WATERMARKING OF MPEG2 VIDEO STREAMS

Ivan Damnjanović

Submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Department of Electronic Engineering
Queen Mary, University of London

January 2005

To My Father

ABSTRACT

The main field of this study is digital watermarking of compressed video sequences. The targeted area in the field of digital watermarking is data embedding and indexing, which can be used in applications such as video indexing and retrieval. Having in mind that the content in video databases is mainly compressed and that video retrieval applications demand real-time capabilities, this work is focused on efficient and real-time watermark embedding and decoding in the compressed domain.

The implemented watermarking technique is based on well-known spread spectrum techniques. The watermark is spread by a large chip factor, modulated by a pseudo sequence and then added to the DCT coefficients of an MPEG2 sequence. Detection probability was increased by a new **block-wise random watermark bits interleaving**.

The watermark must be embedded in such a way that it does not introduce visual artefacts into the host signal, hence the power of the watermarking signal is bounded by perceptual visibility. The perceptual watermark adjustment is done using the information from a corresponding DCT block in the original sequence. A novel adaptation method is based on the Just Noticeable Difference (JND) model with block classification.

Since a transmission channel has its own particular capacity, the bit-rate of a video stream needs to be chosen to comply with the capacity of the channel. Therefore, watermarking of a compressed video bit-stream must not increase its bit-rate. A novel technique for **bit-rate control on the macroblock level** increased the number of watermarked coefficients in comparison with existing schemes.

To boost the capacity, a state-of-the-art error correction coding technique – turbo coding - was employed. The watermarking channel has a small signal-to-noise ratio and a potentially large bit-error rate due to the noise introduced by the host signal and attacks. In such an environment, it was essential to protect the watermark message by introducing redundant bits, which will be used for error correction.

ACKNOWLEDGEMENTS

There is so many people that I would like to thank for their continuous support in my work. Regardless if they are mentioned here or not, I own all of them the same gratitude.

First, this work would not be possible without my supervisor, Professor Ebroul Izquierdo. His guidance and support, advices and criticism, along with his vision and determination were essential to the accomplishment of my Thesis. In addition, I would like to thank to my co-supervisors Dr. Andrea Cavallaro and Dr. Mike Davies for all discussions and advices.

I am grateful for the privilege to work with the "BUSMAN people", especially Dr. Teddy Furon of INRIA, France and Dr. Alan Pearmain and Mr. Damien Papworth of Queen Mary, University of London.

I would like to thank to the whole Electronic Engineering Department at Queen Mary, University of London. To all people in the Multimedia and Vision lab that I have pleasure to share the last few years with. To Mr. Naeem Ramzan for exchanging ideas and the work we did together. To all of them that greeted me when I came to the MMV lab, especially to Miguel, Andres, Winyu, Maria, Andy. To all friends that will still be there when I leave the MMV: Divna, Marta, Nikola, Emilio, Gaban, Toni…

My special gratitude goes to Dr. Veselin Rakočević and Dr. Janko Ćalić for their advices, continious support and friendship.

To all friends and family for bringing constant joy in my life.

Above all, I would like to thank to the three women I love most - my mother Mikica, my sister Saša and my beloved Tratinčica, for their love, support and understanding.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Motivation and Objectives

The rise of digital information technology permanently transformed modern society. People worldwide can easily create, manipulate, store and enjoy a wide range of multimedia data. New devices, such as Personal Computers (PC), Compact Discs (CD), Digital Versatile Disks (DVD), portable personal multimedia devices enabled higher quality while reducing the costs of creation, processing and transmission of digital audio, image and video content. In addition, the Internet as a global network offers an easy way of worldwide delivery and exchange of digital media. However, these advantages have brought a major concern and number of challenges that need to be resolved:

- The ease of creating perfect copies of digital multimedia material raises the issue of copyright protection.

- Using a wide range of available editing software one can easy alter multimedia content actualizing the integrity protection issue.

- An opportunity to make and distribute an unlimited number of high quality copies of copyrighted material stimulated the need for a good copy protection mechanism.

Digital watermarking is one possible solution to these problems. At the beginning of the 1990s, interest in digital watermarking expanded rapidly and digital watermarking is now one of the hottest topics in the signal processing community. Although it all started merely to address the problem of copyrights, this expansion in the interest in watermarking led to a number of other watermarking applications and possible research areas, such as authentication, broadcast monitoring, tracing dishonest users or data embedding.

The major driving force for this research is the use of digital watermarking for data embedding and indexing applications. These applications are strongly connected to other interesting research areas such as video indexing and retrieval. The main idea is to have a unique identification number embedded in a video sequence that can later be extracted and

sent to a metadata server to retrieve additional information about the video. Since the video content is mainly stored in compressed format and because this kind of application requires real-time processing, this research is based on watermarking of compressed video.

Digital video watermarking is still a relatively unexplored area of digital watermarking. Due to its complexity, it still draws less attention than image watermarking. This is in contrast to image watermarking where powerful and very robust techniques and solutions already exist. There are still a number of questions in video watermarking research regarding tradeoffs between robustness and computational costs as well as robustness, imperceptibility and capacity.

Having this in mind, the main objective is to make advances in video watermarking technology with a qualitative study of watermarking in the compressed domain. Therefore, the research started with the thorough analysis of applications, their requirements and state-of-the-art techniques that have emerged in digital watermarking. In parallel, the thorough study of the MPEG-2 compression standard and techniques proposed for watermarking of MPEG-2 video sequences was conducted. The targeted application - data embedding and indexing of huge video databases - has its own requirements that needed to be identified. Through the research of state-of-the-art digital watermarking, we came to the conclusion that a novel spread spectrum watermarking technique is arguably most suitable for a given application and its requirements. However, it was obvious that spread spectrum technology needs to be studied in detail and that significant improvements are necessary. In that way, an additional set of objectives was specified:

- Watermark embedding in MPEG-2 sequences needs to be studied in detail with the special attention to the limitations created by the requirements of the MPEG-2 standard. Novel solutions are needed in order to increase the amount of embedded information with respect to, for example, the required bit-rate of a sequence.

- A thorough study of perceptual adjustment and novel solutions are needed to maximize the capacity and robustness of embedded information at the same time minimizing degradations introduced into the video picture.

- Given the limited watermark power due to the compression and imperceptibility requirements, some form of error correction coding is required to protect the binary watermark message.

1.2 Contribution of the Thesis

Following the guidelines given above, a spread spectrum watermarking technique for MPEG-2 video streams has been developed. Using this technique, it is possible to embed medium capacity messages (few hundred bits) into a 5 seconds video segment, in a highly imperceptible way, that is also robust to a wide range of non-intentional attacks. To achieve this, the technique was studied in its every detail and almost every part of the technique introduces novelty:

- **Block-wise random watermark bits interleaving** – this technique was introduced in the watermark composition scheme. Instead of spreading bits one by one and having a watermark bit spread through a part of the frame, the bits are spread randomly through the frame. In that way, the bits are evenly spread through the textured, edge and plain area of the video frame, giving equal detection probabilities for each watermark bit. This technique can be adapted to improve the detection characteristics of any other watermarking system (image, video or audio) that uses spread spectrum and perceptual adjustment methods.

- **Bit-rate control on the macroblock level** – An important characteristic of a MPEG-2 video stream is its bit-rate that must be preserved after any operation that the sequence might undergo. Contrary to the previous models that control bit-rate by comparing VLC codes of watermarked and original coefficients one by one, this technique compares sizes of the watermarked and original macroblock. The percentage of altered coefficients is significantly increased. This study is beneficial to techniques that intend to alter streams with similar bit-rate preserving requirements.

- **Perceptual adjustment based on the Just Noticeable Difference model with block classification** – The widely used Watson DCT model [Wat93] and an improved model based on block classification [Zha05] are thoroughly analyzed. Based on this research, a new model, capitalizing on good characteristics of these two models, has been proposed. The perceptual quality of the new model is comparable with the previous model, while at the same time the model showed significant capacity improvements. The model is applicable to all DCT based techniques.

- **Duo-binary protection of watermarking channel** – Capacity enhancement and protection of the watermark message using two state-of-the-art turbo coders have been studied. Employing a duo-binary turbo code doubled the capacity of the spread spectrum watermarking technique. Duo-binary protection gave near optimal utilization of the available signal to noise ratio in the watermarking channel. To the best knowledge of the author, this is the first study of the application of duo-binary turbo coders to watermarking.

The spread spectrum technique described in the thesis with all improvements, its implementation and evaluation issues, has been presented at a number of conferences. Beyond that, it contributed to the successful completion of the EU IST "BUSMAN" project. The list of author's publications is given on the page 123.

1.3 Structure of the Thesis

The thesis is divided into eight chapters. After the short introduction to the problem that is presented in this chapter, a short summary of other chapters is given as follows:

Chapter 2 gives an overview of digital watermarking. We start with the statement of the problem and a definition of digital watermarking. To highlight the importance of digital watermarking, possible applications and their requirements are studied in detail. In addition, a survey of techniques that are used independently of the embedding domain, is given to show the tools that can be used to tackle different problems that one might experience in digital watermarking.

Chapter 3 begins with a general introduction to the MPEG-2 compression standard. It provides an overview of MPEG-2 coding techniques used for removing spatial and temporal redundancies as well as an overview of the layers of MPEG-2 video streams. In the second part of Chapter 3, state-of-the–art MPEG-2 watermarking techniques were studied, based on the embedding domain, computational costs and robustness.

Chapter 4 brings methods that are used, in particular technical aspects of bit spreading and adding to the DCT coefficients of an MPEG-2 video sequence. This chapter brings detailed description and evaluation of **block-wise random watermark bits interleaving** as a novel method for bit spreading. Important aspects of embedding in the DCT domain are described, followed by a novel method for bit-rate preservation called **bit-rate control**

on the macroblock level. The chapter ends with a detailed mathematical analysis of watermark decoding, error rate and capacity boundaries.

Detailed description of considered perceptual adjustment models is given in Chapter 5. It starts with an analysis of the Human Visual System (HVS) and its imperfections that can be exploited to facilitate the perceptual adjustment. The Watson JND model in DCT domain and an improved JND model based on block classification are studied in detail, followed by a description of the proposed model that takes advantage of the good characteristics of these two models. Finally, detailed evaluation and comparison in terms of imperceptibility and capacities is given at the end of the chapter.

Chapter 6 presents in depth analysis of error correction coding (ECC) applicability to the given watermarking model. Achievable capacity rates using ECC for a given signal-to-noise ratio in the channel are first theoretically analysed. After that, a description of parallel-concatenated convolutional coder is presented as well as its implementation issues. In addition, the protection of the watermarking channel with a state-of-the-art duo-binary turbo coder is observed and the robustness of the technique to transcoding is evaluated.

Chapter 7 gives a description of the BUSMAN multimedia system and aspects of the implementation of the presented watermarking technique into the BUSMAN system. Robustness of the implemented technique in the context of the BUSMAN system had been tested to a wide range of attacks and results are presented at the end of the chapter.

Chapter 8 provides the conclusion of the thesis. A summary of implemented techniques and experimental results are presented. Finally, given the results and conclusions of this work, directions and ideas for future research are indicated. In appendices, the derivations of capacity boundary formulas are given, followed by the lists of authors publications and references.

Chapter 2 Digital watermarking

2.1 Introduction

The huge success of the Internet in the past two decades has introduced several issues regarding security. The glorious idea of having a global network, which enables exchange of information and easy access to content databases, also makes possible unauthorized copying and distribution of the content and violation of the content owners' rights. Concerning fraud related to copying and alteration, conventional analogue distribution systems have been equipped with a characteristic built-in protection mechanism. With every new copy, an unavoidable amount of noise is introduced and the quality of the duplicated content is degraded accordingly. In contrast to analogue media, digital multimedia material is completely susceptible to unlimited copying without any degradation and quality losses. The ease of the distribution of high-quality digital material over the Internet brought major concerns about possible financial losses it could cause to copyright holders. At the beginning of the 1990s, the introduction of the Digital Versatile Disc (DVD) was delayed since several media companies refused to provide DVD material in the absence of an adequate solution for the copy protection issue [Haj00].

Several encryption methods were introduced to address this problem [Blo99]. Encryption can provide secure delivery of digital content from the content provider to the consumer. However, once the content is received and decrypted at the consumer side, it is identical to the original data and no longer protected. A malicious consumer is then able to access, edit, copy and distribute an unlimited number of perfect reproductions of digital data, causing significant financial losses to the owner and the content provider. This was not acceptable and it brought major concerns about the protection of digital content, forcing the copyright owners to insist on additional protection. Searching for the technique that will complement encryption, researchers turned to digital watermarking.

Digital watermarking is a technique for embedding a secret signal, a watermark, direct to the digital content in a robust, secure and imperceptible way. Ideally, the embedded watermark should be robust to any processing of the watermarked content, it should be

secure and able to resist intentional attempts to remove it and should not introduce any perceptibly distinguishable difference to the original data. However, robustness and imperceptibility are conflicting in their nature and it is impossible to accomplish simultaneous maximization of both of them. Preserving fidelity and quality of an image means introducing minimal distortion, that is keeping the amount of embedded information minimal and consequently very susceptible to corruption or total removal. On the other hand, maximization of robustness leads to the introduction of large distortions in the image or video and consequently strong perceivable image changes.

The main goal of watermarking research is to maximize the tradeoffs among the basic conflicting properties of data hiding. It comprises maximization of robustness to non-intentional removal by any kind of image or video process as well as to hostile attacks, while keeping the distortion introduced to a minimum in the sense that the amount of change to the document remains indistinguishable to the human eye. In addition, it comprises maximization of the capacity in the sense of the amount of information that may be hidden and recovered from the watermarked material.

Intensive research on this subject and the progress that has been made gave the opportunity to think about using digital watermarking, not only for copyright protection and copy prevention, but for a number of other interesting applications, such as authentication, broadcast monitoring, indexing and data hiding.

2.2 Watermarking Applications

In order to design a good watermarking system, one needs to start with the identification of its application and the requirements that need to be fulfilled. To stress the potentials of digital watermarking, we present brief description of some watermarking applications followed by the requirements for each of them.

Copyright protection/Proof of ownership - Watermarking in the copyright protection application is used for protection of intellectual properties. The owner can embed watermark, which will represent the copyright information. This watermark can prove his ownership in the legal battle in the court of law when someone is questioning his copyrights on the particular material.

The watermark algorithm used for copyright protection must be highly robust. First of all, it needs to be robust against any editing process the watermarked data may undergo, such as compression, editing for quality enhancement purposes, digital-analogue conversion, rotation, scaling, translation, etc. Beyond that, in the copyright protection malicious attacks are inevitable. An adversary will try various attacks in order to forge the ownership or at least to destroy or corrupt the watermark, so it cannot be used as an evidence of copyright. In such an adverse environment, since it is hard to predict what will be used in an attempt to destroy the watermark, it will be hard to make the watermark definitely robust or at least to evaluate the level of its robustness.

Such a watermark needs to be imperceptible too, and it must not introduce any noticeable distortion to the original work. The copyright owner wants to preserve the quality of its digital material, since any quality loss will reduce the value of the work. Fortunately, to compensate for high imperceptibility and high robustness a trade-off on behalf of the capacity can be made. For the protection of the copyrights, it is reasonable to embed some form of identification number such as ISBN-International Standard Book Numbering, which consists of ten digits. In addition, the watermark should contain information about the year of copyright, the permission granted and the rating for the work [Kut99]. In that way, we can assume that amount of information that needs to be embedded in the host data should be 60 [Fri99] to 70 [Kut99] bits. Finally, in this kind of application, it is desirable to use complex techniques for embedding and detection such as extensive search, since real-time embedding and detection is not important, while it is highly important to unambiguously identify the owner of the data.

Copy protection - For copy protection purposes, a watermark can contain information that can control digital recording devices [Lan98]. That information can be a copy-prohibit bit that will forbid or let the system to record and make another copy of the data. For example, digital audio recording systems such as DAT or minidisk are equipped with a copy protection system SCMS (Serial Copy Management System) [IEC89]. This system sets a "copy" bit in all copies preventing further copying of the digital material.

The capacity of the watermarking system used for copy protection must be at least one bit, but it is desirable to be up to few hundreds bits, since in that way special options such as time-stamps may be included [Lan96]. The watermark needs to be extremely robust against

removal, since attacking rate is very high. In addition to this, imperceptibility requirements need to be met.

Finally, embedding could be off-line and it is preferable to use a computationally expensive model for embedding to make the watermark as robust as possible. On the other hand, watermark detection must be real-time and without the original unwatermarked material.

Integrity Authentication - The goal of the watermarking system used for authentication is to assure that data have not been changed intentionally. In other words, one can expect that an image will be transformed during its life (compressed, filtered, etc.), but must protect image actual semantic meaning [Rey02]. It is crucial to assure that the image is authentic if for example it is used as a proof in a trial, printed in a journal, or as part of the medical records of a person applying for a life insurance.

A fragile watermark system can be used to determine whether the data has been altered by a malicious manipulation, but at the same time, it needs to be tolerant to lossy compression or any other non-intentional alteration. It should be able to locate any alteration of the original data. Desirably, the watermark should contain information that will be used for reconstruction of altered or corrupted regions in order to allow at least partial knowledge of the original information in the corrupted area. To accommodate these requirements, the capacity of the watermark needs to be relatively high.

An asymmetric watermarking algorithm has to be applied. For the authentication purposes, only the author of the work can embed a watermark in order to protect it, but any other interested party should be able to check its integrity. In addition, the embedded authentication data should not be perceivable under normal viewing conditions [Wu98]. Although there is no need for real-time embedding or detection, the watermark extraction has to be performed without the original data, since the availability of the original work will obviously make integrity checking useless.

Broadcast monitoring – An organization dealing with the intellectual property rights would like to monitor the use of its members' content. A broadcast surveillance can check how many times a station broadcasts particular content and charge the station according to the findings. Similarly, an advertiser that purchases the airtime from a broadcaster needs to be sure that their commercials are aired according to an agreement [Cox00]. A web spider can

report the sites using the works of an artist. Afterwards, he can check whether royalty fees were paid.

A watermarking system can be used to insert a unique watermark in a piece of multimedia content prior to broadcasting. A monitoring station can then automatically extract watermarks from the received broadcasts and notify when and on which channel each peace of the content was broadcast. Such a watermarking system needs to embed watermarks in an imperceptible way, but the quality of the material can be degraded if introduced distortions do not exceed distortions introduced by broadcasting process. The watermark should be robust against compression, analogue-digital conversion, other processes used in broadcasting and various forms of noise in the broadcasting channel.

The complete watermarking process is done in a professional environment and there is trust relationship between interest groups involved, so no adversary and intention to remove the watermark should be assumed. Finally, since a watermark is extracted automatically from a broadcast, detection needs to be done in real-time and without the original material.

Fingerprinting – The idea behind digital fingerprinting is to trace dishonest users who have access rights for a piece of content, but may use it for unintended purposes such as redistribution [Tra03]. A serial number or ID of the user can be embedded in the content. When there is illegal copy of the content, the copyright owner is able to retrieve watermark and to identify the dishonest user responsible for this leakage.

The pre-releasing of free copies of new movies for marketing, reviewing and similar purposes is common practice of production companies. These copies are then sent to trusted users who can review the movie and give back their comments. However, a number of movie hits in the recent years became available for free downloading from the Internet even before its release date. To find the place where leakage occurs, a movie company can embed a digital watermark with the records of the user into a copy sent to him, download illegal copy from the Internet and extract the information about the dishonest user.

A fingerprint must not introduce any visible distortion to a picture. It needs to be highly robust against attacks, since it is reasonable to expect that a pirate would try to remove the watermark before selling and distribution of the media. In order to embed the data about

the copy, such as a time and a place of a transaction, user id etc, high capacity of the watermark is needed.

The watermark insertion can be performed in a consumer device, such as pay-per view recording module, and it needs to be done in real-time. On the other hand, decoding process can be performed off-line and desirably using the original work for extraction.

Data embedding and indexing – Production companies, content providers, broadcasters work with a large audio-visual data-bases of annotated material and often exchange a huge amount of the multimedia content. In order to exchange annotations together with the content, it is desirable to use watermarking to embed annotation data into the content. Annotations that are "travelling" with the content can be extracted from the content at the user site, cannot be lost or connected to the wrong content and does not need extra storage capacity.

However, inserting annotations into the content require high capacity watermarking technique. In addition, since this application targets professional users, the embedded watermark has to be imperceptible. Even though intentional attacks are not to be expected and a watermarking algorithm needs to resist only against typical signal processing manipulations, it is still hard to accomplish requirements for extremely high capacity and high imperceptibility.

To reduce the size of a message that needs to be embedded, it is better to embed an index, which will be used by search engines to connect the content with its annotations. An index would be also easier to extract, which makes real-time decoding possible. In addition, embedded annotations cannot be edited without re-embedding them, which would also need a real-time embedding algorithm. In contrast, index that is once embedded does not need to be changed and annotation can be change on the annotations server. Finally, in this application, either annotations or an index needs to be extracted without the original data.

2.3 Watermarking Requirements

In the previous section, we described some common watermarking applications. It was shown that every application has it own requirements. Therefore, there is no set of requirements to be met by all applications, but some general directions can be given and they will be described in this section.

Perceptibility - We saw that in most, if not in all, applications the watermark must be imperceptible. That means the watermark should not affect the quality of the original material. At the same time, the watermarking amplitude needs to be as high as possible for high capacity and high robustness [Har99]. To achieve these opposite goals, the watermarking scheme should exploit imperfections of the Human Visual System (HVS) in the way that humans cannot distinguish the original data from the data with the inserted watermark [Swa98a]. If we directly compare the original material with the one that is watermarked, even the smallest change may become apparent. However, since users normally do not have access to the original material, they cannot perform this comparison. In some applications, it should be sufficient that the modifications go unnoticed as long as the data is not compared with the original [Voy98].

Several objective measures have been proposed to evaluate perceived distortions [Kut99], [Wan04], [Pin04]. These measures however are still imperfect and cannot replace a subjective evaluation where a group of human observers evaluates if introduced distortion is perceivable. Usually, a blind testing is used for subjective evaluation purposes. Images with and without embedded watermark are presented to human observers that are asked to choose the image with higher perceptual quality. As a testing measure, a Just Noticeable Distortion (JND) can be defined as a level of distortion perceivable in fifty percent of the experimental trials [Cox01].

In addition, a watermark must remain unperceivable even after any editing processes on the watermarked material. As an example, the watermarked image can sharpened or undergo any other high-pass filtering, which can amplify the watermark embedded in high frequencies of the image spectrum. Other operations such as zooming, contrast or brightness adjustment can make visible otherwise imperceptible watermark.

Capacity (Data payload) - The watermark capacity or data payload refers to the amount of information, that is the number of bits that can be stored within a unit of time or within a unit of an original material (i.e. within one frame of a video) [Cox01].

Required watermark capacity depends on the application and can vary widely. A capacity as small as just one bit may be enough for copy protection application - to instruct a recording device to prevents further copying of the protected material. A copyright protection operation would need 60-70 bits to embed a digital object identifier, while for fingerprinting purposes capacity needs to be increased to include information such as user

identification or timestamp. Finally, for embedding annotations capacity of few hundreds to few thousand bits would be needed.

The maximum capacity that can be embedded will highly depend on the way that embedding algorithm adopts the watermark to the host signal. Moreover, embedding large amount of information decreases the imperceptibility and weakens the robustness severely [Lan00]. Therefore, the balance of capacity, invisibility and robustness is essential part to be controlled.

Robustness - Every application has its own requirements for watermark robustness. Authentication needs fragile watermark, which will show the data were altered. In that case, the watermark does not have to be, on contrary should not be robust against any processing technique or intentional alternations of the data. However, for other applications it is desirable to be able to detect watermark after common signal processing operations. There are two groups of operations (attacks) that might be expected, intentional and non-intentional attacks. List of operations that watermark needs to be robust on, depends on the application and the environment in which watermarking will be used. For example, in broadcast monitoring, the watermark needs only to survive the transmission process.

The watermark should be robust against common processes that cannot be avoided during image storage and transmission, such as compression for efficient storing, filtering for higher quality, digital-analogue and analogue-digital conversions. Some watermarking algorithms can achieve very strong robustness against these attacks. For instance, a watermark can resist JPEG compression with a quality rate as low as 4%, if it is inserted into the low-frequency DCT coefficients of an image [Fri98]. The second class of video processing methods or attacks includes geometric distortion. This kind of attack generally includes image rotation, scaling and translation (RST), cropping, swapping and mirroring, etc. They affect the synchronization between the watermark and the watermarked image when conventional spread spectrum techniques are used. Even very simple geometric distortions can impair the watermark detection [Pet98].

In applications like copyright protection or fingerprinting, we can expect that a dishonest person will have intention to remove watermark, so robustness must be taken much more seriously into consideration. One can try just to mask watermark, so it will not be detectable by the decoder or one can try to remove watermark with techniques such as

collusion attack. If the attacker has several different copies, each with different watermark, he could combine them to produce a copy without watermark. In general, the watermark should be able to resist and should not be removed or altered without degradation of the perceptual quality that will make the material unusable.

Detection method - An important characteristic of watermarking is the restriction of the detection method. In some applications, such as proof of ownership, decoding is usually done by the owner of the original material. He can provide the detector with an unwatermarked version of the material and by that, improve detector performance. This kind of detection method is known as informed detection.

However, it is not always possible to have the original work to aid the detection process. For instance, in copy protection decoding must be done in every consumer's recording device. It is evident that availability of an unwatermarked copy in decoder would undermine the purpose of watermarking system. Decoding which has to be done without knowing the original is usually called blind or public decoding.

As mention before, it is impossible to meet all requirements, especially if we consider computational costs and real-time processing that is desired in some applications. It is important to bear in mind that the above requirements are strongly related one to another. In proof of ownership application, a very robust watermark is required. This can be achieved by making many large modifications to the original data. However, these large modifications will be noticeable, so requirement for high imperceptibility is not met. Hence, when designing any watermarking application it is very important to consider this trade-off between watermarking requirements so an optimal watermarking model can be obtained.

2.4 General digital watermarking model

Watermarking can be seen as communication of a message from embedder through the noisy channel to the watermarking receiver. The noise in the channel is originated by a number of intentional and non-intentional attacks that might occur between the embedding and extraction process. The most general watermarking model is depicted in Figure 1.

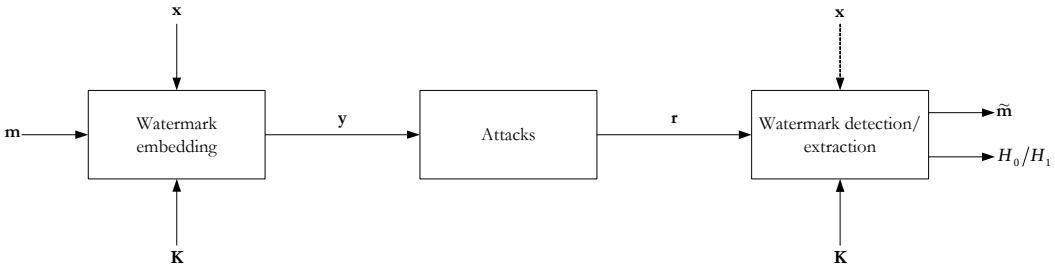

Figure 1. Principle watermarking scheme

The input for the embedding process is the watermark, the host signal and one or more secret cryptographic keys. The watermark message can be text, number, an image or similar. The input host signal may be compressed or uncompressed depending on the application. Embedding in the uncompressed data allows more sophisticated alterations of the host data, but when the host data are already compressed and real-time embedding is required, it is inevitable to insert the watermark in the compressed domain.

It is possible to make distinction between informed and blind embedding [Cox01]. The Blind watermark embedder ignores the fact that the host data **x** is known, while informed embedder exploits the side information that is possible to extract from the host data. The watermarking signal **w** to be embedded is created by function f_0 and usually depends on the watermark message **m** and a secret key **K**, but can also, in the case of informed embedding, depend on the host signal **x** (image, video, audio, text) to ensure that the watermark is well adapted to the host:

$$\mathbf{w} = \begin{cases} f_0(\mathbf{m}, \mathbf{K}) & \text{blind embedding} \\ f_0(\mathbf{m}, \mathbf{K}, \mathbf{x}) & \text{informed embedding} \end{cases} \tag{2.4.1}$$

Using the embedding function f_1, the watermark signal **w** is inserted into the host data **x** yielding the watermarked signal **y**:

$$\mathbf{y} = f_1(\mathbf{x}, \mathbf{w}) \tag{2.4.2}$$

To ensure the imperceptibility, watermark must be embedded in such a way that quality difference between the watermarked data **y** and the original data **x**, $d(\mathbf{x}, \mathbf{y})$ is small. Commonly used distance measure is squared Euclidean metric:

$$d_E(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2 \tag{2.4.3}$$

Although this measure is not computationally expensive and can help in some scenarios, it fails to quantify the complexities of human perception, like masking and threshold effects [Jay93].

The watermarked data can undergo further editing processes or even intentional attempts to remove it. These processes produce the attacked data **r** by introducing some acceptable level of distortion $d(\mathbf{x}, \mathbf{r})$. Having defined distortion functions, it is possible to define a set of feasible embedding algorithms and a set of feasible attacks that satisfy distortion constrains:

$$d(\mathbf{x}, \mathbf{y}) \le D_{emb} \tag{2.4.4}$$

$$d(\mathbf{x}, \mathbf{r}) \le D_{att} \tag{2.4.5}$$

Embedding D_{emb} and attacking distortions D_{att} are subject of the game theory introduced in [Mou03] to determine capacity of the watermarking channel and widely used in the works of many authors, such as [Gue03], [Coh02]. Two simple measures are also useful to determine capacity:

Watermark-to-noise ratio: $$WNR = \frac{D_{emb}}{D_{att}} \tag{2.4.6}$$

Watermark-to-document ratio: $$WDR = \frac{D_{emb}}{\|\mathbf{x}\|^2} \tag{2.4.7}$$

The attacked data **r** is input to the watermark receiver where watermark is detected or extracted with a secret key **K** (Figure 1). The watermark extraction (decoding) denotes

retrieving watermark message $\tilde{\mathbf{m}}$ using the secret key **K**, while watermark detection means the hypothesis test between [Egg02]:

- Hypothesis H_0: the received data **r** is not watermarked using the key **K**
- Hypothesis H_1: the received data **r** is watermarked using the key **K**

The original data **x** might be available to the receiver in the case of *informed decoding*. However, more often the receiver performs *blind decoding* without access to the original data:

$$\tilde{\mathbf{m}} = \begin{cases} g(\mathbf{r}, \mathbf{K}) & \text{blind decoding} \\ g(\mathbf{x}, \mathbf{r}, \mathbf{K}) & \text{informed decoding} \end{cases} \tag{2.4.8}$$

2.5 Digital watermarking techniques review

Many techniques and methods have been developed for data hiding in the last fifteen years. These techniques can be classified by nature of the host signal (text, audio, image, video), domain of insertion (spatial, DFT, DCT, wavelet etc.) or perceptibility of the watermark (visible, invisible). This work is focused on embedding invisible watermark in DCT coefficients of an MPEG-2 video stream, therefore review of the techniques specifically used for MPEG-2 video watermarking will be given later in the Chapter 3. The aim of this section is to provide comprehensive review of general techniques that are used for watermarking regardless of the nature of the host signal or the embedding domain. Good reviews of the state-of-the-art in digital watermarking could be also found in recent books and papers on digital watermarking, such as [Haj00], [Cox01], [Egg02], [Mou04], [Lan00], [Moh99].

2.5.1 Least Significant Bit watermarking

Least Significant Bit (LSB) watermarking is one of the earliest techniques for data embedding. This technique is applicable to the host signal $\mathbf{x} = \{x_1, x_2, ..., x_N\}$, where each sample x_i can be represent using *b* bits and takes integer values between 0 and $2^b - 1$. The sequence of *N* bits made of the least significant bits for every value is called the LSB plane. The LSB plane can be changed without introducing significant quality distortions, so it can be used to insert an information sequence.

One of the first methods that utilizes embedding in to the LSB plane was proposed by Turner in [Tur89] for inserting identification code of a Compact Disk. The watermark was embedded into the least significant bits of randomly selected words. To extract the watermark, LSBs of the selected words were compared to corresponding LSBs from the original unwatermarked material. The major drawback is that method has low robustness. The watermark can be easily removed by erasing or randomizing the LSB plane.

For image watermarking, Tirkel et al. [Tir93] proposed an LSB plane method for embedding an "undetectable" watermark to 512x512 8-bit grey scale images. Prior to embedding, the cover image is compressed to 7 bits through histogram manipulation in order to gain full access to the LSB plane with low distortion. The watermark extraction is straightforward, since the watermark in LSB plane is separated from the cover image

information that is in the other seven planes. A modified version of this algorithm is given in [Sch94].

Aura in [Aur96] used the LSB plane of an image for secret communication. The idea was to embed secret encrypted message in to selected bits using pseudo-random permutation function. The authors also argued that the original image needs to be destroyed, otherwise an adversary can easily find which bits are used to embed the message by comparing the watermarked file with the original. Because of intensive use of a random function, three times per hidden bit, the algorithm proves to be computationally expensive.

The advantage of LSB techniques is that they introduced minimal degradation to the cover signal. Although, these techniques can effectively insert enormous amount of data by rejecting the host interference, they are also not robust to even modest attacks, since LSB plane can be easily erased or replaced by random bits without perceivable distortions.

2.5.2 Patchwork watermarking techniques

In order to make a watermark more robust to attacks such as cropping, patchwork embedding can be employed. The patchwork approach is to embed a watermark repeatedly into pseudo randomly selected areas of an image (or patches). Each of the selected areas contains the watermark, so if the watermarked image is cropped, it will still be possible to extract the watermark from the areas that are not cropped [Joh98].

The patchwork approach was firstly used and named by Bender et al in [Ben96]. They proposed to choose pseudo randomly two patches **A** and **B**. The brightness of pixels in the patch **A** is increased, while the brightness of the pixel in the patch **B** is decreased for the same amount. As an example, if there are N pixels in every patch and if pixels a_i in the first patch are increased by one and pixels b_i in the second patch are decreased by one, provided that N is large, expected value of the sum of all differences between pixels in the patches is:

$$S = \sum_{i=1}^{N} (a_i - b_i) = \begin{cases} 2N & \text{watermarked} \\ 0 & \text{unwatermarked} \end{cases} \tag{2.5.1}$$

The major drawback of the technique is the low embedded data rate. The authors also reported robustness on cropping and JPEG compression.

