

UNIVERSITE CATHOLIQUE DE LOUVAIN
Ecole Polytechnique de Louvain
ICTEAM Institute
Communications and Remote Sensing Lab.



Towards a combining scheme for Compression and Watermarking for 3D Stereo Images

Rony DARAZI

Thesis presented for the Ph.D. degree
in Electrical Engineering

PhD committee

Benoît MACQ	UCL/ICTM - Supervisor
Chaker LARABI	Université de Poitiers, France
Jean-Jacques QUISQUATER	UCL/ICTM, Crypto Group
Christophe DE VLEESCHOUWER	UCL/ICTM
Gaël ROUVROY	INTOPIX, Belgium
Patrice RONDAO ALFACE	Alcatel-Lucent, Bell Labs, Belgium
Christophe DE VLEESCHOUWER	UCL/ICTM - Chairman

August 2011

*Say not, "I have found the truth," but rather, "I have found a truth."
Say not, "I have found the path of the soul." say rather, "I have met the soul
walking upon my path." For the soul walks upon all paths.
The soul walks not upon a line, neither does it grow like a reed.
The soul unfolds itself, like a lotus of countless petals.*

Gibran Khalil Gibran

Acknowledgements

A PhD is a great challenge that I would not been able to achieve without the aid of lot of people. I would like to express first, my deepest gratitude to my supervisor Benoît Macq for his support and encouragement during my years at UCL. I have benefited not only from his guidance but also from his confidence and sincerity.

I would like to thank my jury members for their advices and their comments during this research: Chaker Larabi, Jean-Jacques Quisquater, Christophe De Vleeschouwer, Gaël Rouvroy and especially Patrice Rondao Alface for his encouragement and support along all my research.

This thesis has been funded by ARC-Cryptomark, BCRYPT and TRACEA projects for which I am very grateful.

I would like also to thank Fady Fadel, Antoine Rajeh and Joe Bou Jaoudé from the Antonine University for ceaselessly encouraging me to pursue my PhD.

I would like to express my special appreciation to Annabelle Gouze who has been a precious guide at the beginning of my research and a collaborator during this thesis. My special thanks go to people with whom I have collaborated on the topics of this thesis, Pilar Callau, Haoji Hu, Charles de Brabandere, Amit Kumar K.C, Mireia Montanõlla, Jonathan Bonnevie and Li Weng. Exchanging ideas and sharing experiences with them has greatly influenced my research.

I cannot forget the days I have spent with all my colleagues, friends during the last five years in Image Group, ICTEAM institute. Many thanks for people with whom I have appreciate intersecting discussions and their emotional support.

Very special thanks to my good friends, family in Belgium that made my life away from my homeland so incredibly easy and with whom I have shared wonderful moments. Françoise, Jean-Pierre, Anne-Marie, Nicolas, Elie, Gaby and Christelle.

Last, I would like to thank my parents and my two brothers for their unconditional support, I am also very thankful for my future parents-in-law for their love and support. And finally and most importantly, thank you to Marie-Joe to whom I dedicate this thesis for her care, her deepest love and her endless patience.

Abstract

Reproducing a natural and real scene as we see in the real world everyday is becoming more and more popular. Stereoscopic and multi-view techniques are used for this end. However, the fact that more information is displayed requires supporting technologies such as digital compression to ensure the storage and transmission of the sequences. In addition, the fast development of communication technologies and the exponential increase of internet users have made the protection of intellectual property rights and the prevention of reproducing original contents a very challenging issue.

This thesis is devoted to compression and protection of Intellectual Property Rights associated with 3D Stereo Images.

Based on a predictive framework, the proposed scheme combines compression and watermarking for 3D stereo images. An adaptive scheme for stereo image coding is proposed. Our goal is to design an efficient transform that reduces the existing redundancy in the stereo pair. This approach is inspired by the Lifting Scheme (LS). Aiming to reduce the residual error, we first design an optimized filter for inter prediction, a local spatial filtering is incorporated after disparity compensation. Second, directional inter prediction is investigated. The predictor is adapted by exploiting local orientation information denoting to the optimal way of prediction for each pixel. Finally, a residual LS is proposed to fit with the particularities of the transformed residue tuned by the disparity map.

The second aim is to design a watermarking scheme for 2D, 3D stereo image which can be integrated in the DCinema decoding pathway. We have proposed an adaptive exhibition watermarking method during the decoding phase. This method is QIM-based and is designed to fit in the JPEG2000 decoding pipeline. Based on the properties of the DWT related to the human visual system, we define a pixel-wise masking vector that will tune the watermarking scheme taking into consideration a number of constraints. An extension of the 2D exhibition watermarking scheme is applied to stereo images. A disparity guided watermarking technique with an improved perceptual mask is extended to fit with the particularities of stereo.

Symbols and Notation

HVS	Human Visual System
DWT	Discrete-Wavelet-Transform
DCT	Discrete-Cosine-Transform
DCinema	Digital Cinema
DCI	Digital Cinema Initiatives
HD	High Definition
IPR	Intellectual Property Rights
DE/DC	Disparity Estimation / Disparity Compensation
ICC	Illumination Change Compensation
DCD	Disparity Compensated Difference
DPCM	Differential Pulse Code Modulation
DV	Disparity Vector
DM	Disparity Map
QTD	Quadtree Decomposition
LS	Lifting Scheme
P	Prediction step
U	Update step
QIM	Quantization Index Modulation
STDM	Spread Transform Dithering Modulation
FGA	Fixed Gain Attack
JND	Just Noticeable Difference
PSNR	Peak Signal to Noise Ratio
SSIM	Structural Similarity
ICTEAM	Institute of Information and Communication Technologies, Electronics and Applied Mathematics at UCL

