaa and Morpheus. The movie and music industry are particularly keen to develop any system which will stop users copying the digital media. Especially after the introduction of Internet sharing technologies which allow users from the entire planet to share any kind of digital media between them (like Napster, Gnutella, Morpheus and others) the record labels are trying to stop this trend by virtually any method possible. The cheapest and most effective ways in the long term are the non-technical methods like endless threats and law suites and using their huge influence to promote harsher copyright protection laws. When everything else fails, the only remaining alternative are the technical methods in the form of various copyright protection techniques.

The perfect example is the case of the VCR (Video Cassette Recorder). Probably not too many people know that when the VCR was marketed, the record labels tried to stop the technology by filling a law suit, on the grounds that the VCR technology could be used to copy protected material. Fortunately for us and for the technical development which led to technologies like CD-R and DVD-R they didn't succeed, but they managed instead to impose legal taxes on the blank recording media (VCR tapes, CD-R's, CD-RW's and others), taxes which are included in the final price paid by the user. Although this system is not currently implemented in UK, it is in force in most of the European countries and USA. Since in this instance the legal way failed, the technical approach was the only alternative left. The result was the development of the Macrovision copy-protection system which proved to be quite efficient against the casual VCR piracy.

Even if the user in fact pays for the right of copying digital materials, the record labels recently introduced a copyright protection system for the audio CD's which actually tries to stop the users from copying their legitimate CD's and even playing the CD's on a computer. Actually, this protection system developed by the Israeli company Midbar, deliberately introduces during the fabrication process a substantial number of errors on the disk, in fact so many, that even the powerful error correction capability of the computer drives is defeated. This is a rather "sad" method which destroys the very core of the digital technology, lowering not only the quality, but also the reliability of the disk. In fact the legitimate buyers were so upset, that the record labels had to withdraw the disks from the market, and as a result Philips who holds the rights for the CD-ROM standard won't allow the record labels to use the CD logo on this kind of protected CD's. These methods are rather obtrusive and have the "quality" of angering the legitimate customers, and even more, they are apparently illegal in those countries in which the customers are paying levies on the recording media.

Unlike these "crude" methods, digital watermarking is an unobtrusive way of protecting such material and for audio, images and video it operates by hiding a perceptually invisible signal into the host signal.

1.1 Historical Roots

The roots of watermarking as an information hiding technique can be traced in the ancient Greece as *Steganographia* or *steganography* as we know it now. The origin of the word steganography comes from the Greek στεγανός – "steganos" γϱαφειν – "graphein" which literally means “covered writing”. Many dictionaries are not even mentioning the word and few of them which are including the word are wrongly explaining it as cryptography. In fact although these two notions are related they are quite different. While cryptography focuses on encrypting a message so it can be read only by its intended recipient, steganography, on the other hand, keeps the message secret by hiding the fact that the message exists at all.

The historical roots of steganography and the beginnings of watermarking are well described in the literature [Singh, 1999], [Kobayashi, 1997], [Swanson et al, 1998-1], [Hartung et al, 1999-2], [Wolfgang et al, 1997 and 1999], [Langelaar et al, 2000]. Maybe one of the most well known example of a steganographic technique which is still widely used even today are the omnipresent watermarks which could be found in virtually any bank note, in different official documents and even in some stamps.

From the perspective of copyright protection, most of the researchers are making a clear distinction between steganography and watermarking from the robustness point of view. One important property of watermarking which is not characteristic to the steganography is the robustness to attacks, by attacks understanding virtually any technique which tries to modify/alter/remove/destroy the watermark.

The first efficient analogue copyright protection system appeared in the early eighties, soon after the VCR made its public debut. The system called Macrovision, basically adds some “parasite” pulses to the video signal during blanking periods, in such a manner that the TV sets are not affected. These pulses are instead seriously perturbing the sync circuitry within any modern recorder with an analogue input, making impossible to copy a protected tape. Although the system is not too difficult to defeat, it has been proven to be an excellent tool against casual piracy. As a result, the Macromedia copyright protection was embedded into the DVD standard in order to protect the analogue outputs of the DVD players.

The debut of the digital watermarking techniques was made in the early nineties [Tanaka et al, 1990]. The actual term “watermarking” was introduced in [Tirkel et al, 1993]. Although the word has its roots in maritime terminology, a more appropriate translation is that of transparent or invisible marking. The term “watermarking” survived in spite of few other alternatives like: labelling, stamping and tattooing.

The watermarking passed more or less unnoticed, until in the mid nineties when the digital distribution of media content started to emerge on a larger scale and the content providers together with the copyright owners started to become very interested in copyright protection technology in order to reduce or stop the major piracy threat involved by digital media distribution. The year 1995-1996, marked a real boom in watermarking research, as **Table 1-1** shows.

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
2	0	2	1	8	21	59	106	224	259	258	258

Table 1-1 The number of publications on digital watermarking during the years according to [Peter Meerwald, 2002].

1.2 Classification of the Watermarks

It is not an easy job to classify the watermarks and the watermarking techniques. They could be classified from so many perspectives that an exhaustive classification is almost out of question. The classification presented here will be limited to the cases relevant to this thesis.

Classifying the watermarking techniques function of the media to be marked is maybe the first one worth to be mentioned. There can be text, audio, image and video watermarking. Some attempts were done to watermark the software as well. It has to be mentioned from the beginning that this thesis is mainly dealing with video watermarking and occasionally with image watermarking, in order to compare it with other schemes.

To see exactly where we stand, **Figure 1-1** presents the main classes of watermarks and highlights the case presented during this thesis. So from now on, the word "watermark" refers to the robust and invisible case.

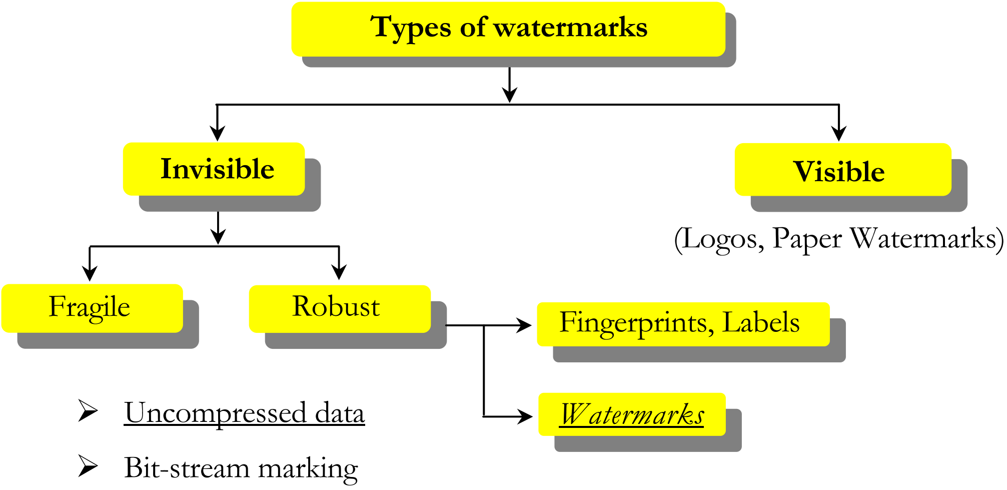


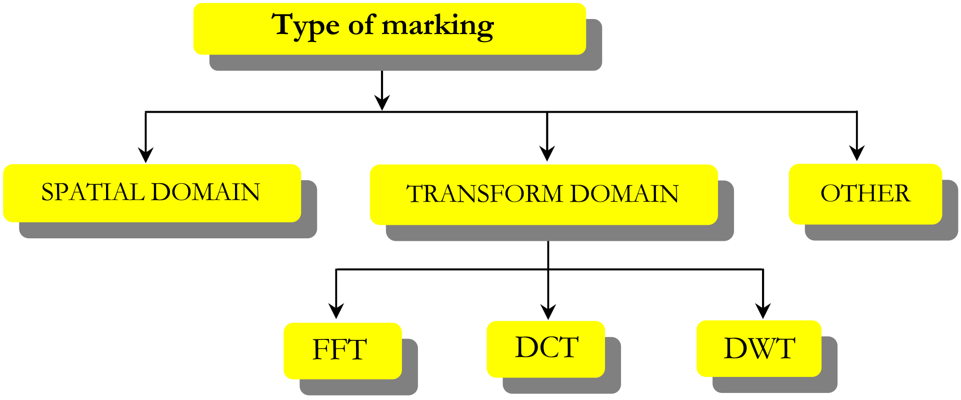
Figure 1-1 Classification of copyright protection marking