Pitas and Kaskalis proposed an extension of this approach called signature casting on digital images [Pit96], [Pit98]. A digital watermark **S** is a binary pattern, which is the same size as the host image **I**, and consists of the same number of ones and zeros. The image pixels are divided into two subsets:

$$\begin{aligned} \mathbf{A} &= \{x_{nm} \in \mathbf{I}, s_{nm} = 1\} \\ \mathbf{B} &= \{x_{nm} \in \mathbf{I}, s_{nm} = 0\} \end{aligned} \tag{2.5.2}$$

The image is watermarked by changing the pixels in the subset **A** by a positive factor k, giving the watermarked image:

$$\mathbf{I'} = \mathbf{C} \cup \mathbf{B}, \quad \text{where } \mathbf{C} = \{x_{nm} + k, x_{nm} \in \mathbf{A}\} \tag{2.5.3}$$

The watermark is detected by hypothesis testing using the testing statistic q, which is defined as normalized difference between the pixel mean in the subset **C** and the subset **B** :

$$q = \frac{\bar{c} - \bar{b}}{\sigma_C^2 + \sigma_B^2} \tag{2.5.4}$$

Where σ_C^2 and σ_B^2 are the sample variances of the two subsets. A watermark is present if the test statistic is bigger then a certain threshold. The method, as reported by authors, is robust against additive noise attack and against JPEG compression to ratios up to 4:1.

Langelaar et al proposed an improved model [Lan96], [Lan97] based on the work of Pitas and Kaskalis and called iterative spatial labelling method. They argued that dividing the image into blocks and searching for an optimal embedding level k for each block, rather then using the same k for the whole image, can improve the capacity and the robustness of the watermarking model.

They first divided the image into blocks that have width and height as multiples of 8. A block **B** is pseudo randomly selected from the luminance plane of the image and a pseudo random pattern **P** of the same size that consists of ones and zeros, is generated. To embed a watermark bit, **P** is added to the block **B**, if the value of the bit is 1, or subtracted form the block otherwise:

$$\mathbf{R} = \begin{cases} \mathbf{B} - k \times \mathbf{P} & \text{label} - \text{bit } 0 \\ \mathbf{B} + k \times \mathbf{P} & \text{label} - \text{bit } 1 \end{cases} \tag{2.5.5}$$

Let $D_{high} = I_1 - I_0$ be the difference of the mean of all pixels in the block **R** for which a corresponding pixel in **P** is one (I_1) and the mean of all remaining pixels (I_0). Similarly, difference is calculated $D_{low} = \hat{I}_1 - \hat{I}_0$ after the JPEG compression with the predefined quality factor Q. If a '0' is to be embedded, pattern **P** is iteratively subtracted from the block until both differences D_{high} and D_{low} are not below the zero. If a '1' is to be embedded, the pattern is iteratively added until both differences are not above the certain threshold T. To extract an embedded bit the same pattern is generated in the decoder and the means I_0 and I_1 and the difference $D = I_1 - I_0$ are calculated. If the difference is positive the embedded bit is one, otherwise zero.

In this scheme, the visibility of the watermark is highly dependent on the number of ones in the pattern. Patterns are forced to have five times more zeros then ones. This makes the watermark less visible, but at the same time less robust. The authors also reported that the optimal size of the block is 32x32, concerning the trade-off between capacity and robustness. If a 3x3 edge-enhancement filter is used prior to decoding, it was shown that the watermark can resist JPEG compression attack to the quality factor of 75 (5% or less errors).

2.5.3 Spread spectrum and correlation based watermarking

Regarding a relatively weak robustness against compression attack of the first watermarking techniques, it has been argued that a watermark needs to be inserted in perceptually significant components [Cox96]. However, any stronger alteration of these components leads to significant degradations of the cover signal. To overcome this, Cox et al [Cox96] proposed to use a well-known technique from communication theory called spread spectrum, where a narrow band signal is transmitted via a wide band channel.

At the same year, similar ideas were proposed by Tirkel et al. in [Tir96] and Smith and Comiskey in [Smi96]. In the latter, the authors argued that, as in RF communications, signal-to-noise ratio could be traded for bandwidth. In other words, the watermark signal energy can be spread over a wide range of frequencies (or pixels in spatial domain) with low SNR so it will be difficult to detect, intercept or remove. Techniques that utilize pseudo-random watermark sequence and correlation based detection can be also found in some earlier works [Tan90], [Mat94], [Car95], but these techniques failed to acknowledge and to exploit the findings of the communication theory.

In the Cox et al. approach [Cox96], [Cox97], the watermark consists of a sequence of real numbers $\mathbf{w}=w_1,\ldots,w_n$, derived from the normal distribution with zero mean and variance equal 1 that is $\mathbf{w}\sim N(0,1)$. The number n controls spreading of the watermark: as the number of alerted components of the image increases, the amplitude of the alteration decreases. This sequence is inserted to the host data $\mathbf{x}=x_1,\ldots,x_n$ using additive, multiplicative or exponential rule and the watermarked data $\mathbf{y}=y_1,\ldots,y_n$ are obtained:

$$\begin{aligned} y_i &= x_i + \alpha w_i \\ y_i &= x_i \cdot (1 + \alpha w_i) \\ y_i &= x_i \cdot e^{\alpha w_i} \end{aligned} \qquad (2.5.6)$$

A scaling parameter α controls the extent of the host data alteration. The additive rule is the most commonly used, but authors argued that it may not be appropriate if the host data vary widely, such as the case of transform domain watermarking. They proposed to use multiplicative rule with an empirically derived scaling factor $\alpha = 0.1$ to modify 1000 most significant DCT coefficients (excluding the DC component) of the host image. The extraction is done using the components from the original and the watermarked image and inverting the multiplicative formula to obtain possibly corrupted watermark signal **w'**. The watermark is present in the image if the similarity measure $sim(\mathbf{w'},\mathbf{w})$ is bigger then the given threshold T:

$$sim(\mathbf{w'}, \mathbf{w}) = \frac{\mathbf{w'} \cdot \mathbf{w}}{\sqrt{\mathbf{w'} \cdot \mathbf{w'}}} > T \qquad (2.5.7)$$

The authors claim that one needs to embed $O(\sqrt{n / \ln n})$ similar watermarks to have any chance of destroying the watermark, where n is number of altered image components. In addition, they showed that the technique is very robust against rescaling, JPEG compression, dithering, clipping, printing/scanning and collusion attacks. Unfortunately, since the original image is needed in the detection process, the usefulness of this technique is limited to the private watermarking applications, where the original image is available. The other disadvantage of the technique is the low watermark rate. This can be improved by placing different marks to the image, but that will decrease robustness.

To make the spread spectrum technique useful for blind detection, Piva et al. [Piv97], [Piv98] proposed slightly different multiplicative embedding rule. To embed the watermark into the host image **x**, the DCT coefficients are order in zig-zag fashion and N DCT

coefficients are chosen to generate a vector $\mathbf{t}=t_L,\ldots,t_{L+N}$, starting from a coefficient L in order to exclude the low frequencies where degradations are most visible :

$$t'_{L+i} = t_{L+i} + \alpha|t_{L+i}|w_i \tag{2.5.8}$$

The modified vector is reinserted and inverse DCT is performed to obtain the watermarked image **y**. They also exploited a form of visual masking to enhance the robustness of the technique. A pixel-by-pixel weighting factor β_{ij} was introduced to take into account Human Visual System (HVS) sensitivity to the noise in the different regions of the image:

$$y_{ij} = \beta_{ij} y_{ij} + (1-\beta_{ij})x_{ij} \tag{2.5.9}$$

The weighting factor β_{ij} is derived calculating the normalized sample variance for a square block with a centre at pixel y_{ij} with respect to the maximum of all block variances. The presence of the watermark is detected by extracting the vector $\mathbf{t}^*$ in the same manner as in the embedding process, correlating it with the watermark sequence **w** and comparing the correlation z with a predefined threshold S_z:

$$z = \frac{1}{N}\sum_i t_i^* \cdot w_i > S_z \tag{2.5.10}$$

Various different perceptual models for spread spectrum watermarking have been proposed in the literature. Considering that the estimation of the watermark for a decoding purpose amounts to a de-noising problem, stochastic criteria have been derived in [Vol99] for content adaptive watermarking. The texture masking function is related to a so-called *noise visibility function NVF*, which is in turn used to adapt the strength of the mark in the form of $\alpha_i = S_0.(1\text{-}NVF)+S_1.NVF$, where S_0 and S_1 are strength factors respectively in textured and flat regions.

Kutter et al. [Kut97], [Kut98] proposed a method in the spatial domain which inserts the watermark into the blue image component, in the RGB colour space, to maximize the watermark strength while keeping visual artifacts minimal. This concept is applicable to any spread spectrum watermarking in the spatial domain. The authors reported that the method is robust against both classical attacks, such as filtering, and geometrical attacks.

Wolfgang, Podilchuck, Zeng and Delp in [Pod98] and [Wol99] proposed two watermarking techniques for digital images that are based on utilizing visual models, which have been developed for image compression. They used the visual models to determine the upper bounds for watermarking amplitudes. To embed the watermark, additive embedding rule was used in block-based DCT domain and multiresolution wavelet domain. To determine a scaling factor α_i, they used the Watson et al. Just Noticeable Difference (JND) models developed for DCT domain image compression [Wat93] and wavelet-based image compression [Wat96]. In that way, the watermark is embedded with the maximum strength possible to remain transparent and to be extremely robust to common image processing and editing such as JPEG compression, rescaling, and cropping. Further dissemination of the perceptual adjustment model in DCT domain given in [Pod98] and comparison with an improved model is given in the Chapter 5.

The spread spectrum watermarking is arguably the most widely used technique in digital watermarking. Marvel et al. in [Mar99] for steganography purposes combine spread spectrum with image restoration and error correction techniques. Hartung and Girod proposed a spread spectrum watermarking technique in spatial domain for embedding the watermark into an video sequence [Har98]. Spread spectrum watermarking in DCT domain can be found in the works described above but also in [Fri98], [Koh00] and [Mit02]. The authors in [Zhu98], [Rua96] and [Won00] suggest using spread spectrum technique for image watermarking in the wavelet, Fourier transform and log-2-spatio domain respectively.

2.5.4 Watermarking as a communication with side information

The host signal is usually seen as noise from the watermarking point of view. However, since the host signal is known at transmitter it is possible to extract information from the host data used to model the watermark in a way that resolution of the watermark decoding process can be independent of the host data. This approach to digital watermarking was first recognized independently by Chen and Wornell [Che99] and Cox, Miller and McKellips [Cox98]. In the latter paper, authors discussed general concepts based on an early channel with side information studied by Shannon [Sha58]. On the other hand, in [Che99] authors argued that their previously proposed watermarking scheme based on quantization index modulation (QIM) [Che98] can be seen as a special case of an almost-forgotten work by Costa [Cos83].

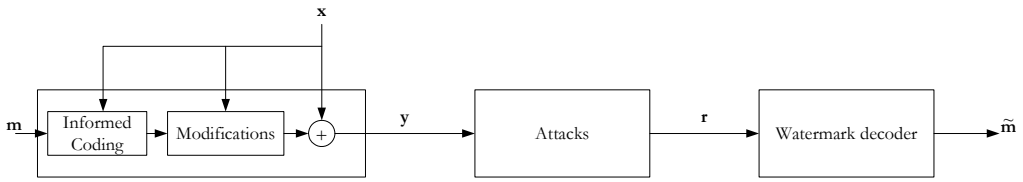

Figure 2. Principal scheme of informed embedding and coding

Cox in [Cox01] distinguished two ways of using the host data as side information: *informed embedding* and *informed coding* (Figure 2). In the informed embedding, the original data are examined in order to modify and prepare the message for embedding. One of the earliest methods that can be recognized as informed embedding is watermark perceptual modelling. More advanced form of informed embedding is optimization of detection process, such as maximizing robustness while maintaining fixed embedding distortion or vice versa.

Informed coding denotes using side information in message coding to select between alternative code vectors mapped to the watermark message, the one that introduce least embedding distortion. Costa in his analogy to writing on a dirty paper [Cos83] showed that if the noise known to the encoder and the channel noise are independent, identically, distributed Gaussian, the capacity is determined only by the power constrains (i.e. amount of ink in Costa analogy or embedding distortion in watermarking) and the channel noise (i.e. attack distortion). This rather surprising result, that the capacity is not influenced by the original data, was derived using a coding method in which a message can be represented by one of the many dissimilar code vectors. Costa's method, also known as Ideal Costa Scheme, is not realistic since it requires large codebooks. However, his result leads to suboptimal, but yet practical approaches, such as dither modulation [Che98] or Miller's dirty paper codes [Mil01].

Miller showed that dirty paper codes, based on suboptimal and finite code book, can improve performance of watermarking, but he also pointed out that dirty paper codes are not suitable for practical applications such as watermarking of MPEG-2 video streams in compressed domain. Extraction of side information from the video stream needs an extra pass through the video, which introduce additional overheads and is opposite to the requirement for fast embedding.

In the context of compressed domain real-time watermarking more appealing are the quantization-based methods, since they are using sample based side information extraction. The basic idea presented in [Che98] is to use the one of two dithered quantizers to quantize the original data sample depending on the embedding bit:

$$y = \begin{cases} Q_0(x) = Q(x + \frac{\Delta}{4}) - \frac{\Delta}{4} & , m = 0 \\ Q_1(x) = Q(x - \frac{\Delta}{4}) + \frac{\Delta}{4} & , m = 1 \end{cases} \tag{2.5.11}$$

Decoding is done by using same quantizers on the received data sample **r** and finding the quantizer point that is closest to **r**. This scheme works perfectly if the channel noise amplitude is smaller then $\frac{\Delta}{4}$, but performs poorly if this level is exceeded. To overcome this problem, Chen and Wornell defined method called distortion compensation [Che01]:

$$y = \begin{cases} Q_0(\alpha x) + (1-\alpha) \cdot x & , m = 0 \\ Q_1(\alpha x) + (1-\alpha) \cdot x & , m = 1 \end{cases} \tag{2.5.12}$$

where $\alpha \in (0,1]$ is a parameter that needs to be optimized according to trade-off between robustness and embedding distortion.

These ideas directed the path of the present watermarking research and are widely studied [Egg00, 02, 03] and discussed [Mou04], [Sta03], [Bar03] in the recent years. However, even though non-effective in theory spread spectrum techniques still outperform host-rejection methods if a StirMark [Pet98], [Pet00] benchmark is used. On the other hand, the authors in [Kov04] argued that the capacity of a spread spectrum technique can be comparable to a quantization-based method at high WNR regime, if a proper stochastic model of the host signal at the encoder is used to shape the watermark. They showed that if Contrast Sensitivity Function (CSF) is used to adapt the watermark spectrum, host interference is significantly weakened leading to a noticeable capacity improvement. Following these ideas about perceptual adjustment, we based our research on improving spread spectrum using a novel perceptual model and state-of the-art error-correction techniques.

Chapter 3 Watermarking of MPEG2 video streams

3.1 MPEG-2 Standard

Since the work presented in this Thesis analyses watermarking of MPEG-2 sequences, this section provides an overview of the MPEG-2 video coding algorithms and standards. The Moving Picture Experts Group (MPEG) is a working group established in 1988 by ISO/IEC [Bha97]. MPEG is in charge of the development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio and their combination. The basic concepts and techniques relevant in the context of the MPEG-2 video compression standard are reviewed in this chapter. A more detailed description of the MPEG standard can be found in [Mit96] as well as ISO/IEC MPEG-2 video standard itself [MPG96].

3.1.1 Coding Techniques

Video sequences contain a significant amount of statistical and subjective redundancy within and between frames. The goal of video source coding is to reduce the bit-rate for storage and transmission by exploring both statistical and subjective redundancies and to encode a minimal amount of information using entropy-coding techniques. The performance of video compression techniques depends on the amount of redundancy contained in the image data as well as on the actual compression techniques used for coding. With practical coding schemes a trade-off between coding performance (high compression with sufficient quality) and implementation complexity is targeted.

We can distinguish two different coding techniques: "lossless" and "lossy" coding of the video data. The aim of "lossless" coding is to reduce image or video data for storage and transmission while retaining the quality of the original images - the decoded image quality is required to be identical to the image quality prior to encoding. In contrast the aim of "lossy" coding techniques is to meet a given target bit-rate for storage and transmission. In these applications, high video compression is achieved by degrading the video quality - the decoded image objective quality is reduced compared to the quality of the original images

prior to encoding. The smaller is the target bit-rate, the higher is the necessary compression of the video data and usually more coding artefacts become visible. The ultimate aim of lossy coding techniques is to optimise image quality for a given target bit rate subject to objective or subjective optimisation criteria.

3.1.2 Spatial and Temporal Redundancy

The MPEG digital video coding techniques are statistical in their nature. Video sequences usually contain statistical redundancies in both temporal and spatial directions. The basic statistical property upon which MPEG compression techniques rely is inter-pixel correlation, including the assumption of simple correlated motion between consecutive frames. Thus, it is assumed that the magnitude of a particular image pixel can be predicted from nearby pixels within the same frame (using Intra-frame coding techniques) or from pixels of a nearby frame (using Inter-frame techniques). The MPEG compression algorithms employ Discrete Cosine Transform (DCT) coding techniques on image blocks of 8x8 pixels to efficiently explore spatial correlations between nearby pixels within the same image. However, if the correlation between pixels in nearby frames is high, i.e. in cases where two consecutive frames have similar or identical content, it is desirable to use Inter-frame DPCM coding techniques employing temporal prediction (motion compensated prediction between frames). In MPEG video coding schemes an adaptive combination of both temporal motion compensated prediction followed by transform coding of the remaining spatial information is used to achieve high data compression (hybrid DPCM/DCT coding of video).

3.1.3 MPEG Chrominance Sampling

Generally, video coding techniques use subsampling and quantization prior to encoding. The basic concept of subsampling is to reduce the dimension of the input video (horizontal dimension and/or vertical dimension) and thus the number of pixels to be coded prior to the encoding process. It is worth noting that for some applications video is also subsampled in temporal direction to reduce frame rate prior to coding. At the receiver, the decoded images are interpolated for display. This technique may be considered as one of the most elementary compression techniques, which also makes use of specific characteristics of the HVS and thus removes subjective redundancy contained in the video data - i.e. the human eye is more sensitive to changes in brightness than to chromaticity changes. Therefore, the MPEG coding schemes first divide the images into YUV

components (one luminance and two chrominance components). Next the chrominance components are subsampled relative to the luminance component with an Y:U:V ratio specific to particular applications (i.e. with the MPEG-2 standard a ratio of 4:1:1 or 4:2:2 is used).

3.1.4 Motion Compensated Prediction

As it was pointed in section 3.1.2, a key element of the inter frame prediction is motion compensation. The concept of motion compensation is based on the estimation of motion between video frames. If there is a motion in a video scene, pixels can be described by a limited number of motion parameters, that is by motion vectors, and difference from pixels in nearby frames.

Usually both prediction error and motion vectors are transmitted to the receiver. However, encoding one motion information with each coded image pixel is generally neither desirable nor necessary. Since the spatial correlation between motion vectors is often high, it is sometimes assumed that one motion vector is representative for the motion of a "block" of adjacent pixels. To this aim images are usually separated into disjoint blocks of pixels (i.e. 16x16 pixels in MPEG-1 and MPEG-2 standards) and only one motion vector is estimated, coded and transmitted for each of these blocks.

In the MPEG compression algorithms the motion compensated prediction techniques are used for reducing temporal redundancies between frames and only the prediction error images - the difference between original images and motion compensated prediction images - are encoded. In general, the correlation between pixels in the motion compensated Inter-frame error images to be coded is reduced compared to the correlation properties of Intra-frames due to the prediction based on the previous coded frame.

3.1.5 Transform Domain Coding

Transform coding has been studied extensively during the last two decades and has become a very popular compression method for still image coding and video coding. The purpose of Transform coding is to de-correlate the Intra- or Inter-frame error image content and to encode Transform coefficients rather than the original pixels of the images. The input images are split into disjoint blocks of pixels **p** (i.e. of size *N*x*N* pixels). The transformation can be represented as a matrix operation using a *N*x*N* Transform matrix **T** to obtain the

*N*x*N* transform coefficients **c** based on a linear, separable and unitary forward transformation.

$$\mathbf{c} = \mathbf{T} \cdot \mathbf{p} \cdot \mathbf{T}^{\mathrm{T}} \tag{3.1.1}$$

Here, $\mathbf{T}^{\mathrm{T}}$ denotes the transpose of the transformation matrix **T**. Note that the transformation is reversible, since the original *N*x*N* block of pixels **p** can be reconstructed using a linear and separable inverse transformation

$$\mathbf{p} = \mathbf{T}^{\mathrm{T}} \cdot \mathbf{c} \cdot \mathbf{T} \tag{3.1.2}$$

Upon many possible alternatives the Discrete Cosine Transform (DCT) applied to smaller image blocks of usually 8x8 pixels has become the most successful transform for still image and video coding [Ahm84]. In fact, DCT based implementations are used in most image and video coding standards due to their high decorrelation performance and the availability of fast DCT algorithms suitable for real time implementations. VLSI implementations that operate at rates suitable for a broad range of video applications are commercially available today.

A major objective of transform coding is to make as many Transform coefficients as possible small enough so that they are insignificant (in terms of statistical and subjective measures) and need not to be coded for transmission. At the same time, it is desirable to minimize statistical dependencies between coefficients with the aim to reduce the amount of bits needed to encode the remaining coefficients.

Figure 3 depicts the variance (energy) of an 8x8 block of Intra-frame DCT coefficients based on the simple statistical model assumption already discussed in equation 3.1.1. The figure depicts the variance distribution of DCT-coefficients "typically" calculated as average over a large number of image blocks. The variance of the DCT coefficients was calculated based on the statistical model used in equation 3.1.1. Coefficients with small variances are less significant for the reconstruction of the image blocks than coefficients with large variances. As may be depicted from Figure 3, on average only a small number of DCT coefficients need to be transmitted to the receiver to obtain a valuable approximate reconstruction of the image blocks. Moreover, the most significant DCT coefficients are concentrated around the upper left corner (low DCT coefficients) and the significance of

the coefficients decays with increased distance. This implies that higher DCT coefficients are less important for reconstruction than lower coefficients.

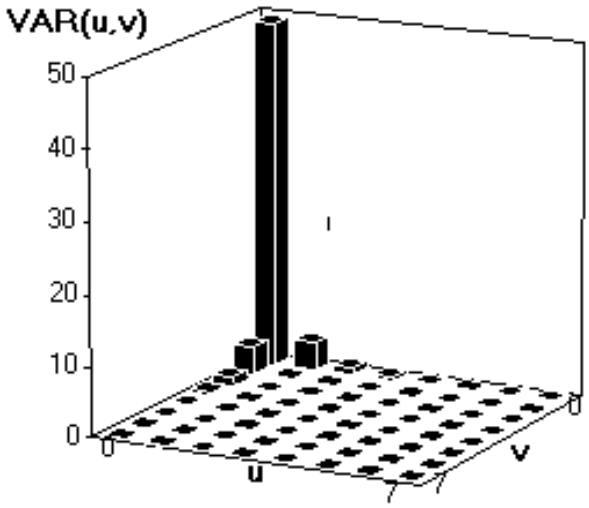

Figure 3. Variance distribution of DCT-coefficients [Sch95]

The DCT is closely related to Discrete Fourier Transform (DFT) and it is of some importance to realize that the DCT coefficients can be given a frequency interpretation close to the DFT. Thus, low DCT coefficients relate to low spatial frequencies within image blocks and high DCT coefficients to higher frequencies. This property is used in MPEG coding schemes to remove subjective redundancies contained in the image data based on human visual systems criteria. Since the human viewer is more sensitive to reconstruction errors related to the low spatial frequencies than to the high frequencies, a frequency adaptive weighting (quantization) of the coefficients according to the human visual perception (perceptual quantization) is often employed to improve the visual quality of the decoded images for a given bit rate.

The combination of the two techniques described above - temporal motion compensated prediction and transform domain coding - can be seen as the key elements of the MPEG coding standards. A third characteristic element of the MPEG algorithms is that these two techniques are processed on small image blocks (of typically 16x16 pixels for motion compensation and 8x8 pixels for DCT coding). To this reason, the MPEG coding algorithms are usually referred to as hybrid block-based DPCM/DCT algorithms.

3.1.6 MPEG-2 video stream

An MPEG-2 stream consists of video, audio and system layer. The system layer wraps around the other two layers and provides, among other things, synchronization of video and audio streams. From the video watermarking point of view, the most interesting is the video stream, which will be briefly described in this section.

A video sequence (Figure 4) always starts with a sequence header and ends with a *sequence end code*. Between this two there is one or more groups of pictures (GOP). Another sequence headers can be inserted anywhere in sequence layer in order to, for example, assist in video editing. Inside a group of pictures, a GOP header consists of the 32-bit *group start code* followed by the time and reference information about the GOP. In addition to the header, the group of pictures must have at least one picture. More often, a GOP consists of an intra picture (I) followed by forward (P) and forward-backward (B) predicted pictures. A picture starts with a picture header that bears information about its temporal reference, coding type (I, P, B) and forward and backward motion vectors scale and precision (in the case of P and B frames). One or more slices follow a picture header.

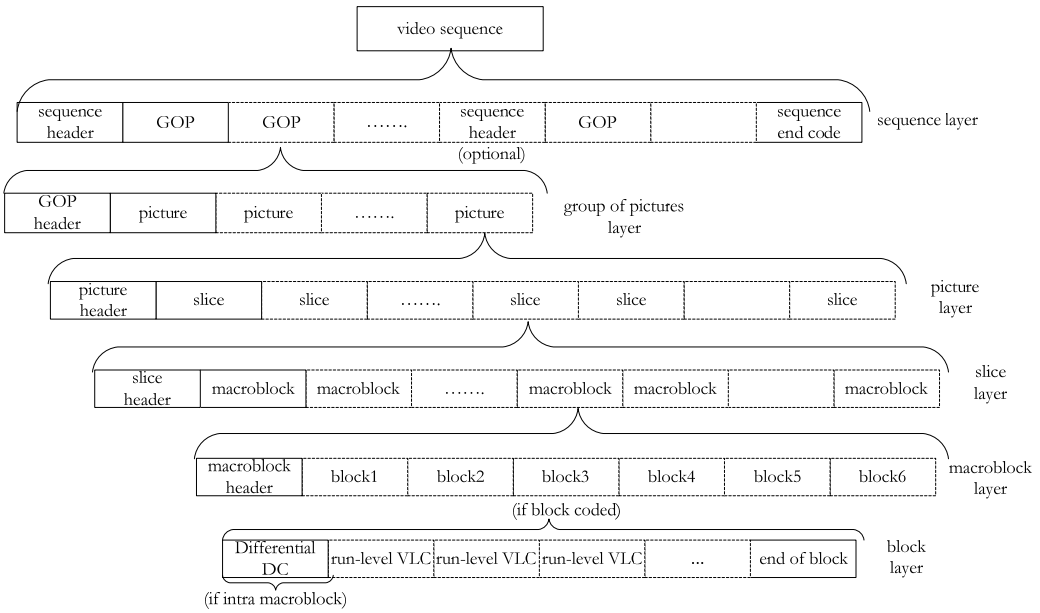

Figure 4. The layers of a video stream [Mit96]

Every picture is divided in to slices. In the beginning of every slice there is necessary data about quantization scale followed by optional extra information about slice and coded macroblocks. The macroblock header provides the information about the macroblock position relative to the position of the previous macroblock in the slice. It also includes the type of the macroblock, quantization scale, the motion vectors (if inter-coded macroblock) and information about which blocks are actually present in the macroblock. Usually, a macroblock consists of four luminance and two chrominance blocks. If a block does not contain any information, it is simply skipped.

A block contains variable-length codes (VLC) of coefficients of an 8x8 DCT block. If a macroblock is intra-coded, the first code in a block is difference between DC coefficient of the block and the prediction made from the DC coefficient of the pervious coded block. In addition, a block may include variable length codes of run-level pairs of AC coefficients, where run is number of zero coefficients that precedes the coefficient and level is quantized value of the coefficient followed by its sign. Since a number of VLC codes may vary, every block must ends with the *block end code*.

3.2 MPEG-2 watermarking techniques – State-of-The-Art

Several techniques have been proposed in the literature aiming at watermarking in the MPEG-2 watermarking. Depending on watermarking application requirements, various domains have been chosen for inserting the watermark (Figure 5). If the watermark needs to be robust against rotation, scaling, translation and diverse intentional attacks, one can choose to decompress video frame and to embed a robust watermark in spatial domain. However, using these techniques, it is hard if not impossible to meet often required real-time watermark embedding and detection. On the other hand, a watermarking application in professional environment may require only moderate robustness (e.g. against transcoding and some video editing), but at the same time may need both embedding and decoding done in the real-time. For these purposes, it is better to decompress a bitstream only partially and to embed the watermark before computationally expensive operations such as inverse DCT and motion compensation. This increase in the speed of the watermarking process is traded for its robustness, since robust algorithms that embed the watermark in different frame statistics and transform domains (DFT, DWT etc) are discarded.

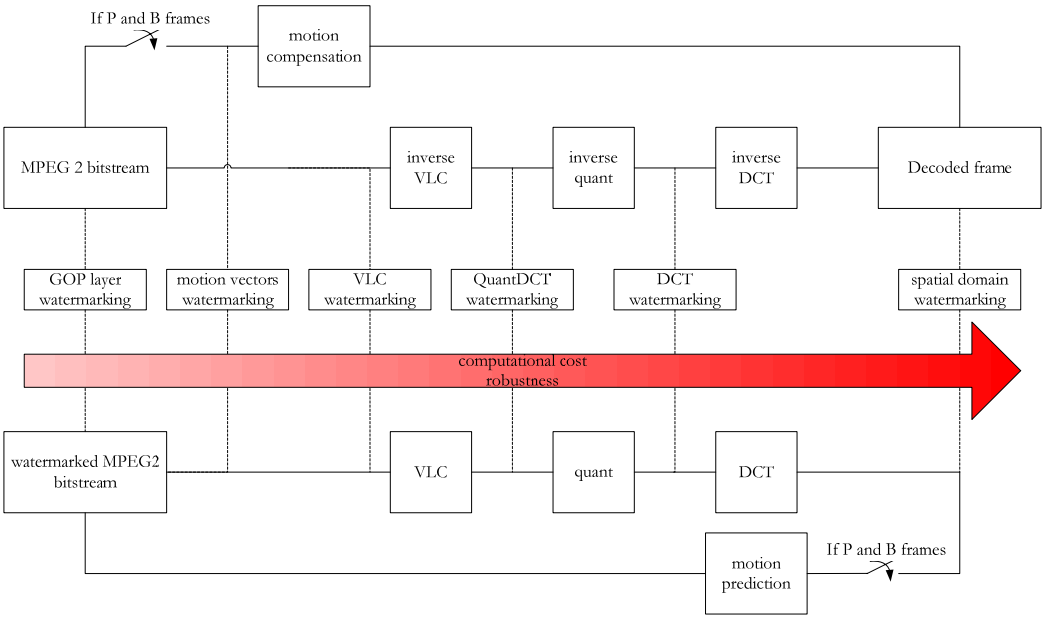

Figure 5. MPEG-2 watermarking- embedding domains

For spatial domain watermarking, the host MPEG-2 stream must undergo the whole decompression process involving inverse variable length coding, inverse quantization, inverse discrete cosine transform and motion compensation. After decompression, any image watermarking technique, for example LSB modification, but resulting watermarked frame must undergo the complete time consuming compression again.

In such a way, Hartung and Girod employed straightforward spread spectrum approach to embed an additive watermark [Har98]. The watermark consisted of a pseudo-random sequence, which is the same size as the video frame and modulated by watermark bits. Depending on a spreading factor, they reported different, but moderate, levels of robustness, such as robustness on further block-wise DCT compression, Gaussian and impulse noise addition and low-pass filtering. In addition, they proposed an extension of their technique for watermarking in compressed domain.

More sophisticated method was proposed by Swanson et al. in [Swa97] and [Swa98c]. They first segment uncompressed video sequence into scenes. Each scene then undergoes temporal wavelet transform giving low-pass and high-pass temporal frames. A perceptually shaped pseudo-sequence, which uniquely represents the author of the sequence, is embedded into each of the temporal wavelet components. Hence, the watermark components that are embedded in the high-pass temporal frames will change rapidly with

time, while the ones in the low-pass frames will change slowly or not change at all. This makes the watermark highly robust against attacks, such as frame averaging, frame dropping, or even detectable from a single frame without knowledge of its location in the sequence. Furthermore, the method is robust against severe additive noise attack and MPEG video compression. However, the low point of the technique is its complexity, which includes forward and backward wavelet transform and complex HVS model based on block-wise DCT.

Deguillaume et al. in [Deg99] and [Deg00] proposed another method that exploits both spatial and temporal dimension of a video sequence. A spread spectrum watermark is added to the coefficients of a 3-D DFT transformed video block. In addition, a template grid is embedded to detect and invert different forms of geometrical and other severe attacks. The authors reported robustness against MPEG-2 compression (watermark was detectable after compression to 3 Mbps), frame-rate changes and aspect-ratio changes. Robustness against frame averaging attack is also achieved by ensuring minimal temporal redundancy of the watermark.

Kalker et al. [Kal99] have developed a video watermarking method that embeds watermark pattern of size 128x128 drawn from Gaussian distribution and repeated to fill the video frame. To avoid visible artifacts, the watermark amplitude is scaled with a scaling factor computed using a Laplacian high-pass filter. For the watermark detection, a correlation detector is used after applying matched filtering to boost the detection performance. To make the watermark shift invariant, cycling convolution in the FFT domain is performed. The watermark can convey up to about 35–50 bits/s and is claimed to be robust against MPEG-2 compression down to 2 Mbps, format conversion, scaling and addition of noise. Following this work, Kalker and Haitsma [Hai01] proposed to insert the watermark by exploiting temporal axis only and changing the mean luminance of a frame. They were targeting threat of illegal copying by handheld video cameras. The watermark was detectable even after severe attack conditions such as handheld camera, recording from different angles, slightly rotated, with zoom changes and with people passing between the camera and the screen.

The more recent spatial domain technique, proposed by Niu et al. [Niu02], aims to robustness against rotation, scaling and translation. Two watermarks are embedded along with the time-axis template to recover after potential geometric distortions. The watermark

information is protected using Hamming error correction coding and Direct Sequence Code Division Multiple Access is introduced to generate the watermarks. The authors reported robustness against the RST attacks, bending and shearing of frames, MPEG-2 lossy compression, color-space conversion, and frame dropping attacks.