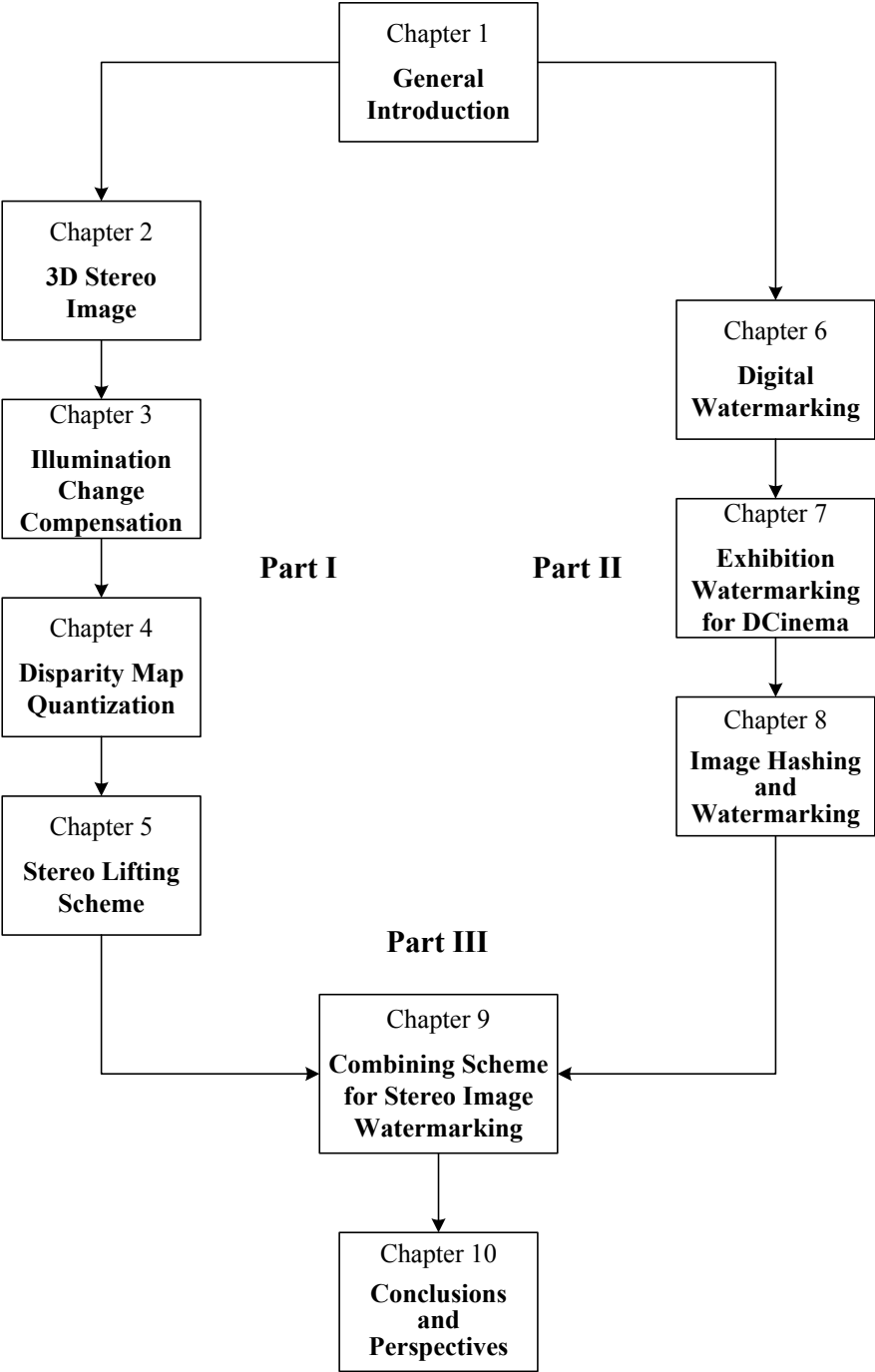


Figure 1: Overall organization of thesis

Contents

Acknowledgements	v
Abstract	vii
Symbols and Notation	ix
1 Introduction	1
1.1 Context of the work	1
1.2 Motivation and Problem Formulation	3
1.2.1 Stereo Image Compression	3
1.2.2 2D, 3D Stereo Image Watermarking	4
1.3 Main contribution	5
1.4 Thesis outline	8
 Part I 3D Stereo image Compression	 11
2 3D Stereo image	15
2.1 Brief history of 3D understanding	16
2.2 Stereo modeling	16
2.3 Stereo image coding	19
2.3.1 Disparity Estimation and Compensation	19
2.4 Predictive coding scheme	21
2.4.1 Idea of Disparity Compensated Difference Transform . . .	21
2.4.2 Directional Disparity Compensated Transform, 2D Direc- tional Transforms	23
2.5 Background on lifting scheme	23
 3 Illumination Change Compensation	 27
3.1 Introduction	27
3.2 Global illumination compensation	28
3.2.1 Illumination-reflectance model and homomorphic filtering	28
3.2.2 Histogram equalization	29
3.2.3 Implementation	30
3.3 Block-based and local ICC	30
3.4 ICC performance and transmission cost, Global vs. Local approach	31
3.5 Implementation of local ICC	31
3.6 Conclusion	34