As **Figure 1-1** shows, there are two types of watermarks: the visible ones, like different logos either on paper or on a TV screen and the most important one, the invisible or transparent watermarks, which cannot be perceived by the human sensory system. An invisible watermark can be either robust or fragile. The use of a fragile watermark is important when one wants to verify if the protected media was tampered with or not. This type of watermark is especially designed to be as fragile as possible, so even the slightest modification of the marked media will destroy it, indicating that someone tampered with the media in question. This type of watermark is like a CRC (cyclic redundancy code).

The main class of watermarks – the robust ones – can be classified function of the purpose of the watermark, and therefore this could be an indication of the length of the watermark. Usually the fingerprints and labels are in fact serial numbers, typically quite short (e.g. 64 bits), which will uniquely identify the marked media (of course they could carry additional information if desired). If the embedded mark is more than a label, or if is quite long or has other purposes, then is specified by the general term: "watermark".

Additional to this classification, in the specific case of video watermarking, depending of where exactly the watermark is embedded during the distribution chain, we could have a watermark embedded in the uncompressed data or in the MPEG2 bit stream (without decoding and re-encoding the video). The case described during this thesis is the watermarking of the RAW uncompressed data.

Figure 1-2 shows another useful classification of the watermarking techniques,



- ➢ Non Perceptual
- ➢ Perceptual – VIP – Video Independent, Perceptual
  – VDP – Video Dependent, Perceptual

Figure 1-2 Classification of digital watermarking techniques function of the domain where the watermark embedding is performed.

function of the domain where the watermarking is embedded. The first attempts to watermark an image/video sequence were done in the spatial domain. This is quite simple, quick and obviously has DSP implementation advantages, but suffers from the lack of a good visual model. Generally speaking, embedding a watermark in the transform domain is more attractive because of the higher degree of freedom, and because it is naturally suited for perceptual marking based upon the *Human Visual System* (HVS). Visual models were specifically developed for this domain, mainly in the context of JPEG and MPEG2 compression. The usual transforms considered are the DFT/FFT (Discrete Fourier Transform/Fast Fourier Transform), DCT (Discrete Cosine Transform) and DWT (Discrete Wavelet Transform). The DFT/FFT can be a very attractive choice, offering the possibility of phase modulation and some useful invariance properties. Its main disadvantage is that in order to obtain a real image it is necessary to maintain complex conjugate symmetry, which effectively halves the potential marking capacity. Another disadvantage is the lack of good HVS models. Probably the most popular choice is the DCT transform, and this may remain so especially when watermarking MPEG2 compressed video, since the DCT is the basis of MPEG2. A more advanced choice is DWT. The wavelet transform is structurally very close of the way in which is thought the HVS is working, being by itself a HVS model, so the need of a complex HVS model is more reduced. The DWT marking is a lot more flexible than both FFT and DCT and has major advantages compared with these two, as will be shown in Chapter 6.

As **Figure 1-2** suggests, another classification of watermarking techniques could be with respect to the perceptual algorithm used at the embedding of the watermark. The first watermark schemes were non-perceptual and as a direct consequence not too robust. Embedding the watermark according to a perceptual algorithm brings the advantage of reduced visibility and increased robustness. Depending of the HVS model used, this could be image/video independent for the simplest ones or preferably image/video dependent in the case of a more advanced model, which therefore can adapt itself to the particular image/video.

1.3 The Applications of Digital Watermarking

As a general definition, digital watermarking can be regarded as hiding a message signal into a host signal without any perceptual distortion of the host signal. In theory, watermarking should protect a file permanently, like an invisible tattoo which cannot be removed without significantly damaging the protected data. This makes it the ideal tool for copyright protection, file tracking and monitoring. Therefore most of the research was carried out with respect to these applications. Taking a closer look at the range of possible applications, we can identify:

- ✓ *Copyright protection:* To protect its intellectual rights, the owner of the digital media uses a watermark which carries copyright information in order to be able to prove later on (in a court of law) that a third party infringed its copyrights. In other words, the embedded watermark is used to show ownership of the digital media.

- ✓ *Fingerprinting:* A fingerprint can be compared with a serial number, and makes it easy to trace the source of illegal copies. The owner embeds a different fingerprint in each copy of the media sold to a different customer. In this way is very easy to establish who have broken the licence agreement by supplying the media to a third party and therefore to detect unauthorised duplication.

- ✓ *Broadcast monitoring:* By watermarking the broadcasted media prior to transmission, one can continuously monitor all the broadcasting channels, and identify all the intellectual property violations. This application is of major interest to broadcast corporations, who are very keen to protect their broadcasted material, especially since some of it has a huge market value, making it particularly prone to piracy. One eloquent example is the News, where even a photo or a short amateur video sequence can worth a huge amount of money. A slightly different application could be an automatic registration and monitoring of broadcasted radio and TV programs such that the appropriate royalties are automatically paid to the right owners.

- ✓ *Copy protection and usage control:* This is another major application of watermarking particularly wanted by the record labels. The information contained in the watermark directly controls the digital recording devices and can allow or prohibit the recording/viewing or even the storage of the media. The watermark could indicate for example "never copy" or "copy allowed". DVD access control could be one of the candidates who will benefit from this technology. By inserting a dynamic watermark which includes a number indicating how many copies of the media can be allowed, and by decreasing that number every time when the media was copied, one can restrict the copying of the media to a certain predefined number. This could be useful in electronic media distribution.

- ✓ *Authentication and integrity verification:* Using a fragile watermark, one can establish if the watermarked media is the original or if someone tampered with it. In this way is possible to authenticate the media or to validate its integrity.

Although the main interest at the moment is in the copyright protection area, mostly due to the strong involvement of the industry and record labels, there could be many more applications suitable for digital watermarking, most of them being less demanding in terms of constraints compared with the typical copyright protection applications:

- ✔ *Hidden content labelling and indexing:* A watermark could be used to provide subsidiary information about an image or a video, such as who is the subject, the date of production and any additional comments. This information could be used for searching and indexing purposes, making a lot easier to find and organise digital media.

- ✔ *Medical safety:* Closely related with the previous, this application could be a very useful safety measure, by allowing insertion of patient specific data into the medical images. By its nature, this application requires that the watermark should alter as little as possible the original data, or in other words a very light marking.

- ✔ *Media enhancements:* The watermarking could be used for adding different enhancements to the existing media. An example is the adding of various information to audio and video files, like details about the singer, a biography of the actors, subtitles and virtually anything else. All this "extras" are "free", in the sense that they could be applied to any existing format while keeping the backward compatibility and they do not require any additional storing space, and therefore the dimension of the media is preserved.

- ✔ *Data hiding and secret communications:* Watermarking techniques can be certainly used for transmission of secret, private messages exactly like the steganography was used during the course of the history for non digital media. With governments trying to restrict the encryption techniques, sooner or later this will be the ideal tool for secret communications. This is probably the worse nightmare of the secret services and various other governmental agencies who want to be able to intercept and control everything, since watermarking techniques are certainly not making this easier for them.