All these techniques that demand complete recompression showed impressive results for robustness, but for many applications they are prohibitively expensive. On the other side of the spectrum of video watermarking techniques, are low-complexity methods, which however, are modeled to resist only some weak non-intentional attacks that are specific for the environment in which a method tends to be used. Such a technique was proposed by Linnartz et al. [Lin98] to embed information in the GOP structure of the MPEG-2 compressed video. In MPEG-2 standard, GOP consists of an intra-coded I frame, coded by exploiting only spatial redundancy, followed by a sequence of P and B frames, frames predicted from previous frames and frames predicted from previous and successive frames, respectively. The picture type is signaled in the picture header. In theory, for a GOP of N frames there are 2^{N-1} possible mixtures of P and B frames giving a possibility to embed N-1 bits in to the GOP. Linnartz et al. propose a scheme where they embed 6 bits of information per GOP of 12 frames yielding few bytes per second. The watermark must be embedded during compression, since after compression the GOP structure is already fixed. Also any recompression will remove the watermark completely. Advantage of this technique is however, extremely low complexity.

Another method with negligibly low complexity has been proposed by Jordan et al [Jor97] aiming at watermarking of compressed video. The idea was to embed information by slightly modifying motion vectors that point to flat areas in pseudo-random manner. Since the modified vectors point to blocks that are very similar to blocks pointed by original vectors, the method does not introduce any visible artifacts. As long as the video is in compressed format, the embedded information can be completely retrieved from the motion vectors, but the watermark can be also retrieved after decompression by first recompressing the video. A probability that the motion vectors will be the same, after recompression, is high enough to statistically recover the watermark.

In order to find the proper trade-off between complexity and robustness, Langelaar et al. propose two different information embedding schemes for compressed video [Lan98]. The first method adds the label directly into variable length codes of DCT coefficients. This

method can be viewed as an LSB method that works in the compressed domain. In the MPEG-2 variable-length code tables it is possible to find pairs of codes which represent the same run and levels that deviate only by one from each other. Depending on watermarking bit, one of the two codes in the pair is chosen. The authors reported that they were able to embed up to 7 Kbps into TV resolution video compressed at 8 Mbps by watermarking only I frames and 29 Kbps if inter-coded frames are watermarked too. However, they admit that the label can be removed easily by decompression and recompression with preserving video quality.

The second method proposed by Langelaar et al., named Differential Energy Watermarking (DEW), is more complex, but also more robust. For each information bit to be embedded, a set of $n \in [16,64]$ 8x8 blocks is pseudo-randomly taken from the frame and divided into two subsets of equal size. The energy of the high-frequency content is reduced by removing high-frequency coefficients in one of the subset depending on the watermarking bit. The method has relatively low complexity, since it requires only partial decoding of the bit-stream. The reported capacity is up to 420 bps for TV resolution video encoded at 8 Mbps. The method is relatively robust to MPEG-2 compression. If the watermarked video is compressed to 6 Mbps, only 7% of watermark bits are wrongly decoded, but at 4 Mbps bit-error rate is increased to 38 %. In addition, to avoid visible blurring due to removal of high-frequency content, the embedding parameters must be adjusted carefully. In [Lan01], the authors also proposed to use Hamming error correction coding to improve robustness of the method. It is also worth of mentioning that Hefei Ling et al proposed slightly different technique called Differential Number Watermarking (DNW) [Hef04]. The authors proposed to change number of high-frequency coefficients rather then the energy in the high-frequency content. They claim that their method is less complex, more robust and that capacity is doubled.

Lu et al. proposed a robust real-time technique in VLC domain based on watermarking with side information [Lu02]. For *N* macroblocks in the video stream, they define *u(i)* as a mean value of levels in the i-th macroblock and n_i as a number of VLC codes in i-th macroblock. A watermark bit is embedded into the normalized difference mean value of the macroblock. They also define $\overline{u}(i)$ as a mean filtered value obtained from *u(i)*. The watermark bit is embedded into normalized difference between *u(i)* and corresponding mean filtered value $\overline{u}(i)$. The authors reported relatively good robustness against compression, sharpening, additive noise, frame averaging, frame-rate changing, mean

filtering and I frame dropping. The watermark was detected in 12 seconds for a video that needs 182 seconds to be decoded.

Many authors recognized DCT and quantized DCT domain watermarking as a good trade-off between watermarking complexity and robustness of a watermark. Discrete Cosine Transform (DCT) and inverse DCT are most complex operations in the MPEG-2 encoding and decoding chain. Therfore, embedding a watermark in DCT coefficients can still be performed in real-time, while comparing to VLC domain, more complex manipulations that produce more robust watermark can be used.

Busch et al. [Bus99] based their video watermarking scheme on the Zhao-Koch still-image DCT watermarking algorithm [Koc95]. Although Busch method was actually made to embed the watermark into the luminance component of the uncompressed video by performing DCT transform, we include it here since it can be easily modified to embed watermark into compressed domain before inverse DCT. The watermark is embedded by forcing a predefined relation between a pair of DCT coefficients from appropriate sub-band of a pseudo-randomly chosen DCT block. To minimize perceptual impact of the watermark, especially at edges, blocks are selected according to the block activity. Depending on the sequence, the authors reported bit error rates between 0 and 50% if a 64-bit watermark is embedded into each frame of ITU-R 601 video and retrieved after MPEG-2 compression at 4–6 Mbps. To improve the robustness in critical sequences, they proposed to repeat watermark in 25 or 50 consecutive frames. Presented results show that integrating over 25 frames is sufficient for high accuracy.

Hsu and Wu extended their image watermarking method [Hsu96] to compressed video watermarking [Hsu98]. To embed the watermark, they modify middle-frequency DCT coefficients in relation to neighboring blocks. In I frames modification is done relative to preceding neighboring blocks from the same frame, while in P and B frames relative to neighboring blocks from the previous frame. The watermark signal is a logo consisting of binary pixels. The watermark logo is scrambled prior embedding, so it can be detected from the cropped video. A major drawback of this method is non-blind detection, so the original unwatermarked video is needed in the detection process.

Agung and Sweeney proposed to embed watermark into DC coefficients of image blocks [Agu04]. They argued that the DC coefficient is much larger then the AC components of a block and therefore easier to modify, and much more robust against the standard image

and video processing. Multiplicative rule was used to embed watermark with a scaling factor chosen according to a block texture. In addition, they proposed to encode the watermark with a Reed-Solomon code prior embedding to increase robustness. The results show relatively low robustness against MPEG-2 compression, but relatively good robustness against frame dropping and averaging when non-oblivious detection is used.

One of the pioneering, and arguably most cited, works aiming at watermarking in the DCT domain is the Hartung and Girod extension of their spatial domain technique mentioned earlier [Har98]. The watermark, consisted of a sequence of bits {-1,-1}, is spread by a large chip-rate factor (the authors used different values from 10^{+4} to 10^{+8}), modulated by pseudo-random sequence and then organized into frames, which are the same size as video frames. The watermark sequence can be also amplified by a factor α, which can be adjusted according to local properties of the frame (authors used constant values from 1 to 5). To embed watermark to a DCT block of the video frame, an 8x8 block from the same relative position in the watermark frame is DCT transformed and added to the block. If watermarking of an AC coefficient yields a Huffman code that is longer then from the original unwatermarked coefficient, watermarking is discarded for that coefficient in order to preserve the bit-rate of the sequence. Due to the bit-rate preserving, percentage of watermarked non-zero DCT coefficients is reported to be 10-20%. The watermark signal is embedded into all frames of the video sequence (I, P, B frames). In order to suppress a propagation of the embedding distortion through the frames due to temporal prediction, the authors introduced drift compensation signal that is the difference of the motion compensated predictions from the unwatermarked and the watermarked sequence. Since this technique was considered as a started point of this research, further dissemination of the technique is given in the following chapter.

This technique had the major impact on the digital video watermarking research. Many authors were influenced by the work of Hartung and Girod: [Chu98], [Sim02], [Amb01], [Are00], [Pra04]. Chung et al. proposed to use direct sequence spread spectrum method during MPEG-2 compression process. They argued that if a watermark is embedded during compression, drift compensation is unnecessary and amount of information is increased, since there is no need for bit-rate preserving. The authors introduced a new model for perceptual adjustment based on block classification in DCT domain according to its edginess and texture energy. Although the model gives quite promising results, some of the parameters, such as the mean value of block activity, need to be calculated on the whole

video frame making the model more computationally expensive then the model that can perform the embedding process at one block in time. The results showed the detection rate above 97% when 120 bits is embedded into CCIR601 recommended video sequences (704pixels x 480 lines x 30Hz) and almost the same video quality as MPEG-2 video without watermark embedding.

Simitopoulos et al. proposed improved perceptual adjustment model and to add watermark bits to quantized DCT coefficients [Sim02], [Sim04]. They have used Chung classification model [Chu98] to find suitable coefficients for embedding and Watson JND model in DCT domain to adjust the amplitude of the watermark. The watermarking scheme was reported to be able to withstand attacks such as transcoding, and filtering, and even geometric attacks, if methods for reversing such attacks are incorporated.

Ambroze et al. [Amb01] also based their model on the Hartung-Girod technique and examined the capacity improvement provided by forward error control (FEC) coding and perceptual adjustment based on Watson JND model. Watermarking itself was carried out in the DCT domain using additive rule. They found that using multiple parallel concatenated convolutional codes (3PCCCs) can give an order improvement in capacity for compressed video, and typically gives 0.5 Kbps capacity under a combined compression geometric attack.

To reduce the bit-error rate of the Hartung-Girod technique, Arena et al. [Are00] implemented interleaved encoding technique and some form of spatial and frequency masking. Interleaved encoding is applied by spreading each watermarking bit in an integer number of non adjacent macroblocks. In order to control the watermark invisibility, a DCT block is first partitioned in seven logarithmically spaced subbands in the frequency domain. The activity factor is computed as the power of the signal in the subband in which the coefficient lies. To get the watermarking amplitude, the activity factor is then elevated proportionally to the quantization factor in the MPEG quantization matrix to account for different frequency sensitivity, the absolute value of the DCT coefficient of the image, in order to exploit local frequency masking and the visibility factor for the different gray level sensitivity. Using these improvements, the authors reported that they were able to embed 36 bits per I frame with a bit-error rate of 1.6% after transcoding to 2 Mbps.

Pranata et al. were examining bit-rate control mechanism in Hartung-Girod technique [Pra04]. They proposed bit-rate control scheme that evaluates the combined bit lengths of a

set of watermarked VLC codewords, and successively replaces watermarked coefficients that introduced the largest increase of the set length with the corresponding original coefficients until a target bit length is achieved. Their experimental results showed that this bit-rate control scheme improved the robustness against transcoding attacks.

Depending on the point of the watermark insertion described state-of-the-art methods range from complex and computationally expensive, such as spatial domain methods, to low complexity methods [Lin98]. However, in general higher complexity also means that the embedded watermark will be more robust. Having this in mind, embedding domain needs to be carefully chosen. For watermarking of compressed video watermarks can be embedded in the DCT coefficients [Har98], in VLC codewords [Lan98], in the motion vectors [Jor97], or in the GOP structure [Lin98]. Finally, after consideration of the embedding domain for particular complexity requirements, techniques, such as amplitude adjustment, need to be carefully employed to give an appropriate trade-off between imperceptibility, capacity and robustness.

Chapter 4 MPEG-2 compressed domain data embedding

4.1 Introduction

In this chapter, methods that are used are described in particular technical aspects of bit spreading and embedding into the DCT coefficients of an MPEG-2 video sequence. Targeted watermarking application is pointed out and set of requirements that needs to be fulfilled with the respect to the application and embedding domain is defined. With a careful consideration of these requirements and watermarking techniques presented in the previous chapters, spread-spectrum watermarking was seen as a technique that can be adjusted to meet the requirements of the chosen application.

As a starting point, the compressed watermarking spread spectrum technique proposed by Hartung and Girod [Har98] is described and its strong and weak points are pointed out. To meet given requirements, it was essential to find novel solutions that will improve weak parts of the technique. We proposed to do both embedding and watermark extraction in DCT domain. This significantly lowered computational cost and gave opportunity for implementation of a low computational method for watermark adjustment in DCT domain.

In section 4.5, we are proposing a novel method for watermark bit spreading called **block-wise random watermark bits interleaving**. Implementation of the method resulted in similar detection probability for each watermarking bit, since the bits are spread evenly through textured, edge and plain area of the video image. In addition, with the interleaved bit spreading, distribution of detection values can be approximated with normal distribution giving much easier way for theoretical analysis of the watermarking system performance.

Important aspects of embedding in the DCT domain are described, followed by a novel approach to bit-rate preservation called **bit-rate control on the macroblock level**. The chapter ends with a detailed mathematical analysis of watermark decoding, error rate and capacity boundaries.

4.2 Targeted watermarking application and requirements

As a starting point for any watermarking research, the application that technique will be used for and its requirements need to be properly set up. This Thesis is based on data embedding and indexing application. The initial aim was to embed 64-bits Digital Item Identifier (DII) directly to MPEG2 stream in efficient, imperceptible and robust way. The DII is a serial number that uniquely identifies a piece of content and can be used as a pointer to metadata that describes the content. The reason for using the DII is that the capacity needed for embedding the DII is smaller then the capacity needed for embedding the metadata. In addition, unlike the metadata, the DII does not change according to contextual environments and thus it is much more suitable for long-term information hiding. Watermark decoding must be fast and needs to be carried out in the compressed domain. This requirement is important because the video bit-stream will only be partially decoded during the watermark extraction.

The next requirement refers to the minimum size of the video segment from which it will be possible to extract the DII. This minimum duration of the watermarking video segment is often defined as 5 seconds (The European Broadcasting Union video-watermarking requirements [Che02]). For the PAL television standard this can be seen as 125 frames (5s x 25 f/s). Since typical GOP size for MPEG-2 compression is 15 frames and every GOP have one I frame, the minimum size of the watermarking segment can be defined as 8 I frames. In this way only 8 frames need to be processed in watermark decoder, so extraction can be performed in real-time. On the other hand, due to temporal compression, the embedding space in inter-frames is considerably lower, so this can be seen as reasonable trade-off between the capacity and the computational cost.

In a particular scenario, we are targeting professional users and professional environment, so intentional attacks are not expected, but preserving the video quality and the high reliability of detection are strongly desired. Because of this, other technical challenges that need to be tackled are:

- Embedding and decoding of watermark with high imperceptibility.

- Robustness to video editing processes, e.g., colour adjustment, logo insertion, bit-rate changing, cross-fade, etc.

- High decoding rate.

To meet these requirements, we decided to start with an implementation of a spread spectrum technique proposed by Hartung and Girod in [Har98] as it can be regarded as a cornerstone of the video watermarking research. In addition, several improvements in watermark composition, strength adjustment, bit-rate preserving are described in following sections.

4.3 Spread spectrum watermarking in the compressed domain

Spread spectrum technique proposed in [Har98] introduces watermarking of uncompressed video and then adapts the technique to the compressed domain. This technique uses well-established methods from spread spectrum communication theory, where a narrow-band signal is transmitted via a wide-band channel by frequency spreading. In the case of watermarking, the idea is to transmit the watermark signal as a narrow-band signal using the video signal that acts as the wide-band channel.

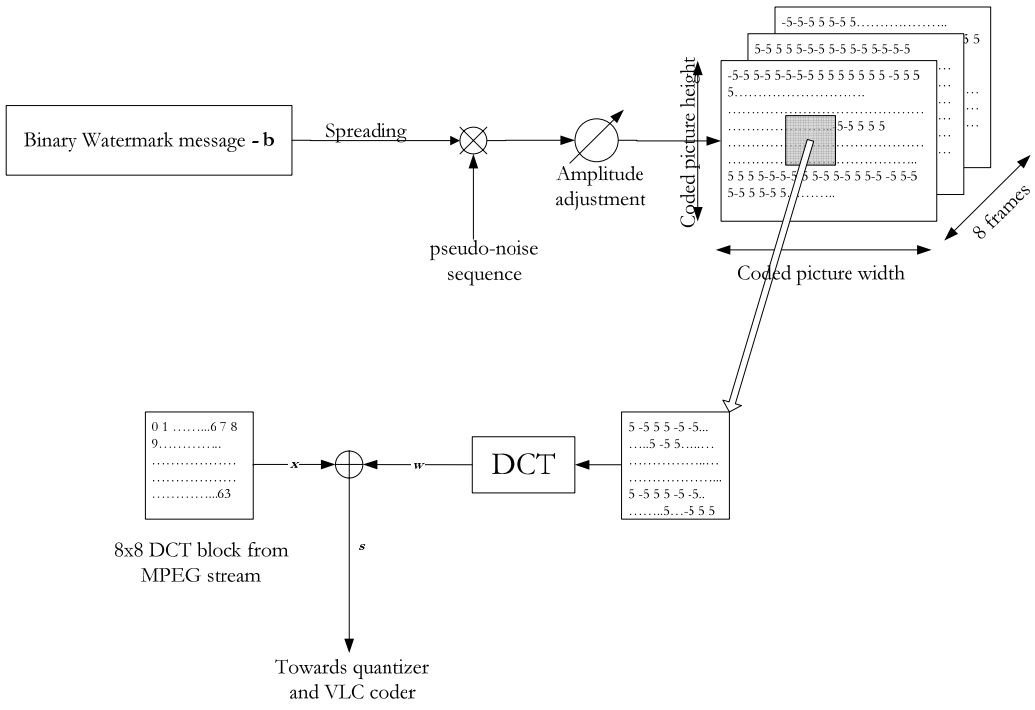

Figure 6. Spread spectrum watermarking scheme proposed in [Har98]

In the uncompressed domain, the watermark signal is usually spread using a large chip factor, amplified afterwards, modulated by a pseudo-noise sequence and finally added

directly to the video pixels. This scheme is efficient for watermarking before compression of the video. However, when large archives of already compressed video content are available, embedding in the compressed domain is required. Due to the performance constraints in terms of computational efficiency and degradation, re-encoding and watermarking in the pixel domain is not feasible. Therefore, watermarking need to be done in the compressed domain. For this purpose, the authors in [Har98] proposed to add a similarly prepared watermark signal directly to DCT coefficients of the compressed video. The scheme (Figure 6) can be divided in the following steps:

- Spreading every watermark bit with chip factor c_r (suggested value is 633600 so the watermark data rate is roughly 16 bits per second), modulating by binary pseudo-noise sequence and amplifying by constant amplitude.

- Arranging this signal into the frames having the same size as the video frames.

- For each 8x8 block DCT of the video (both intra and inter coded blocks), corresponding 8x8 block of the watermark signal is DCT transformed and added to the block.

- Each coefficient of the watermarked block is written to the new bit-stream if its Variable Length Code (VLC) is smaller or equal to the VLC of the original coefficient. Otherwise, the original coefficient is written, so the bit-rate is preserved (i.e. not increased).

- All other parts of the original video stream (system and audio layer, headers, motion vectors, etc) are copied to the new bit-stream.

Hartung and Girod embedded 16 bits per second and were able to decode watermark bits with Bit Error Rate (BER) of 5×10^{-3}. They also reported that the watermark embedded with amplitude 5 is hardly visible (only in direct comparison with the original). These results do not meet the requirements described at the beginning of this chapter in terms of BER and high imperceptibility. In the following sections, we are proposing several adjustments that improve the performance of the spread spectrum watermarking.

4.4 Decoding in DCT domain

The technique presented in the previous section was developed to closely match the authors' scheme in uncompressed domain. They wanted to embed the watermark in a compressed MPEG-2 video sequence that will be possible to extract with their decoder made for uncompressed domain scenario, i.e. to decode the watermark in the uncompressed domain. One can notice that this is not the ideal solution in the terms of computational costs. The scheme that uses decoding in DCT domain, directly from DCT coefficients will be more efficient, since there will be no need for IDCT in the decoder. The encoder efficiency is also increased, since DCT transformation of watermarking signal is avoided (Figure 7).

The watermark is embedded in AC coefficients in order to avoid visual artefacts in the mean luminance of the block. In that way, the DC coefficient of the block can be regarded as noise from watermarking point of view and can be expelled from correlation in the decoder. This has similar effect on the decoder performance as the high-pass filtering proposed in [Har98].

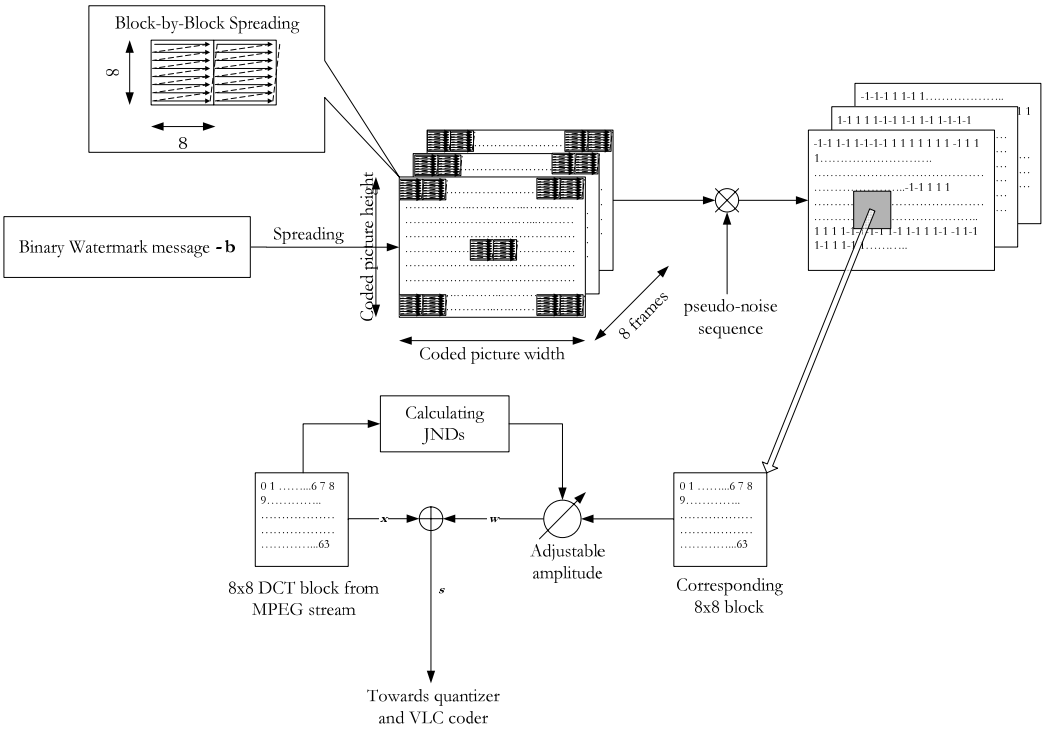

Figure 7. Proposed watermarking scheme

Another positive outcome of this solution is that watermark amplitude adjustment is performed in the DCT domain. The amplitude of the watermark signal is locally adjusted using only information from corresponding DCT block of the host signal. The most appealing approach for amplitude adjustment is Just Noticeable Difference (JND) modelling that is widely used for calculation of image differences perceptible by human visual system. In the proposed scheme, two models were considered: widely used Watson JND model [Wat93] and its extension with block classification proposed by Zhang et al in [Zha05]. Analysis of the two models, comparison of their performances and a new model that combines good characteristics of these two models are given in Chapter 5.

It is also worth noticing in Figure 7 that the watermarking bits are spread in block-by-block fashion. Consider the watermark signal with the chip factor of four frame widths and spreading in line-by-line fashion. An odd watermark bit will then be spread across mainly low-frequency DCT coefficients, while an even one will be embedded in high frequency areas. Because of compression, there will be many zero DCT coefficients in high frequency content, so even watermark bits will much harder to decode. The block-by-block spreading also gives an easy way to manipulate with watermarking bits and considerable increase in decoding rate using bits interleaving.

4.5 Block-wise random watermark bits interleaving

The spread spectrum scheme in Figure 6 has another major drawback. As already mentioned in the previous section, the watermarking bits are spread in line-by-line fashion. Every watermark bit is embedded in one part of the frame. If the particular area mainly consists of plain blocks, locally adjusted amplitude will be small as well as number of non-zero DCT coefficients, so the bit embedded in that area will be hard to decode. For example, if the bit is embedded in the clear blue-sky area of a landscape video, it will be much harder to decode the bit then the other one embedded in a highly textured grass or wood area.

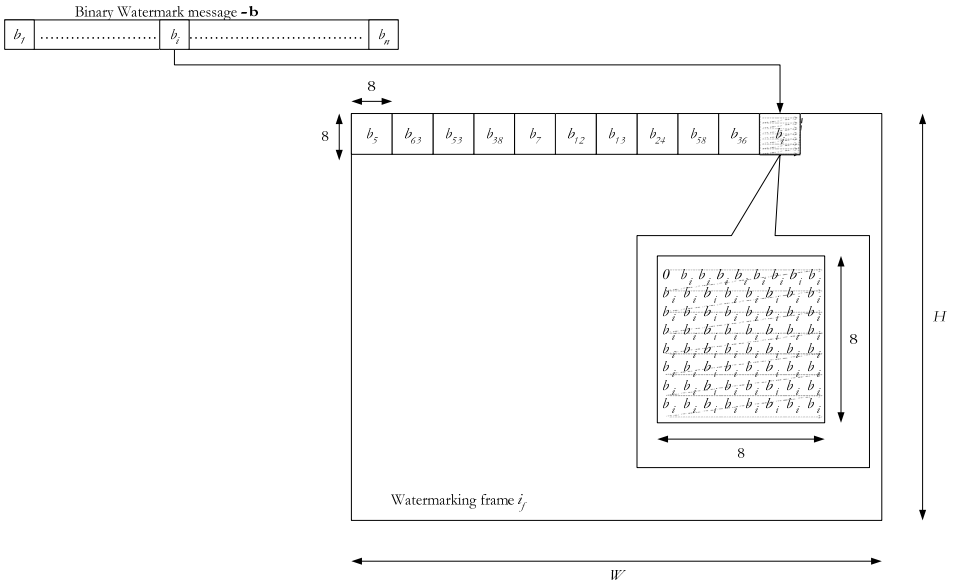

Figure 8. Interleaved watermark bits spreading

To have a similar detection probability, the bits need to be spread evenly through textured, edge and plain area. One way of doing this is to spread watermarking bits randomly through the frame. We are proposing to embed bits as shown in Figure 8, where index *i* is derived by following formula:

$$i = mod\left(rnd(1, \frac{W \cdot H}{64}) + i_f, n \right) \tag{4.5.1}$$

where *W* and *H* are frame width and height, *rnd(a,b)* gives pseudo-random integer between *a* and *b*, $i_f \in [1,8]$ is watermarking frame number and *n* is number of watermarking bits.

Every bit is spread 63 times (63 AC coefficients) in $\frac{W \cdot H}{64 \cdot n}$ blocks in every of eight frames. Frame index i_f is included in (4.5.1) to avoid the same spreading pattern in the frames. In the watermarking decoder, the same pseudo-random generator needs to be used and the result of correlation is added to i-th of n correlation sums, where i is derived by the formula (4.5.1).

The improvements made with pseudo-random interleaving are illustrated in Figure 9. The figure presents distribution of the correlation sums according to their values in the case of interleaved and non-interleaved spreading. The results are given for embedding 512 bits per 8 frames in the "Cactus and Comb" video. All watermarking bits are chosen to be +1 for the sake of clearer presentation, so detection error is registered when the sign of the correlation sum for particular bit is negative.

It can be seen that there were no detection errors if the pseudo-random interleaving is used. All results are closely distributed around the mean watermark amplitude. Without interleaving, some bits are embedded in the textured areas and have high detection values, while others are embedded in the plain areas and do not have enough watermark power to be correctly detected. Figure 9 also shows that in the case of interleaved watermark bits distribution of the detection results can be fitted with normal distribution, which gives a relatively easy way for theoretical analysis of the watermarking system performance.

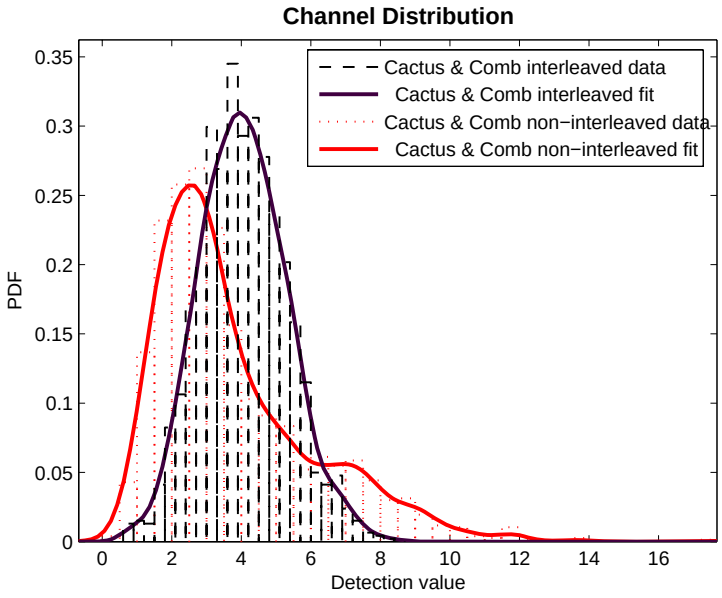

Figure 9. Distribution of detection results for two schemes (512 bits per 8 frames embedded in the "Cactus and Comb" video)

4.6 Bit-rate control on the macroblock level

One important property of a video bit-stream is its bit-rate. A transmitting channel has its own particular capacity, so the bit-rate of the video stream needs to be chosen to comply with the capacity of the channel. Therefore, watermarking or any other process on a compressed video bit-stream must not increase its bit-rate.

Hartung and Girod approach to this problem was to compare sizes of Huffman VLC codes for the watermarked (n_1) and the original coefficient (n_0). If $n_1 \leq n_0$ then the VLC of the watermarked coefficients is written in the new bit-stream, otherwise the original is written. Consequently, a number of altered DCT coefficients will be drastically lowered. Authors reported that typically 10-20% of DCT coefficients are altered, depending on GOP structure, bit-rate and the video-sequence itself.

The MPEG-2 video bit-stream is divided into packets. Every packet has the size of the packet written in the packet header. Concerning the bit-rate, we can change the coefficients as much as we like as long as the size of the packet is not altered. Hence, DCT coefficients that have $n_1 > n_0$ can be written in the new bit-stream if they increase the bit-rate as much as the other watermarked coefficients decrease the bit-rate. This can be controlled on the macroblock level with three controlling variables:

- *Pack_Diff* – Controlling the difference between the original and the new packet
- *MB_Rd_Cnt* – The size in bits of the macroblock read from the original stream
- *MB_Wr_Cnt* – The size of the watermarked macroblock that is going to be written to the new stream.

In the encoder, firstly the macroblock is read from the stream. For every AC coefficient of the original luminance block, *MB_Rd_Cnt* is increased by the coefficient VLC length n_0. After the watermark embedding, all coefficients are VLC coded. For every AC coefficient of the watermarked luminance block, its VLC length n_1 is added to *MB_Wr_Cnt* and VLC difference between watermarked and original coefficient (n_1 - n_0) is stored in an array. Before writing watermarked macro block in the new stream, following condition is checked:

$$MB\_Wr\_Cnt \leq MB\_Rd\_Cnt + Pack\_Diff \qquad (4.6.1)$$

If the condition is not satisfied, the watermarked AC coefficient with the biggest VLC difference is swapped with the original one and *MB_Wr_Cnt* is decreased by the coefficient's VLC difference. The condition (4.6.1) is checked again and the process is repeated until it is not satisfied. Macro block is then written in the new stream and *Pack_Diff* is recalculated:

$$Pack\_Diff_{new} = MB\_Rd\_Cnt + Pack\_Diff_{old} - MB\_Wr\_Cnt \qquad (4.6.2)$$

If *Pack_Diff* is bigger then zero at the end of the packet zeros are added to compensate the difference in the packet size.

Percentage of altered AC coefficients depends on the several compressed stream properties. If quantization steps are high, there is possibility that some coefficients will not be altered since watermarked level is smaller then the quantization step. This can be avoided with careful watermark amplitude adjustment. On the other hand, if an altered coefficient has a zero value, the original one needs to be written, since the number of zero coefficients must not be changed because of run level coding (increasing the run of the preceding zero coefficients significantly increase the VLC of a coefficient).

Table 1 depicts improvements achieved with the new method for bit-rate preserving. Six standard test sequences (MPEG-2, PAL, 358 frames, 704x576, 25 fps, GOP structure IBBP, GOP size 12, bit-rate 6 Mbps) were watermarked with same embedding parameters, once with the Hartung and Girod bit-rate preserving and once with bit-rate preserving on the macroblock level. It can be seen that the percentage of altered AC coefficients is almost doubled in comparison with the Hartung and Girod scheme.

Table 1. Percentage of altered AC coefficients in I frames

Video Sequence	Hartung-Girod	Proposed method
Table Tennis	32.26%	61.47%
Cactus and Comb	26.80%	50.46%
Flower Garden	35.74%	65.52%
Mobile and Calendar	27.49%	49.25%
Susie	28.05%	58.21%
BBC3	21.39%	37.06%
Average	28.62%	53.66%

4.7 Watermark decoding and error analysis

Watermark decoding is straightforward using a correlation receiver. If there is no attack, received data **r** is equal to the watermarked one **y**. The data is correlated with the same pseudo sequence **p** used in embedding process resulting in *n* correlation sums:

$$s_j = \frac{1}{N_j} \sum_i y_i \cdot p_i \tag{4.7.1}$$

where elements are summed over blocks given by interleaving formula used in encoder (4.5). N_j is number of non-zero AC coefficients in blocks in which bit b_j is embedded. Substituting y_j with embedding formula 4.3.1 becomes:

$$s_j = \frac{1}{N_j} \sum_i x_i \cdot p_i + \frac{1}{N_j} \sum_i \alpha_i \cdot b_j \cdot p_i^2 \tag{4.7.2}$$

Let all embedded bits b_j be equal +1. As shown in section 4.4, since bits are randomly embedded we can assume that samples are independent and identically distributed (iid), so following the central limit theorem s_j has approximately normal distribution:

$$s_j \sim N\left(\mu_\alpha, \frac{\overline{x^2}}{\overline{N}}\right) \tag{4.7.3}$$

where μ_α is the mean amplitude of the watermark over the embedding window (8 frames). The variance of s_j is given by mean power of non-zero DCT coefficient $\overline{x^2}$ over the embedding window divided by the mean number of non zero coefficients in the blocks $\overline{N} = N_{nz}/n$, since from the first term of the equation 4.7.2 the main source of noise arises from cross-correlation between pseudo-noise sequence with non-zero DCT coefficients. N_{nz} is number of non-zero coefficients in embedding window.