4	Optimal DM quantization using optical flow method	35
4.1	Introduction	35
4.2	Disparity map estimation and optimal quantization	37
4.2.1	Disparity map estimation	37
4.2.2	Optimal quantization of disparity map	37
4.3	Experimental Results	38
4.4	Conclusion	40
5	Stereo Lifting Scheme	43
5.1	Introduction	43
5.2	Proposed Lifting-based structure	44
5.3	Joint Lifting Structure with Optimized Prediction	45
5.3.1	Linear Optimized Inter Predictor	46
5.3.2	Non Linear Directional Inter Predictor	49
5.3.3	Hybrid LS and Edge Map	52
5.3.4	Directional LS and Orientation map	53
5.4	Experimental results	54
5.4.1	Linear Optimized Predictor	55
5.4.2	Non Linear Directional Predictor	55
5.4.3	Residual LS	61
5.5	Conclusion	61
Part II	From 2D to 3D Stereo image Watermarking	65
6	Digital Watermarking	67
6.1	Origins and History: Information Hiding	68
6.2	Applications and Requirements	70
6.2.1	Modern Applications	70
6.2.2	Requirements/Properties	70
6.2.3	Evaluation- Effectiveness watermarking systems	71
6.3	Watermarking Systems	71
6.3.1	Watermarking Models: Informed/Blind	72
6.3.2	Watermarking Media Space/ Marking Space	73
6.3.3	Watermarking Classes and techniques: SS/QIM	74
6.4	Survey on QIM methods	76
6.4.1	Quantization Index Modulation	76
6.4.2	Scalar-QIM	78
6.4.3	Basic Dither Modulation with Scalar-QIM	79
6.4.4	Distortion-Compensated QIM	82
6.4.5	Spread Transform Dither Modulation	83
6.4.6	Distortion Measures	87

7	Exhibition DWT-based Watermarking for D-Cinema	89
7.1	Introduction	90
7.2	Digital Cinema Watermarking requirements	90
7.3	Exhibition Watermarking	91
7.4	Flexible DWT marking space	92
7.5	QIM-based watermarking using perceptual models	93
7.5.1	DWT perceptual modeling	94
7.5.2	Perceptual Mask	96
7.6	Valuometric scaling attacks	99
7.7	Implementation and Results	100
7.7.1	Perceptual Quality	101
7.7.2	Robustness	105
7.7.3	Anti-camcording	105
7.7.4	Intended implementation scenario and perspectives . . .	105
7.8	Conclusion	108
8	Image Content Authentication Using Robust Hashing and Watermarking	109
8.1	Introduction	110
8.2	System requirement and overview	112
8.2.1	The image hash system	112
8.2.2	The watermarking system	114
8.2.3	The attacker	114
8.2.4	Performance evaluation	115
8.3	Image authentication by hashing and watermarking	116
8.3.1	The watermarking algorithm	116
8.3.2	The image hash algorithm	117
8.4	Experiment results and analysis	118
8.5	Conclusion	122
 Part III Towards a combining scheme for 3D Stereo image Coding and Watermarking		123
9	Combining Scheme	125
9.1	Stereo Image Particularity	126
9.1.1	Constraints	126
9.2	Stereo Watermarking Scenario	126
9.2.1	Feedback/Content Analysis from the compression process	127
9.2.2	Depth map guided watermarking for Stereo Images	127
9.2.3	3D-Visual Mask	127
9.2.4	Disparity guided STDM	128
9.3	Results and Evaluation	129
9.4	Conclusions and Perspectives	131

10 Conclusions	133
10.1 Discussion, Back to the main contributions	133
 Publications	 135
 Appendices	 137
A Optical Flow-based Disparity map estimation	139
A.0.1 Disparity map estimation	139
 B Structural Similarity Metric	 141
 C JPEG2000 in DCI	 143
C.1 Digital Cinema system overview	143
C.2 DCI specifications for JPEG2000	144
C.2.1 JPEG2000 overview	145
C.2.2 DCI restrictions for JPEG2000	146
 Bibliography	 147

Introduction

Contents

1.1	Context of the work	1
1.2	Motivation and Problem Formulation	3
1.3	Main contribution	5
1.4	Thesis outline	8

1.1 Context of the work

Reproducing a natural and real scene as we see in the real world everyday is becoming more and more popular. Over the last two decades, efficient representation schemes for visual data, such as image and video, have been actively developed. Communication technologies have matured so fast. More recently, the main development in image-related technologies has been focusing on the addition of perceptual sensations for immersive communication.

After the increasing resolution of images, visual realism has been further enhanced by adding depth information. Human visual system (HVS) is more sensitive to 3D than to 2D images. In general, humans perceive 3D using various 3D cues such as perspective, occlusion and shading. However, those 3D cues alone are not enough to provide realistic 3D. An efficient method for providing depth sensation for images is to use stereoscopic approaches.

Stereovision is associated with the fact that humans perceive a scene in 3D by simultaneously viewing a scene from slightly different positions. Providing well-composed stereo pairs simultaneously for each eye, makes the scene perceived in three dimensions by the human brain.

3D film industry has been attracting considerable attention over the last few years. Furthermore, recently introduced technologies for autostereoscopic displays and HD 3D displays/projectors using active/passive glasses continue contributing to the widespread usage of stereo techniques.

Related applications have encountered a significant success and are since very

popular. Such applications are varied and goes from the entertainment industry (Digital Cinema, 3D-Tv, Video games, Attraction parks in mixed or virtual reality,...) to the technical or scientific world (medical surgical imaging, geographical information systems...).

However, the fact that more information are displayed requires supporting technologies such as "*digital compression*" in order to ensure the storage and transmission of these sequences. In addition, the fast development of communication technologies and the exponential increase of internet users have made the "*protection of intellectual property rights*" (IPR) and the "prevention of reproducing original contents" a very challenging issue.

The context of this thesis is related to *compression and protection of Intellectual Property Rights associated with 3D Stereo Images*. Note that, during this thesis we have targeted Digital Cinema as a potential application for our research, however the results of this work can also be subject to other type of applications.