1.4 Requirements of a Video Watermarking System

The requirements for a watermarking system are obviously different for different applications, but even so, referring to the copyright protection context, it is possible to identify several common requirements for a watermarking system:

✔ *Perceptual transparency:* One of the most important requirements in most applications is to embed a watermark in a perceptually transparent manner. This means that one cannot make a difference between the marked and the original media under typical viewing conditions.

✔ *Robustness to attacks:* The aim of any watermarking system is to embed the mark in such a way that it cannot be removed, unless the media will be severely degraded during this process. Therefore the watermark should be able to withstand a wide range of attacks some of them unintentional, but most of them intentional, e.g. designed to modify/alter/remove/destroy the watermark. This is usually one major requirement in many watermarking applications, with the exception of fragile watermarking, which requires exactly the opposite: the watermark has to be as fragile as possible.

✔ *Blind recovery:* Due to the dimensions of video materials, especially in the uncompressed case presented in this thesis, video watermarking techniques specifically require blind or oblivious recovery of the mark. This means that the original cannot be used in the recovery process, since due to the huge dimensions involved it is impossible (or at least extremely expensive) to maintain a database with all the originals. This requirement is paramount.

✔ *Bit rate of data embedding algorithm (data payload of the watermark):* Depending on the application, this requirement can range anywhere between 1 data bit (this type of scheme is only able to tell if the image was marked or not with a specific watermark) and few hundreds or more data bits. Closely related with the bit rate of the embedding algorithm is watermark granularity, which represents the minimum segment of data containing a unit of watermark. For example, the typical requirement for a broadcast monitoring system is at lest 64 data bits per a video segment of 1 second (25 frames for PAL).

✔ *Security:* The embedding procedure must be secure or in other words, an unauthorised user should not be able to detect and remove the watermark. Most of the schemes are embedding the watermark according to a secret key which controls the insertion of data in the host signal. This respects the Kerckhoff's principle (1883) from cryptography which states that the security should lie in a secret key rather than in the algorithm's secrecy. Even if one is familiar with the scheme but doesn't know the secret key, an unauthorised detection/removal of the watermark should be impossible in a reasonable amount of time.

✔ *Copyright protection and ownership deadlock:* When the watermark is used to establish ownership of the media, the scheme should be able to resolve the rightful ownership even when multiple ownership claims are made. This is still an open problem, as today no scheme can unambiguously determine the ownership of a marked media if it does not use the original media or another copy in the detection process, which is evidently not the case in video watermarking due to the blind recovery requirement. The problem was first described in [Craver et al, 1997]. The deadlock arises when a pirate simply adds his watermark to an already marked media and then claims the ownership. The problem is to establish (in the absence of the original) who watermarked the media first.

All these requirements are related to each other and quite contradictory. Probably visibility versus robustness is the most common example. A very robust watermark assumes a more heavily marked video, which obviously leads to increased visibility. Therefore the necessity of a trade-off is obvious. The other requirements are weakening even more a watermarking system. The blind recovery and the necessity of embedding many bits in a limited minimum segment are clear examples of that. Security for example is closely related with robustness: if the watermark is not secure, the system is not robust at all.

Several other application dependent requirements can be added to those already mentioned here: the algorithm should work in real time (hardware) and should be as simple and efficient as possible, preferably cheap to implement in hardware and easy to interface with the existing electronic devices and of course a reliable detection of the watermark. Depending of the application, there could be additional economical and technical requirements.

1.5 Structure of the Thesis

The thesis is organised as follows:

Chapter 1 offers a basic introduction in digital watermarking and its applications. The historical roots of the watermarking are traced back in time and the watermarks are then classified, emphasising the case discussed in this thesis. An overview of the possible applications of digital watermarking, together with the major players in this field is also provided. Finally, the main requirements of a video watermarking system are presented and briefly analysed.

Chapter 2 is an overview of the existing image/video watermarking methods. From spatial domain to wavelet-based watermarking, this is the place where the most important methods and algorithms relevant to the thesis are underlined. The basis and starting hypotheses of the thesis are set here, emphasising the reasons for choosing one alternative rather than another. The reader is then introduced into the specific problem of video watermarking. This chapter also provides a specific introduction to the particularities of attacks in the video watermarking context, in close relationship with the EBU's recommendations. The chapter ends by highlighting the contributions of the thesis.

Chapter 3 presents in detail the structure and the algorithm of a spatial domain spread spectrum video watermarking system. This chapter establishes the foundations of the spread spectrum watermarking system. The main structure and many of the components described here are used for building the DCT scheme from Chapter 5. Starting with the basic uniform marking, the system is gradually improved by embedding the watermark in a more efficient way (e.g. HVS dependent) and by using a sliding correlator for watermark recovery. The effects of pre-filtering and various block sizes on the performance of a spatial domain watermarking scheme are also investigated.

Chapter 4 presents some basic communication principles related to channel capacity and forward error correction (FEC) codes. The Shannon's channel capacity theorem and its practical implications are also discussed, as well as the ways of getting closer to this limit. It is shown that in order to achieve better performance in a communication system, e.g. to get closer to the Shannon's limit, one can use the new state-of-the-art FEC codes (e.g. Turbo Codes). A short introduction to Turbo codes and their characteristics and performance is also provided in this chapter. The watermarking is seen by the information theory perspective and therefore by applying this theory and by using Turbo coding, the performance of the watermarking system is greatly improved (as Chapter 5 and Chapter 6 will show).

Chapter 5 discusses the case of watermarking in the DCT domain, together with several methods of increasing the capacity/robustness of the system. To achieve this goal, the system uses both advanced HVS models for watermark embedding and state-of-the-art FEC (Turbo codes) in order to protect the watermark. The casting of the watermark and other alternative modulation techniques are also analysed. In order to improve the system even further, 3-D marking replaces the usual frame by frame approach (2-D marking) by taking into account the temporal dimension. This increases the "local" chip rate leading to better cross-correlation results (wider cross-correlation area) and caters for frame dropping/duplication attacks.

Chapter 6 begins with an introduction to the wavelet transform, with the accent on the 2-D DWT (Discrete Wavelet Transform) case. The multiple advantages of the DWT transform are discussed and compared with the traditional FFT/DCT transforms, taking into account the specific framework of digital watermarking. Choosing a proper basis constitutes an important step which will be also discussed. Due to major advantages of the DWT, the wavelet coefficients are one of the most suitable places to insert a watermark. The proposed watermarking system is described in detail during this chapter, including the HVS aspects of the scheme and error correction. The performance of the system will be then analysed for both image watermarking (in order to compare the results with the existing image watermarking schemes described in the literature) and video watermarking.

Chapter 7 discusses one of the most difficult problems in digital video watermarking: watermark recovery in the presence of geometric attacks like frame shift, cropping, scaling, rotation, and change of aspect ratio, especially when some of these are combined together. Re-establishing the synchronisation in a reasonable time is capital, and this is the object of this chapter. After a short introduction of the available techniques for combating geometrical attacks, this chapter proposes the use of an additional spatial watermark which combined together with image registration techniques will counteract the geometric attacks. The proposed technique, its implications and its performance are extensively analysed. The specific problems for video watermarking are also highlighted and measures to counteract these problems are proposed. As the results suggest this technique proves itself to be very successful, leading to a highly robust, high capacity blind watermarking system.

Chapter 8 will conclude the work presented in this thesis and suggest further research directions. The system's performance is summarised and compared with EBU's requirements. This comparison is presented in a tabular form, which fully illustrates the capabilities of the proposed system. In most of the cases, the proposed system meets or even exceeds (by far) the requirements of the EBU. Further research directions are also suggested during this chapter. This chapter also provides the complete list of author's publications.

Chapter

2

> "I never found the perfect quote. At best I have been able to find a string of quotations which merely circle the ineffible idea I seek to express."
>
> Caldwell O'Keefe

A Review of Watermarking

This chapter establishes the basis and specifies the requirements for the video watermarking system described in this thesis. The starting hypotheses of this project are detailed in close relationship with the EBU's video watermarking recommendations for a typical broadcast monitoring system. A brief overview of the existing watermarking techniques relevant to this thesis it is then presented, ending up by highlighting the contributions of this thesis.