The mean amplitude μ_α is calculated in encoder after quantization of watermarked coefficients in order to include the quantization effect on the watermarked amplitude:

$$\mu_\alpha = \frac{1}{N_{nz}} \sum_i \left| \lfloor y_i \rfloor - \lfloor x_i \rfloor \right| \tag{4.7.4}$$

The performance of the watermarking system in the terms of the decoding reliability can be measured by bit error rate. The bit error rate (BER) is the number of measured wrongly decoded bits divided by the number of embedded bits. The bit-error probability can also be theoretically derived for the specific stochastic model of the watermarking process.

It was already shown in the section 4.4 that detection value s have normal distribution over the embedding window, so the bit rate error can be defined as probability:

$$BER^e = P(s < 0) = P\left(\frac{s - \mu_\alpha}{\sqrt{\overline{x^2}/\overline{N}}} < -\sqrt{\frac{N_{nz}}{n} \cdot \frac{\mu_\alpha^2}{\overline{x^2}}}\right) = P(Z < -\sqrt{SNR}) \qquad (4.7.5)$$

$$P(Z < -\sqrt{SNR}) = \Phi(-\sqrt{SNR}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{SNR}} e^{t^2/2}\, dt \qquad (4.7.6)$$

where BER^e is estimated bit error rate, $\Phi(\cdot)$ is standard normal cumulative distribution function and SNR is a signal-to-noise ratio. Since the SNR is highly dependent on characteristics of the video segment, the BER^e will take wide range of values. For that reason, let us first look at how BER^e can be estimated on longer video sequences.

The video sequence used in this experiment is consisting of 11 standard test sequences (Table Tennis, Cactus and Comb, Flower Garden, Mobile and Calendar, Susie, BBC3, Balloon-pops, Tempete, Football, Sailboat, City). Watermark embedding process is repeated L=35 times (i.e. 35 x n bits is embedded in the whole sequence, where n is the number of watermarking bits embedded in the 8-frame window).

Distribution of detection values is shown in Figure 10. Since characteristics of the video sequence are changed with time and watermark bits are embedded in different time slots, assumptions about iid samples no longer stands, so detection values are not normally distributed. However, distribution in Figure 10 can be modelled as a mixture of L normal distributions with mean equal to locally derived mean watermark amplitude (over embedding window) and variance equal to locally derived mean power of non-zero DCT coefficient divided by the number of them.

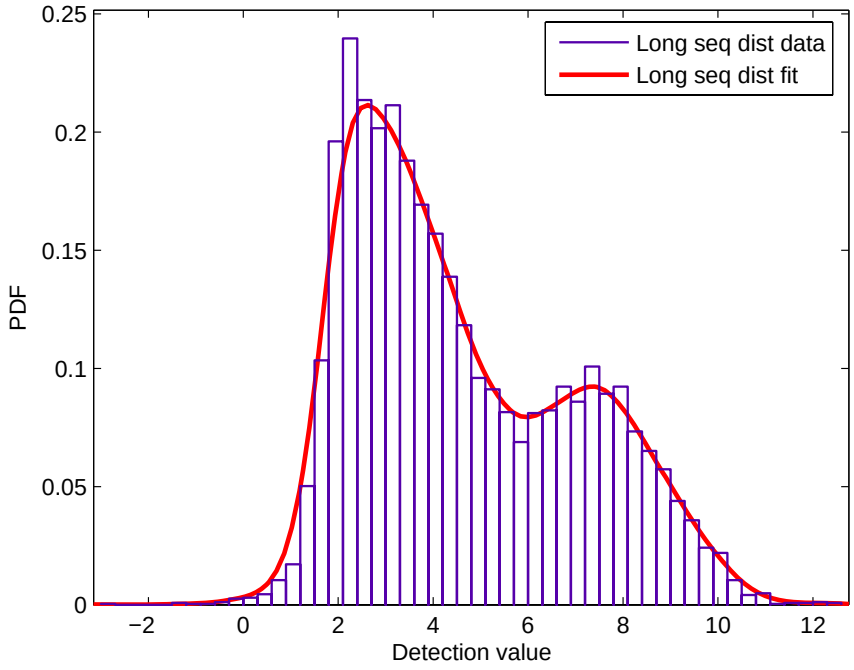

Figure 10. Distribution of detection for long sequence (256 embedded in 8 frames 35 times)

Probability density function $f(s)$ of the detection values distribution can be mathematically formulated as:

$$f(s) = \sum_{l=1}^{L} p_l \frac{1}{\sqrt{2\pi\sigma_l^2}} e^{\frac{(s-m_l)^2}{2\sigma_l^2}} \tag{4.7.7}$$

where:

$$m_l = (\mu_\alpha)_l \quad , \quad \sigma_l = \sqrt{\left(\frac{\overline{x^2}}{N_{nz}}\right)_l} \tag{4.7.8}$$

Since there is a same number of bits embedded in every embedding window, mixing probabilities are equal:

$$p_l = \frac{1}{L} \quad , (l = 1,..., L) \tag{4.7.9}$$

Given the probability density function (4.7.7), the bit error rate can be estimated with:

$$BER^{e} = \int_{-\infty}^{0} f(s,\theta)ds = \frac{1}{L}\sum_{l=1}^{L}\int_{-\infty}^{0}\frac{1}{\sqrt{2\pi\sigma_l^2}}e^{\frac{(s-m_l)^2}{2\sigma_l^2}}ds \tag{4.7.10}$$

$$BER^{e} = \frac{1}{L}\sum_{l=1}^{L}\Phi(-\sqrt{SNR_l}) \tag{4.7.11}$$

where SNR_l is signal-to-noise ratio in l-th video segment. The bit error rate can be also measured at the decoder side:

$$BER^{m} = \frac{e}{L \cdot n} \tag{4.7.12}$$

where e is number of wrongly detected bits and n is number of bits in one embedding window. Comparison between estimated and measured bit error rate depending on n is given in Table 2 and Figure 11.

Estimated and measured BER values for more then 256 bits embedded per 8 frames show close matching between the model and the detection values distribution. Small differences can be explained with assumption made that variance of mean amplitude over the embedding window is much smaller then the noise arises from cross-correlation between pseudo-noise sequence with non-zero DCT coefficients. For 64 and 128 bits per embedding window, departures are bigger as a consequence of the small number of observed bits (in first case 35x64=2240 and in latter 4480), so measured BER values have relatively low accuracy (4.46×10^{-4} and 2.23×10^{-4} respectively).

Table 2. Estimated and measured bit error rates values

n	Estimated BER	Measured BER
64	6.474×10^{-7}	≈ 0
128	9.692×10^{-5}	2.232×10^{-4}
256	1.462×10^{-3}	1.671×10^{-3}
512	6.806×10^{-3}	7.860×10^{-3}
1024	1.909×10^{-2}	2.022×10^{-2}
2048	4.618×10^{-2}	5.116×10^{-2}
4096	9.695×10^{-2}	1.082×10^{-1}

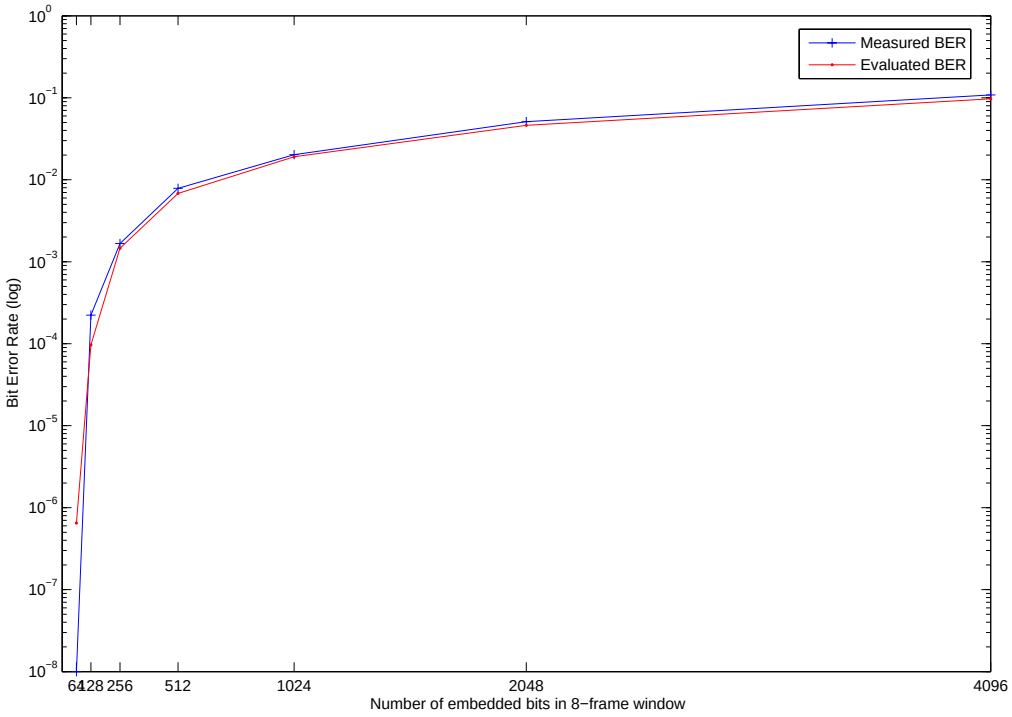

Figure 11. Bit error rate curves – measured and estimated

From Figure 11, it is clear that for the given watermarking system, maximum 64 bits can be embedded in 8 I frames of the video sequence in order to have reliable detection ($BER<10^{-5}$). However, there should still be room for the capacity improvement.

In this section, we showed how BER can be estimated for long video sequences. This experiment gave us opportunity to analyze wide range of embedding windows. It was spotted that relatively high BER values are mainly due to embedding in "BBC3" video sequence. This sequence consists of edge and plain (Figure 12) blocks, leading to a low watermark power and at the same time it has relatively high mean power of non-zero DCT coefficients, giving low SNR ratio. It is evident that in the given setup maximum number of bits that can be embedded is determined by the "BBC3" sequence. Hence, in the following sections all experiments concerning BER and watermark capacity will be conducted on this particular sequence.

Figure 12. Frame from "BBC3" video sequence

4.8 Capacity boundaries

The watermark capacity is bounded by imperceptibility requirements. Given the watermark power embedded with amplitude adjustment method used in the previous section (Zhang JND method with w=0.5 as described in section 5.4), we were able to embed 128 bits in a 5 seconds video segment with *BER* ~10^{-4}. However, before experimenting with the amplitude adjustment methods and error correction coding to increase capacity, it would be useful to find out what is the actual capacity boundary given the embedded signal power.

Watermark embedding can be regarded as communication over a noisy channel. The noise in the channel is originated by the video itself and by the attacks and is independent of the watermark message **b**. If there are no attacks, watermarking message is communicated via Gaussian channel with the noise originated from non-zero DCT coefficients $\mathbf{x} \sim N(0, \overline{x^2})$. Let us first assume that input watermark signal has Gaussian distribution $\mathbf{b} \sim N(0, \mu_a^2)$, which maximizes mutual information between the input and the output signal.

For such a communication system, Claude Shannon stated in his famous theorem [Sha48] that maximum channel capacity is given by the following formula (Appendix A):

$$C_{GAUSS} = \max_{EX^2 \le S} I(\mathbf{X}; \mathbf{Y}) = \frac{1}{2} log_2(1 + \frac{S}{N}) \tag{4.8.1}$$

The maximum channel capacity is given in bits/dimension, where the dimension in the given watermarking channel is a number of non-zero DCT coefficients in the 8-I frame window N_{nz} divided by the number of watermarking bits *n*. Signal-to-noise ratio can be expressed as:

$$\frac{S}{N} = \frac{N_{nz}}{n} \cdot \frac{\mu_\alpha^2}{\overline{x^2}} \tag{4.8.2}$$

where μ_α is mean watermarking amplitude and $\overline{x^2}$ is mean power of DCT coefficients over the embedding window.

Shannon's main theorem also asserts that communication with error probabilities as small as desired is possible if the transmission rate is smaller then the given channel capacity. This can be achieved by an appropriate encoding and decoding operation. However, Shannon theory does not say anything about how to achieve this limit.

As already stated, maximum capacity in (4.8.1) is achievable if an input signal is a Gaussian signal. However, in the given watermarking system we are restricted to BPSK (Binary Phase Shift Keying) signalling, since watermarking bits are taking only two values either +1 or -1. In that way, since the input signal is no longer ideal for the given channel, mutual information between the input and the output signal will not be maximal leading to a lower maximum achievable capacity (Appendix B):

$$C_{BPSK} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-5}^{+20} e^{-\frac{1}{2}(y - \sqrt{\frac{S}{N}})^2} \cdot log_2(1 + e^{-2y\sqrt{\frac{S}{N}}}) dy \tag{4.8.3}$$

Equation (4.8.3) is solvable numerically and represents the actual capacity boundary of our watermarking system. Two capacity boundaries are given in Figure 13.

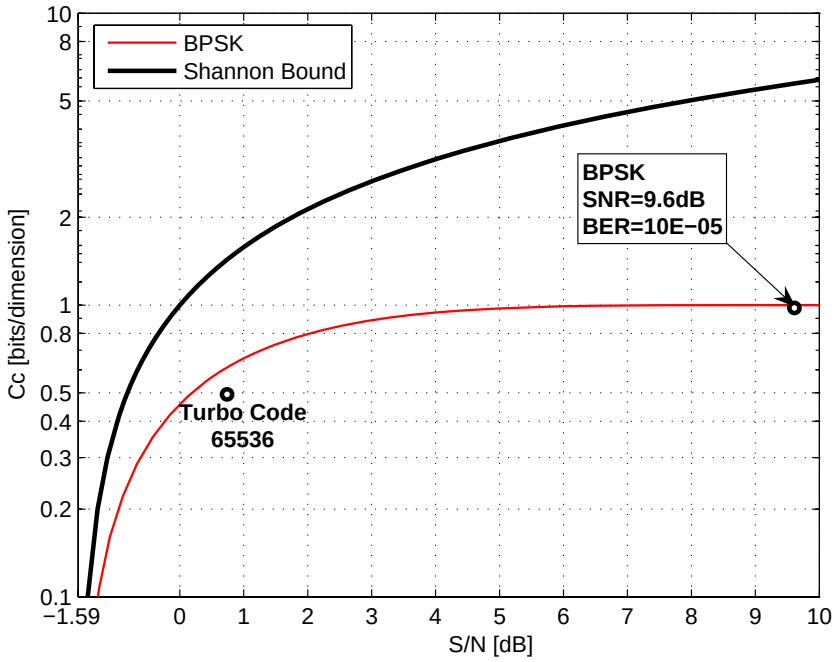

Figure 13. Capacity of AWGN channel for BPSK and Gaussian signalling

As can be seen from the Figure 13, it is possible to achieve bit error rate of 10^{-5} with BPSK signalling and signal-to-noise ratio of 9.6 dB. It will be shown in the chapter 5 that indeed with the proposed method for amplitude adjustment we were able to embed a message 192 bits long into the BBC sequence with an SNR of 9.5 dB and no errors on 10^5 embedded bits.

On other hand, to approach the capacity limit error correction coding techniques can be used. All rates for given SNR levels bellow BPSK curve on Figure 13 are achievable using an appropriate coding technique. Turbo codes [Ber93] with iterative decoding come close to the Shannon limit capacity. Hence, turbo coder with rate ½ and frame size of 65536 bits achieved BER of 10^{-5} with SNR of only 0.7 dB and longer turbo codes can come even closer to the capacity limit. However, since SNR is decreasing rapidly with the number of embedded bits (equation 4.8.2), we are restricted to shorter frame sizes (~300 bits) and suboptimal performance of an error correction coding technique. Further discussion on the coding gain that can be achieved with turbo coding is given in chapter 6.

Chapter 5 Perceptual Adjustment in DCT domain

5.1 Introduction

One of the main watermarking requirements in a professional environment is high imperceptibility of the watermark. The watermarked material has to preserve the quality and at the same time to have high-fidelity, i.e. must be indistinguishable from the original. Having this in mind, the local amplitude of the watermark needs to be carefully chosen in order to maximize the watermark capacity and at the same time to minimize perceptual distortions.

In this chapter the main characteristics of the Human Visual System (HVS) and methods to exploit its imperfections such as sensitivity to the intensity changes at different frequencies will be presented. Watson visual model developed for JPEG compression quantization matrices [Wat93] is widely used in DCT-based watermarking [Pod98], [Sim02] and [Amb01]. The model estimates the perceptibility of changing coefficients of an image's 8x8 DCT blocks. It consists of a sensitivity function and two masking components based on luminance and contrast masking. Recently, Zhang et al proposed an improved estimation of Just Noticeable Distortion (JND) [Zha05] in terms of luminance and contrast masking. We compared performances of these two models in our compressed domain watermarking technique and finally proposed a new model, which combines good characteristics of these two models.

Proper subjective evaluation of the perceptual distortions is hard to accomplish. As pointed out in [Cox01], imperceptibility is often claimed in the watermarking literature, but rigorous quality and fidelity tests with human observers are rare in the evaluation of watermarking models. Many of these claims are based on judgment by single observer with insufficient number of trails. There is still lack of good objective measure for perceptual evaluation and comparison of watermarking techniques. Measures based on mean square error (such as peak signal-to-noise ratio) are typically used regardless to the fact that they do not take into account imperfections of the HVS. In order to give better evaluation of perceptual

distortions we also included most degraded frames (lowest PSNR) for subjective evaluation.

5.2 Fundamentals of Human Perception

The Human Visual System is the subject of a wide range of research and experimentation. Most of the visual perceptual models try to determine how lower level psychology of visual system limits visual sensitivity [Pap00]. They are trying to determine masking thresholds - the level of distortion which can be introduced and still not be noticeable by human observer. In psychophysics studies, these thresholds are also defined as a level of distortion that can be perceived in 50% of experimental trials and is often referred as just noticeable difference (JND) [Cox01].

The JND models used in this work tend to exploit three basic types of phenomena: non-uniform frequency response of human eye (contrast sensitivity function), sensitivity to the different brightness levels (luminance masking) and sensitivity to one frequency component in the presence of another (contrast or texture masking).

5.2.1 Contrast Sensitivity Function

A human eye perceives differently luminance changes at different frequencies. This spatial frequency response can be described as Contrast Sensitivity Function (CSF) [Man74], or can be derived empirically and given in a table [Saf89]. One of the models for CSF is presented in Figure 14. [Man74]. It can be observed that HVS is most sensitive to the changes at mid range frequencies and that sensitivity decreases at lower and higher frequencies.

The spatial frequency is given in a number of cycles per degree of visual angle, which means that visibility of details at given frequency is a function of viewing distance. When an observer moves away from the image, details in the image remain the same and take up fewer degree of visual angle. On the other hand, moving closer enables viewer to perceive fine details that have risen above visibility threshold. For example, an art critic will always get closer look to the painting to see painter's fine brush strokes.

Image quality metrics take this into account by specifying a minimum viewing distance and evaluating distortion perceptibility at this point. There are several standard minimum viewing distances specified for subjective tests and used also for objective measurements.

For standard definition television quality measurements, this is six times image heights and three times for high definition television.

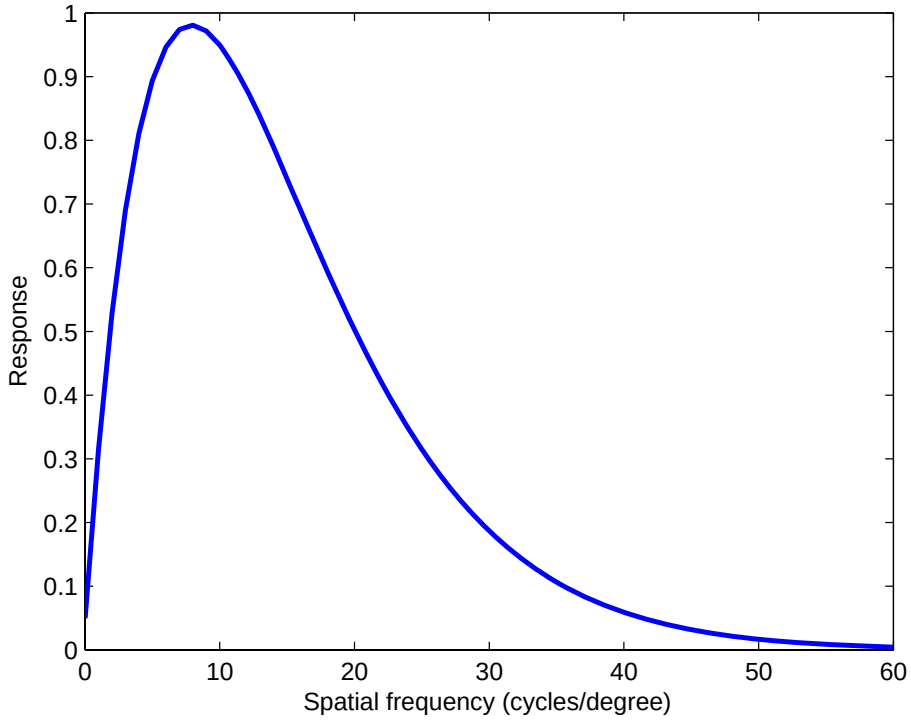

Figure 14. Contrast Sensitivity Function based on Mannos and Sakrison model [Man74]

It is also worth to notice that Mannos-Sakrison model is given only for single orientation. It has been shown that the eye sensitivity is also dependent on the orientation of the frequency pattern [Cox01]. The eye is most sensitive to horizontal and vertical luminance changes and that is least sensitive to the lines and edges with 45 and 135 degrees orientation.

5.2.2 Luminance Masking Effect

The human visual system's sensitivity to variations in luminance is dependent on the local mean luminance. This is called luminance adaptation or luminance masking [Wat93] to emphasize the similarity to contrast masking. The luminance of the original image masks the introduced distortion and this masking is a function of the local luminance.

In the basic Ahumada-Peterson model [Ahu92], luminance masking is accounted as a linear luminance adaptation based on Weber-Fechner's law and Watson in [Wat93] proposed simplified power law version of Weber-Fechner's law. Zhang et al. in [Zha05] argued that Watson model over-simplifies the viewing conditions for practical images. They stated that gamma-correction of the display tube and ambient illumination falling on the display partially compensate effect of Weber-Fechner's law and as a result give higher visibility thresholds in either very dark or very bright regions, which Watson model fails to acknowledge (Figure 15).

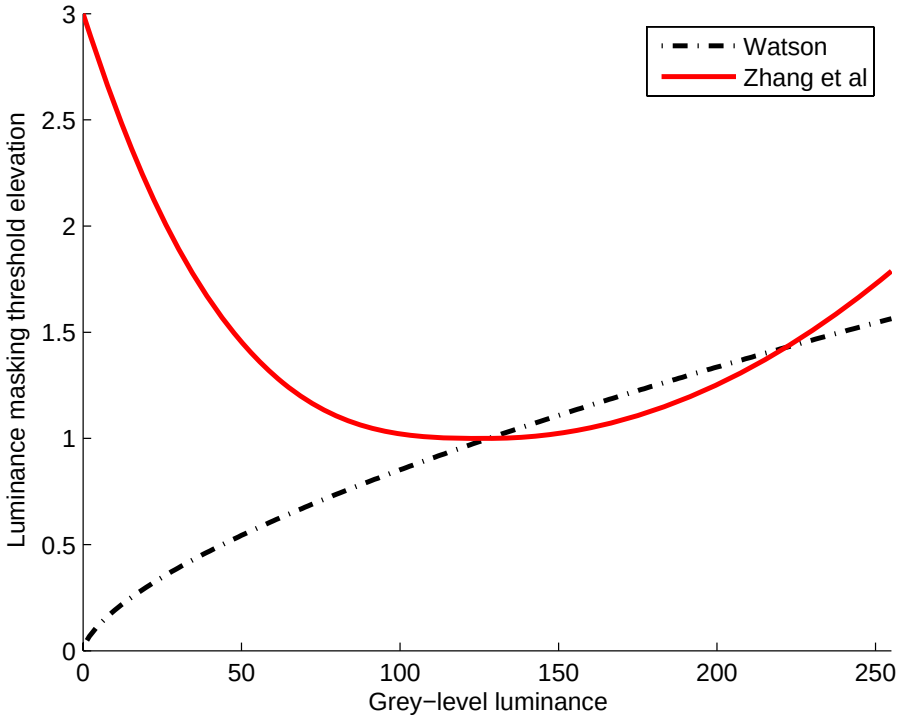

Figure 15. Comparison of Watson and Zhang luminance adaptation models

5.2.3 Contrast Masking Effect

The presence of one image component can mask the visibility of another. One texture can be easily seen in isolation, but might be unnoticeable when it is added to a highly texture image. This masking effect is strongest when both components have the same frequency, location and orientation. For example, distortion may occur only in the single frequency and it can be masked by a single signal of the same or possibly different frequency. In a more general case, more appropriate for image and video processing, distortion will occur

in a frequency bandwidth, which should be masked by frequency bandwidth of image components.

The term contrast masking is often used to denote the masking effect when both components have the same frequency and the texture masking is used to refer to the more general case. Distinction between intra-band (as contrast masking) and inter-bend (texture) masking can be also found in the literature. The majority of metrics model only contrast masking, for example Watson model [Wat93] uses only intra-band masking effect to determine masking threshold. Moreover, it overestimates contrast masking effect near the edges. Tong and Venetsanopoulos [Ton98] a proposed model based on block classification into texture, edge and plain blocks, which gives fair evaluation of distortion perceptibility near edges, but still fail to include CSF and intra-band masking effect in relatively plain blocks. Zhang et al [Zha05] used Tong-Venetsanopoulos block classification model to improve Watson DCT model by avoiding overestimation of the contrast masking effect.

Figure 16. Effects of Contrast and Luminance masking: Noise is invisible in trees texture area (contrast masking) and in the white surface of the roof (luminance masking)[1]

[1] Courtesy of M.Sc. Ivan Pavlovic, Laboratory of Biocybernetics, Faculty of Electrical Engineering, University of Ljubljana, Slovenia.

5.3 Watson DCT-based Visual Model

To adjust the amplitude of the watermark in a DCT domain one needs to see how above mentioned characteristics of HVS are converted to visibility thresholds for DCT coefficients. Watson developed his DCT model [Wat93] to derive the perceptually optimal quantization matrix for purposes of the DCT-based image coders, such as JPEG. His model that computes visibility thresholds for DCT coefficients is based on Ahumada-Peterson-Watson model [Ahu93] for baseline contrast sensitivity thresholds and improved by luminance and contrast masking. The model's JND threshold can be expressed as:

$$t_{JND}(n_1, n_2, i, j) = t_{CSF}(n_1, n_2, i, j) \times t_l(n_1, n_2, i, j) \times t_c(n_1, n_2, i, j) \qquad (5.3.1)$$

where threshold t_{CSF} accounts for spatial contrast sensitivity function, t_l for background luminance adaptation and threshold t_c accounts for contrast masking. Indices n_1 and n_2 show the position of the 8x8 DCT block in the image or the video frame, while i and j represent position of the coefficient in the DCT-block.

5.3.1 Sensitivity Thresholds

The Contrast Sensitivity Function (CSF) models the effect of a spatial frequency on the human visual system sensitivity. The visibility threshold t_{CSF} is a function of the total luminance L and spatial frequency f. In Ahumada-Peterson model [Ahu92], [Ahu93] logarithm of a threshold $T(i,j)$ is approximated by a parabola in log spatial frequency:

$$\log T(i, j) = \log \frac{s \cdot T_{min}}{r + (1 - r)\cos^2 \theta_{i,j}} + K(\log f_{i,j} - \log f_{min})^2 \qquad (5.3.2)$$

where T_{min}, K and f_{min} are functions of the total luminance L. T_{min} is the minimal luminance threshold that occurs on frequency f_{min} and K determines steepness of the parabola given by following formulas:

$$T_{min} = \begin{cases} 0.142 \times \left(\frac{L}{13.45}\right)^{0.649} & , \quad L \le 13.45\, cd/m^2 \\ \frac{L}{94.7} & , \quad \text{otherwise} \end{cases} \qquad (5.3.3)$$

$$f_{min} = \begin{cases} 6.78 \times \left(\frac{L}{300}\right)^{0.182} & , \quad L \le 300\, cd/m^2 \\ 6.78 & , \quad \text{otherwise} \end{cases} \tag{5.3.4}$$

$$K = \begin{cases} 3.125 \times \left(\frac{L}{300}\right)^{0.0706} & , \quad L \le 300\, cd/m^2 \\ 3.125 & , \quad \text{otherwise} \end{cases} \tag{5.3.5}$$

The spatial frequency f_{ij} associated with *i, j* DCT coefficient is given by:

$$f_{ij} = \frac{1}{2N}\sqrt{\frac{i^2}{\omega_x^2} + \frac{j^2}{\omega_y^2}} \tag{5.3.6}$$

where ω_x and ω_y are horizontal and vertical sizes of a pixel in degrees of visual angle and can be calculated from viewing distance λ and display pixel size Λ:

$$w_x = 2\,arctan\left(\frac{\Lambda_x}{2\lambda}\right) \quad \text{and} \quad w_y = 2\,arctan\left(\frac{\Lambda_y}{2\lambda}\right) \tag{5.3.7}$$

In the equation 5.3.2, the parameter $s \in (0.0,1.0)$ is to account for summation of distortions over a spatial neighbourhood.

As already mentioned in section 5.2.1, HVS is most sensitive in horizontal and vertical direction (*i*=0 or *j*=0) and less sensitive in other directions. In order to account for reduced sensitivity due to obliqueness in other directions, the model includes a factor $r + (1-r)\cos^2\theta_{i,j}$. The magnitude of obliqueness is determined by $r \in (0.0,1.0)$, and angle $\theta_{i,j}$ is given by:

$$\theta_{i,j} = arcsin\left(\frac{2 f_{i,0} f_{0,j}}{f_{i,j}}\right) \tag{5.3.8}$$

A visual threshold *T(i,j)* measured in luminance, needs to be converted in gray level units prior it is used for determine JND threshold (5.3.1). If *G* is number of gray levels in the image/video (i.e. *G*=256 in MPEG-2 Standard) and L_{max} and L_{min} are display luminances that correspond to maximum and minimum gray level then:

$$t_{CSF}(n_1, n_2, i, j) = \frac{G}{\alpha_i \alpha_j (L_{max} - L_{min})} T(i, j) \tag{5.3.9}$$

where

$$\alpha_x = \begin{cases} \sqrt{\frac{1}{N}} & x = 0 \\ \sqrt{\frac{2}{N}} & x \neq 0 \end{cases} \tag{5.3.10}$$

accounts for the normalization constant of the DCT.

The HVS sensitivity can also be derived empirically and given in contrast sensitivity table. One such threshold table is given in Table 3 [Pap00]. The numbers in the table present contrast sensitivity threshold levels for 8x8 block of DCT coefficients. This table is produced for the viewing distance of six image heights as defined for testing standard definition television. Since we are targeting application in professional environment and obliged to strict viewing conditions (HDTV), we are eager to use thresholds in Table 4 given by Cox et al in [Cox01] for Watson model in watermarking applications.

Table 3. Y channel contrast sensitivity table given for viewing distance of six image heights

5	3	4	7	11	16	24	34
3	4	4	6	8	12	18	25
4	4	8	9	11	15	20	28
7	6	9	14	16	20	26	33
11	8	11	16	26	28	34	42
16	12	15	20	28	41	46	54
24	18	20	26	34	46	63	71
34	25	28	33	42	54	71	95

Table 4. Y channel contrast sensitivity table given by Cox et al [Cox01]

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

5.3.2 Luminance Based Adjustment of Thresholds

Distortion detection threshold depends on the mean luminance of the local image region and it is higher in the brighter area [Wat93]. To adjust sensitivity threshold Watson model suggests the use of a scaling power function of average block luminance, i.e. DC component of DCT block $C(n_1,n_2,0,0)$. The luminance-masked threshold is given by formula:

$$t_l^{WAT}(n_1,n_2,i,j)=\left(\frac{C(n_1,n_2,0,0)}{\overline{C(0,0)}}\right)^{a_T} \tag{5.3.11}$$

where $\overline{C(0,0)}$ is the average of the DC coefficients in the image and can be set to the expected mean value, i.e. 128 for our MPEG-2 watermarking model. The constant parameter a_T is taken from Ahumada-Peterson model [Ahu92] (equation 5.3.3) with a suggested value of 0.649. This parameter controls the amount of luminance masking in the threshold model. Watson also suggested that it could be used for controlling display Gamma by multiplying it with Gamma exponent.

The Watson model for luminance adaptation is illustrated by dot-dashed line in Figure 15. This power function model should be adequate for brighter regions of the image, but fails to approximate accurately luminance adaptation in darker regions where Weber-Fechner's law is not valid [Ton98], [Zha05].

5.3.3 Contrast Masking

The Watson contrast masking model considers only masking within a block and a particular DCT coefficients, since the contrast masking effect is strongest when two components have the same frequency, orientation and direction. In this model contrast masking threshold $t_C(n_1,n_2,i,j)$ is a function of DCT coefficient $C(n_1,n_2,i,j)$ and thresholds t_{CSF} and t_l :

$$t_C^{WAT}(n_1,n_2,i,j) = max\left\{1 \quad , \quad \left(\frac{|C(n_1,n_2,i,j)|}{t_{CSF}(n_1,n_2,i,j)\cdot t_l(n_1,n_2,i,j)}\right)^{w(i,j)}\right\} \quad (5.3.12)$$

where $w(i,j)$ is a constant that lies between 0 and 1, which may differ for each frequency, but usually has a constant value. Figure 17 is an example of masking threshold dependence on magnitude of DCT coefficient for typical parameters values (i=5, j=5, $C(n_1,n_2,0,0)$=230, $w(i,j)$=0.7).