In this introductory chapter, we first present the motivations for stereo image compression, and watermarking in two distinct sections. We give a global insight on the problem and specify the challenges related to each subject. We then describe at a glance the main contributions of our research. Therefore the research work consists of three main parts: Part I: 3D Stereo Image Compression, Part II: From 2D to 3D Stereo Image Watermarking and Part III: Towards a combining scheme for 3D Stereo image Compression and Watermarking

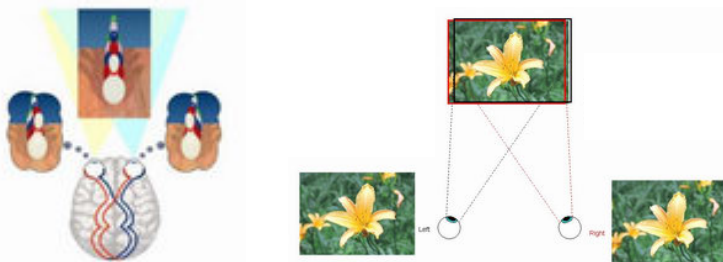


Figure 1.1: Stereovision principle, Binocular vision

1.2 Motivation and Problem Formulation

1.2.1 Stereo Image Compression

1.2.1.1 Motivation: Stereo Image Compression

A consequence of the increased realism related to stereo, is the double amount of volume information, compared to the mono channel case. Raw stereo images contain an immense amount of data. Communication and storage capabilities are limited and expensive. One of the main approaches to deal with the increasing data information is the use of efficient compression techniques. However, in the case of stereo images, an additional source of redundancy has to be addressed, this is related to the strong "binocular redundancy" between two images in a stereo pair.

A lot of interest has been paid to stereo image compression over the last decade. Several stereo image coding schemes have been proposed so far. They can fall into two main categories: the independent stereo image coding techniques and the dependent stereo image coding techniques. In the first approach, the two images of the stereo pair are coded independently using existing 2D-image coders, the binocular dependency is not exploited. As for the second category, one image is considered as a reference image, and the other one as the target image. The disparity between the reference image and the target image is estimated (Disparity Estimation), then the residual image which is the difference between the reference image and the compensated target image (Disparity Compensation) is computed. So the reference image, the residual image and the disparity map are coded [Frajka 2002].

Exploiting the binocular dependency allows achieving higher compression ratios. In this research work, we focus on designing a lossless-to-lossy transform that reduces the inter and intra-image redundancy in the stereo pair. Our objective is to improve the exploitation of the high correlation in the stereo pair. This work addresses an efficient transform specialized to generate the disparity compensated residual image.

1.2.1.2 Problem formulation: Stereo Image Compression

In dependent coding schemes, much of the literature focuses on improving the disparity estimation/compensation (DE/DC) process [McNerney 2007]. Much less work has been done on the design of a transform to generate the residue from the disparity compensated image. Traditionally, the binocular dependency has been exploited by the DE/DC , and the quantization process

Usually, the predictive coding system includes (*DE/DC*), Transform, Quantization and Entropy Coding. Therefore, the overall coding performance can be controlled at various stages of the pipeline. This is illustrated in Figure 1.2.

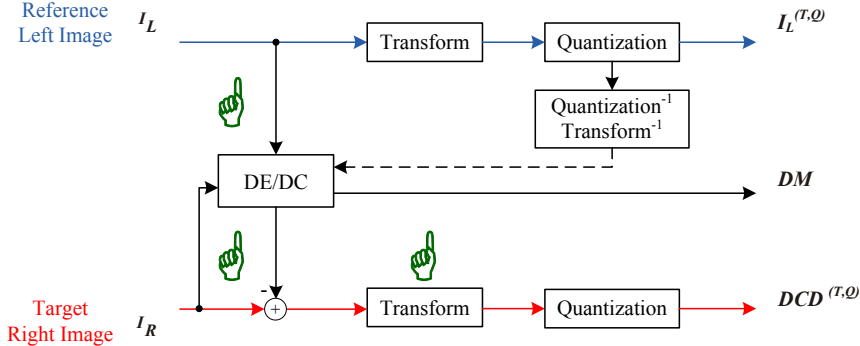


Figure 1.2: Canonical disparity compensation based compression

In 2D compression, especially with high resolution images, the Discrete Wavelet Transform (*DWT*) has shown better performance and flexibility than the Discrete Cosine Transform (*DCT*). JPEG2000 [Rabbani 2002], a DWT-based coder not only outperform the DCT-based JPEG coder but also offers scalability in reconstruction quality and spatial resolution for many applications. Moreover, JPEG2000 is the adopted standard by the Digital Cinema Initiatives (DCI). In this research, we will focus on the design of a DWT-based predictive coding scheme. The "hand sign" in Figure 1.2 highlights at which level the contributions take place in the first part of this work.

1.2.2 2D, 3D Stereo Image Watermarking

1.2.2.1 Motivation: 2D, 3D Stereo Image Watermarking

The wide distribution of Digital Media is more and more suffering from the lack of content protection. Digital Watermarking is a widely proposed solution to identify and prove the ownership of digital data. It is mostly used to detect malicious tampering of a digital data and to track illegal copies of copyrighted material.

In a Digital Cinema (DCinema) system, piracy can occur at various points in content distribution, in particular at the end of a given distribution, decoding pipeline, where the content is decrypted, decoded, and displayed. The privacy holes are illustrated in Figure 7.1. The final stage of the distribution chain, the exhibition, has been identified by members of the Motion Picture Association

(MPA) as the most susceptible and damaging source of theft and unauthorized distribution of motion picture data [Lubin 2003].

1.2.2.2 Problem Formulation: 2D, 3D Stereo Image Watermarking

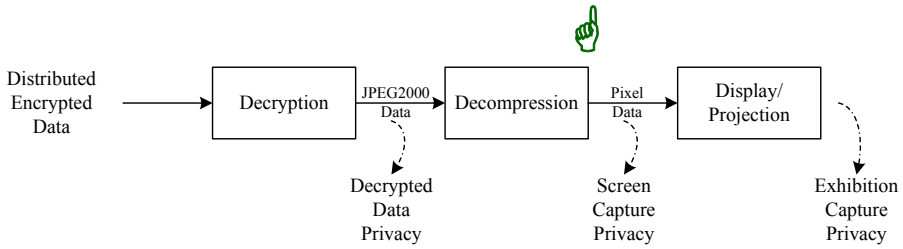


Figure 1.3: Privacy holes in the decoding flow in DCinema.