2.1 A Basic Watermarking System

After a comprehensive survey of the existing techniques, by carefully analysing and weighting all the advantages and disadvantages of different methods a decision was reached to further pursue a spread spectrum technique, considering that spread spectrum has more potential compared with other methods, in spite of being so sensitive to the de-synchronization attacks.

Spread spectrum radio techniques have been developed for military applications, since mid 1940's for their anti-jamming and low-probability-of-intercept properties. They allow the reception of radio signals that are over 100 times weaker than the atmospheric noise.

In particular, the spread spectrum techniques are offering a good flexibility and are very suitable for watermarking due to the similarities between the watermarking and spread spectrum communications. Watermarking can be seen as a communication problem, in which the original image plays the role of the channel noise and attackers may try to disrupt the transfer of information. In both cases the channel is a very difficult one characterised by high levels of noise. The large bandwidth required by a spread spectrum technique is not a problem, since usually the video sequences are quite big, offering a large number of coefficients and therefore the chip rate is sufficiently high for obtaining a robust watermarking system. The noise like spread spectrum signal is very difficult to detect/intercept and jam and is obviously spread in the entire video sequence, therefore suggesting a good robustness to certain attacks and a very secure system. Furthermore, the system can be relatively easy implemented, the watermark embedding and retrieving are based on secret keys and the system doesn't require the presence of the original video for watermark retrieving. A general block diagram for a video watermarking system based on the DSSS (Direct Sequence Spread Spectrum) is presented in **Figure 2-1**. The secret key is used for generating the same PN sequence for both embedding and retrieving. The spreading is achieved by multiplying this PN sequence with the data payload. As a result each watermark data bit is randomly spread in the entire video sequence, with a chip rate c_r. Typical for a video watermarking system, the recovery of the mark is blind, e.g. without resorting to the original video. The watermark is recovered by using

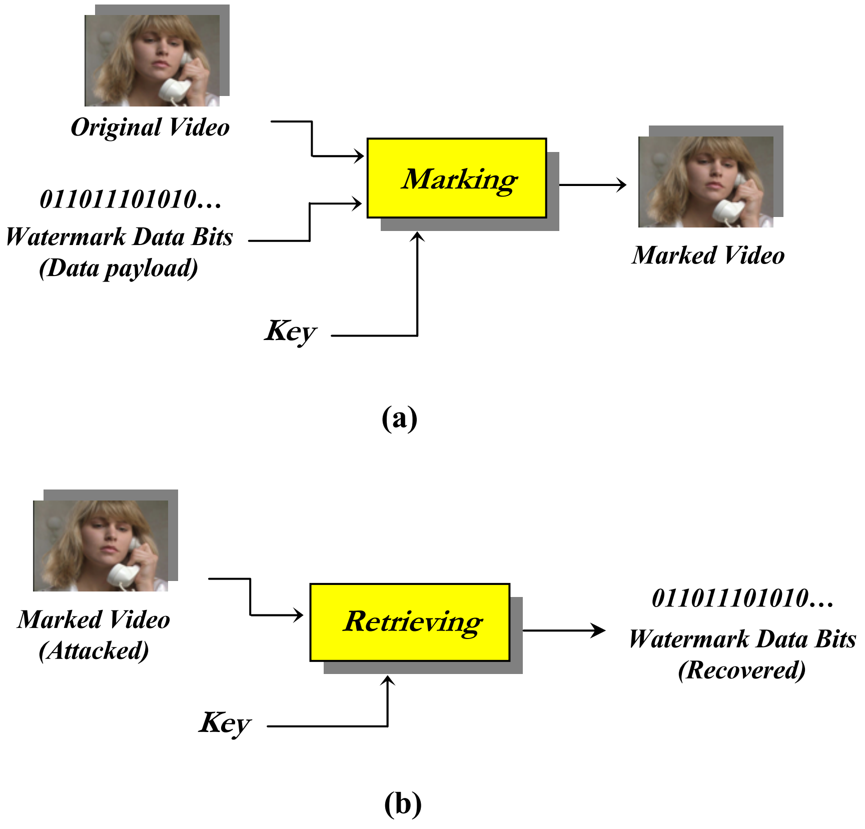


Figure 2-1 A basic video watermarking system: watermark embedding (a) and recovery (b).

cross-correlation methods, in the form of a matched filter (correlation receiver), following the principle of optimum reception. The SS technique will be described in detail in Chapter 3.

2.2 The Video Format

All the work done in this thesis was carried out in the context of uncompressed video, as found in TV studios and described by the ITU-R 601 (ITU-T BT.656) standard. To be more specific, the video sequences supplied by the BBC are in the raw Y-C_B-C_R format, with all the components separate and without any file header. The first half of the file represents the luminance component Y, followed by the chrominance components C_B and C_R.

The chrominance components are not robust at all, because they can be easily discarded, without affecting the video quality in any other way except the resulting black and white picture. Therefore all the algorithms described during the thesis are marking only the luminance component. Anyway marking the chrominance components has several other disadvantages. The human eye is much more sensitive to slight colour changes compared to slight luminance changes. As a result, these components have to be more lightly marked (with reduced amplitude) and from this reason are less robust compared with the luminance. Moreover, the complexity of the algorithm which uses the chrominance components is more than double, while the gain is quite small and it could be even zero if an attacker decides to discard the chrominance components. This is a strong enough reason to avoid the marking of chrominance components. Maybe in the applications where the real time requirement is not important and the cost can be tolerated one could use them in order to get a bit more robustness.

Finally, one more remark: the sequences supplied by the BBC are special test sequences, well known in the video processing field, and therefore they were especially selected to be more difficult to mark, and to show more easily any possible artefacts. The appropriate remarks for each particular sequence will be made at the right time.

2.3 Possible Attacks in the Video Watermarking Context

Generally, the attacks can be classified as intentional and un-intentional ("friendly"). For video watermarking this classification is of a special interest, due to the particularities of the video processing chain. For example attacks like MPEG2 compression or slight spatial and/or temporal frame shifts can be considered un-intentional since they can appear during the video editing chain and their intent is not to destroy the watermark. In order to characterise an attack as intentional or unintentional, one has to carefully analyse the context of the application. Some attacks could easily fall in both categories, the difference being made by the intensity of the attack.

Although generally speaking each author has more or less his view when classifying the attacks, usually they can be divided in 3 main classes, as follows:

Signal processing attacks, which are probably the most usual category of attacks, contains mainly:

- signal enhancements: brightness, contrast, sharpening, blurring, etc
- digital and analogue, linear and non-linear filtering
- addition of noise and/or noise reduction
- digital-analogue (D/A) → analogue-digital (A/D) conversion and re-sampling
- data compression: MJPEG, MPEG1, MPEG2, DV, digital recording, etc
- PAL coding and analogue recording (VHS)
- colour space conversion / grey-scale conversion

Geometric attacks are one of the most efficient and demanding attacks against many watermarking systems. The most usual geometric attacks are:

- line/column cut and/or duplicate
- frame cut and/or duplicate
- frame rate conversion: 24 Hz ↔ 25Hz ↔ 30Hz
- picture aspect-ratio conversion: 4:3 ↔ 16:9
- line-scan conversion: progressive ↔ interlaced
- cropping, shifting (translation), rotation, scaling and possible others

Collusion and collusion-like attacks are attacks that use several copies of the same host media with different embedded watermarks, with the express purpose of removing the watermark. These copies are averaged together and a new data set (media) is created by using several different algorithms [Cox et al, 1995 and 1998], [Kilian et al, 1999], [Stone, 1996], [Craver et al, 1996, 1997 and 1998].