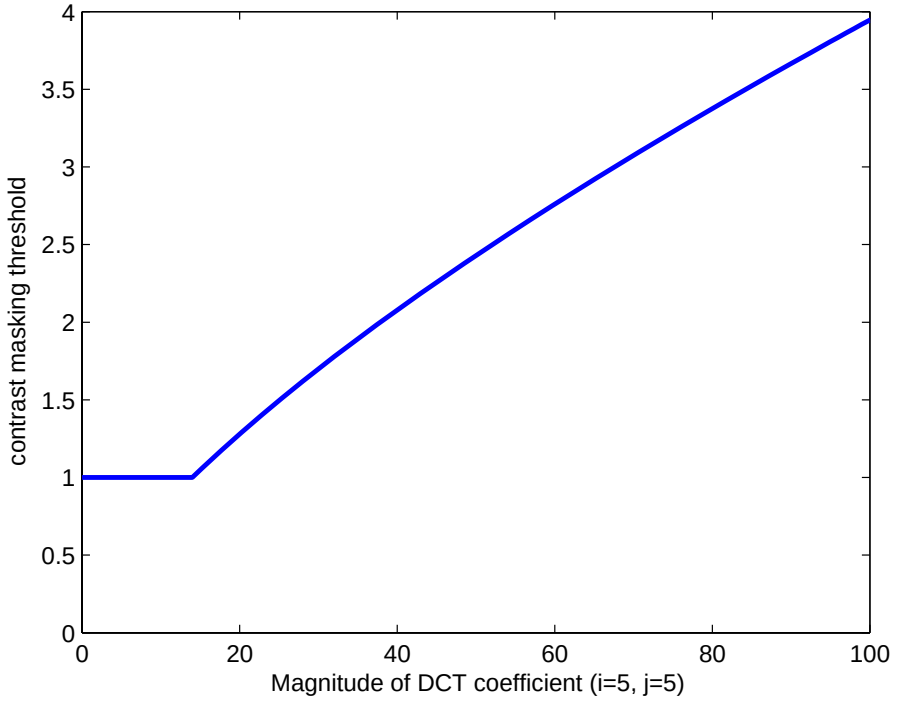

Figure 17. Watson contrast masking model - threshold dependence on DCT coefficient (i=5, j=5)

5.4 Improved Just Noticeable Difference Model

Watson luminance masking model, as mentioned in [Wat93], does not apply for dark sections of the image. On the other hand, contrast masking model considers only masking by the component of the same frequency. Zhang et al in [Zha05] proposed an improved JND estimation model and claim that it outperforms the Watson model. Their model uses the same contrast sensitivity function, given by Ahumada and Peterson, but proposes a new formula for luminance masking and the incorporation of block classification in the contrast masking model.

5.4.1 Zhang Luminance Masking Model

Techniques for luminance adaptation based on Weber-Fechner's law [Ahu92], [Ahu93], [Wat93] over-simplify the viewing conditions. Weber-Fechner's law can be applied to medium and high luminance range, where just-noticeable luminance difference ($\Delta L / L$) remains approximately constant. On contrary, in the low range it is not consistent with the findings of subjective testing [Cho95], [Saf89], [Jay93].

Beside background luminance, other factors mentioned in section 5.2.2 will also affect luminance adaptation by compensating the effect of Weber-Fencher's law. This will result in higher thresholds in very dark and very bright regions. Zhang et al [Zha05] approximate luminance thresholds with two functions, for low region ($L \leq 128$) and for high region of luminance ($L>128$). If $C(n_1,n_2,0,0)$ is DC component of a DCT block and average luminance of the image is 128 then:

$$t_l^{ZHA}(n_1,n_2,i,j) = \begin{cases} k_1\left(1-\dfrac{C(n_1,n_2,0,0)}{128}\right)^{\lambda_1}+1 & \text{if} \quad C(n_1,n_2,0,0) \leq 128 \\ k_2\left(\dfrac{C(n_1,n_2,0,0)}{128}-1\right)^{\lambda_2}+1 & \text{otherwise} \end{cases} \tag{5.4.1}$$

where k_1 =2, k_2=0.8, λ_1 =3, λ_2 =2.

From the comparison with Watson model (Figure 15) it can be observed that Zhang luminance model better approximate findings of subjective tests of HVS's luminance adaptation.

5.4.2 Contrast Masking based on Block Classification

To evaluate the effect of contrast masking more accurately, it is essential to classify DCT blocks according to their energy. It is well known fact that noise is less visible in the regions where texture energy is high and it is easy to spot in smooth areas. On the other hand, HVS is sensitive to the noise near a luminance edge in an image, since the edge structure is simpler then texture one and a human observer have better idea about edge look. Hence, block classification will lead to better adaptation of the watermark to different part of the image.

DCT blocks are assigned in one of three classes: TEXTURE, EDGE and PLAIN. According to the block class and its energy contrast masking threshold is derived. First, an 8x8 block is divided into four areas shown in Figure 18 and the absolute sums of the DCT coefficients in the areas are denoted with DC – mean block luminance, LR – low frequency region, ER – edge region and HR for high-frequency region.

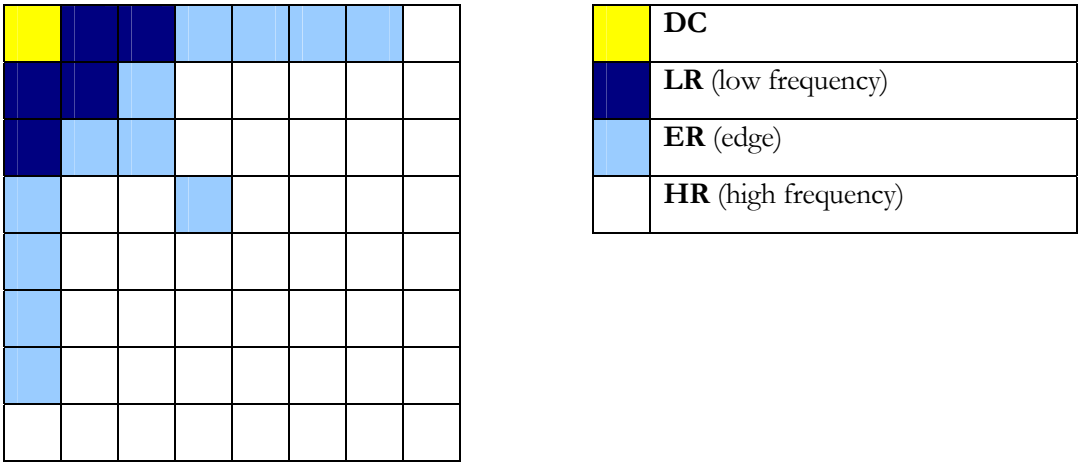

Figure 18. DCT block classification

The texture energy of the DCT block can be approximated by:

$$E_{TEX} = E + H \tag{5.4.2}$$

where L, E and H represent the sums of the absolute values of DCT coefficients in LR, ER and HR regions respectively.

Since information about edges is reflected in LR and ER portions and texture is mainly reflected in HR portion, it was determined that high magnitudes of the following two ratios indicate the presence of an edge:

$$E_1 = \frac{\overline{L} + \overline{E}}{\overline{H}} \tag{5.4.4}$$

$$E_2 = \frac{\overline{L}}{\overline{E}} \tag{5.4.5}$$

where $\overline{L}, \overline{E}$ and $\overline{H}$ denote mean energies in low-frequency, edge and high-frequency blocks respectively.

According to $E_{TEX,}$ E_1 and E_2 blocks can be classified using block classification algorithm given in Figure 19. Using five conditions in the classification procedure a block is assigned in the one of three classes:

- **Condition 1**: if $E_{TEX} \le \mu_1$ block is assigned to PLAIN class.

- **Condition 2**: For the blocks with high spatial activities ($E_{TEX} > \mu_3$) EDGE thresholds needs to be lowered and these thresholds are given in **Condition 3**. Otherwise, **Condition 4** is used.

- **Condition 3,4**: Two ratios E_1 and E_2 indicate presence of edges. If $E_1 \ge \gamma$ block is assigned to EDGE class. Similarly, if $max(E_1, E_2) \ge \alpha$ and $min(E_1, E_2) \ge \beta$ block is assigned to EDGE class. In **Condition 3** thresholds (α, β) are set to lower values (α_1, β_1), otherwise thresholds are set to be (α_2, β_2). If **Condition 3** has not been met block is assigned to TEXTURE class.

- **Condition 5**: If **Condition 4** is not satisfied, condition $E_{TEX} \le \mu_2$ is tested and if it is true then block is assigned to PLAIN class and otherwise to TEXTURE class.

Zhang et al derived following set of thresholds for edge-block detection: α_1=0.7, β_1=0.5, α_2=7, β_2=5, γ=16.

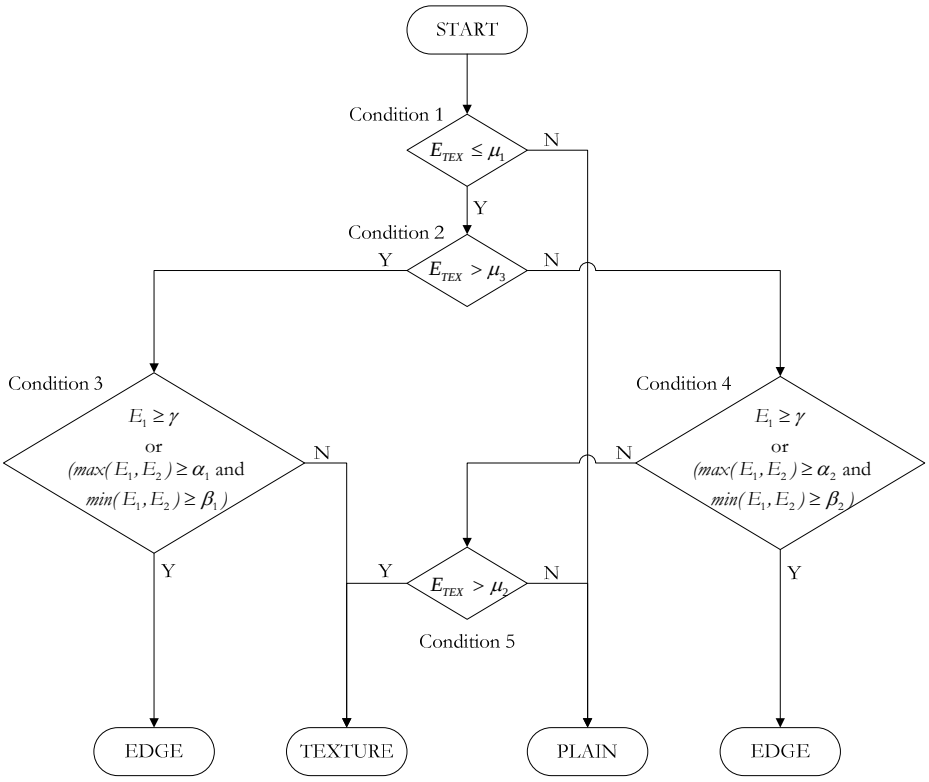

Figure 19. Block classification algorithm

According to the block class and its texture energy, inter-band elevation factor is derived using following formula:

$$\xi(n_1,n_2) = \begin{cases} 1 + 1.25 \cdot \dfrac{E_{TEX}(n_1,n_2) - \mu_2}{2\mu_3 - \mu_2} & \text{for TEXTURE block} \\ 1.25 & \text{for EDGE block and L + E > 400} \\ 1.125 & \text{for EDGE block and L + E} \le 400 \\ 1 & \text{for PLAIN block} \end{cases} \quad (5.4.6)$$

To include intra-band masking effect, Zhang et al used similar formula to Watson's one (5.3.12):

$$t_C^{ZHA}(n_1,n_2,i,j) = \begin{cases} \xi(n_1,n_2) & \text{for } (i,j) \in L \cup E \text{ in EDGE block} \\ \xi(n_1,n_2) \cdot max\left\{1, \left(\dfrac{|C(n_1,n_2,i,j)|}{t_{CSF}(n_1,n_2,i,j) \cdot t_l(n_1,n_2,i,j)}\right)^{w(i,j)}\right\} & \text{otherwise} \end{cases} \quad (5.4.7)$$

In this model for intra-band masking $w(i,j)$=0.36. Low (L) and edge (E) coefficients of an EDGE block are not elevated to avoid over-estimation of thresholds near edges.

Experimenting with two models in the given watermarking setup indeed shown that Zhang improved model gives better result concerning imperceptibility. However, Zhang model also tends to underestimate JNDs in edge blocks giving low decoding rates in "edgy" video sequences. Hence, there is a need to identify which parts of two models to adopt and incorporate in the watermarking amplitude adjustment model.

5.5 "Tuning-up" the model

As can be seen from the equation 5.3.1, just noticeable difference is modelled as a product of three thresholds: contrast sensitivity, luminance masking and contrast masking. In that way, we will analyze our watermark perceptual adaptation model.

First, we decided to use Contrast Sensitivity Table (CST) given in Table 4 [Cox01] to incorporate HVS's sensitivity to changes at different frequencies. Comparing the CST used by Cox in his watermarking application with CST given in Table 3 for standard definition, first one can notice that Cox thresholds are considerably lower. Moreover, it is worth noticing that thresholds raise more rapidly in Table 3 then in Table 4, for example ratio between lowest and highest frequency CST(7,7)/CST(1,0) is 95/3=31.67 in Table 3 and 21.5/1.01=21.29 in Table 4. The main reason for this is that in professional environment, high definition screens with smaller pixel sizes are used, leading to higher visibility of fine high frequency details comparing to standard definition television used in derivation of CST in Table 3.

Concerning the HVS's sensitivity to variations in luminance, Zhang luminance masking model have been employed in our watermark perceptual adaptation. As already mentioned in section 5.4.1, this model adapts well to the findings of subjective luminance sensitivity tests [Zha05]. The major difference, in comparison with Watson model, can be seen in lower grey level range (L<128), where Weber-Fechner's law is no longer valid.

Finally, the most appealing part for fine-tuning of perceptual adaptation is the contrast masking model. Block classification is a crucial part of this model, since it gives opportunity to separately analyze and to better adapt contrast masking thresholds. The algorithm used for block classification is shown in Figure 19.

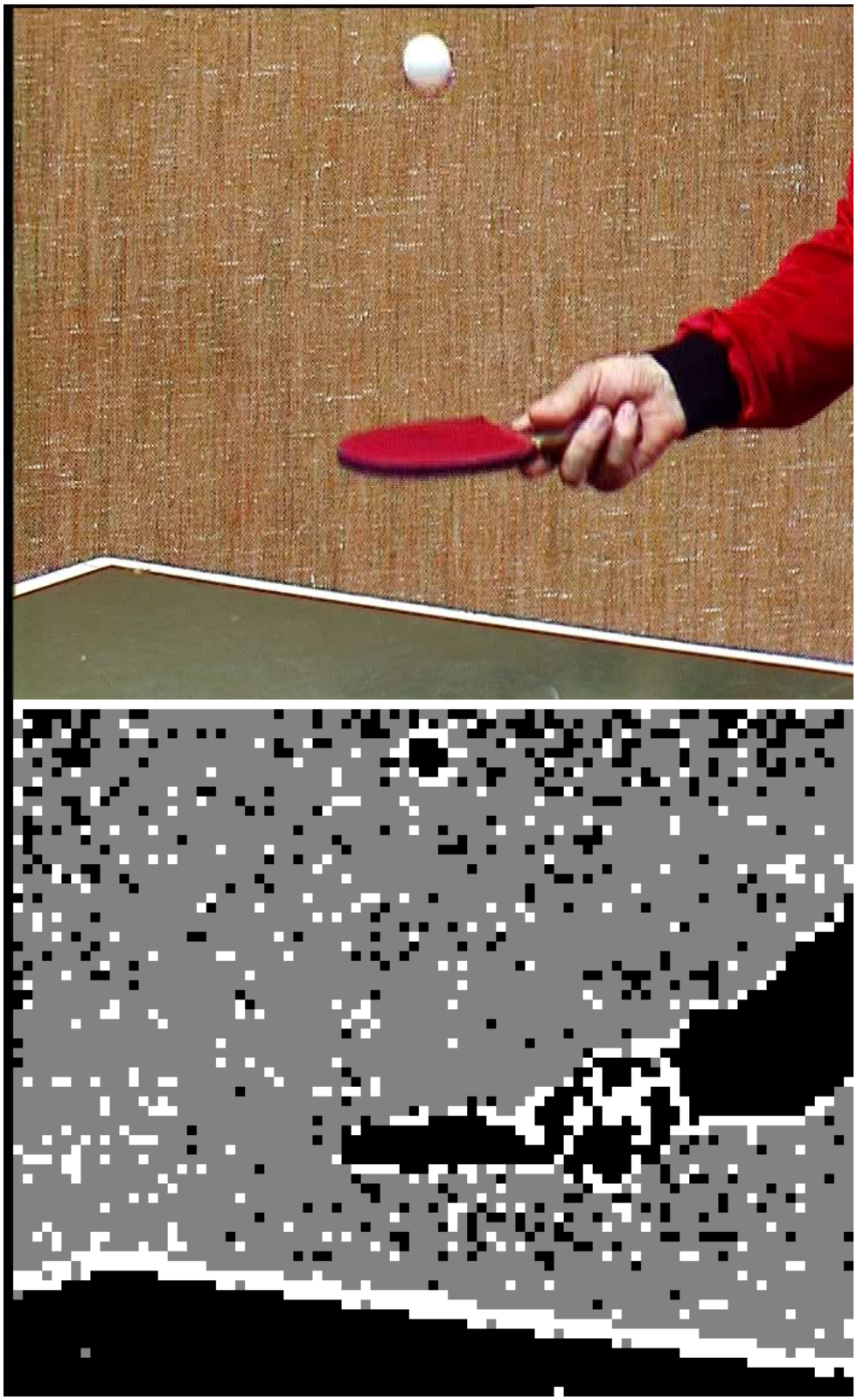

Figure 20. Table tennis sequence block classification (α_1=0.7, β_1=0.5, α_2=7, β_2=5, γ=16) – PLAIN (black), EDGE (white) and TEXTURE (grey) blocks

Result of block classification on a frame from the "Table Tennis" sequence is shown in Figure 20. This frame has been chosen, since it has proportion of plain, edge and texture blocks. For tuning the block classification algorithm, we were testing different sets of thresholds, starting with the one given by Tong and Venetsanopoulos in [Ton98] for energy levels in *L*, *E* and *H* areas of a block to the Zhang et al one given in [Zha05] for the mean energy levels in a block. It was observed that Zhang set of thresholds (α_1=0.7, β_1=0.5, α_2=7, β_2=5, γ=16) gives the best results in terms of smallest number of wrongly classified blocks, especially when distinguishing between texture and edge blocks.

In further experiments, Watson contrast masking model was under consideration. We examined the image quality dependence on the parameter w in the equation 5.3.12. It was noticed that Watson proposed value 0.7 indeed overestimates threshold round edges. Lowering the parameter, we came to conclusion that it needs to have value 0.36 proposed by Zhang et al. Detail of the edge from the "Table Tennis" sequence is present in Figure 21 for comparison.

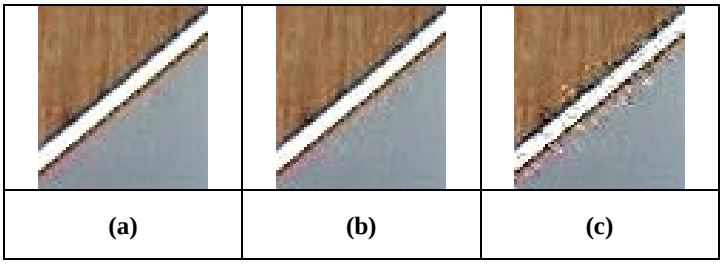

Figure 21. Edge detail from Table Tennis sequence: original **(a)** and watermarked with Watson model **(b)** w=0.36 and **(c)** w=0.7

On the other hand, in the Zhang model, as can be seen from equation 5.4.7, low (*L*) and edge (*E*) coefficients of an EDGE block are not elevated for intra-band masking, but only with inter-band elevation factor $\xi \in \{1.125, 1.25\}$. This tends to underestimate JNDs in EDGE blocks and gives a low watermark power (comparing to Watson method) in sequences that are mainly consisted of EDGE blocks (e.g. "BBC3"). In addition, Zhang method elevates high (*H*) coefficients with both intra and inter band elevation factor, so if there is a high-frequency noise already present around edge, Zhang method will amplify it and make it more visible and annoying. It was observed that using only Watson intra-band

masking for all coefficients in an EDGE block performs better then Zhang method concerning both embedding power and visibility of the artefacts.

Having this in mind, we are proposing a combined method that use the Watson model with w=0.36 for EDGE and PLAIN blocks and Zhang contrast masking elevation model in TEXTURE blocks. The proposed combined model, as it will be seen from experiments, outperforms both models by exploiting good characteristics of the Watson model in EDGE blocks and the Zhang model in Texture blocks. Concerning the image fidelity, it will be shown that it is still high and comparable with other two methods.

5.6 Performance of perceptual adjustment models in compressed domain watermarking

5.6.1 Perceptibility evaluation

Objective measures for perceptual image quality attempt to quantify visibility of differences between original and distorted image. The most common measure is Peak Signal to Noise Ratio (PSNR), widely used in describing quality of communication channel. However, this method does not take in account specifics of Human Visual System (HVS) and in that way might not match subjective perception of image quality. For this reason, frames with smallest PSNR are included for the sake of subjective evaluation.

The simplest method to evaluate distortions introduced by channel, i.e. the difference between transmitted and received signal are the mean square error (MSE) and the root mean square error (RMSE), computed by averaging the squared intensity differences of transmitted data chunks (in the case of video original and edited video pixels):

$$E_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - x_i^{'}\right)^2} \tag{5.6.1}$$

where:

x - original signal (image, video frame) and

x' - distort signal

An extension of Root Mean Square Error is peak signal-to-noise ratio (PSNR):

$$PSNR = 20 \times log\left(\frac{255}{E_{RMS}}\right) \tag{5.6.2}$$

These measures are appealing, because they are easy to calculate and have clear physical meanings. On the other hand, since RMSE and PSNR metrics treat equally the differences between pixels of the original and the distorted image, they might not match subjective perception of the image difference. In other words, these methods might calculate that the difference between original and watermarked video is high, but this will not be perceived by a human, since a distortion occurs in a part of the image where human eye is less sensitive, i.e. in the textured area that will mask introduced distortion.

The presented techniques were evaluated using five typical MPEG2 sequences (Table Tennis, Flower Garden, Mobile and Calendar, Suzy and BBC3). All sequences were 375 frames long, PAL (704x576, 25 fps), with GoP IBBP structure, size 12 and bit-rate 6 Mbps. These sequences are used since they have good proportion of plain (Suzy), edge (BBC3) and texture details (Flower Garden, Mobil and Calendar).

The result of the PSNR test is given in Table 5. A watermark is embedded in each of the five sequences and they are compared with the originals frame by frame. The table shows minimal PSNR values, maximal PSNRs and average PSNR for a whole sequence. The most interesting is the minimal value, which presents the most degraded I frame in a sequence. From the given results, it is possible to see that in most degraded frames a difference in PSNR between the proposed method and the other two is never bigger then 1.5 dB. The minimal PSNR value of 35.82 dB confirms that high fidelity is preserved and that watermarked frames are indistinguishable from originals.

It is worth noticing that sequences that have higher percentage of texture blocks (Flower Garden – 17.98% of texture blocks, Mobile and Calendar – 17.01%) are more degraded then the one consisted mainly of plain and edge blocks (Suzy – 0.05% of texture blocks and BBC3 – 1.76 %). This is consistent with the findings on sensitivity of Human Visual System that the noise is less visible in highly textured areas.

Table 5. Peak Signal to Noise Ratio comparison of three methods

		Perceptual adjustment method		
	PSNR	Zhang	Watson	Proposed
Table Tennis	Min	36.99	36.96	36.79
	Avg	39.62	40.09	39.58
	Max	55.77	53.99	53.1
Flower Garden	Min	37.26	37.01	36.54
	Avg	39.78	39.36	38.88
	Max	47.65	46.26	45.82
Mobile and Calendar	Min	36.31	37.06	35.82
	Avg	40.86	41.79	40.05
	Max	49.55	50.17	47.64
Suzy	Min	44.46	44.85	44.15
	Avg	48.8	48.99	48.31
	Max	56.9	56.99	55.61
BBC3	Min	40.39	39.61	38.98
	Avg	47.84	46.26	45.53
	Max	55.53	53.73	53.01

Following figures present further dissemination of the perceptual quality and fidelity. Frame by frame Peak Signal to Noise Ratios for "Mobile and Calendar", "Suzy" and "BBC3" sequences are given in figures 22, 25 and 28 respectively. In addition to the frame by frame PSNR, the most degraded frame per sequence is singled out and presented with its amplified embedded watermark signal (figures 23, 26 and 29) as well as original and watermarked frames for comparison (figures 24, 27 and 30). It can be observed that the watermark is not perceivable nor in the overall most degraded frame (frame 345 from "Mobile and Calendar"), nor in the most degraded frames from the plain (frame 0 from "Suzy") and the edge sequence (frame 0 from "BBC3).

Since the watermark is embedded in I frames local minimum PSNR for every GOP is on I frame and maximum PSNR is on the last B frame in the GOP. Moreover, it is possible to see that in sequences with static camera (like in first 170 frames in "Suzy" sequence Figure 25) following P and B frames in GOP are strongly relying on I frame. As a consequence, there is more drift of degradations through the frames comparing with sequences that have constant movement more changes (Figure 22 – frame by frame PSNR for "Mobile and Calendar" sequence).

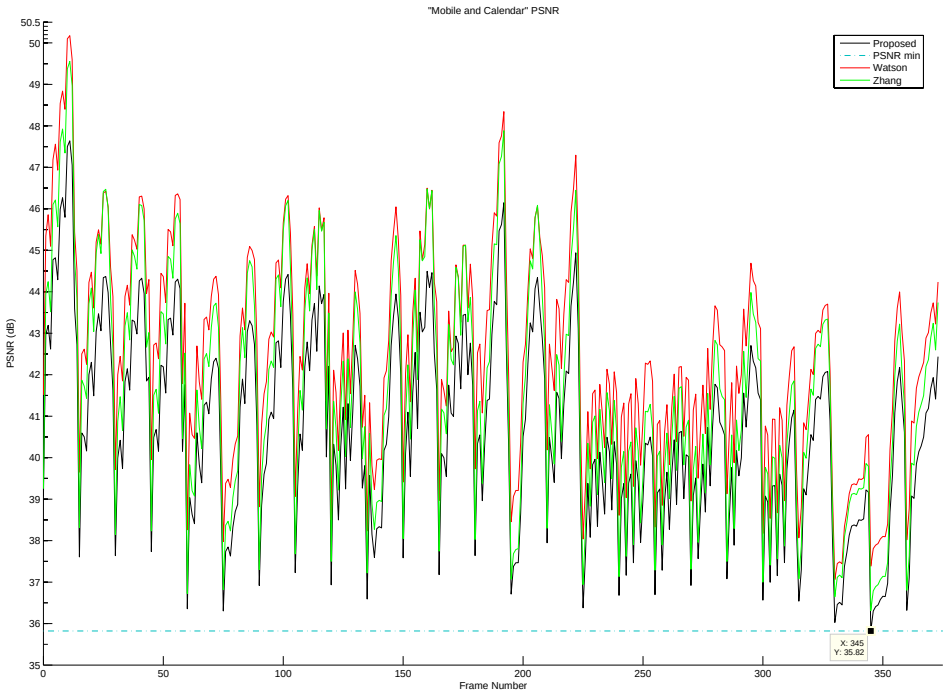

Figure 22. Peak Signal to Noise Ratio for "Mobile and Calendar" sequence

Figure 23. Amplified difference of the original and the watermarked frame 345 from "Mobile and Calendar" sequence

(a)

(b)

Figure 24. The most degraded frame (345) from "Mobile and Calendar" sequence: original (a) and watermarked (b)

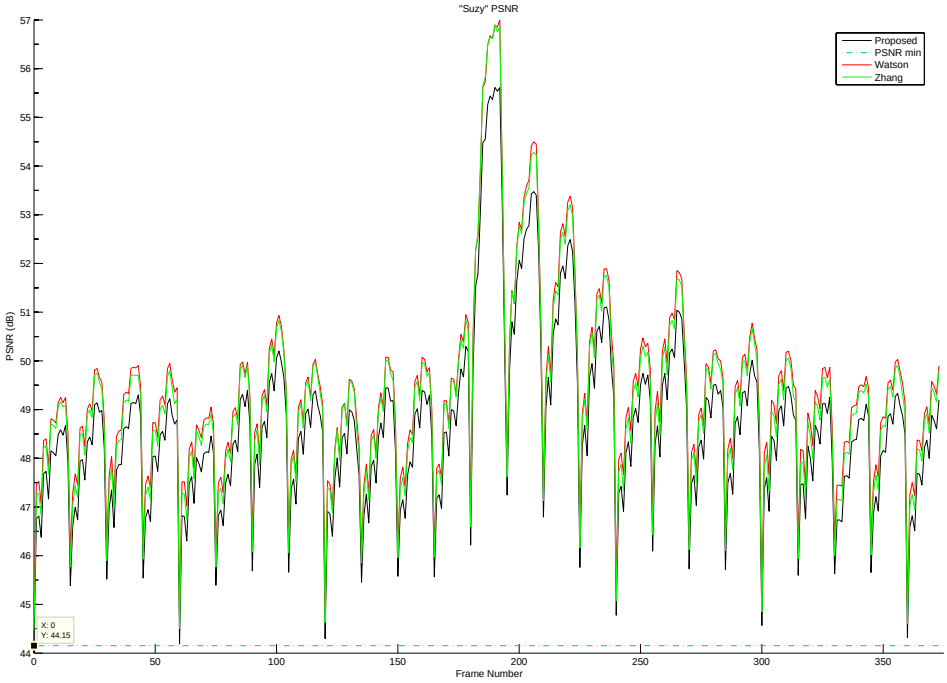

Figure 25. Peak Signal to Noise Ratio for "Suzy" sequence

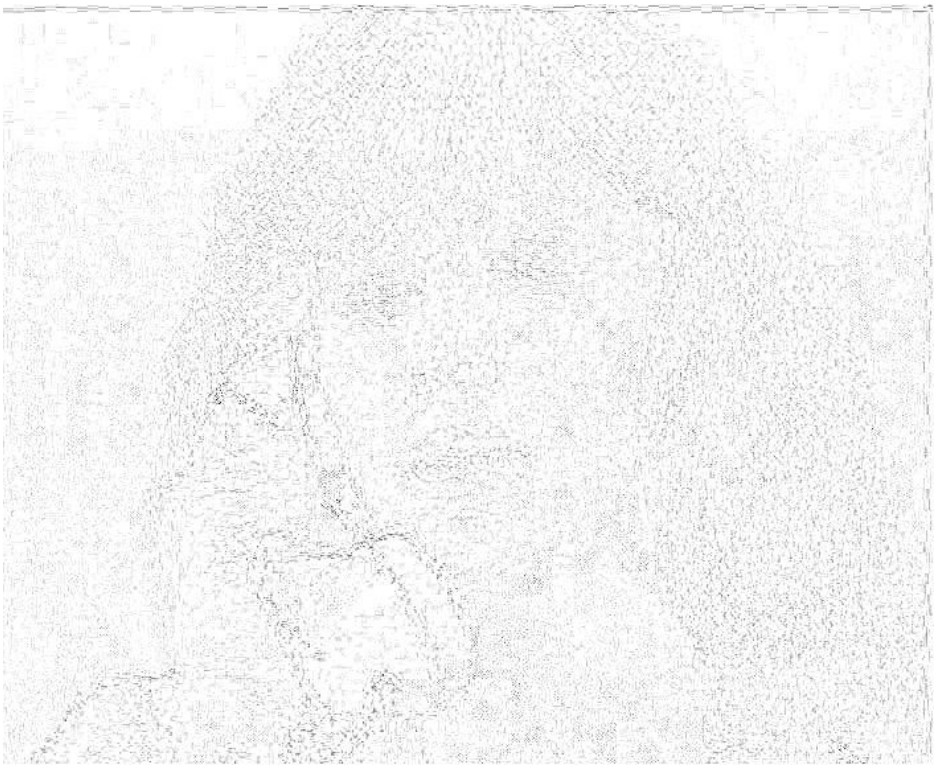

Figure 26. Amplified difference of the original and the watermarked frame 0 from "Suzy" sequence

(a)

(b)

Figure 27. The most degraded frame (0) from "Suzy" sequence: original (a) and watermarked (b)

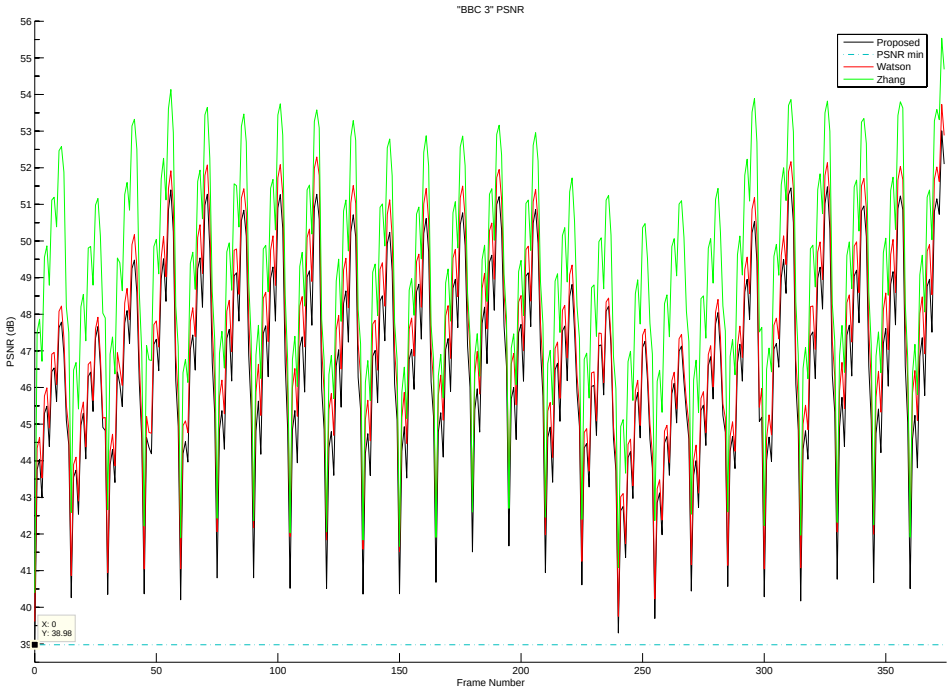

Figure 28. Peak Signal to Noise Ratio for "BBC 3" sequence

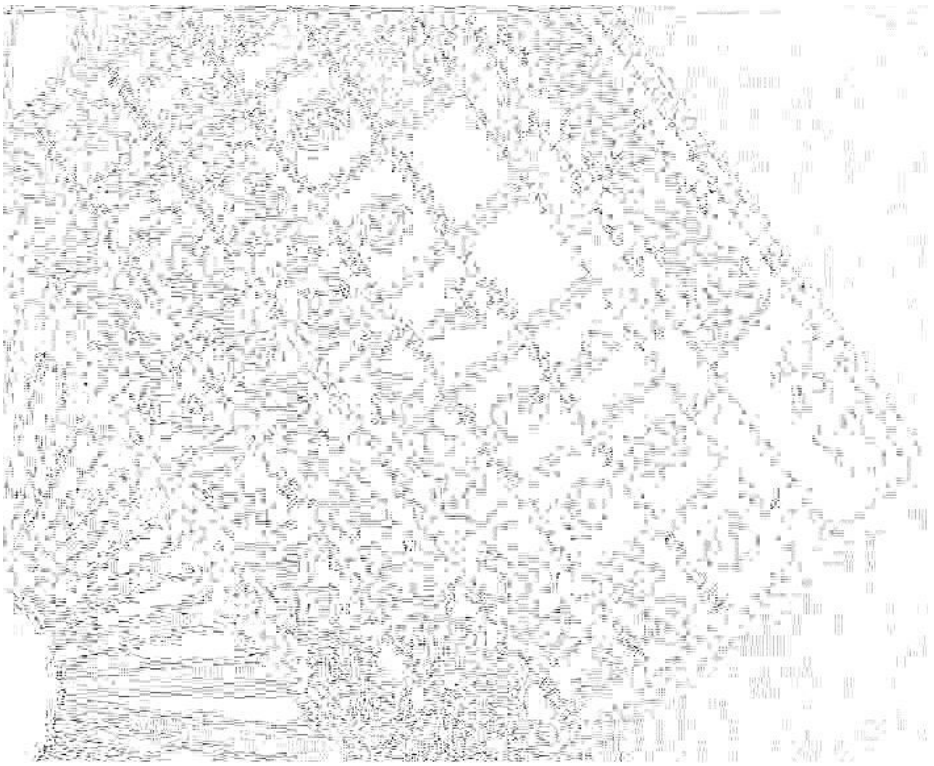

Figure 29. Amplified difference of the original and the watermarked frame 0 from "BBC 3" sequence

(a)

(b)

Figure 30. The most degraded frame (0) from "BBC 3" sequence: original (a) and watermarked (b)

5.6.2 Capacity and decoding rates

The results of perceptual evaluation showed that proposed method is comparable to other two methods and has high imperceptibility. However, the main advantage of the proposed combined method can be seen through a watermark to host ratio (*WHR*) that is watermark signal-to-noise ratio in an embedding window:

$$WHR = N_{nz} \cdot \frac{\mu_\alpha^2}{\overline{x^2}} \tag{5.6.1}$$

Table 6 shows embedding statistics when using three methods. There are three 8-frame embedding windows (EW) in any of tested sequences. For every embedding window, mean AC coefficients power ($\overline{x^2}$), number of non-zero AC coefficients (N_{nz}), mean embedding amplitudes (μ_α) and watermark to host ratios (WHR*)* are given. The WHR values for three methods show that the proposed method outperforms the other two methods by a large margin.