In a DCinema environment, the playback system should be able to insert the watermark identifying the time and location of the projection at the same rate when the motion pictures are decoded. The practice of embedding the watermark before the projection or the display of the content is referred to "Exhibition Watermarking". Such digital solution can be used to fill the privacy gap for digital cinema and can be similarly employed to video download.

Since the decoding flow in digital cinema is executed in real time, the exhibition watermarking system should meet the real-time constraint, and hence the complexity of the watermarking algorithm should be as low as possible. The functional encode and decode flow in the DCinema is illustrated in Figure 1.4 [DCI].

1.3 Main contribution

The primary aim of this work is (i) to design a efficient transform for stereo image coding within a predictive framework that can be extended to JPEG2000. "Efficient" in this dissertation means reducing the rate while maintaining a small distortion. The traditional disparity compensated coding scheme in Figure 1.2 is modified in order to optimize the system at various stages that are presented in Figure 1.5:

- Basically, the prediction is only related to DE/DC process. **Illumination Change Compensation (ICC)** in stereo images has been investigated in this research. A global approach based on a simple homomorphic

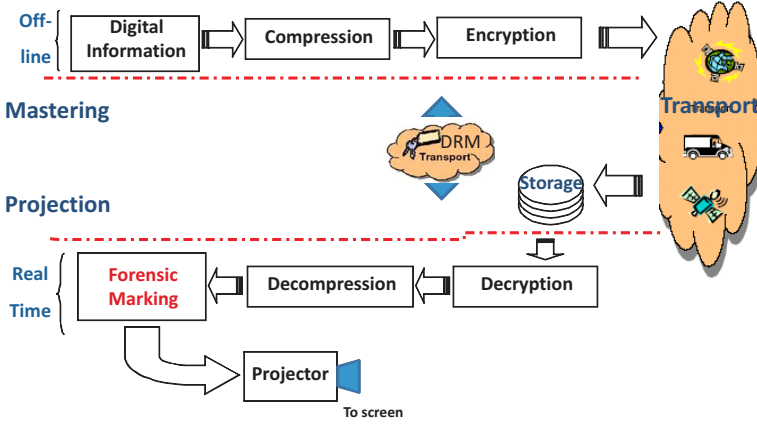


Figure 1.4: Functional encode and decode flow in the DCinema part of DCI specifications.

filtering based illumination compensation has been studied. Prior to that, a local block-based approach has been explored and implemented in a predictive scheme.

- We have proposed a **dense disparity map optimal quantization** technique. The proposed method directly implements JPEG2000 encoder in lossy mode to compress dense disparity map and computes corresponding residue and selects the compression ratio for which the total rate is minimum.
- **Edge-based Filter.** A one-dimensional optimized filter has been implemented to the disparity compensated image so as to smoothen the edges in resulting residual image. Though it leads to better results, it exploits the smoothing in only horizontal direction. Moreover, the idea has been extended to two-dimension and an optimized 2D filter has been designed. Experiments have shown that this filter reduces the required bit rate by around 0.5 bpp.
- In the proposed **lifting-based structure, directional inter prediction** has been investigated. The signal is adapted by exploiting local orientation information. In order to discriminate between different geometrical information, the system makes use of a criteria, denoting to the optimal way of prediction.
- We have proposed an **adaptation of the Lifting Scheme (LS) for the residual images** by using the disparity map. Two adaptation methods are described: Hybrid Lifting Scheme and Directional Lifting Scheme. In

Hybrid LS, the adaptation is related to the length of the filter used in the prediction and update steps. This choice is tuned by an edge map generated from the disparity map. As for the Directional LS, the adaptation is related to the choice of prediction direction or neighborhood. This is done with the help of the orientation map, that is also computed from the disparity map.

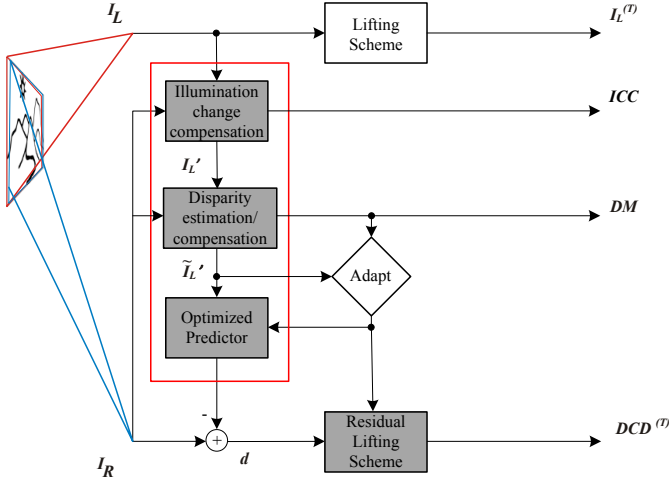


Figure 1.5: Lifting structure in a block diagram

(ii) design a watermarking scheme for 2D, 3D stereo image which can be integrated in the DCinema decoding pathway.

- We have proposed an **adaptive exhibition watermarking method** in the DWT-domain during the decoding phase. This method is designed to fit in the JPEG2000 decoding pipeline, where the embedding is done after entropy decoding and de-quantization and prior to inverse-DWT. The watermarking technique is QIM-based technique.
- Based on the properties of the DWT related to the HVS, we define a **pixel-wise masking vector that will tune the watermarking scheme** taking into consideration a number of constraints. Moreover, we have investigated the impact of the Fixed Gain attack on STDM and we have proposed a modified STDM that withstands this FGA. We have considered an implementation scenario, that is compliant to DCI.
- A **robust image authentication technique** using non-interfering Hashing and Watermarking has been proposed.

- An extension of the 2D exhibition watermarking scheme was applied to stereo images. We have proposed a **disparity guided watermarking technique** where the **perceptual mask is extended to fit with the particularities of stereo**. This was integrated in the predictive coding scheme that was investigated in part I of the thesis.

N.B: An additional research work that was investigated during this thesis is related to 3D mesh watermarking. This work is not highlighted in this dissertation for the reason that it cannot be integrated in the same framework. However for more details the reader can refer to the related publications.