The ownership ambiguity attacks are trying to create an ambiguity about who's the real owner of the watermarked media. The so called "IBM attack" was first described in [Craver et al, 1996, 1997 and 1998] who suggests few techniques for combating this attack, like time-stamps and non-invertible watermarks based on one way hashing functions (OWHF). At the moment these attacks are more or less forgotten, being considered as of second interest, especially in the video context; the EBU recommendations are not even mentioning this kind of attack.

These attacks and the robustness requirements for each class of attacks in the context of a broadcasting monitoring system are described in the EBU's watermarking recommendations [Cheveau et al, 2000]. It is worth mentioning here that these recommendations were published in March 2001, together with the test results of four watermarking systems provided by leading industry players: Lucent, Philips, Tektronix and Thomson, and this was the result of a "Watermarking Call for Systems" issued by EBU in May 2000. The conclusion of the tests was that none of the systems satisfied all the robustness requirements and only two of them complied with the 64 data bits capacity requirement. This shows once more that is not easy at all to comply with these requirements. Even at this time, none of the available systems can satisfy all these recommendations. To show once more the complexity of the problem, the conclusions drawn from the tests were that further development is needed together with a possible reduction of the data capacity from 64 to 48 data bits, in order to improve the robustness of the system [Cheveau et al, 2000].

2.4 Overview of the Existing Watermarking Techniques

In contrast with the steganography which has a very long history, the digital watermarking is a relatively young field, less than 10 years old. Practically, in the context of image/video watermarking the first papers appeared in the 1994, the debut being made in [Matsui et al, 1994] and [Schyndel at al, 1994]. They set the basis for the so called LSB (Least Significant Bit) watermarking. While this technique works in a noise free environment, it is totally useless when comes to robustness. Indeed, one can set the LSB either to 0 or to 1, without affecting the image quality too much and obviously the watermark is completely removed. In spite of this drawback quite a few variants were developed during 1994 and 1995.

The real breakthrough came in 1995/1996 with the introduction of spread spectrum watermarking [Cox et al, 1995 and 1996]. This can be regarded as the real starting point of robust watermarking. As expected, the first watermarking techniques were developed in the spatial domain, immediately followed by the DCT domain watermarking. Beside image watermarking, which dominates the watermarking literature, we can find few uncompressed video watermarking techniques and several MPEG2 bit-stream watermarking techniques.

During the years several trends can be identified, starting with the spatial domain watermarking, closely followed by the DCT based watermarking. Developing of robustness benchmarking tools and watermarking algorithms which can withstand the attacks produced by these tools is another well represented trend in the watermarking research. The need of more robust systems led to introduction of perceptual watermarking, where the watermark is embedded according to HVS models. Few other research areas can be identified: the use of different transforms like FFT, DWT and several others, developing of RST (Rotation, Scaling and Translation) invariant watermarking techniques and finally the most recent trend is to regard the watermark as a hidden communication channel which is therefore protected by different error correction codes. Closely related with the communication theory and statistical modelling of the channel is the use of optimum detection theory in order to improve the reliability of the watermark detector. Hypothesis testing was extensively used in the detection process. Since the watermarking literature is quite large, only those papers directly relevant to this thesis are presented, describing briefly the algorithm for the most important ones.

2.4.1 Spatial Domain Watermarking Techniques

This is the first and the most straightforward way to add a watermark in an image/video. Ignoring the early schemes based on LSB, several pre-spread-spectrum image watermarking techniques can be mentioned: the "Patchwork" method in which randomly selected pairs of pixels are used to hide 1 bit by increasing the value of one pixel and decreasing the other one and the "Texture Block Coding" which embeds the watermark by copying one image texture block to another area in the image with a similar texture. Both methods were proposed in [Bender et al, 1995]. A similar "Patchwork" type technique, which this time divides the image into two equal sets was proposed in [Pitas et al, 1995 and 1996]. An improved version was later proposed in [Langelaar et al, 1996 and 1997]. Many other variations can be mentioned here but since they are not directly relevant to this thesis, the interested reader could find them in one of these watermarking overviews: [Kobayashi, 1997], [Swanson et al, 1998-1], [Hartung et al, 1999-2], [Wolfgang et al, 1997 and 1999] and [Langelaar et al, 2000].

Spread-spectrum and watermarking

As stated before one of the best way to insert a watermark is to use a technique based on spread-spectrum. The spread spectrum watermark is embedded in the media by "amplitude modulation". This is simply done by adding the watermark to the luminance values of the pixels. The entire mechanism is detailed in Chapter 3. Due to the simplicity and flexibility of the algorithm there are many watermarking schemes based on this principle.

One of the first and most well known representatives is the uncompressed video watermarking scheme developed by Hartung [Hartung et al, 1996 and 1998], who transposes the idea of Cox [Cox et al, 1995 and 1996] in spatial domain. The main advantage of Hartung's method is the blind recovery of the watermark. Another important difference is the multi-bit nature of the watermark, compared with Cox's scheme. The scheme uses a binary pseudo-sequence and uniform marking, e.g. in absolute value, each pixel is modified with the same value. Therefore the algorithm is non media dependent and non perceptual, which is indeed a significant drawback. In the latter paper, Hartung improves its algorithm by filtering the video sequence prior to cross-correlation and analyses the recovery of the watermark in more detail.

The "Watercast" system

Although it is widely accepted that the spatial domain is not the best place to cast a highly robust watermark [Ramkumar et al, 1998-2], [Fei et al, 2001], from the reasons discussed in the further chapters, one of the best video watermarking schemes available at this moment, Watercast [op de Beeck et al, 2001], [Kalker et al, 1999-1 and 1999-2] developed by Philips, uses the spatial domain.

Watercast is an improved version of JAWS (Just Another Watermarking System), adapted to the broadcast monitoring applications [Kalker et al, 1999-1 and 1999-2]. JAWS was initially developed for DVD copy-protection [Maes et al, 2000].

For the sake of maintaining low complexity, both watermark embedding and detection are performed in the spatial domain. The embedded watermark consists of watermark patterns of size 128x128 with Gaussian distribution, which are repeated (tiled) to fill the whole video frame. In order to avoid visible artefacts, the watermark is scaled on a pixel-by-pixel basis, with a scaling factor which is derived from an "activity measure". The "activity measure" is in fact an empirical HVS model, computed using a Laplacian high-pass filter. This is quite a common method in digital watermarking. More details about this type of HVS marking can be found in Chapter 3. The same watermark is embedded into several consecutive video frames.

For watermark detection, a correlation detector is used after applying a spatial pre-filter [Depovere et al, 1998] that reduces cross-talk between video signal and watermark. Since the watermark must be detected even in the presence of spatial shifts, a search over all possible shifts is performed. Because the watermark signal is generated by tiling of a smaller watermark pattern, only 128 x 128 positions have to be searched, according to the size of the watermark pattern. In order to reduce complexity, the search and correlation is done in the FFT domain. Further, only the phase information of the FFT is used in the correlation. This method of detection has been previously proposed for pattern recognition and is referred to as Symmetrical Phase Only Matched Filtering (SPOMF) [Kuglin et al, 1975], [Chen et al, 1994] and [Pech-Pacheco et al, 199x].

In order to embed a sufficiently large multi-bit watermark, the system has to use several different basic watermark patterns on top of each other. The information is encoded in the choice of the basic patterns and their relative positions, leading overall to a quite complex system. The watermark can convey just enough information to comply with the EBU's requirements.

Other systems

Coming back to the image watermarking, several other authors could be mentioned: [Swanson et al, 1996-1], [Nikolaidis et al, 1998], [Kutter, 1999-2] and [Queluz et al, 2000]. Few other techniques can be found in [Kobayashi, 1997], [Swanson et al, 1998-1], [Hartung et al, 1999-2], [Wolfgang et al, 1999] and [Langelaar et al, 2000].

Blind versus non-blind recovery

Browsing the literature, it can be remarked that initially most of the watermarking techniques were non-blind, requiring the use of the original in the detection process. Later these techniques were adapted to work with blind recovery. Most of the watermarking techniques used the entire image/frame without dividing it into small blocks, later evolving into block based schemes.