Table 6. Mean amplitudes and signal-to-noise ratio for different embedding windows

				Zhang		Watson		Proposed	
	EW	$\overline{x^2}$	N_{nz}	μ_α	*WHR*	μ_α	*WHR*	μ_α	*WHR*
TT	1	1675.23	859741	2.95	4466.19	2.91	4345.61	3.31	5622.75
	2	2067.58	1163959	4.13	9602.30	4.16	9722.13	4.46	11198.12
	3	2090.74	1170604	4.21	9923.71	4.28	10268.14	4.54	11540.42
FG	1	2700.77	1117868	4.39	7976.86	4.9	9937.911	5.23	11321.56
	2	2873.88	1040038	4.33	6785.10	4.87	8582.98	5.23	9898.83
	3	3170.28	835122	3.87	3945.24	4.39	5076.69	4.7	5818.99
M&G	1	3339.83	846145	2.85	2057.83	2.24	1271.20	3.56	3210.85
	2	3269.65	864405	3.02	798.40	2.47	1612.90	3.75	3717.73
	3	3055.7	930484	3.54	1077.95	2.99	2722.32	4.34	5735.58
Suzy	1	347.05	980522	2.15	13059.97	2.07	12106.15	2.23	14049.96
	2	348.76	925484	1.96	10194.23	1.91	9680.74	2.03	10935.39
	3	309.4	995247	2.03	13255.70	1.96	12357.27	2.1	14185.65
BBC3	1	4919.6	609132	2.31	660.70	3.02	1129.26	3.72	1713.43
	2	3603.53	648995	2.78	1391.88	3.01	1631.72	3.83	2641.86
	3	4939.23	613720	2.3	657.30	3.17	1248.61	3.81	1803.68

As expected, the lowest mean embedding amplitudes are observed in "Suzy" sequence, which mainly consists of plain blocks (93.95%). However, watermark to noise ratios in this sequence are quite high due to the low mean power of DCT coefficients and relatively high number of non-zero DCT coefficients. The most demanding sequence, as mentioned before, is the "BBC3" sequence. Consisting mainly of low-frequency transitions from black to white and vice-versa, this sequence contains relatively small number of DCT coefficients with high values describing strong edges with high luminescence changes. Although the mean embedding amplitudes for the "BBC3" sequence are higher then in the "Suzy" sequence, the WHR values for the "BBC3" are the smallest in the given set of testing sequences.

Considering the low number of TEXTURE blocks (1.76%) and relatively high number of EDGE blocks (35.42%) in the "BBC3" sequence, the Zhang method fails to compete with the other two methods and gives disappointingly small mean embedding amplitude and WHR values. On the other hand according to the results presented in Table 6, proposed method that incorporates Zhang method for luminance masking and Watson contrast masking method in EDGE blocks overwhelmingly outperforms the other two methods.

For the further dissemination of the three amplitude adjustment methods, we will focus on the first embedding window of the "BBC3" sequence, which has lowest WHR values. To estimate the maximal capacity of the watermarking message, we define signal-to-noise ratio per watermarking bit:

$$\frac{S}{N}[\text{dB}] = 10 \cdot log_{10} \frac{WHR}{n} \tag{5.6.2}$$

where n is the number of embedded bits.

The SNR curves, showing dependence of SNR on the used embedding method and number of bits per embedding window, are given in Figure 31. In the section 4.8, it was shown that it is possible to achieve BER ~10^{-5} when sending a BPSK signal through an AWGN channel with an SNR of 9.6dB. Having that in mind and from the given SNR curves, we conclude that in order to achieve bit error rate as low as 10^{-5}, we can embed maximum 72 bits using the Zhang method, 123 bit using Watson method or 188 bits using the proposed combined method for the watermark amplitude adjustment.

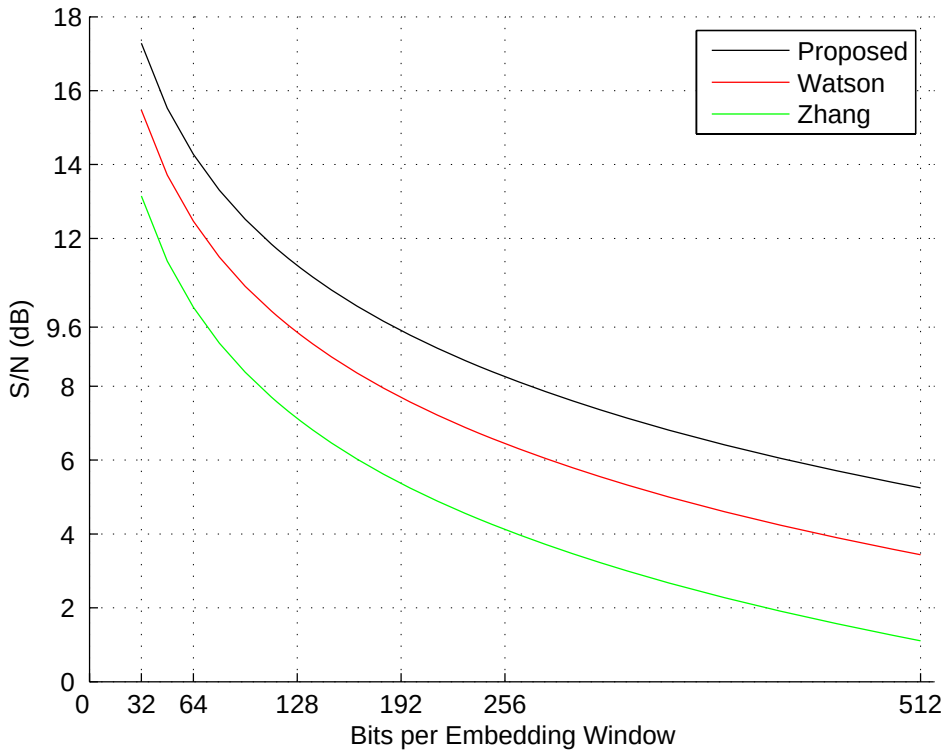

Figure 31. Signal-to-noise Ratio for the first embedding window in the "BBC 3" sequence for three methods

This maximal capacity rates were evaluated by measuring bit error rate. We were embedding 10^{+5} bits with different embedding message sizes (64-256 bits per 8 I frames) in the first embedding window of the "BBC3" sequence using three methods. Messages were extracted and compared with original ones. Measured bit error rates are shown in the Figure 32. It can be seen that measured values are comparable with estimated ones. We were able to decode 64-bit messages embedded with Zhang method without errors. Using Watson model 128-bit messages were decoded without errors, while as expected the best results are achieved with the proposed method. Here, 192-bit messages were extracted and no errors were observed, which is in consistent with estimated maximal capacity of 188-bits.

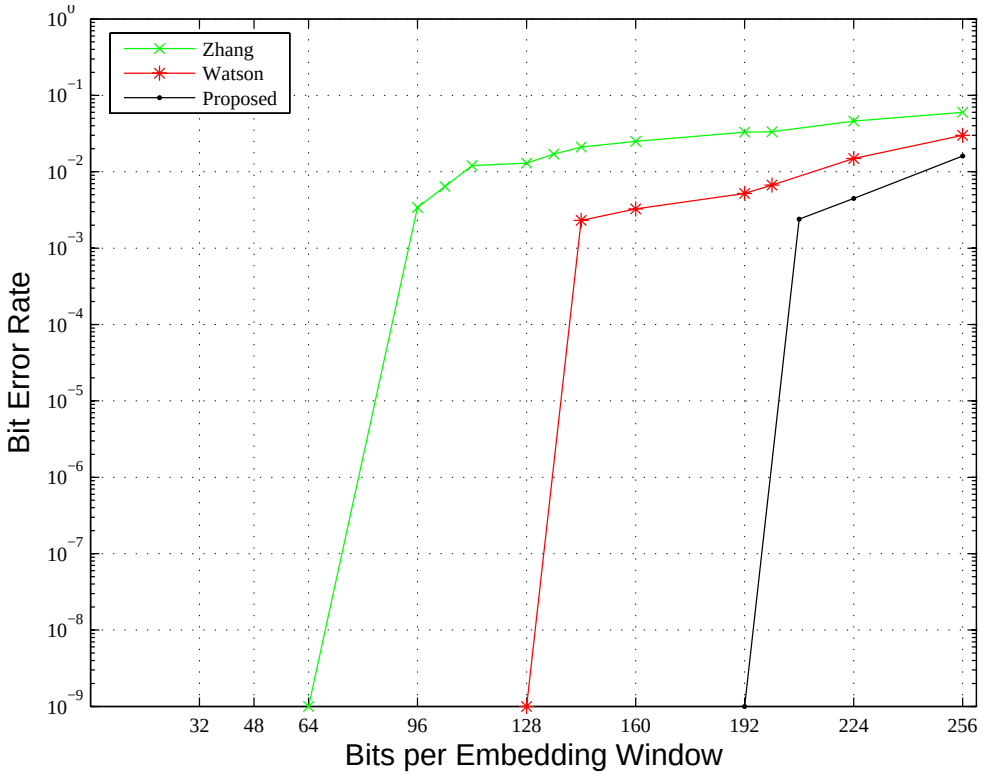

Figure 32. Bit Error Rates for three methods (Z- Zhang, W- Watson, P- proposed): no attack – 6Mbps and transcoding attack – 2Mbps.

Chapter 6 Capacity enhancement using Error Correction Coding

6.1 Introduction

The latest generation of watermarking techniques models the process as communication through a noisy channel. The channel noise is originated by two different sources. The video itself does not carry any useful information regarding the watermark message and from a watermarking point of view can be considered as noise. As it was shown in the chapter 4 using **block-wise watermark bits interleaving**, we can approximate this noise with Gaussian distribution. The other is noise originated by attacks and it is as well usually modelled as Gaussian white noise in the evaluations of watermarking systems.

We have already seen that the power of the watermarking signal is bounded by perceptual visibility. The watermark must be embedded in such way that it does not introduce visual artefacts to the host signal. In compressed domain watermarking, the second limiting factor is that the video bit-rate must remain the same, so the number of watermarked coefficients is smaller, and consequently the power of the watermarked signal is limited.

Probably, the best approach to overcome the noise introduced by the host signal is watermarking with side information [Mil01]. However, the complexity of this approach makes it less useful in the compressed domain. Extraction of side information from video needs one extra pass through the video, which introduces additional overheads, increasing the computational costs opposing the requirements for fast watermark embedding.

Another approach uses error correction coding to correct errors due to channel low signal-to-noise ratio. In 1993, C. Berrou, A. Glavieux, and P. Thitimajshima made a major breakthrough in channel coding theory with their pioneering work introducing Turbo coders, which enable near Shannon limit capacity [Ber93]. This technique is widely used in communication via low SNR channels, such as mobile communication, deep space communication and more recently in watermarking. To boost the capacity of the spread spectrum scheme proposed in chapter 4, we introduced turbo coding.

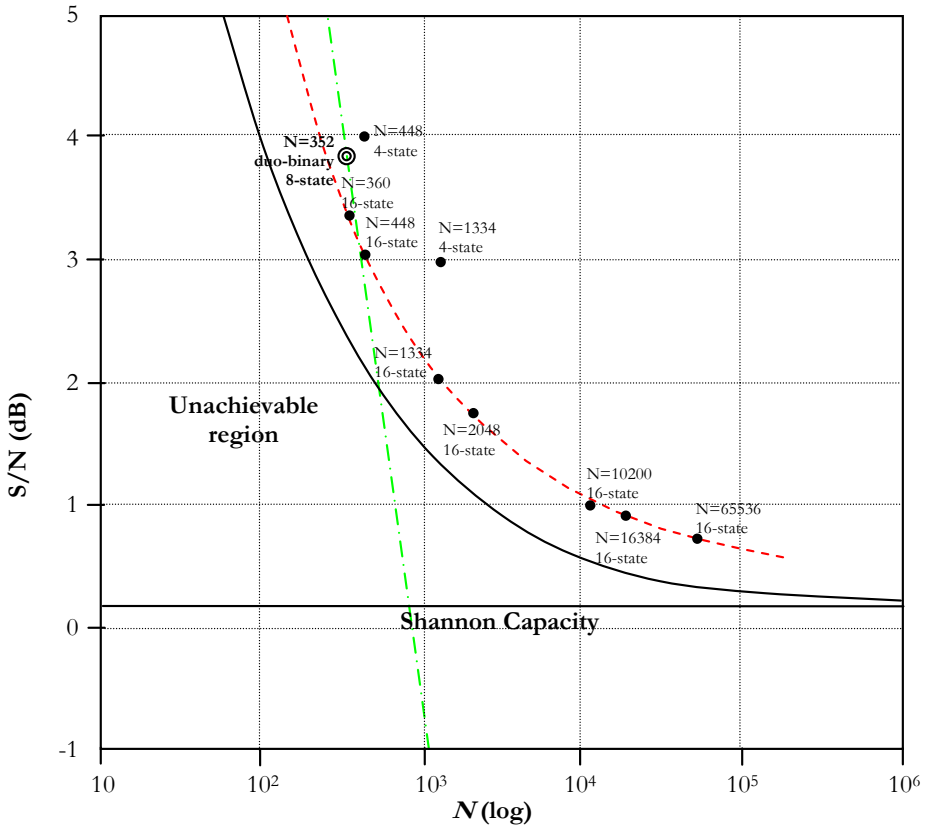

Figure 33. The sphere-packing bound and performance of different turbo codes depending on the code length *N*

The watermarking channel, as stated previously, has a small signal-to-noise ratio and a potentially large bit error rate due to the noise introduced by the host signal and attacks. In such an environment, it is essential to protect the watermark message by introducing redundant bits, which will be used for error correction. As pointed out by Ambroze in [Amb01], turbo coding and soft output Viterbi decoding, as usual maximum likelihood decoder for the AWGN channel, lead to significant gain in capacity.

Therefore, the question is – what capacity gain can be achieved if error correction coding is introduced in the watermarking system? Before turbo codes were introduced, there was the wide spread conviction that the Shannon capacity limit is achievable only if near infinite complexity is introduced in the decoder [Sch04]. It was argued that prohibitively large codes are required to approach this limit. Indeed, Shannon et al. [Sha67] derived the lower bound on probability of the codeword error P_B, known as sphere packing bound:

$$P_B > 2^{-N(E_{sp}(R)+o(N))}; \quad E_{sp}(R) = \max_{q} \max_{\rho>1} (E_0(q,\rho) - \rho R) \tag{6.1.1}$$

where $E_0(q,\rho)$ is Gallager exponent that depends on the symbol probability distribution q and the optimization parameter ρ and *R* is code rate. It is worth of noticing that there is exponential dependence of the lower bound on code length *N*. Schlegel and Perez in [Sch99] plotted this bound for rate 1/2 and BPSK signalling, together with different turbo codes (Figure33).

It is possible to see that turbo codes with code length of 10^{+4}-10^{+5} give near optimal performance and longer codewords can introduce only small gains. For the code lengths smaller then 10^{+4}, sub-optimality in the performance of an error-controlling scheme is inevitable. However, from the equation 5.6.2 we can see that signal-to-noise ratio is inversely proportional to the number of embedded bits, which is depicted in Figure 33. with green dash-dot line for the "BBC3" sequence and the proposed perceptual model (section 5.5). It is clear that in the given set-up only sub-optimal gains can be achieved with error correction coding. To embed more bits and to achieve better performance, we need either to embed with stronger amplitudes, which will introduce perceptual degradations, or to embed the bits in the longer video segment, which is opposite to the minimum watermarking segment requirement. Still, even with sub-optimal performance due to short code length, we were able to almost double the number of embedded bits. Using duo-binary turbo codes presented in the section 6.3, we embedded and decoded 352 bits without errors.

6.2 Parallel concatenated convolutional codes

The watermarking channel has a small signal-to-noise ratio and a potentially large bit error rate due to noise introduced by the host signal and attacks. Hence, it is essential to protect the watermark message by introducing redundant bits, which will be used for error correction. We were first experimenting with a classical turbo coder proposed in [Rya98]. This turbo coder is a parallel concatenation of two binary rate 1/2 Recursive Systematic Convolutional (RSC) encoders that are separated by an interleaver. The overall TC rate is 1/3 without puncturing (Figure 34). We used a puncturing mechanism to reduce the number of bits that needs to be embedded in the sequence.

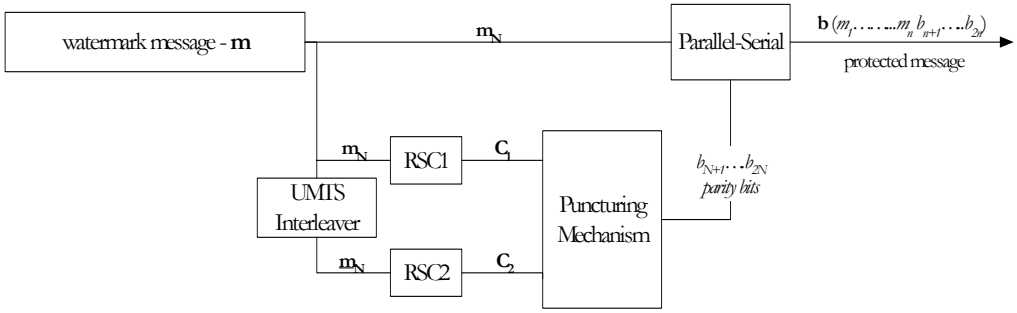

Figure 34. Parallel-concatenated convolutional coder

Redundant bits are produced by two Recursive Systematic Coders (RSC) (Figure35). The RSC encoder is rate 1/2 with 3 memory registers and generator polynomial $\mathbf{G} = \{g_1, g_2\} = \{17,15\}$, where feedback g_1 and the output g_2 are given in octal notation. The role of RSCs can be seen through their infinite impulse response, meaning that if a '1' in a data sequence is followed by a series of '0's, due to feedback an RSC will not turn to all-zero state. Why is this important?

The quality of an error control code is determined by the minimum number of bit positions in which any two codewords differ, called the minimum Hamming distance. This can be determined by comparing each codeword with the all-zeroes one. The minimum Hamming distance is then determined by the minimum codeweight (the minimum number of '1') that can occur. The important fact is that at high SNR, bit-error rate performance will depend on the minimum Hamming distance and will fall slowly with SNR, if the minimum Hamming distance is low. This slow rate of fall is called "error floor". However, at low SNR the whole distance spectrum needs to be taken into account.

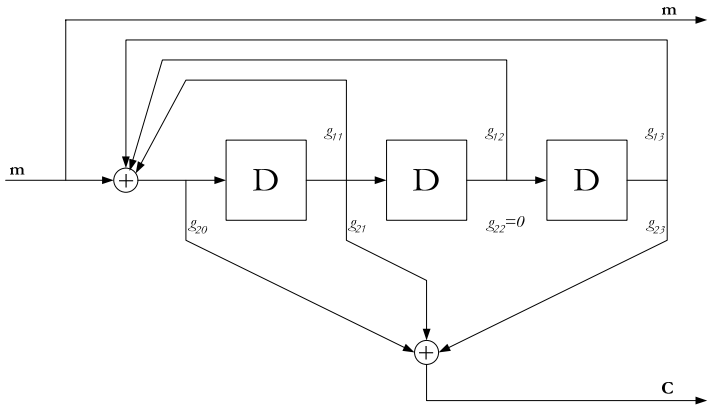

Figure 35. Recursive Systematic Coder $\mathbf{G} = \{g_1, g_2\} = \{17,15\}$

So, to achieve good performance of an error correction code, it is crucial to have a good distance spectrum. We have already seen that the infinite impulse response of an RSC code will tend to generate high weight sequence from the low weight information sequence. The inputs of two RSCs are separated by the interleaver. Typically used pseudo-random interleavers ensures that the input sequences of two RSCs $\mathbf{m}_N$ and $\underline{\mathbf{m}}_N$ are different, although the same weight yielding two codewords ($\mathbf{C_1}$ and $\mathbf{C_2}$) of different weights at the outputs of the encoders. In that way, the resulting composite codeword will, with high probability, have a high codeweight.

This low multiplicity of low codewords in distance spectrum makes turbo codes perform very well at low SNR. However, because of the pseudo-random nature of the interleaver, there will be a few pairs of input sequences that will result in a low weighted codeword at the outputs of both RSCs. Consequently, turbo coders that are using pseudo-random interleavers and have short input sequences will have small minimum Hamming distance and high error floor at high SNR. Thus, interlever design is crucial especially when turbo coders are used for protection of short messages. In our experiments, we were using the interleaver proposed for the UMTS turbo code implementation, since it was proved to have good performance when working with short block sizes [3G99]

To achieve overall rate R=1/2, RSCs outputs $\mathbf{C_1}$ and $\mathbf{C_2}$ are punctured and added to watermarked massage **m** to produce protected message **b**, which is then embedded to the video sequence. Puncturing is done by including only even bits from the output $\mathbf{C_1}$ and odd bits from the output $\mathbf{C_2}$ in the protected message. Having the overall rate 1/2 instead of 1/3, there is less bits to embed into the video sequence giving a higher SNR ratio per bit of the message. However, discarding half of protection bits leads to unavoidable degradation in turbo code performance.

At the decoder side, detection values from the watermark extractor, for the received message $\mathbf{m^r}$ and the received parity bits $\mathbf{b^r_1}$ and $\mathbf{b^r_2}$, are handed over to an iterative decoder as shown in Figure 36. The iterative turbo decoder consists of two Maximum A Posteriori decoders MAP_1 and MAP_2 dedicated to decoding of sequences from RSC_1 and RSC_2. First decoder MAP_1 takes as its input received bits of the message $\mathbf{m^r}$ and the received parity values $\mathbf{b^r_1}$ belonging to the first RSC to produce a sequence of soft estimates $\mathbf{E_{12}}$ of the transmitted bits **m**. The sequence $\mathbf{E_{12}}$ is called *extrinsic* data since it does not contain any information given to the decoder MAP_1 by the MAP_2. This information is interleaved and passed to the MAP_2 together with interleaved message data and the received parity bits

from the second RSC. The output data of MAP_2, the soft estimates $\mathbf{E_{21}}$ of the transmitted bits, are then de-interleaved and passed to MAP_1 as the extrinsic information formed without the aid of parity bits from the first RSC.

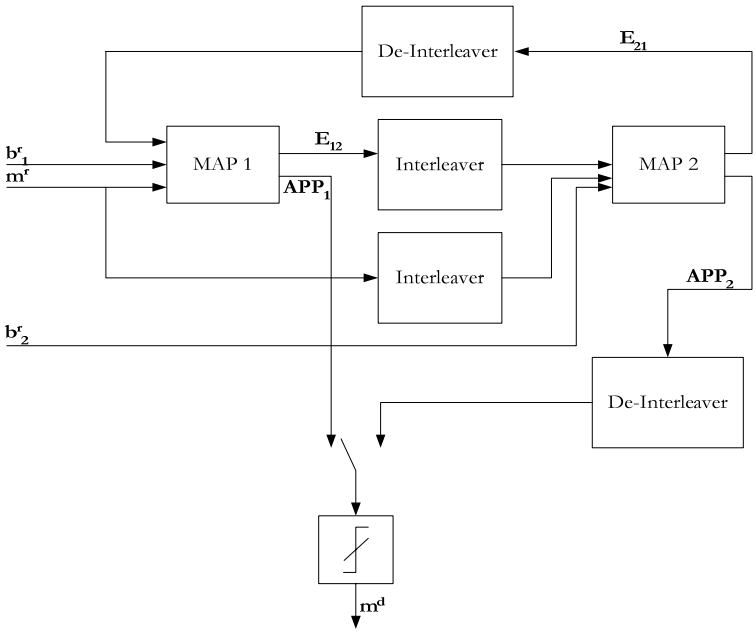

Figure 36. Iterative turbo decoder

After several iterations, estimated message bits from two decoders will tend to converge and the final decision can be made either by the first or second decoder. The estimates present the *a posteriori probabilities* (APP) of a transmitted '0's and '1's based on the systematic information and parity bits from both RSCs. More negative values represent a high probability that transmitted bit is '0' and more positive represent a high probability that '1' was transmitted. Finally, simple threshold operation is performed to produce the final hard decision $\mathbf{m^d}$ about the transmitted message bits.

MAP decoders are based on the Bahl, Cocke, Jelinek and Raviv (BCJR) algorithm [Bah74]. The BCJR algorithm tends to maximize the probability of correct detection of each input bit. Although quite complex, The BCJR algorithm is primarily used because it gives soft decision on the input bits and, in that way, suitable for an iterative decoder. Less complex algorithms, such as Soft Output Viterbi Algorithm (SOVA) or Max-Log-MAP, can be also used in iterative decoder, but their performance in terms of bit-error rate is inferior to the BCJR algorithm.

To evaluate the performance of the watermarking system protected by turbo coding technique described above, we embedded 10^{+5} bits with turbo coding protection and compared results with embedding unprotected messages. Uncoded and turbo coded messages of different sizes (64-640) were spread through 8 I frames (5 seconds of video sequence) and embedded in sequences.

Table 7. Peak Signal to Noise Ratio comparison – watermarked vs. watermarked and transcoded to 2Mbps

	PSNR	Wat	Wat+T2Mbps
BBC3	Min	38.98	28.05
	Avg	45.53	32.38
	Max	53.01	42.68

We also tested the robustness on transcoding to 2 Mbps. For the transcoding test, ffmpeg software [FFM04] was employed as the fastest available coder and also useful for batch processing that we extensively used in our tests. Since ffmpeg coder mainly compress a sequence by changing quantization steps, it indroduces severe degradations to the sequence. In Table 7, PSNR levels after watermarking and transcoding to 2 Mbps for the "BBC3" sequence are compared with PSNR levels after watermarking of the "BBC3" sequence (Table 5).

A watermark message was extracted from the first 8 I frames (embedding window with lowest *WHR*) and compared with a original message. Bit Error Rate (BER) results with and without turbo coding are given in Figure 37. During experiments, no errors were observed when communication 192-bit messages without turbo coding, while with turbo coding capacity is increased to 256 bits per 8 I frames that is 256 bits in 5 seconds of video, if there is no attack. If a watermarked sequence is transcoded to 2 Mbps, a 96-bit watermarked message embedded without protection can persist. In the case of protection with a classical turbo code with UMTS interlever, given the number of communicated bits (10^{+5}) and no errors observed with 128-bit messages, it can be say with 99% confidence that BER will be low as $4.61*10^{-5}$ [2].

[2] Given no errors observed in N experiments and desired confidence level C% it can be shown that [HFT00]:

$$BER < -\frac{ln(1-C)}{N}$$

From the given analysis, it can be seen that with a classical punctured turbo code with UMTS interleaver, message payload is increased by 32 bits (from 96 to 128 bits), in the presence of severe transcoding attack. This result is rather disappointing and can hardly justify additional computational costs. Due to small watermarking messages and puncturing mechanism, protection is suboptimal and increase in payload is small.

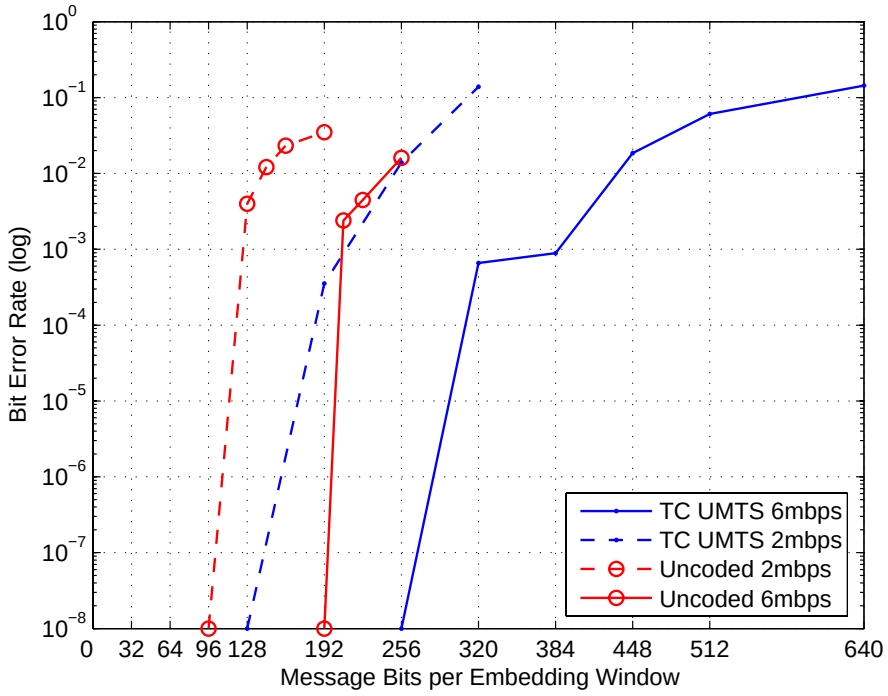

Figure 37. Bit Error Rates for protected (TC UMTS) and unprotected (Uncoded) watermark message: without attack – 6Mbps and with transcoding attack - 2Mbps.

6.3 Duo-binary turbo codes

The turbo coder presented in the previous section is a parallel concatenation of two binary rate 1/2 Recursive Systematic Convolutional (RSC) encoders that are separated by an interleaver. The overall TC rate is 1/3 without puncturing. We used puncturing mechanism to reduce the number of bits that needs to be embedded in the sequence. However, puncturing unavoidably leads to sub-optimal performance of a turbo code. More recently, Berrou et al. [Ber01] argued that non-binary turbo codes based on RSCs with $m \geq 2$ input bits outperforms classical binary turbo coders. In this section, we are experimenting with duo-binary turbo codes. These codes consist of two binary RSC encoders of rate 2/3 and an interleaver of length k. Each binary RSC encoder encodes pair of data bits and produces

one redundancy bit, so desired rate 1/2 is the natural rate of the double binary TC, so no puncturing is needed yielding better protection.

We considered the 8-state duo-binary TC with RSCs that have generator polynomial $\mathbf{G} = \{g_1, g_2\} = \{15,13\}$ as has been adopted by the ETSI (European Telecommunications Standards Institute) standards for Digital Video Broadcasting with Return Channel via Satellite (DVB-RCS) [DVB05] and Digital Video Broadcasting with Return Channel via Terrestrial (DVB-RCT) [DVB02] as shown in the Figure 38. The tail-biting [Bet98] technique, also called Circular Recursive Systematic Convolutional (CRSC) [Ber99], is used to convert the convolutional code to block code that allows any state of the encoder as the initial state. This technique encodes input bit sequence twice, first time the RSC initial state is all-zero state and final state is used to calculate the initial state for the second encoding, which will also be final state after second encoding. Therefore, this technique asure that initial and final state will be the same, so there is no need to tail bits to derive the encoders to the all-zero state.

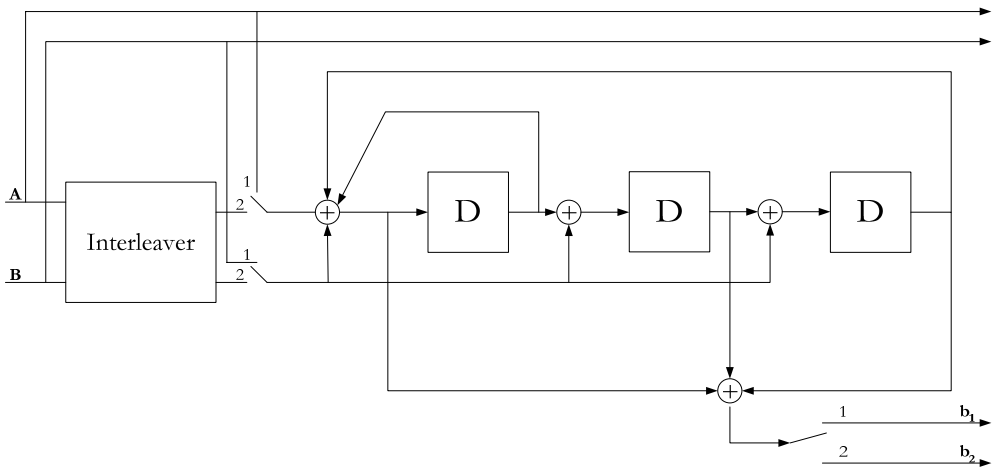

Figure 38. Duo-Binary Turbo Encoder

At low error rates or high signal-to-noise ratio, the performance of the classical turbo coder fluctuates due to the "error floor". The higher minimum distance can reduce the error floor effect at low error rates. Duo-binary turbo coders normally have better performance than classical turbo coders due to larger minimum distance. The minimum distance of turbo codes depends on the interleaver. The interleaver design is a critical issue and the performance of the turbo code depends on how well the information bits are scattered by the interleaver to encode the information by second binary RSC encoder. To get better

performance for the duo-binary code for watermarking channel, the particular block length is selected that behave better in the low error rates. This can be accomplished by using All-zero iterative method [Gar04] to check the performance of the duo-binary turbo code.