1.4 Thesis outline

The overall organization of thesis is presented in Figure 1.

Chapter 2 provides a brief introduction on stereo modeling and coding techniques. Specifically, it lists the requirements of basic stereo correspondence technique; then describes the fundamental predictive coding framework in a nutshell. Following basic stereo modeling, a brief introduction on Lifting Scheme is presented, further to a summary on transforms related to stereo image coding.

Chapter 3 describes the illumination compensation problem. Global compensation and Local compensation are considered. Global compensation is addressed as a homomorphic filtering problem. An image is separated into illuminance and reflectance factors and the illuminance factor of one image is tuned to match the one of another. This chapter presents the theory, optimization, implementation, results and transmission cost for the proposed scheme.

In chapter 4, stereo correspondence and coding are addressed. For disparity estimation, the optical flow based method is used with the total variation as the regularization of the field and L1-norm for the data term. After the estimation step is done, the coding step is explained. The coding is done so as to trade-off the bitrates for disparity field and residual image. Results of above-mentioned technique are provided to justify the technique.

Chapter 5 is devoted for the development and implementation of an optimized prediction filter that minimizes the variance of residual image so that coding is efficient. It also discusses about hybrid wavelet transform method applied to the residual image to exploit the best of 5/3 transform for smooth areas and Haar transform for edges. Directional intra transform is also applied.

In chapter 6, we give an overview on digital watermarking, the main properties, and the potential applications. We also present the related work to our contributions in part II of the thesis.

In the first part of chapter 7, the context of the Digital Cinema application is introduced. The requirements of the watermarking method are defined based on the DCI related specifications. The motivation of the proposed approach is presented. In the second part of this chapter, the proposed watermarking method is described with a particular attention to the improved fidelity, robustness. A perceptual mask to modulate the watermarking technique is defined and applied. The evaluation of the effectiveness of the algorithm is then analyzed.

In chapter 8, we revisit the concept of content authentication by robust hashing and watermarking from a different angle. In our view, the hash algorithm is the essential component that achieves content authentication, watermarking is a tool to deliver the hash value, and there is interaction between the two.

In chapter 9, we integrate the watermarking scheme in a predictive stereo image decoding framework. The embedding process occurs in the JPEG2000 decoding pipeline of the reference (left) image. We design a 3D- visual-mask with respect of the HVS properties related to the DWT of 2D images, and the depth information related to the 3D experience of the stereo images.

In chapter 10, we conclude the thesis with further discussions and perspectives.

A complete view of the work performed during this thesis is represented in Figure 1.6. It describes at the same time how the various contributions are interconnected to produce a combined scheme for compression and watermarking for stereo images.

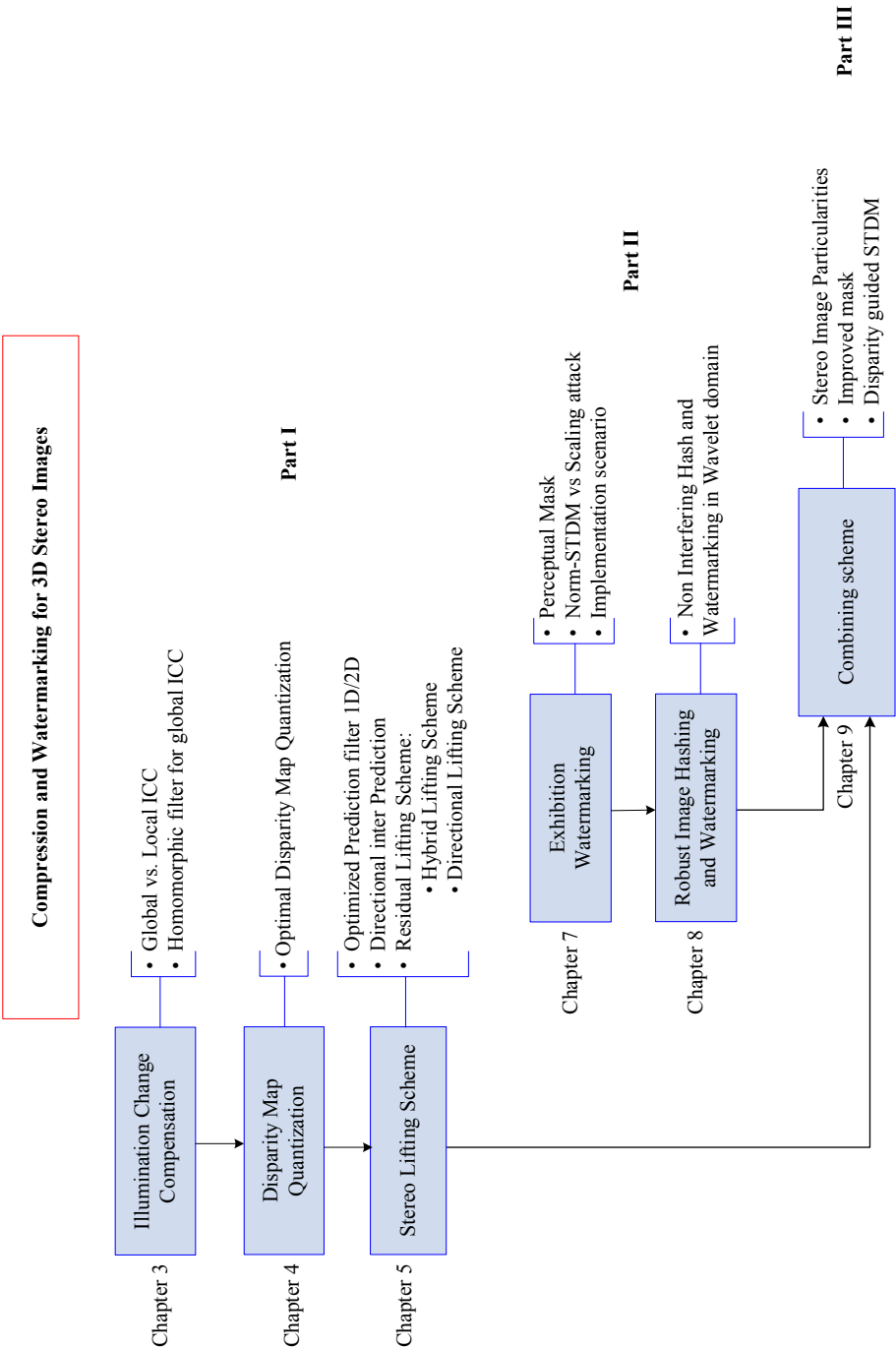


Figure 1.6: Complete pipeline of the contributions with reference to the thesis chapters.