Perceptual considerations

From the perspective of watermark embedding, most of the early techniques relied on uniform marking, with an amplification factor experimentally tweaked for acceptable visibility.

The HVS based marking started to appear in the form of region classification, different edge/gradient detection algorithms and different other empirical measures of the local activity within a block/area. It must be remarked here that all these methods are more or less empirical, due to the lack of HVS models in the spatial domain.

Capacity of the watermarking system

From the capacity point of view, most of the watermarking schemes were capable of hiding only one data bit, or in other words they were only capable to tell if a certain watermark was found or not in the media. Quite late, the multi-bit schemes finally arrived. Another major improvement which surprisingly was adopted quite late was the use of pre-filtering as a mean of reducing the cross-talk between the original media and the watermark.

2.4.2 Watermarking in the DCT Domain

The DCT domain is far the most popular one, from several reasons. One reason is that all the major compression techniques were developed in the DCT domain (JPEG, MJPEG, MPEG1, MPEG2, H26x) and therefore the image processing community was familiar with it. Much research was carried out in developing various perceptual models for the DCT domain, and these models could be easily applied to watermarking, since watermarking and compression are very closely related. Since the compression algorithms are well known, one could compensate for it during the watermark embedding process, making the algorithm robust against compression. Furthermore marking in the frequency domain rather than spatial domain has few advantages: better robustness against certain attacks, higher capacity, more close to the HVS and relatively good frequency localisation of the coefficients. Those who are marking in the bit-stream domain (MPEG2) have the additional advantage of the direct bit-stream marking, without decoding and re-encoding the signal.

Spread-spectrum watermarking

One cannot start talking about DCT domain watermarking without mentioning from the beginning the image watermarking scheme developed by Cox [Cox et al, 1995 and 1996]. As mentioned before, Cox officially introduced the use of spread spectrum in digital watermarking, being one of the most cited authors in the watermarking world. In their scheme, the watermark is inserted in the DCT domain. Cox was probably between the first to realise the importance of perceptual marking, and although they do not actually use a visual model, the watermark is embedded in what they regarded to be the most relevant regions of the signal. Therefore, the DCT transform of the entire NxN image was computed and the watermark was inserted in the first n highest magnitude DCT coefficients, excluding the DC coefficient. Due to this arrangement, the necessity of the original image for watermark detection is obvious, which is a serious drawback. The scheme basically embeds only one bit, which is recovered by computing the similarity (which is in fact another name for normalised cross-correlation) between the original and extracted watermarks.

Full-frame versus block-based embedding

Similar with the spatial domain schemes, one can identify techniques which are embedding the mark in the full-frame DCT coefficients [Cox et al, 1995 and 1996], [Barni et al, 1998], [Bartolini et al, 1998-1], [Piva et al, 1998] or in the block-wise manner [Swanson et al, 1996-1 and 1996-2], [Podilchuk et al, 1997-1, 1997-2 and 1998], [Wolfgang et al, 1999], [Ramkumar et al, 1998-1], [Kim et al, 1999], [Zhu et all, 1996], [Tao et al, 1997] and [Hernandez et al, 1998-2, 1999-1 and 1999-2]. The main reason for adopting the block-based approach is robustness to certain geometrical attacks. It is also helpful that most of the existing HVS models are working on a block-based basis.

HVS-based marking

The watermarking research community realised pretty soon that uniform marking is not the best choice of embedding a watermark, since offers quite an unfavourable robustness/visibility report. Inserting a watermark tacking into account some aspects of human vision led to better results, as more energy can be packed into the media while keeping the visibility of the watermark low. After different trials with some more or less empirical methods, finally more advanced HVS models mostly developed in the context of human/machine vision and image/video compression started to be adapted to the needs of watermarking.

Papers like [Carlson et al, 1980], [Legge et al, 1980], [Girod, 1989], [Peterson et al, 1993], [Ahumada et al, 1992], [Jayant et al, 1993-1 and 1993-2], [Watson, 1993], [Watson et al, 1994], [Zhu et al, 1995], [Chou et al, 1995 and 1996], [Eckert et al, 1998] started to become actual and the models described started to be adapted and applied to watermarking, leading to important robustness gains.

Cox is underlining the importance of perceptual embedding in [Cox et al, 1997]. Some of the most important examples of watermarking systems incorporating HVS models are mentioned below.

The watermarking system described in [Swanson et al, 1996-1, 1996-2 and 1997] and [Zhu et al, 1996] is based on the HVS model described in [Zhu et al, 1995]. This is in fact a modified version of the TEL (Tolerable Error Level) model developed by Girod [Girod, 1989]. This rather complex model is based on both spatial and frequency domain masking models.

Some authors preferred to develop their own algorithms, for example [Tao et al, 1997], [Delaigle et al, 1998] and [Bartolini et al, 1998]. Tao used a regional perceptual classifier, which assigns noise-sensitivity indexes to each DCT block. The algorithm exploits luminance, edge and texture masking effects of the HVS and classifies a block into one of 6 categories [Tao et al, 1997]. Following a similar approach, [Bartolini et al, 1998] uses a combination of several different filters and thresholds to build a perceptual mask.

In [Podilchuk et al, 1997-1, 1997-2 and 1998] and [Wolfgang et al, 1999] the authors use a modified version of the JND (Just Noticeable Distortion) model developed at NASA [Peterson et al, 1993], [Ahumada et al, 1992], [Watson, 1993] and [Watson et al, 1994]. It is largely believed that the JND model (sometimes called Watson's model) is the best DCT based HVS model available. This is due to the highly adaptive nature of this model which takes into account the most important masking effects of the HVS: the MTF (Modulation Transfer Function) of the eye, the luminance and the contrast masking. The model is capable to accurately estimate a JND level for each DCT coefficient within a block in contrast with the other HVS models which usually assign only one threshold for the entire block. [Kim et al, 1999] improves the scheme described in [Podilchuk et al, 1997-1, 1997-2 and 1998] and [Wolfgang et al, 1999] by extending the JND model to account for another HVS masking effect: lateral inhibition masking.

Several other papers describing watermarking systems based on HVS models could be quoted and few other visual models as well, but they are not directly relevant with the work described in this thesis and therefore they will not be presented here. The interested reader could find them in these comprehensive watermarking reviews: [Kobayashi, 1997], [Swanson et al, 1998-1], [Hartung et al, 1999-2], [Wolfgang et al, 1997 and 1999] and [Langelaar et al, 2000]. A review of existing visual models could be found in [Jayant et al, 1993-1, 1993-2] and [Eckert et al, 1998].

Detection (recovery) of the watermark

Many techniques are requiring the original in the detection process [Cox et al, 1995 and 1996], [Swanson et al, 1996-1], [Tao et al, 1997], [Podilchuk et al, 1997-1 and 1997-2], [Wolfgang et al, 1999] and [Kim et al, 1999] while many other are blind [Swanson et al, 1996-2], [Barni et al, 1998-3], [Bartolini et al, 1998], [Piva et al, 1998], [Ramkumar et al, 1998-1], [Wolfgang et al, 1999] and [Zhu et al, 1996].

For example in [Barni et al, 1998-3] the authors use a modified form of Cox's technique which does not require the original for recovery. They insert the watermark in a known fixed location: the coefficients from the $\left(L+1\right)th$ to the $\left(L+M\right)th$ are taken according with the zigzag ordering of the DCT spectrum, where the first L coefficients are skipped to achieve perceptual invisibility of the mark.

Speaking by watermark recovery, much work was carried during the time in order to improve the reliability of the detectors and to develop better detection models and strategies. In particular, many researchers used the existing optimum detection theory developed in the context of communications [Hernandez et al, 1998-2, 1999-1, 1999-2 and 2000-1], [Barni et al, 1998-1 and 1998-2], [Piva et al, 1998 and 2000], [Linnartz et al, 1997], [Robert et al, 2000].