The turbo-decoder is composed of two Maximum A Posteriori (MAP) [Ber01] decoders, one for each stream produced by the singular RSC block as shown in Figure 39. The first MAP decoder receive the two distorted systematic bits ($\mathbf{A}^r_k$, $\mathbf{B}^r_k$) after channel along with the parity $\mathbf{b}^r_{k1}$ for first binary RSC encoder and produce the *extrinsic information* $\mathbf{E}_{12}$ that is interleaved and feed to the second MAP decoder as the *a priori* information. The second MAP decoder produces the *extrinsic information* $\mathbf{E}_{21}$ based on interleaved distorted systematic bits ($\mathbf{A}^r_k$, $\mathbf{B}^r_k$), distorted parity by second binary RSC encoder $\mathbf{b}^r_{k2}$ and *a priori* information from first MAP decoder. Then $\mathbf{E}_{21}$ is used as the *a priori* information of the first MAP decoder. After a certain number of iterations, usually 3 to 10, the *a posteriori probability* (APP) is taken, deinterleaved and performed hard decision to get transmitted information.

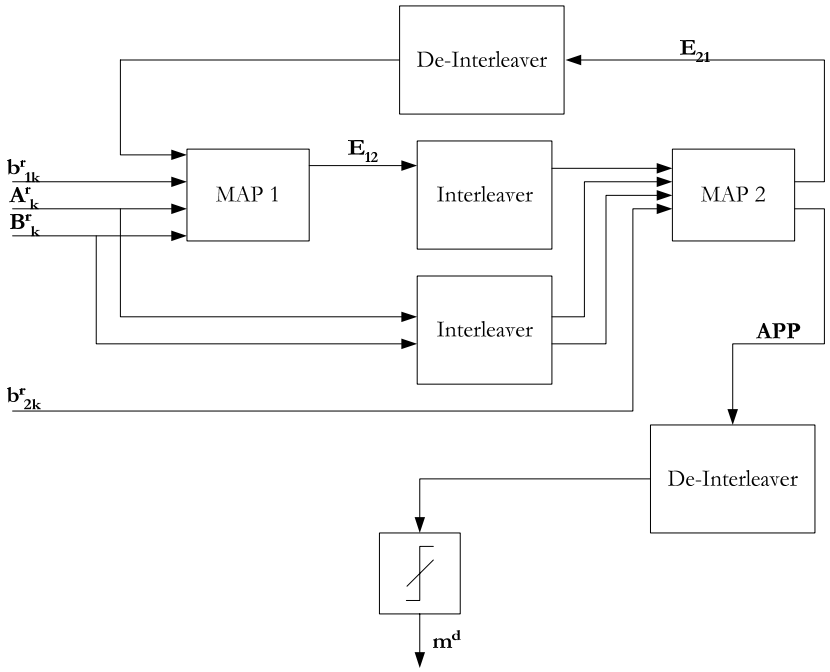

Figure 39. Iterative Turbo Decoding based on MAP algorithm for duo-binary TC

The Bit-Error Rate was measured in four different setups: uncoded without attack, uncoded and transcoded to 2 Mbps, turbo coded .without attack and turbocoded and transcoded to 2 Mbps. We simulated watermarking with different embedding packet sizes (96-496 bits per 8 I frames). To get meaningful results, we embedded ~10^{+5} bits per

simulation. Results are given in Figure 40. The iterative nature of the turbo coding shows more than a double gain in the embedded bits for uncoded watermarking messages at 6Mbps and after transcoding at 2 Mbps. A 352-bit watermark message is separated into two 176-bit sequences that are encoded with duo-binary turbo coder of rate 1/2 and after watermarking channel and turbo decoding, there is no error found. However, in order to resist transcoding watermark message needs to be at most 216 bits long. Again, given the experiment set-up, we can be 99% confident that bit error rate will be lower then $4.61*10^{-5}$.

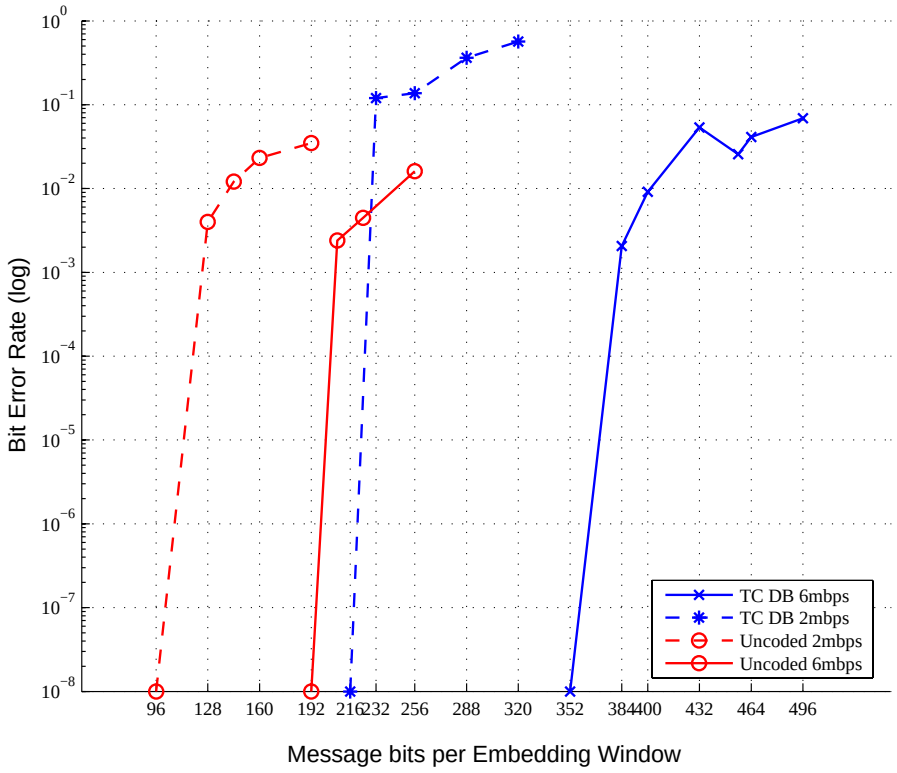

Figure 40. Bit Error Rates for protected (TC DB) and unprotected (Uncoded) watermark message: without attack – 6Mbps and with transcoding attack - 2Mbps.

It was shown that duo binary turbo codes can effectively increase the watermark payload. Duo-binary codes perform better then classical turbo coders in protection of watermarking channel, since they have natural rate of 1/2 and no puncturing is needed. Beyond that, they are computationally less expensive, show better convergence for iterative decoding and have a large minimum distance.

Chapter 7 Applications: a case study in a complex multimedia system

7.1 The BUSMAN project

The spread spectrum watermarking technique presented in this thesis was also a part of a complex project targeting the users of huge unstructured multimedia libraries. In order to realize the full potential of the multimedia libraries, tools to support usage, storage, retrieval and protection of digital visual data as well as intelligent search engines are urgently needed. This need was prime motivation and driving force of several European research institutions and industrial partners to start up the IST BUSMAN project aiming at tackling some of the most critical underlying problems.

The main objective of the project has been to design, implement, validate and trial an efficient system for querying and delivering of video from large databases. The BUSMAN project has been targeting delivery and querying across mobile and fixed networks. Two classes of users were considered: content provider and consumer. To satisfy the provider needs, advanced techniques for video indexing and watermarking have been developed for the semi-automatic generation of metadata, organization and its linking to the video. At the consumer side, efficient browsing and retrieval functionalities are implemented to provide user-friendly access to large video databases.

The project developed techniques to embed watermarks with different levels of robustness and capacity according to different applications. An important objective of BUSMAN was to develop, implement and test watermarking schemas for information embedding into low- and high-resolution content. Two techniques have been developed to deal with insertion of BUSMAN Content Tag (BCT), which is used as a pointer to metadata and will guarantee the availability of such information independently of delivery channel. This work is based on section three of the MPEG-21 standard called Digital Item Identification (DII) [ISO03]. This functionality makes any piece of digital video in a BUSMAN environment

able to establish it source, to obtain description and to express relationships to other video samples.

To achieve efficient and user-friendly indexing and retrieval of video content for seamless open interactive services, BUSMAN selected existing MPEG-7 description schemes (DSs) and descriptors (Ds) to organise the video content in databases [ISO01]. However, The BUSMAN is also addressing the gap in between the high level information that carries semantics of the context and the low level descriptors that can be processed automatically (e.g., color, shape or textures). Targeted video processing technologies are automatic video segmentation and key-frame extraction, semantic annotation and retrieval, advanced search techniques, such as 'query by example' and relevance-feedback.

A prototype system was designed and constructed aimed at demonstration and verification of all functionalities of the BUSMAN concepts. The main system functionality is management MPEG video. During the system design both desktop and mobile users were considered. The whole BUSMAN system is subdivided into three main functional modules: Input unit, Information Server and User Terminal. The implementation of all these three subsystems was conducted in parallel. Additional system software architecture, system integration and testing tasks provided the necessary glue between the three components of the prototype. BUSMAN defined Unique Identifiers strongly influenced by the evolving MPEG-21 standard in order to track the content along its use chain, and to be able to relate it to metadata describing it.

7.2 The BUSMAN system architecture

The BUSMAN system is designed to provide generic end-to-end multimedia content delivery, from content production, through analysis, indexing and watermarking, to delivering through heterogeneous networks and browsing and retrieval by final user [Vil03]. The overall architecture, shown in Figure 41**,** follows popular client-server model.

The server is consisted of two logical elements:

- *Input Unit* that includes modules for uploading content, analysis and feature extraction, indexing and watermarking.

- *Information Server* comprises a metadata server and a content server, which are used to store, maintain and deliver data to end users.

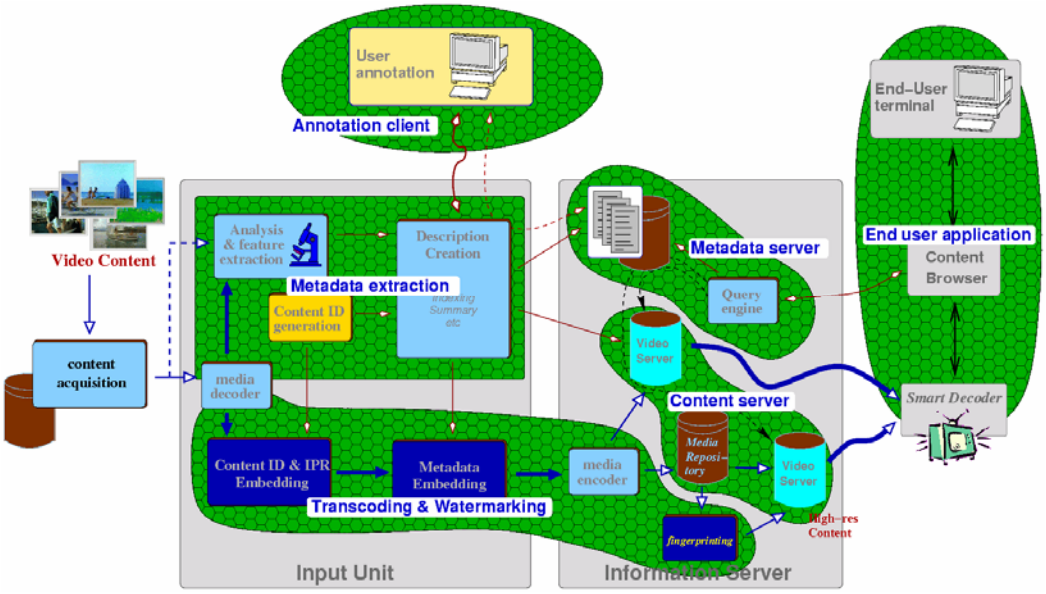

Figure 41. BUSMAN System Architecture diagram

On the client side we can identify two types of terminals:

- *User Annotation Client*, which interacts with the input unit, using advanced video technologies to assist a human annotator (professional user) in the semi-automatic video annotation.

- *End-User Terminal* is an advanced search engine that provides the interface for efficient querying, browsing and retrieving desired information through a fixed and mobile network.

The diagram also shows three different data flows along the system. Filled arrows show the flow of uncompressed media content, empty arrows present the flow of compressed media, while the flow of extracted or annotated metadata is illustrated with lines that have empty diamond arrows. Figure 42 further describes the processing steps associated with each of the three data flows within the server.

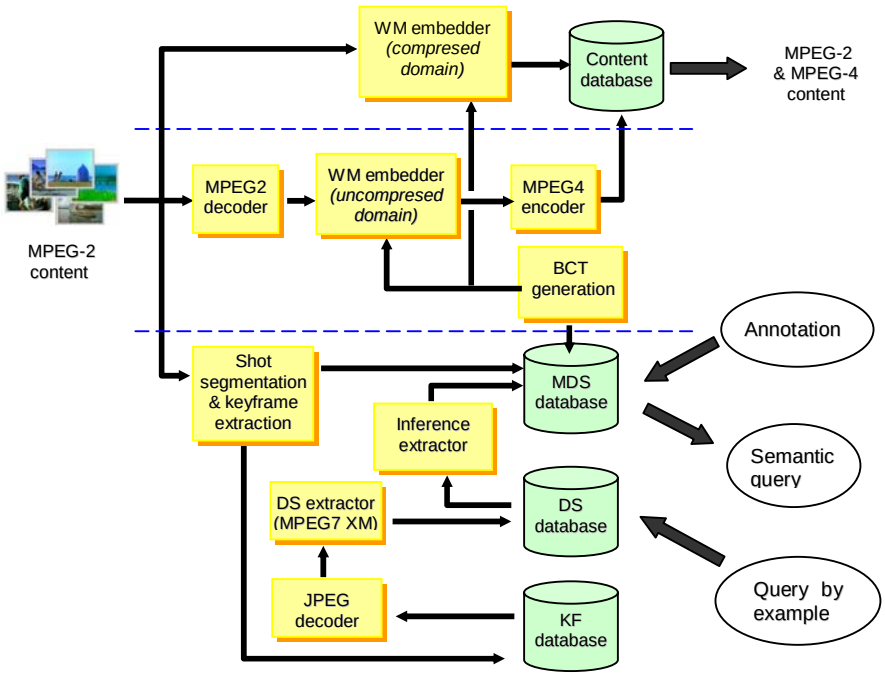

Figure 42. The data flows in the BUSMAN server [Izq04]

The uploaded MPEG-2 compressed video is associated with a unique BUSMAN Content Tag (BCT). In the first flow, the BCT is embedded in the content using watermarking in compressed domain technique and the content is stored in the server in the native MPEG-2 format. For future use in a mobile environment, in the second flow, MPEG-2 content is decoded, the BCT is embedded in the uncompressed stream and then transcoded and stored in various MPEG-4 format.

The most interesting part of the BUSMAN architecture is the third flow, describing the generation, storage and usage of the meta-data in the BUSMAN system. In this flow, the uploaded video material is firstly automatically segmented into shots and key-frames were extracted and stored in the key-frame database. Following the MPEG-7 standard, key-frames are used to extract low-level visual descriptors and stored in descriptor (Ds) database. These descriptors through an inference extractor, together with information from the shot detection and human annotation, as well as the BCT were used to form meta-data description, which is stored in the MetaData's server for a future semantic query and retrieval. The low-level descriptors were also used for measuring similarity metrics in the query by example and relevance feedback features of the system.

7.3 Watermarking techniques in the BUSMAN system

The task of the BUSMAN watermarking work has been the specification of video watermarking techniques dedicated to some relevant applications within the BUSMAN framework. The main objectives were to develop techniques to embed watermarks with different levels of robustness and capacity according to different application scenarios and the overall project objectives, to design efficient solutions for different levels of distortion according to potential attacks, to tailor developed watermarking techniques to user requirements and foreseen application scenarios.

The first goal was to clearly identify what applications are relevant in the framework of BUSMAN:

- In a B2B mode (the professional domain), the watermark carries a tag called BCT (BUSMAN Content Tag) associated with a piece of content in compliance with the MPEG-21 standard. The BCT plays the role of a pointer to metadata stored in professional databases. This technique needs to be robust to classical video editing processing. Moreover, one may wish to embed the watermark or decode the BCT without displaying the content. This supports the need to embed and decode from the video bit stream rather from raw video.

- In a B2C mode (the video is transmitted to a consumer end-user), the watermark carries some metadata (at least the BCT). As the video is transmitted over a heterogeneous network, one is sure that these metadata travel along with the content. The watermark is inserted when the video content is transcoded from MPEG-2 to MPEG-4. Thus, the watermark is embedded in the raw video. It is decoded as the player decompresses the video on the client. Thus, it is also decoded from raw video. It should be robust to some known transcoding processes such as change of bit rate. The payload is optimized with respect to this bit rate.

The watermarking technique developed for B2B scenario is actually the technique presented in this thesis. It aims at decompressing the minimum part of the video bit stream. Basically, it parses the stream, runs the inverse VLC and isolates quantized DCT coefficients of Intra frames. The video does not need to be fully decompressed. This saves computer power and time. The watermark embedding directly in MPEG-2 compressed stream was intended for the professional scenario defined in the project. Therefore, the

technical challenges that need to be met are: high imperceptibility, robustness to video editing processes and fast decoding speed.

The spread spectrum watermarking technique presented in previous chapters together with introduced improvements in terms of random bit spreading, bit-rate control on the macroblock level, perceptual adjustment and turbo coding protection manage to fulfil these requirements. In addition to the tests presented in previous chapters, the technique was also evaluated by one of the partners involved in the BUSMAN project. In these experiments, the bit rate was 6 Mbps. The GOP pattern was an I-frame every 12 frames and 2 B-frames between I and P-frames. The whole testing process consists of decompressing the watermarked file, performing the edit and recompressing again.

From performed tests, the watermark has been able to survive recompressing, greyscaling, pixel level changes (from 0.7 to 1.1), fade-in, cross-fade, logo insertion and titling (the word "BUSMAN" - size of 80% of the width and 20% of the height). However, the decoder failed to extract the watermark after attacks such as flip horizontal, turn 180 degrees, shift left, stretch, horizontal squeeze, 2-stack horizontal and 2-stack vertical.

These results were rather expected, since redundancy, introduced by spread spectrum and error-correction coding, makes the technique quite robust against attacks that degrade the video, but do not change its geometry. On the other hand, the scheme has been proven not to be robust against temporal and geometrical degradations, since no synchronization or RST attacks inversion algorithms were used in the scheme. These algorithms can be incorporated to make the technique more robust and suitable for more demanding applications, but it will seriously affect decoding time.

Chapter 8 Conclusions and Future Work

The thesis presents a thorough study of watermarking of MPEG-2 compressed video sequences. The thesis focused on finding a novel solution for content tagging of compressed video sequences. The main aim was efficient video indexing of huge unstructured MPEG-2 video archives that emerged with easiness of the production, storage and distribution of the video materials in the digital era. With respect to professional users and their requirements for real-time processing, the thesis is targeting embedding in compressed domain.

As a starting point, applications, requirements and techniques for digital watermarking were discussed. The thorough analysis of the digital watermarking techniques, starting from simple LSB to demanding watermarking with side information, gave crucial understanding of what can be achieved with different techniques and at what cost. After an introduction to MPEG-2 compression standard, an exhaustive survey of the state-of-the-art MPEG-2 watermarking techniques was given. Even though an intensive research in past fifteen years gave a qualitative advance in digital watermarking theory and numerous of watermarking techniques, there is still lack of robust and reliable solution for video watermarking.

The work started with implementation of a spread spectrum technique in compressed domain. As this technique did not give desired results in terms of perceptibility, data payload and robustness, we proposed a several improvements to this technique. Firstly, a new watermark composition scheme with **block-wise random watermark bits interleaving** has rapidly enhanced performance of the system. The main result of this scheme is similar detection probability for each watermarking bit, since the bits are spread evenly through textured, edge and plain area of the video image. In addition, with the interleaved bit spreading, distribution of detection values can be approximated with normal distribution giving a much easier way for theoretical analysis of the watermarking system performance.

A novel approach to bit-rate preserving, called **the bit-rate control on the macroblock level,** is given. In past schemes, in order to preserve the bit-rate, a DCT coefficient was

altered only if its new VLC code is smaller or equal to the original one. In the new scheme, we proposed to compare watermarked macroblock size with the original macroblock size increased by the difference left after comparing previous macroblocks. If the watermarked macroblock is bigger, watermarked coefficients with largest increase in VLC size are swapped with the original one until the watermarked macroblock is smaller or equal to the original one. It was shown that the percentage of altered AC coefficients is almost doubled in comparison with previous schemes.

After careful analysis of the capacity and capacity boundaries of the technique, special attention was given to increase of the watermark power by watermark strength adjustment. The method takes advantages of information about local characteristics of the picture, information that can be easily extracted from DCT coefficients. Two state-of-the-art methods for Just Noticeable Difference (JND) estimation were considered and a new model capitalizing on the good characteristics of two models was proposed. The results of experiments with the new model showed PSNR levels comparable to the previous model, but at the same time significant increase in the capacity.

To boost the capacity of our technique we introduced state-of-the-art error correction coding technique – turbo coding. The watermarking channel has a small signal-to-noise ratio and a potentially large bit error rate due to noise introduced by the host signal and attacks. In such an environment, it is essential to protect the watermark message by introducing redundant bits, which will be used for error correction. Performances of two state-of-the art turbo coding techniques were examined. It was shown that by employing a duo-binary turbo code it is possible to double the capacity of the presented watermarking technique. This technique gave near optimal utilization of the available signal-to-noise ratio in the watermarking channel (Figure 33).

In the beginning of this project, the main aim was to embed imperceptible 64-bit watermark into 5 seconds video segment, which can resist typical editing attacks. As the result of this research, with all introduced techniques we were able to embed 216 bits in a robust and imperceptible way. The watermark was able to resist recompression, greyscaling, pixel level changes, cross-fade, titling as well as severe transcoding attack.

However, before considering the presented technique for future commercial deployment more rigorous tests needs to be performed. In the chapter 5, it was pointed out that there is still lack of good objective measure for perceptual degradations of an image. In that way,

proper subjective test needs to be performed. One such subjective test for video quality in multimedia is recently proposed by European Broadcast Union (EBU) [Koz04].

The presented watermark scheme is not robust against geometrical and temporal attacks. Techniques that can answer this problem are more computationally expensive and could hardly be performed in real-time. However, to make the presented watermarking technique suitable for more demanding applications such as copyrights protection, they surely need to be tackled. The future development of the technique should look for compressed domain mechanisms that will introduce robustness to temporal and RST attacks.

In the first stages of our research, we were considering embedding in DC coefficients and in P and B frames of a MPEG-2 compressed video. Initial experiments showed that achievable increase in the capacity could hardly justify introduced distortions and additional computational costs, so we were reluctant to use these two watermarking channels. However, they can be used to embed low capacity temporal and spatial templates that can be used to revert attacks such as frame dropping, frame swapping or in the case of spatial attacks rotation, scaling, translation, frame resolution changes etc.

For example, by changing DC coefficients to force predefined relationship between mean luminance levels of secretly chosen areas of the frame, we can detect if particular frame contain the watermark in AC coefficients or not. Using this mechanism, we can easily find eight frames that contain the watermark and then extract secret information. Similarly, secret predefined patch can be embedded through the frames. By correlating the video frames with the same patch but with different angles and sizes at receiver, it is possible to revert rotation and scaling operations that might be applied to the video.

Video watermarking is a relatively young topic in digital watermarking research. Due to its much bigger complexity comparing to image watermarking, it still draws less attention then image watermarking. However, there is a strong belief that the ideas presented in this thesis together with ideas from other authors focused on video watermarking will attract more research in the area and that the performance of a video watermarking system could be comparable with the performance of an image watermarking system in the near future.

Apendix

A. Capacity of AWGN channel – Gaussian input signal

The capacity of the channel is defined as the maximum of the mutual information between the input signal **X** and the output signal **Y** over all distribution of the input signal. It can be also shown that the Gaussian distribution of the input signal maximizes mutual information [Cov91].

$$C = \max_{p(x):EX^2 \le S} I(\mathbf{X};\mathbf{Y}) \tag{A.1}$$

Mutual information $I(\mathbf{X};\mathbf{Y})$ is given as follows:

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &= H(\mathbf{X}) - H(\mathbf{X}\mid\mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}\mid\mathbf{X}) \\ &= H(\mathbf{Y}) - H(\mathbf{X}+\mathbf{Z}\mid\mathbf{X}) \\ &= H(\mathbf{Y}) - H(\mathbf{Z}\mid\mathbf{X}) \\ &= H(\mathbf{Y}) - H(\mathbf{Z}) \end{aligned} \tag{A.2}$$

where $\mathbf{Z} \sim N(0,N)$ is Gaussian noise independent of the input signal $\mathbf{X} \sim N(0,S)$. Now, we need to find the entropy of signal with Normal distribution:

$$\begin{aligned} H(\mathbf{Y}) &= -\int p(y)\cdot \log_2 p(y)\,dy \\ &= -\frac{1}{\ln 2}\int p(y)\cdot\left(-\frac{y^2}{2\sigma_y^2} - \ln\sqrt{2\pi\sigma_y^2}\right)dy \\ &= \frac{1}{\ln 2}\left(\frac{EY^2}{2\sigma_y^2} + \frac{1}{2}\ln 2\pi\sigma_y^2\right) \\ &= \frac{1}{\ln 2}\left(\frac{1}{2}\ln e + \frac{1}{2}\ln 2\pi\sigma_y^2\right) \\ &= \frac{1}{2}\log_2 2\pi e\sigma_y^2 \end{aligned} \tag{A.4}$$

Same can be applied to $H(\mathbf{Z})$ and we can also conclude that:

$$\sigma_y^2 = E\mathbf{Y}^2 = E(\mathbf{X}+\mathbf{Z})^2 = E\mathbf{X}^2 + 2E\mathbf{X}E\mathbf{Z} + E\mathbf{Z}^2 = S + N \tag{A.5}$$

since **X** and **Z** are independent and $E\mathbf{Z}=0$. Applying A.4 and A.5 to A.2, we obtain:

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &\le \frac{1}{2}\log_2 2\pi e(S+N) - \frac{1}{2}\log_2 2\pi eN \\ &= \frac{1}{2}\log_2\left(1+\frac{S}{N}\right) \end{aligned} \tag{A.6}$$

Hence, maximum information capacity of the Gaussian channel is obtained with the input signal that have normal distribution and it is equal:

$$C = \max_{EX^2 \le S} I(\mathbf{X};\mathbf{Y}) = \frac{1}{2}\log_2\left(1+\frac{S}{N}\right) \tag{A.7}$$

The formula A.7 is famous Shannon bound, given by its version for bandwidth limited channels in [Sha48] as a part of Shannon theorem which asserts that signalling with vanishing error probability is possible if transmission rate R is smaller then C.

B. Capacity of AWGN channel with BPSK signaling

Shannon bound (A.7) for AWGN channel capacity is given for the input signal with normal distribution, which maximizes the capacity of the channel. However, in given watermarking scheme we are using spread spectrum technique with Binary Phase Shift Keying (BPSK), so it would be useful to find the capacity boundary when signalling is restricted to BPSK. In this case, channel can be modelled as:

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}, \quad \mathbf{X} \in \left\{-\sqrt{S}, +\sqrt{S}\right\}, \quad \mathbf{Z} \sim N(0, N) \tag{B.1}$$

where S is signal power and N is noise power in the channel. To show that capacity of BPSK is dependent only on these parameters through their ratio SNR, we can replace $\mathbf{Y}$ with $\mathbf{Y}/\sqrt{N}$ to get a new channel model:

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}, \quad \mathbf{X} \in \left\{-\sqrt{\frac{S}{N}}, +\sqrt{\frac{S}{N}}\right\}, \quad \mathbf{Z} \sim N(0,1) \tag{B.2}$$

to simplify the notation we will set $A = \sqrt{\frac{S}{N}}$. Mutual information is then:

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}) &= H(\mathbf{Y}) - H(\mathbf{Y} \mid \mathbf{X}) \\ &= \sum_{x \in \{\pm A\}} \int_{-\infty}^{+\infty} p(x, y) \cdot \log_2 \frac{p(y \mid x)}{p(y)} dy \\ &= \sum_{x \in \{\pm A\}} \int_{-\infty}^{+\infty} p(x)\, p(y \mid x) \cdot \log_2 \frac{p(y \mid x)}{\sum_{x'} p(x')\, p(y \mid x')} dy \end{aligned} \tag{B.3}$$

The input symbols are equiprobable, so $p(x) = \frac{1}{2}$ and probability of the output signal y given the input signal x is:

$$p(y \mid x = -A) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y+A)^2} \tag{B.4}$$

$$p(y \mid x = +A) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-A)^2} \tag{B.5}$$

Inserting given probability density functions in B.3. we obtain:

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &= 2\cdot\int_{-\infty}^{+\infty}\frac{1}{2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-A)^2}\cdot\log_2\frac{2}{1+e^{-2yA}}dy \\ &= \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-A)^2}\cdot\log_2 2dy-\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-A)^2}\cdot\log_2(1+e^{-2yA})dy \\ &= 1-\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-A)^2}\cdot\log_2(1+e^{-2yA})dy \end{aligned} \tag{B.6}$$

The last equality gives the capacity of AWGN channel with BPSK signalling:

$$C_{BPSK}=1-\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-A)^2}\cdot\log_2(1+e^{-2yA})dy \tag{B.7}$$

This capacity boundary is solvable using numerical methods and mathematical software such as MATLAB. It can be shown that it is not necessary to evaluate the given integral over infinite boundaries and that is enough to evaluate the integral from -5 to 20 [Pur01]:

$$C_{BPSK}=1-\frac{1}{\sqrt{2\pi}}\int_{-5}^{+20}e^{-\frac{1}{2}(y-A)^2}\cdot\log_2(1+e^{-2yA})dy \tag{B.8}$$

Authors publications

[Dam06] I. Damnjanovic, N. Ramzan and E. Izquierdo, "MPEG-2 Watermarking Channel Protection Using Duo-binary Turbo Codes"**,** *accepted for IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP-06*, Toulouse, France, 15-19 May 2006.

[Dam05a]I. Damnjanovic and E. Izquierdo, "Turbo Coding Protection of Compressed Domain Watermarking Channel", *Proc. of IEEE International Conference on Computer as a Tool EUROCON 2005*, Belgrade, Serbia and Montenegro, November 2005.

[Dam05b]I. Damnjanovic, and E. Izquierdo, "Capacity Enhancement of Compressed Video Watermarking Using Turbo Codes", *6th European workshop on Image Analysis for Multimedia Interactive Services, Wiamis'05*, Montreux, Switzerland, 13-15 April 2005.

[Izq04] E. Izquierdo, I. Damnjanovic, P. Villegas, L.-Q. Xu, S. Herrmann, "Bringing user satisfaction to media access: the IST BUSMAN Project", *Proc. of. Eighth International Conference on Information Visualisation IV 2004*, 14-16 July 2004, pp.444 – 449.

BUSMAN project deliverables:

[Fur03] T. Furon, J. Delhumeau, C. Guillemot, S. Pateux, I. Damnjanovic and E. Izquierdo, "Specification of symmetric watermarking techniques", *BUSMAN Project Deliverable D 3.1*, 18. October 2003, IST-2001-35152.

[Fur04a] T. Furon, J. Delhumeau, C. Guillemot, V. Bottreau, I. Damnjanovic and E. Izquierdo, "Specification of side information watermarking techniques", *BUSMAN Project Deliverable D 3.2*, 15. June 2004, IST-2001-35152.

[Fur04b] T. Furon and I. Damnjanovic, "Report on the Peer-Review of Watermarking Specification Deliverables", *BUSMAN Project Deliverable D 7.3.4*, 6. October 2004, IST-2001-35152.

References

[3G99] 3GPP, "Technical Specification, Group Radio Access Network (Multiplexing and Channel Coding for FDD)", Tech. Rep., 3rd Generation Partnership Project, 1999.

[Ahm84] N. Ahmed, T. Natrajan and K. R. Rao, "Discrete Cosine Transform", IEEE Trans. on Computers, Vol. C-23, No.1, December 1984, pp. 90-93.

[Ahu92] A.J. Ahumada and H.A. Peterson, "Luminance-Model-Based DCT Quantization for Color Image Compression", *Proc. of SPIE, Human Vision, Visual and Digital Display III, ed. B. E. Rogowitz*, vol. 1666, 1992, pp. 365-374.

[Ahu93] A.J. Ahumada, H.A. Peterson and A. B. Watson, "An Improved Detection Model for DCT Coefficients Quantization", ", *Proc. of SPIE, Human Vision, Visual and Digital Display IV, ed. B. E. Rogowitz*, vol. 1913-14, 1993, pp. 191-201.

[Amb01] A. Ambroze, G. Wade, C. Serdean, M. Tomlinson, J. Stander and M. Borda, "Turbo Code Protection of Video Watermark Channel", *IEE Proc. Vis. Image Signal Processing*, Vol.148, No.1, February 2001, pp. 54-58.

[Are00] S. Arena, M. Caramma and R. Lancini, "Digital watermarking applied to MPEG-2 coded video sequences exploiting space and frequency masking", in *Proc. of International Conference on Image Processing, ICIP '00*, vol. 1, 10-13 Sept. 2000, pp. 438 – 441.

[Aur96] T. Aura, "Practical Invisibility in Digital Communication", *Proc. Workshop Information Hiding, Lecture Notes in Computer Science,* vol. 1174, Cambridge, UK, May 1996, pp. 265 - 278.

[Bah74] L. R. Bahl, J. Cocke, F. Jelinek and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate", *IEEE Transactions on Information Theory*, vol. 20, March 1974, pp. 284-287.