Part I

3D Stereo image Compression

In the first part of the thesis, we focus on the design of a transform dedicated to Stereo Images. A brief overview on stereo images is presented in chapter 2 along with the related work. In chapter 3 and chapter 4, the problem of illumination change as well as a way of jointly coding the disparity map and the residual image are addressed respectively. In chapter 5, the overall stereo lifting structure is presented, contributions are illustrated and discussed at different levels of the overall scheme.

3D Stereo image

Contents

2.1	Brief history of 3D understanding	16
2.2	Stereo modeling	16
2.3	Stereo image coding	19
2.4	Predictive coding scheme	21
2.5	Background on lifting scheme	23

This chapter provides a brief introduction on stereo modeling and coding techniques. Specifically, it lists the requirements of basic stereo correspondence technique; then describes the fundamental predictive coding framework in a nutshell. Following basic stereo modeling, a brief introduction on Lifting Scheme is presented, further to a summary on transforms related to stereo image coding.

2.1 Brief history of 3D understanding

The idea of stereoscopy and binocular vision dates of long before the invention of photography. In 300 B.C, Euclid deals with the geometric problem of observing a sphere with two eyes. Although there are a number of references to the specific issue of vision with two eyes like with Battista della Porta (1538-1615), Jacopo Chimenti da Empoli (1554-1640) and François d'Aguillon (1567-1617), there is no clear understanding in the prior literature that the two slightly dissimilar left and right images were combined into one image with a new depth sense. The concepts of stereoscopy were therefore well known, but have not led to any public interest, because the lack of a crucial element: the way to reproduce the depth viewing from two images representing the same scene from two views slightly offset. In 1838, Charles Wheatstone discovered the depth sense, and released his theory on binocular vision. Using his invention of the first stereoscope "Wheatstone Viewer" in 1833, Wheatstone turned the problem into a solution. Various people sought to understand the nature of binocular vision and that it took two thousand years to properly formulate the problem. Euclid explained, from a purely geometric point of view, that we can see slightly more of an object, especially a close object, with two eyes than with one. Galen gave a physiological explanation for the singleness of vision. Porta and Kepler offered psychological theories in an attempt to explain this singleness of vision, and Aguilonius was the first to realize that singleness of vision exists not only for the one point in space on which the eyes converge, but also on a locus of points, or a horopter. All these men set the stage for Wheatstone's great discovery of stereopsis, that unique depth sense arises from the binocular nature of our vision [Lipton 1982]. Later in 1839, photography was invented independently by Niepce and Daguerre in France and Talbot in England. In 1844, Brewster invented a stereoscope that uses lenses, much more compact and easier to use than that of Wheatstone's mirror stereoscope. All stereoscopes produced thereafter are only variations of this original concept of Brewster's stereoscope, including the very popular "View- Master" released later in 1940. In 1849, Brewster also conceived the idea of the twin-lens stereo camera. The remainder progress of events related to stereoscopy progress is illustrated in a variable timeline in Figure 2.1.

2.2 Stereo modeling

In stereo photography, two pictures of the same scene are acquired from slightly different perspectives and are called *stereo pair*. The line connecting the focal centres is called baseline b . The plane passing through an object point and the focal centres is the *epipolar plane* [Lipton 1982].