Watermark embedding techniques

Other researchers concentrated their efforts to improve the "modulation" techniques, or in other words searched for better ways of casting the watermark.

For example [Smith et al, 1996] suggested the use of differential modulation techniques in watermarking. On the same note, [Lu et al, 1999 and 2000] proposed a scheme called "Cocktail watermarking" which embeds one data bit in the signs of two coefficients or blocks. They defined four possible types of modulations: $Modu(+,+)$, $Modu(+,-)$, $Modu(-,+)$ and $Modu(-,-)$, where $Modu(+/-,-/+)$ represents a positive/negative transformed coefficient modulated with a negative/positive watermark quantity.

Analysing the influence of a number of attacks in order to see how the coefficients are modified, the authors claim that different attacks are affecting the magnitude of the coefficients in a biased way, and therefore by using the appropriate form of modulation (positive modulation or negative modulation function of the attack) the detector response increases. The embedding is based on a HVS model but the capacity of the scheme is only one data bit. Their technique was applied to both DCT and Wavelet coefficients.

Communication theory and capacity

Most of the early techniques are capable of inserting only one data bit [Cox et al, 1995 and 1996], [Swanson et al, 1996-1], [Tao et al, 1997], [Podilchuk et al, 1997-1 and 1997-2], [Wolfgang et al, 1999], [Barni et al, 1998-1, 1998-2 and 1998-3], [Bartolini et al, 1998] and [Kim et al, 1999] while most of the new techniques are capable of embedding a multi-bit watermark [Swanson et al, 1996-2], [Barni et al, capacity 1999-1], [Perez-Gonzales et al, 2001], [Ramkumar et al, 1998-2 and 1999] and [Hernandez et al, 2000-2].

In 1996, Smith and Comisky raised the capital question: How many bits can we hide into an image? They showed for the first time [Smith et al, 1996], using communication theory and in particular Shannon's channel capacity that the maximum capacity of an image is quite large. This is remarkable bearing in mind that the watermarking was only at its beginnings.

Error correction codes and watermarking

Several years had to pass, until the watermarking research community pushed by the need of more robust watermarks started to see the watermarking as a hidden communication channel which therefore can be protected by FEC (Forward Error-correction Codes) in order to improve the robustness [Mittelholzer, 1999], [Ramkumar et al, 1998-2 and 1999], [Cox et al, 1999], [Perez-Gonzales et al, 2001], [Hernandez et al, 1998-1, and 2000-2], [Barni et al, 1999-1], [Baudry et al, 2001], [Moulin et al, 2001] and [Fei et al, 2001]. The choice of the codes ranges from the most basic ones to the powerful Turbo codes. The use of coding led to significantly better results. Once with the introduction of FEC, the performance of the watermarking schemes begun to be measured in terms of BER (Bit Error Rate).

Video watermarking

Excepting the MPEG2 bit stream watermarking, which does not make the object of this investigation, and in spite of few papers having the word "video" in their title but which in fact are not dealing with the video at all, the DCT-based video watermarking schemes are almost inexistent.

The exception from the rule is the object-based scheme described in [Swanson et al, 1997] which uses segmentation algorithms and HVS models, in a block-based approach. The technique used in their scheme is similar with the MPEG motion tracking. The HVS model is the rather complex TEL model developed by Girod [Girod, 1989]. The results of the scheme are not very bad, but the scheme has a major drawback: it is capable to embed only one single data bit.

Another exception is [Busch et al, 1999] where the authors applied an already known DCT block-based technique developed for still-image watermarking to video. The scheme embeds the watermark only in the luminance component. In order to improve the invisibility of the watermark, the embedding is performed only in those blocks qualified as appropriate by a block activity measure. This is quite a poor choice of HVS model, and as a result the performance of the scheme is rather weak. The authors suggest the use of frame averaging in order to increase the robustness of their scheme, but even so, overall the scheme gives only poor results. Even when 64 data bits are embedded in 50 consecutive frames, the scheme still yields quite a high BER, and the authors suggest using an even longer watermark segment.

2.4.3 Fourier Domain Watermarking

Obviously the DFT (Discrete Fourier Transform) was used for watermarking as well. Unlike in the spatial or the DCT domains, the number of papers dealing with DFT marking is quite low. Each transform domain has its own advantages and disadvantages and the DFT is a particularly good example.

The DFT is shift (translation) invariant, or in other words cyclic shifts of the image in the spatial domain do not affect the magnitude of the DFT and therefore a watermark embedded in the DFT domain will be shift invariant. Of course this is a highly desirable property.

On the other hand, as [O'Ruanaidh et al, 1996] shows, due to its complex nature, the DFT offers the possibility of watermarking the magnitude (amplitude) or the phase (at least in theory). The phase is far more important than the magnitude of the DFT values for the intelligibility of an image, so embedding a watermark in the most important component of an image is very good since any attempts of removing the watermark will lead to heavy artefacts. Moreover, as known from the communication theory, the phase modulation often possesses superior noise immunity in comparison with amplitude modulation.

Based on these observations, in [O'Ruanaidh et al, 1996] the authors decide to mark the phase rather than the magnitude of the DFT coefficients. Well, this idea was quickly dropped, and never pursued again, without explaining the reason. Experiments show that probably one of the reasons was the sensitivity of the phase to JPEG and MPEG attacks.

On the other hand, another major disadvantage of both phase and magnitude marking is the fact that in order to obtain a real image after IDFT, the following symmetry conditions must be fulfilled: changes in magnitude must preserve the positive symmetry of the Fourier coefficients and changes in phase must preserve the negative symmetry of the Fourier coefficients. These symmetry requirements are basically halving the watermarking space and therefore the capacity, being a serious drawback. Furthermore, the lack of HVS models in the Fourier domain is another drawback of the FFT-based watermarking.

If the phase marking is almost inexistent in the literature, the magnitude marking instead has a better faith [Licks et al, 1999 and 2000], [Solachidis et al, 1999], [Piva et al, 2000], [Ramkumar et al, 1998-1 and 1999], [Deguillaume et al, 1999].

In an attempt to exploit the DFT properties, papers like [Solachidis et al, 1999] propose to embed the watermark in a circularly symmetric manner. The obvious advantage is the robustness to certain geometrical attacks like cropping, shifting and rotation for example, but the capacity of the scheme is quite low.

In a rare approach, described in [Deguillaume et al, 1999], the authors propose to embed a spread spectrum watermark into 3-D DFT blocks of video, by employing a 3-D DFT and adding the watermark to the transform coefficients. Additionally they embed a template which is easy to detect even under geometric attacks, but overall the scheme gives only modest results.

Although a shift in the spatial domain does not affect the magnitude of the Fourier coefficients, it will instead affect the phase: a shift in spatial domain is equivalent with a phase shift in the Fourier domain. Furthermore, according to the convolution theorem, cross-correlation in spatial domain is equivalent with multiplication in the FFT domain, and vice versa. As a result, the FFT transform is often used for implementing fast cross-correlators. Since a sliding correlator (e.g. a cross-correlator which is able to search for the right position of the watermark in an attacked image) is very computationally expensive (section 3.2.2), the efficiency of the FFT correlators is particularly welcomed. An example of such a correlator is SPOMF (Symmetrical Phase Only Matched Filter), which was already mentioned and which will be used in Chapter 7.

Another particular case which involves the use of Fourier transform is represented by the RST (Rotation, Scaling and Translation) invariant watermarking schemes [Kutter, 1998], [O'Ruanaidh et al, 1998], [Deguillaume et al, 1999] and [Lin et al, 2000]. This case will be discussed in Chapter 7.