[Bar03] M. Barni, "What is The Future for Watermarking? (Part II)", *IEEE Signal Processing Magazine*, vol. 20, No. 6, November 2003, pp. 53-59.

[Ben96] W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," *IBM Syst. J.*, vol. 35, no. 3–4, 1996, pp. 313–336.

[Ber93] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo Codes," *Proc. Int. Conf. Comm.,* 1993, pp.1064-1070.

[Ber99] C. Berrou, C. Douillard and M. Jézéquel, "Multiple parallel concatenation of circular recursive systematic convolutional (CRSC) codes," *Annals of Telecommunications*, vol. 54, no. 3 – 4, March-April 1999, pp. 166 – 172.

[Ber01] C. Berrou, M. Jézéquel, C. Douillard and S. Kerouédan, "The advantages of non-binary turbo codes," *Proc. Information Theory Workshop*, Cairns, Australia, Sept. 2001*,* pp. 61-63.

[Bet98] C. Bettstetter, "Turbo decoding with tail-biting trellises", Diplomarbeit, Technischen Universitat Munchen, Jul. 1998. http://www.bettstetter.com/publications/bettstetter-1998-diplomarbeit.pdf

[Bha97] V. Bhaskaran, and K. Konstantinides, "Image and Video compression standards – Algorithms and Architectures," *Kluwer Academic Publishers,* 1997, 0-792-39952-8.

[Blo99] J.A. Bloom, et al., "Copy Protection for DVD video", *Proceedings of IEEE*, vol.87, No. 7, July 1999, pp. 1267-1276.

[Bus99] C. Busch, W. Funk, and S. Wolthusen, "Digital watermarking: from concepts to real-time video applications," *IEEE Computer Graphics and Applications*, vol.19, iss. 1, Jan/Feb 1999, pp. 25-35.

[Cal01] Calic J., Izquierdo E., "Towards Real-Time Shot Detection in the MPEG Compressed Domain", *3rd Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2001)*, Tampere, Finland, 16-17 May 2001, pp. 1-5

[Cal02] Calic J., Izquierdo E., "Temporal Segmentation of MPEG Video Streams", *EURASIP Journal on Applied Signal Processing*, Issue 6, pp. 561—565, 2002/6/26

[Cal02] Calic J., Izquierdo E., "A Multiresolution Technique for Video Indexing and Retrieval", *2002 International Conference on Image Processing (ICIP 2002),* Rochester, NY, September 22-25, 2002, Volume 1, pp. 952—955

[Cal02] Calic J., Sav S., Izquierdo E., Marlow S., Murphy N., O'Connor N., "Temporal Video Segmentation For Real-Time Key Frame Extraction", *27th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002).* Orlando, FL, 13-17 May 2002, Volume 4, pp. 3632-3635.

[Car95] G. Caronni, "Assuring ownership rights for digital images," in *Proc. Reliable IT Systems VIS'95*, Germany, Vieweg Publishing Company, 1995, pp. 251-263.

[Che01] B. Chen and G.W. Wornell, "Quantization Index Modulation Methods: A Class of Provably Good Methods for Digital Watermarking and Information Embedding", *IEEE Trans. On Information Theory*, vol. 47, No. 4, May 2001, pp. 1423-1443.

[Che02] L. Cheveau, "Choosing A Watermarking System for Digital Television – The Technology and The Compromises", *IBC2002* (www.broadcastpapers.com/asset/IBCEBUWatermarking03.htm)

[Che98] B. Chen and G.W. Wornell, "Digital watermarking and information embedding using dither modulation", *Proc. of IEEE Workshop on Multimedia Signal Processing (MMSP '98)*, Redondo-Beach, CA, USA, December 1998, pp. 273-278.

[Che99] B. Chen and G.W. Wornell, "Achievable performance of digital watermarking systems", *Proc. of IEEE Int. Conference on Multimedia Computing and Systems (ICMCS '99),* vol. 1, Florence, Italy, June 1999, pp. 13-18.

[Cho95] C.H. Chou and Y.C. Li, "A Perceptually Tuned Subband Image Coder Based on The Measure of Just-Noticeable Distortion Profile", *IEEE Trans on Circuits and Systems for Video Technology*, vol. 5, No. 6, April 1995, pp. 467-476.

[Chu98] T. Chung, M. Hong, Y. Oh, D. Shin, and S. Park, "Digital watermarking for copyright protection of mpeg2 compressed video," *IEEE Trans. on Consumer Electronics*, vol. 44, iss. 3, Aug 1998, pp. 895-901.

[Coh02] A.S. Cohen and A. Lapidoth, "The Gaussian Watermarking Game", *IEEE Trans. on Information Theory,* vol. 48, No. 6, June 2002, pp. 1639-1667.

[Cos83] M.H.M. Costa, "Writing on Dirty Paper", *IEEE Trans. On Information Theory*, vol. 29, No. 3, May 1983, pp. 439-441.

[Cov91] T.M. Cover and A.J. Thomas, "Elements of Information Theory", *John Weiley and Sons Inc,* Second Edition, 1991, 0-471-06259-6.

[Cox96] I.J. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for images, audio and video", *Proc. of International Conference on Image Processing, 1996*, vol. 3, 16-19 Sept. 1996 pp. 243 – 246.

[Cox97] I.J. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia", *IEEE Trans. on Image Processing*, vol. 6, no. 12, Dec. 1997, pp. 1673 - 1687

[Cox00] I. J. Cox, M. L. Miller, J. A. Bloom, "Watermarking Applications and Their Properties", *Proceedings of the International Conference on Information Technology: Coding and Computing - ITCC2000,* Las Vegas, USA, 2000, pp. 6-10

[Cox01] I. Cox, M. Miller, and J. Bloom, "Digital Watermarking," *Morgan Kaufmann Publisher*, 2001, 1-55860-714-5.

[Cox97] I. J. Cox, J. Kilian, T. Leightom, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Proc.*, vol. 6, no. 12, Dec. 1997, pp. 1673–1687.

[Cox99] I.J. Cox, M.L. Miller, and A.L. McKellips, "Watermarking as communications with side information," *Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information,* vol.87, No. 7, July 1999, pp. 1127-1141.

[Deg00] F. Deguillaume, G. Csurka and T. Pun, "Countermeasures for unintentional and intentional video watermarking attacks", in *Proc. of IS&T/SPIE's 12th Annu. Symp., Electronic Imaging '00: Security and Watermarking of Multimedia Content II*, San Jose, California USA, vol. 3971, 23-28 January 2000.

[Deg99] F. Deguillaume, G. Csurka, J. ´O Ruanaidh, and T. Pun, "Robust 3D DFT video watermarking," in *Proc. of IS&T/SPIE's 11th, Electronic Imaging '99: Security and Watermarking of Multimedia Contents*, San Jose, CA, vol. 3657, Jan. 1999.

[Dor02] Dorado A., Izquierdo E., "Fuzzy Color Signatures", *IEEE Proceedings International Conference on Image Processing (ICIP 2002)*. Rochester, New York, 22-25 September 2002, Volume 1, pp. 433-436.

[Dor03] Dorado A., Izquierdo E., "Semantic Labeling of Images Combining Color, Texture and Keywords", *IEEE Proceedings 10th International Conference on Image Processing (ICIP 2003)*. Barcelona, Catalonia, 14-18 September 2003, pp. 9-12.

[Dor03] Dorado A., Izquierdo E., "Semi-Automatic Image Annotation Using Frequent Keyword Mining", *IEEE Proceedings 7th International Conference on Information Visualisation (IV 2003)*. London, England, 16-18 July 2003, pp. 532-535.

[DVB02]"Digital Video Broadcasting (DVB); Interaction channel for Digital Terrestrial Television (RCT) incorporating Multiple Access OFDM", ETSI EN 301 958, V1.1.1, March 2002, pp. 28-30.

[DVB05]"Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems", ETSI EN 301 790, V1.4.1, September 2005, pp. 23-26.

[Egg00] J.J. Eggers, J.K. Su and B. Girod, "Robustness of a Blind Image Watermarking Scheme", *Proc. of IEEE International Conference on Image Processing*, vol. 3, October 2000, pp. 17-20.

[Egg02] J. Eggers and B. Girod, "Informed Watermarking", *Kluwer Academic Publishers*, 2002, 1-4020-7071-3.

[Egg03] J.J. Eggers, R. Bauml, R. Tzschoppe and B. Girod, "Scalar Costa Scheme for Information Embedding", *IEEE Trans. on Signal Processing*, vol. 51, No. 11, April 2003, pp. 1003-1019.

[FFM04] FFMPEG Multimedia Systems – version: ffmpeg-0.4.9-pre1 http://ffmpeg.sourceforge.net/index.php

[Fri98] J. Fridrich, "Combining low-frequency and spread spectrum watermarking", *Proc. of the SPIE Conference on Mathematics of Data/Image Coding, Compression and Encryption*, vol. 3456, 1998, pp. 2-12.

[Fri99] J. Fridrich and M. Goljan, "Comparing Robustness of Watermaking Techniques", *Proc. of Electronic Imaging `99, The International Society for Optical Engineering, Security and Watermarking of Multimedia Contents,* vol.3657, San Jose, CA, Jan. 25-27, 1999, pp. 214-225.

[Gar04] R. Garello and A. Vila, "The all-zero iterative decoding algorithm for turbo code minimum distance computation", *Proceedings of IEEE Int. Conf. Commun. (ICC'04)*, Paris, France, June 2004, pp. 361–364.

[Gue03] G. Le Guelvouit and S. Pateux, "Wide Spread Spectrum Watermarking with Side Information and Inference Cancellation", *Proceedings of the SPIE,* vol.5020, June 2003, pp. 278-289.

[Hai01] J. Haitsma and T. Kalker, "A watermarking scheme for digital cinema", in *Proc. of International Conference on Image Processing, ICIP '01*, vol. 2, 7-10 Oct. 2001 pp. 487 – 489.

[Han98] Hanke M., Izquierdo E., März R., "On Asymptotics in Case of Linear Index-2 Differential-Algebraic Equations", *SIAM Journal on Numerical Analysis* 1998, Volume 35, Issue 4, pp. 1326-1346.

[Haj00] A. Hajnalic, G.C. Langelaar, P.M.B. van Roosmalen, J. Biemond, and R. L. Lagendijk, "Image and Video Databases: Restoration, Watermarking and Retrieval", *Elsevier Science B. V.,* 2000, 0-444-50502-4.

[Har98] F. Hartung, and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Processing*, vol. 66, no. 3, 1998, pp. 283-302.

[Har99] F. Hartung, M. Kutter, "Multimedia watermarking techniques", *Proc. of the IEEE*,vol. 87, no. 7, July 1999, pp. 1079-1107.

[Hef04] L. Hefei, L. Zhengding and Z. Fuhao, "New real-time watermarking algorithm for compressed video in VLC domain", *Proc. of International Conference on Image Processing, ICIP '04*, vol. 4, 24-27 October 2004, pp. 2171 – 2174.

[HFT00] "HFTA-05.0: Statistical Confidence Levels for Estimating BER Probability", *Application Note 1095*: Oct 26, 2000 http://dallassemiconductor.com/appnotes.cfm/appnote_number/1095

[Hsu96] C. -T. Hsu and J. –L. Wu, "Hidden signatures in images", in *Proc. of International Conference on Image Processing, ICIP '96*, vol. 3, 16-19 Sept. 1996, pp. 223 – 226.

[Hsu98] C. -T. Hsu and J. –L. Wu, "DCT-Based Watermarking for Video", *IEEE Trans. Consumer Electronics*, vol. 44, no. 1, February 1998, pp. 206-216.

[IEC89] International Electrotechnical Commission, Geneva, Switzerland. *Digital Audio Interface, IEC 60958*, February 1989.

[ISO01] ISO/IEC JTC1/SC29/WG11, 2001 Information Technology – Multimedia Content Description Interface – Part 3: Visual, ISO/IEC IS 15938-3

[ISO03] ISO/IEC JTC1/SC29/WG11, 2003. Information Technology – Multimedia framework – Part 3: Digital item identification, ISO/IEC IS 21000-3

[Izq95] Izquierdo E., Ernst M., "Motion/Disparity analysis for 3D-Video-Conference Applications", *1995 International Workshop on Stereoscopy and 3-Dimensional Imaging (IWS3DI 1995)*. Santorini, Greece, September 1995.

[Izq98] Izquierdo E., Kruse S., "Image Analysis for 3D Modeling, Rendering, and Virtual View Generation", *Elsevier Journal Computer Vision and Image Understanding,* 1998, Volume 71, Issue 2, pp. 231-253.

[Izq98] Izquierdo E., "Stereo Image Analysis for Multi-viewpoint Telepresence Applications", *Elsevier Journal Signal Processing: Image Communication,* Volume 11, Issue 3, pp. 231-254.

[Izq99] Izquierdo E., Ghanbari M., "Nonlinear Gaussian filtering approach for object segmentation", *IET Proceedings - Vision, Image and Signal Processing,* Volume 146, Issue 3, pp. 137-143.

[Izq99] Izquierdo E., Feng X., "Modeling Arbitrary Objects Based on Geometric Surface Conformity", *IEEE Transactions on Circuits and Systems for Video Technology,* Volume 9, Issue 2, pp. 336-352.

[Izq00] Izquierdo E., Ohm J., "Image-based rendering and 3D modeling: a complete framework", *Signal Processing: Image Communication,* Volume 15, Issue 10, 2000, pp. 817-858.

[Izq02] Izquierdo E., "Using Invariant Image Features for Synchronization in Spread Spectrum Image Watermarking", *Springer EURASIP Journal on Advances in Signal Processing,* Issue 4, pp. 412-419.

[Izq02] Izquierdo E., Ghanbari M., "Key Components for an Advanced Segmentation System", *IEEE Transactions on Multimedia,* Volume 4, Issue 1, pp. 97-113.

[Izq03] Izquierdo E., Guerra V., "An Ill-Posed Operator for Secure Image Authentication", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume 13, Issue 8, pp. 842-852.

[Izq03] Izquierdo E., et al., "Advanced Content-Based Semantic Scene Analysis and Information Retrieval: The Schema Project", *4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2003*), World Scientific Publishing, London, England, 9-11 April 2003, pp. 519-528.

[Izq04] E. Izquierdo, A. Pearmain, P. Villegas, L.-Q. Xu, S. Herrmann, "BUSMAN: A system for digital video content management and delivery", *IBC 2004 http://busman.elec.qmul.ac.uk/publications/publications.htm*

[Jay93] N. Jayant, J. Johnston and R. Safranek, "Signal Compression Based on Models of Human Perception", *Proc. of IEEE*, vol. 81, No. 10, October 1993, pp. 1385-1422.

[Joh98] N.F. Johnson and S. Jajodia, "Steganalysis of Images Created using Current Steganography Software" , *Proc of Workshop on Information Hiding Proceedings*, Portland, Oregon, USA, 15 - 17 April 1998, pp. 273-289.

[Jor97] F. Jordan, M. Kutter, and T. Ebrahimi, "Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video", *ISO/IEC Doc. JTC1/SC29/WG11 MPEG97/M2281*, July 1997

[Kal99] T. Kalker, G. Depovere, J. Haitsma and M. Maes, "A video watermarking system for broadcast monitoring", in *Proc. SPIE IS&T/SPIE's 11th Annual Symposium, Electronic Imaging '99: Security and Watermarking of Multimedia Contents*, Sans Jose, CA, USA, vol. 3657, Jan.1999, pp. 103-112.

[Kay01] Kay S., Izquierdo E., "Robust Content Based Image Watermarking", *3rd Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2001)*, Tampere, Finland, 16-17 May 2001, pp. 53-56.

[Kli05] Kliegr T., Chandramouli K., Nemrava J., Svatek V., Izquierdo E., "Combining Image Captions and Visual Analysis for Image Concept Classification", *9th International Workshop on Multimedia Data Mining*, Las Vegas, NV, pp. 817.

[Koc95] E. Koch and J. Zhao, "Towards robust and hidden image copyright labelling," *Proc. 1995 IEEE Workshop on Nonlinear Signal and Image Processing,* IEEE Computer Society Press, Jun 1995, pp.452-455.

[Koh00] T. Kohda, Y. Ookubo and K. Shinokura, "Digital watermarking through CDMA channels using spread spectrum techniques", *IEEE Sixth International Symposium on Spread Spectrum Techniques and Applications '00, v*ol. 2, 6-8 Sept. 2000, pp. 671 – 674.

[Kov04] O. Koval, S. Voloshynovskiy, F. Deguillaume, F. Perez-Gonzalez and Thierry Pun, "Spread spectrum watermarking for real images: is everything so hopeless?", *In Proceedings of 12th European Signal Processing Conference, EUSIPCO 2004*, Vienna, Austria, September 6-10 2004. http://vision.unige.ch/publications/postscript/2004/KovalVoloshynovskiyDeguillaumePerezGonzalezPun_EUSIPCO2004.pdf

[Koz04] F. Kozamernik, V. Steinmann, P. Sunna and E. Wyckens, "SAMVIQ – A new EBU methodology for video quality evaluations in multimedia", ibc2004 http://www.broadcastpapers.com/ibc2004/ibcpapers04.htm

[Kut97] M. Kutter, F. Jordan and F. Bossen, "Digital signature of color images using amplitude modulation," in *Proc. of SPIE storage and retrieval for image and video databases*, San Jose, USA, no. 3022-5, February 13-14, 1997, pp. 518-526.

[Kut98] M. Kutter, F. Jordan and F. Bossen, "Digital signature of color images using amplitude modulation", *Journal of Electronic Imaging*, vol. 7, no. 2, April, 1998, pp. 326-332.

[Kut99] M. Kutter and F. Petitcolas, "A fair benchmark for image watermarking systems", in *Proc. SPIE IS&T/SPIE's 11th Annual Symposium, Electronic Imaging'99: Security and Watermarking of Multimedia Contents*, Sans Jose, CA, USA, vol. 3657, 25-27 Jan. 1999.

[Lan96] G.C. Langelaar, J.C.A. van der Lubbe, and J. Biemond, "Copy protection for multimedia data based on labeling techniques", *In Proc. of the 17th Symposium on Information Theory in the Benelux*, Enschede, The Netherlands, 1996, pages 33-39.

[Lan97] G. C. Langelaar, J. C. A. ven der Lubbe, and R. L. Lagendijk, "Robust labeling methods for copy protection of images", in *Proc. Electronic Imaging*, San Jose, CA, vol. 3022, February 1997, pp. 298–309.

[Lan98] G. C. Langelaar, R. L. Lagendijk and J. Biemond, "Real-time labelling methods for MPEG compressed video", *Proc of 18th International Symposium on Information Theory in the Benelux*, Veldhoven, Netherlands, May, 1998.

[Lan00] G.C. Langelaar, I. Setyawan and R.L. Lagendijk, "Watermarking digital image and video data. A state-of-the-art overview", *IEEE Signal Processing Magazine,* vol. 17, No. 5, September 2000, pp. 20-46.

[Lan01] G. C. Langelaar, and R. L. Lagendijk, "Optimal differential energy watermarking of DCT encoded images and video," *IEEE Trans. on Image Processing*, vol.10, no.1, Jan 2001, pp.148-158.

[Lin98] J.-P. Linnartz and J. Talstra, "MPEG PTY-Marks: Cheap Detection of Embedded Copyright Data in DVD-Video", *Proc. of the 5th European Symposium on Research in Computer Security, Lecture Notes in Computer Science,* vol. 1485, 1998, pp. 221 – 240.

[Lu02] C. Lu, J. Chen, H. M. Liao, and K. Fan, "Real-time MPEG2 video watermarking in VLC domain," *Proc. IEEE 16th International Conference on Pattern Recognition 2002*, vol. 2, 2002, pp. 552-555.

[Man74] J. L. Mannos, D. J. Sakrison, "The Effects of a Visual Fidelity Criterion on the Encoding of Images", *IEEE Transactions on Information Theory*, vol. 20, No 4, 1974, pp. 525-535.

[Mar99] L.M. Marvel, C.G. Boncelet Jr. and C.T. Retter, "Spread spectrum image steganography", *IEEE Transactions on Image Processing*, vol. 8, no. 8, Aug. 1999, pp. 1075 – 1083.

[Mat94] K. Matsui and K. Tanaka, "Video-steganography: how to secretly embed a signature in a picture," *Journal of the Interactive Multimedia Association Intellectual Property Project*, vol. 1, no. 1, Jan. 1994, pp. 187–205.

[Mil01] M.L. Miller, "Watermarking with dirty paper codes", *Proc. of IEEE International Conference on Image Processing*, vol.2, October 2001, pp. 538-541.

[Mit96] J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, and D. J. LeGall, "MPEG Video Compression Standard," *Kluwer Academic Publishers,* 1996, 0-412-08771-5

[Mit02] M.P.Mitrea, F.Preteux, A.Vlad and N.Rougon, "Spread spectrum watermarking method for image databases", *Proc. of IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SPPRA'02),* Crete, Greece, June 2002, p. 444-449.

[Moh99] S.P. Mohanty, "Digital Watermarking: A Tutorial Review", 1999. (http://citeseer.ist.psu.edu/mohanty99digital.html)

[Mou03] P. Moulin and J.A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding", *IEEE Trans. On Information Theory,* vol.49, No.3, March 2003, pp.563-593

[Mou04] P. Moulin and R. Koetter, "Data-Hiding Codes", October 2004. (http://www.ifp.uiuc.edu/~moulin/paper.html)

[MPG96]MPEG-2 Video Standard: ISO/IEC 13818-2: Information Technology – generic coding of moving pictures and associated audio information: Video (1996) (aka ITU-T Rec. H-262 (1996))

[Mra04] Mrak M., Abhayaratne C., Izquierdo E., "On the influence of motion vector precision limiting in scalable video coding", *IEEE Proceedings 7th International Conference on Signal Processing (ICSP 2004)*. Beijing, China, 31 August - 4 September 2004, Volume 2, pp. 1143-1146.

[Mra05] Mrak M., Sprljan N., Izquierdo E., "Motion estimation in temporal subbands for quality scalable motion coding", *IET Electronics Letters*, Volume 41, Issue 19, pp. 1050-1051.

[Mró04] Mrówka E., Dorado A., Pedrycz W., Izquierdo E., "Dimensionality Reduction for Content-Based Image Classification", *IEEE Proceedings 8th International Conference on Information Visualisation (IV 2004)*. London, England, 14-16 July 2004, pp. 435-438.

[MSSG] MPEG Software Simulation Group (MSSG), (http://www.mpeg.org/MPEG/MSSG/)

[Niu02] X. Niu, M. Schmucker and C. Busch, "Video Watermarking Resisting to Rotation, Scaling and Translation", *In Proc. of SPIE Security and Watermarking of Multimedia Contents IV*, vol. 4675, January 2002, pp. 512-519.

[OCo03] O'Connor N., Sav S., Adamek T., Mezaris V., Kompatsiaris I., Lui Z., Izquierdo E., Bennström C., Casas J., "Region and Object Segmentation Algorithms in the Qimera Segmentation Platform", *3rd International Workshop on Content-Based Multimedia Indexing (CBMI 2003)*, Rennes, France, 22-24 September 2003, pp. 1-8

[Pap00] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation", *Handbook of Image and Video Processing (A. C. Bovik, ed.), Academic Press,* 2000, pp. 669-684.

[Pet98] F. A. Petitcolas, R. J. Anderson, and M.G.Kuhn, "Attacks on copyright marking systems," in Proc. *Second International Workshop on Information Hiding*, Portland, OR, Apr.15-17, 1998, pp. 219-239.

[Pet00] F. A. P. Petitcolas, "Watermarking schemes evaluation," *I.E.E.E. Signal Processing*, vol. 17, no. 5, September 2000, pp. 58–64.

[Pin00] Pinheiro A., Izquierdo E., Ghanhari M., "Shape Matching using a Curvature Based Polygonal Approximation in Scale-Space", *IEEE Proceedings International Conference on Image Processing (ICIP 2000).* Vancouver, BC, 10-13 September 2000, Volume 2, pp. 538-541

[Pin04] M. Pinson and S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality," *IEEE Trans. on Broadcasting*, vol. 50, No. 3, September, 2004, pp. 312-322.

[Pit96] I. Pitas, "Method for Signature Casting on Digital Images", *Proc. of IEEE International Conference on Image Processing (ICIP'96)*, vol. 3, September 1996, pp. 215-218.

[Pit98] I. Pitas, "A method for watermark casting on digital image", *IEEE Trans. on Circuits and Systems for Video Technology,* vol. 8, no. 6, October 1998, pp. 775 – 780

[Piv97] A. Piva, M. Barni, F. Bartolini and V. Cappellini, "DCT-based watermark recovering without resorting to the uncorrupted original image", *Proc. of International Conference on Image Processing '97*, vol. 1, 26-29 Oct. 1997 pp. 520 -

523

[Piv98] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "Threshold selection for correlation-based watermark detection," in *Proc. COST 254 Workshop on Intelligent Communications*, L'Aquila, Italy, Jun. 1998, pp. 66–72.

[Pod98] C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Journal on Special Areas in Communications*, vol. 16, no. 4, May 1998, pp. 525–539.

[Pra04] S. Pranata, V. Wahadaniah, Y. L. Guan and H. C. Chua, "Improved Bit Rate Control for Real-Time MPEG Watermarking", *EURASIP Journal on Applied Signal Processing,* vol.14, 2004, pp. 2132–2141.

[Pur01] Jessica Pursley, "Turbo Product Codes and Channel Capacity," *the 2001 IEEE Southeast Conference (SECON),* April 2001. http://www.clemson.edu/sure/projects/jpursley_secon01.pdf

[Ram05] Ramzan N., Wan S., Izquierdo E., "Joint Source-Channel Coding for Wavelet-Based Scalable Video Transmission Using an Adaptive Turbo Code", *EURASIP Journal on Image and Video Processing,* Vol 7, pp. 1-12.

[Ram05] Ramos J., Guil N., González J., Zapata E., Izquierdo E., "Logotype detection to support semantic-based video annotation", *Journal Signal Processing: Image Communication,* Volume 22, Issue 7-8, pp. 669-679.

[Rey02] C. Rey and J.L. Dugelay, "A survey of watermarking algorithms for image authentication", *EURASIP Journal on Applied Signal Processing,* vol.6, 2002, pp. 613-621

[Rua96] J. Ruanaidh, W. Dowling, and F. Boland, "Phase watermarking of digital images," in *Proc. Int. Conf. on Image Processing*, Lausanne, Switzerland, vol. 3, Sep. 1996, pp. 239–242.

[Rya98] William E. Ryan, CA Turbo Code Tutorial," 1998. (http://telsat.nmsu.edu/~wryan/turbo2c.ps)

[Saf89] R.J Safranek, and J.D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression", *Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP-89*, vol. 3, 23-26 May 1989, pp.1945–1948

[Sch04] C. B. Schlegel and L.C. Perez, "Trellis and Turbo Coding", *IEEE Press*, 2004, 0-471-22755-2.

[Sch94] R.G. Schyndel, A.Z. Tirkel and C.F. Osborne, "A Digital Watermark", *Proc. of IEEE International Conference on Image Processing,* vol. 2, Austin, TX, USA, November 1994, pp. 86-90.

[Sch95] R. Schafer and T. Sikora, " Digital video coding standards and their role in video communications", *Proceedings of the IEEE*, vol. 83, no. 6, June 1995, pp. 907 – 924.

[Sch99] C. B. Schlegel and L.C. Perez, "On Error Bounds and Turbo-Codes", *IEEE Communications Letters*, vol. 3, no. 7, July 1999, pp. 205-207.

[Sha48] C.E. Shannon, "A mathematical theory of communications", *Bell Syst. Tech. Journal,* vol. 27, July 1948, pp. 379-423.

[Sha58] C.E. Shannon, "Channels with side information at the transmitter", *IBM Journal of Research and Development,* 1958, pp. 289-293.

[Sha67] C. E. Shannon, R. G. Gallager and E. R. Berlekamp, "Lower bounds to error probabilities for coding on discrete memoryless channels", *Information and Control*, vol.10, 1967, part I – no. 1, pp.65-103, and part II – no.5, pp. 522-552.

[Sim02] D. Simitopoulos, S. Tsaftaris, N. V. Boulgouris, and M. G. Strintzis, "Compressed-domain Video Watermarking of MPEG Streams", *IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland, August 2002, pp. 569 -572.

[Sim04] D. Simitopoulos, S. Tsaftaris, N. V. Boulgouris, A. Brissouli and M. G. Strintzis, "FastWatermarking of MPEG-1/2 Streams Using Compressed-Domain Perceptual Embedding and a Generalized Correlator Detector", *EURASIP Journal on Applied Signal Processing,* vol.8, 2004, pp. 1088-1106.

[Smi96] J. R. Smith and B. O. Comiskey, "Modulation and information hiding in images," in *Proc. Int. Workshop on Information Hiding*, *Springer-Verlag Lecture Notes in Computer Science*, vol. 1174, Cambridge, UK, May 1996, pp. 207–226.

[Spr05] Sprljan N., Mrak M., Abhayaratne C., Izquierdo E., "A Scalable Coding Framework for Efficient Video Adaptation", 6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005). Montreux, Switzerland, 2005, 13-15 April 2005, pp. 1-4

[Sta03] M. Staring, J. Oostveen and T. Kalker, "Optimal Distortion Compensation for Quantization Watermarking", *Proc. of IEEE International Conference on Image Processing*, vol. 2, September 2003, pp. 727-730.

[Str95] M. Stricker and M. Orengo, "Similarity of color images," *Proceedings of Storage and Retrieval for Image and Video Databases III*, vol. 2420, pp. 381-392, 1995.

[Swa97] M.D. Swanson, Bin Zhu, B. Chau and A.H. Tewfik, "Multiresolution video watermarking using perceptual models and scene segmentation", *Proc. of International Conference on Image Processing, ICIP '97*, vol. 2, 26-29 Oct. 1997, pp. 558 – 561.

[Swa98a] M. D. Swanson, M. Kobayashi and A. H. Tewfik, "Multimedia data-embedding and watermark technologies," *Proc. IEEE,* vol. 86, June 1998, pp. 1064-1087.

[Swa98b] M.D. Swanson, B. Zhu, A.H. Tewfik and L. Boney, "Robust audio watermarking using perceptual masking," *IEEE Trans. Signal Proc.: Special Issue on Copyright Protection and Control*, vol. 66, no. 3, pp. 337–355, 1998.

[Swa98c] M.D. Swanson, B. Zhu and A.H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models", *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, May 1998, pp. 540 – 550.

[Tan90] K. Tanaka, Y. Nakamura, and K. Matsui, "Embedding secret information into a dithered multi-level image," in *Proc. Military Communications Conference*, vol. 1, 1990, pp. 216–220.

[Tir93] A.Z. Tirkel, G.A. Rankin, R.G. van Schyndel, W.J. Ho, N.R.A. Mee and C.F. Osborne, "Electronic Water Mark", *Proc. of DICTA'93 - Digital Image Computing, Technology and Applications*. 1993. pp. 666-673

[Tir96] A.Z. Tirkel, C.F. Osborne and R.G. Van Schyndel, "Image watermarking-a spread spectrum application", *Proc. of IEEE 4th International Symposium on Spread Spectrum Techniques and Applications '96*, vol. 2, 22-25 Sept. 1996, pp. 785 - 789

[Ton98] H.H.Y. Tong, A.N. Venetsanopoulos, "A Perceptual Model for JPEG applications based on Block Classification, Texture Masking and Luminance Masking", *Proc of IEEE International Conference on Image Processing (ICIP'98)*, vol.3, October 1998, pp. 428-432.

[Tra03] W. Trappe, M. Wu, Z. J. Wang and K. J. Ray Liu, "Anti-collusion Fingerprinting for Multimedia", *IEEE Transaction on Signal Processing*, vol. 51, No. 4, April 2003 pp. 1069-1087.

[Tur89] L. F. Turner, "Digital data security system." Patent IPN WO 89/08915, 1989.

[Vil03] P. Villegas, S. Herrmann, E. Izquierdo, J. Teh and L. Q. Xu, "An Environment for Efficient Handling of Digital Assets", *4th European Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2003*, London, UK, Apr 2003, 981-238-355-7.

[Vol99] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," in *Proc. Int. Workshop on Information Hiding*, Dresden, Germany, Sep. 1999, pp. 211–236.

[Voy98] G. Voyatzis, N. Nikolaides, and I. Pitas, "Digital watermarking: An overview," *Proc. IX European Signal Processing Conference (EUSIPCO),* Rhodes, Greece, Sept. 1998, pp. 13-16.

[Wan04] Z. Wang et al., "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Trans. On Image Processing,* vol.13, No. 4, April 2004, pp. 600-611.

[Wan03] Wang Y., Izquierdo E., "High-Capacity Data Hiding in MPEG-2 Compressed Video", *9th International Workshop on Systems, Signals and Image Processing (IWSSIP 2002), World Scientific*, Manchester, England, 7-8 November 2002, pp. 212-218.

[Wat93] A.B. Watson, "DCT quantization matrices visually optimized for individual images", *Proc. of SPIE, Human Vision, Visual and Digital Display IV, ed. B. E. Rogowitz*, vol. 1913-14, 1993, pp. 202-216.

[Wat96] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visual thresholds for wavelet quantization error," in *Proc. SPIE Human Vision and Electronic Imaging,* vol. 2657, 1996,pp. 381–392.

[Won00] P.H.W. Wong, O.C. Au and J.W.C. Wong, "Image watermarking using spread spectrum technique in log-2-spatio domain", *Proc. of the IEEE International Symposium on Circuits and Systems, ISCAS '00*, Geneva, Switzerland, vol. 1, 28-31 May 2000, pp. 224 – 227.

[Wu98] M.Wu and B. Liu, "Watermarking for image authentication", in *Proc. IEEE International Conference on Image Processing*, Chicago, Ill, USA, vol. 2, October 1998, pp. 437–441.

[Zga05] Zgaljic T., Sprljan N., Izquierdo E., "Bitstream Syntax Description based Adaptation of Scalable Video", *2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT 2005).* London, England, 30 November - 1 December 2005.

[Zha05] X.H. Zhang, W.S. Lin and P. Xue, "Improved Estimation for Just-noticeable Visual Distortions", *Signal Processing*, vol. 85, no. 4, 2005, pp. 795-808.

[Zhu98] W. Zhu, Z. Xiong, and Y.Q. Zhang, "Multiresolution watermarking for images and video: a unified approach," in *Proc. Int. Conf. on Image Processing*, Chicago, IL, 1998, vol. 1, pp. 465–468.