2.4.4 Watermarking in the Wavelet Domain

As the watermarking literature suggests and this thesis confirms, the watermarking in wavelet domain is one of the best choices available. The wavelet transform is starting to become more and more popular, as the researchers are starting to be aware of the multiple advantages offered by this transform. These advantages are extensively presented in Chapter 6. Although some authors suggested the use of wavelet watermarking techniques few years before this trend started to appear in the research community, as most of the ideas proposed ahead of their on time, they passed almost unnoticed.

Like any other watermarking scheme, a DWT-based scheme can be characterised by blind/non-blind recovery, one bit/multi-bit watermark and HVS/non-HVS watermark embedding. In the DWT context, the HVS-based embedding is not as critical as in the other domains, due to the fact that the wavelet transform is almost a HVS model by itself. For example the human eye is less sensitive to noise in high resolution DWT bands (level 1) and especially in the DWT bands having an orientation of 45^o (i.e., HH bands).

Another motive for watermarking in wavelet domain is the JPEG 2000 standard, based on the so called embedded zero-tree wavelet coding (EZW), and which will replace sooner or later the old JPEG standard due to its reduced visibility artefacts and better compression. One major advantage compared with its closest competitor - the DCT compression/watermarking methods - is the absence of those very annoying blocking artefacts characteristic to all DCT compression/watermarking schemes. The reason for this big advantage is that DWT is not a block based transform. Furthermore, the DWT is easier to compute than the DCT. Another advantage related with the JPEG 2000 is the ability of tweaking the watermarking algorithm in such a way that becomes robust to JPEG 2000 compression.

The first paper to propose the use of wavelet transform was [Boland et al, 1995] followed by [Podilchuk et al, 1997-1 and 1998], [Swanson et al, 1998-2], [Inoue et al, 1998], [Wolfgang et al, 1999], [Kundur et al, 1997 and 1998], [Lin et al, 1998], [Jayawardena et al, 2000], [Lumini et al, 2000], [Loo et al, 2000], [Dugad et al, 1998], [Pereira et al, 2000], [Barni et al, 1999-2], [Lee et al, 2000], [Xia et al, 1998], [Wang et al, 1998] and [Tsekeridou et al, 2000].

HVS-based marking

It was mentioned before that using HVS models in wavelet domain is not as critical as in the other domains due to the similarity of the DWT with the HVS. Indeed research into human perception indicates that the retina of the eye splits an image into several frequency channels each spanning a bandwidth of approximately one octave. The signals in these channels are processed independently. With its multi-resolution nature, the wavelet transform separates the image into bands of approximately equal bandwidth on a logarithmic scale [Kundur et al, 1997].

In fact some authors suggested that a HVS model in not even necessary for DWT-based watermarking. Evidently this is not true, since a visual model can pack more energy and reduce the visibility of the watermark which translates to increased capacity and robustness of the watermarking system, but some authors may be right to say that usually in the wavelet domain a simpler model may suffice.

Due to these considerations and because the wavelet transform is much younger than its counterparts and therefore the number of compression methods existent in the wavelet domain is quite low compared with those existing in the DCT domain, the choice of HVS models is much reduced.

The most important wavelet HVS models are those developed in [Lewis et al, 1992] and [Watson et al, 1996 and 1997]. Unlike the advanced JND model designed for the DCT by the same author [Watson, 1993], [Watson et al, 1994], the wavelet model is quite simple, giving only one quantisation factor for each wavelet sub-band. In contrast the Lewis model is much more adaptive and therefore much more complex. [Podilchuk et al, 1997-1, 1997-2 and 1998] was the first paper to introduce a proper HVS model. While the authors employ the simpler Watson model, other papers [Barni et al, 1999-2] are using the more complex Lewis model.

Watermark recovery and system's capacity

From the capacity perspective, amazingly enough, the vast majority of existing techniques are only capable of embedding one single bit. The only schemes capable of embedding a multi-bit watermark are those described in [Loo et al, 2000] and [Pereira et al, 2000]. In fact few schemes are so primitive that they are not even worth mentioning.

Most of the techniques are blind [Inoue et al, 1998], [Lumini et al, 2000], [Loo et al, 2000], [Dugad et al, 1998], [Pereira et al, 2000], [Barni et al, 1999-2], [Wang et al, 1998] and [Tsekeridou et al, 2000] but quite a few of them are requiring the original for detection [Podilchuk et al, 1997-1, 1997-2 and 1998], [Swanson et al, 1998-2], [Xia et al, 1998] and [Lee et al, 2000].

It is quite unbelievable that most of the existing wavelet techniques are either non-blind or are capable of embedding only one data bit. This aspect is even more puzzling considering that most of these techniques appeared quite recently. But the worse is still to come.

Video watermarking schemes

Uncompressed video watermarking, is a rarity in wavelet domain. Basically excluding those papers having "video" in their title just for creating an impression, the video watermarking in wavelet domain resumes to [Swanson et al, 1998-2] and [Lee et al, 2000].

The technique described in [Swanson et al, 1998-2] is quite unique. The authors employ a temporal wavelet transform along the video frames, and watermark the wavelet coefficients in a similar manner with their DCT-based scheme presented in section 2.4.2. Unfortunately the wavelet scheme is even more complex than their DCT scheme and doesn't really take advantage of all the opportunities offered by the wavelet transform. The scheme uses both wavelet transform and the DCT transform of the wavelet coefficients, segmentation algorithms and the complex Girod HVS model, therefore being highly inefficient in terms of computing cost. But the real problem is that the scheme is both non-blind and capable of embedding only one data bit into the video sequence, rendering it more or less useless.

In a more recent scheme [Lee et al, 2000], the watermark is adaptively embedded using a HVS model. Moreover, in order to take advantage of the temporal dimension the algorithm take into account region complexity and motion information. Unfortunately this scheme is capable of embedding only one data bit and even worse, requires the presence of the original in the watermark detection process. The detection method used is the one proposed by Cox in his original image watermarking scheme. Therefore the scheme is highly inappropriate to any practical use and is pretty weak in terms of robustness as even the authors acknowledge.

As the case of image watermarking in the wavelet domain wasn't bad enough, even the video watermarking schemes are non-blind and capable of casting only one data bit, transforming them into a publishing exercise rather than a scheme which could be used in practice.

2.5 Contributions

The work presented in this thesis was partly fuelled by a comprehensive literature survey, performed at various stages during the PhD. An integral part of this work was the investigation of different existing watermarking techniques for spatial, DCT and DWT domain watermarking. One of the main conclusions of the literature survey was the importance of regarding and analysing the watermarking system from the communication perspective. Following this path, the watermarking channel is seen as a communication channel and the performance of the watermarking scheme is significantly improved by using powerful error correction codes (Turbo codes). Another main point drawn from this investigation was the importance of the HVS models in any watermarking system. Therefore, various existing visual models developed in the context of image compression were investigated. Simplifying, improving and adapting these models to the requirements of blind digital video watermarking - for both DCT and DWT systems - was an important step in achieving the final result.

This work begun with the developing of two spatial domain, blind video watermarking schemes used as an initial environment appropriate for investigating the effects of pre-filtering, block sizes and HVS models on the performance of a spatial domain watermarking scheme. By using this knowledge and the conclusions drawn from these initial spatial domain watermarking schemes, this work established the foundations for the following high performance, transform domain watermarking systems.

The main contributions of the thesis can be summarised as follows:

- Developing a blind, robust DCT-based video watermarking scheme based on an advanced HVS model, and incorporating state of the art FEC, 3-D marking and 3-D sliding correlation.
- Investigating the use of wavelet transform in digital watermarking and developing a high capacity, robust wavelet-based blind video watermarking system which takes advantage of the properties of the wavelet transform. The performance of the system is further improved by using HVS model and advanced FEC.
- Investigating different techniques against geometrical attacks and developing a system based on image registration techniques which combined with the existing wavelet scheme leads to a highly robust, high capacity, blind video watermarking system capable to withstand a wide range of geometrical attacks. This illustrates a new class of application for image/video watermarking: "blind video registration".