In conventional stereo matching, we assume that the camera is comprised

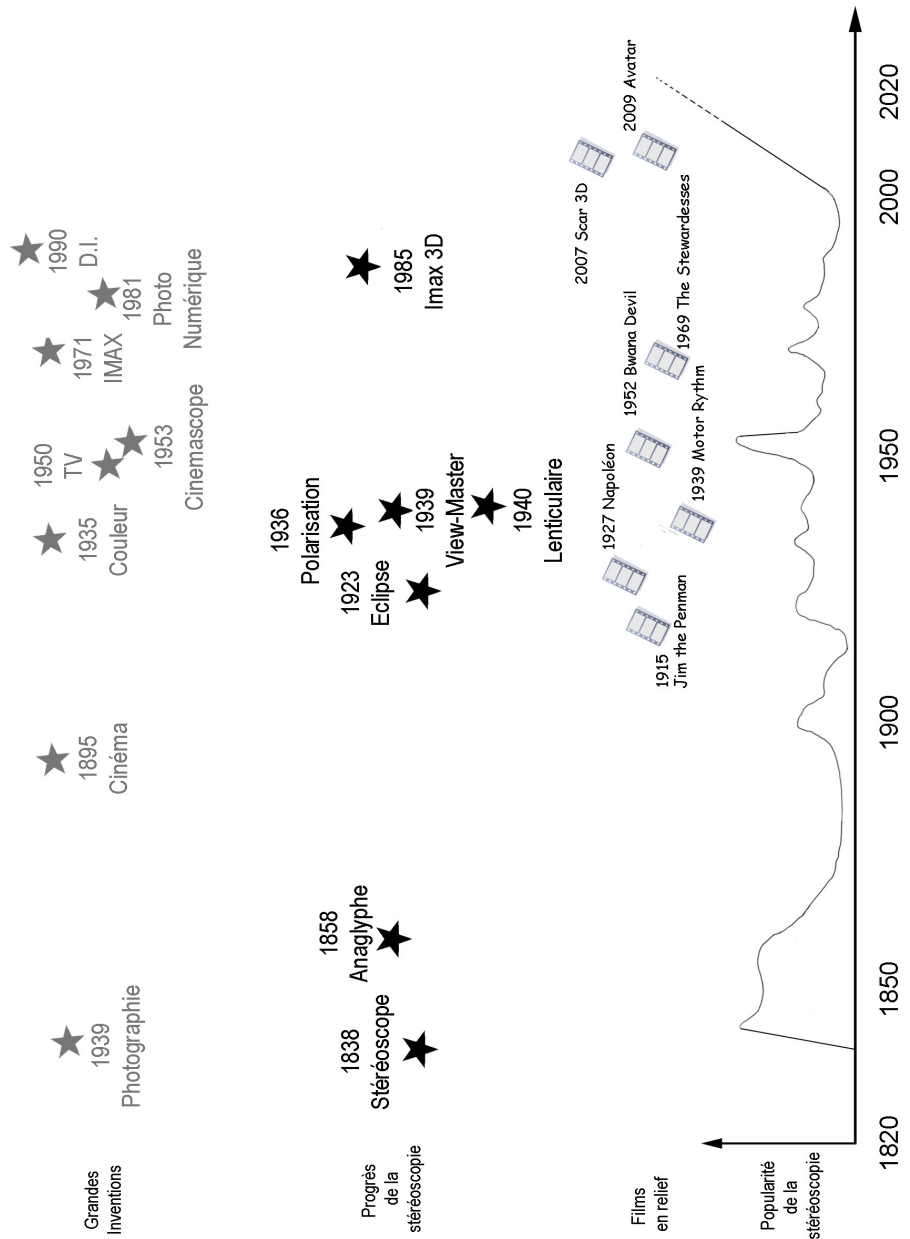


Figure 2.1: Stereoscopy in a variable timeline, courtesy of Benoît Michel

of a thin lens and that the parallel axis constraints are satisfied; that is, the focal rays of the two cameras are parallel and perpendicular to the baseline. The two image planes are coplanar and the scan line is parallel to epipolar line. Also, it is generally assumed that pixels corresponding to each point in the scene have the same brightness. If these constraints are satisfied, then the correspondence is restricted to 1D and the process can be accelerated. Finally, the 3D coordinates can be obtained by triangulation as:

$$x = \frac{b(x_l + x_r)}{2(x_l - x_r)} \quad y = \frac{b(y_l + y_r)}{2(x_l - x_r)} \quad z = \frac{bf}{2(x_l - x_r)} \quad (2.1)$$

From above relationship, we can see that depth of any point is inversely proportional to its disparity. We can also observe that we can increase the accuracy of distance measure by increasing baseline b even though the common area in the images is decreased.

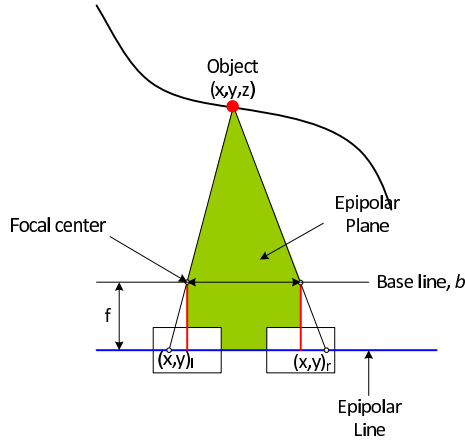


Figure 2.2: Camera geometry for stereo photography

There are different factors that make disparity estimation a difficult task, such as depth discontinuities, photometric variations, lack of texture etc. In order to overcome these ambiguities and make the estimation problem more tractable, a variety of constraints are imposed. Few of them are:

- *Epipolar geometry*: It reduces the stereo correspondence problem into one-dimensional one by requiring that match points always lie on epipolar lines. If cameras meet epipolar constraint, a particular object will appear in the two images with only a horizontal shift between its respective position. The epipolar constraint implies that focal rays of the two cameras are parallel and perpendicular to the stereo baseline.

- *Intensity constancy*: It requires that the corresponding pixel on other image should have same intensity value.
- *Uniqueness*: It imposes that one pixel can have only one corresponding pixel on the other image.
- *Ordering*: It constraints the order of points along epipolar lines remain the same.
- *Smoothness*: It requires a continuous and smooth variation in the uniform areas of disparity field.

2.3 Stereo image coding

Several stereo image coding schemes have been proposed so far. They can fall into two main categories: the independent stereo image coding techniques and the dependent stereo image coding techniques. In the first approach, the two images of the stereo pair are coded independently using existing 2D-image coders (e.g: JPEG, JPEG2000 etc.) [Rabbani 2002], the binocular dependency is not exploited. As for the second category, one image is considered as a reference image, and the other one as the target image. The disparity between the reference image and the target image is estimated, then the residual image which is the difference between the reference image and the compensated target image is computed. So the reference image, the residual image and the disparity map are coded [Frajka 2002].

2.3.1 Disparity Estimation and Compensation

Generally, the dependent stereo coding methods rely on Disparity Estimation, Compensation (DE/DC) techniques. *DE* is a key issue for exploiting the cross redundancies in the stereo pair. This problem has been extensively studied in the literature [Woo 2000], and hence, the disparity can be estimated using several techniques. These methods are presented in three categories: *Feature-based* methods, *Energy-based* approaches like optical flow methods and *Area-based* methods the most popular, like Block Matching Algorithms (BMA) [McNerney 2007].

In block, area-based technique, one image is divided into non-overlapping blocks of size $b_x \times b_y$ and for each block, corresponding block in another image is found within a search area. The disparity vector $\mathbf{v} = (v_x, v_y)$ for a block is obtained by minimizing a distance/dissimilarity criterion as follows:

$$\mathbf{v}(x, y) = \arg \min_{\mathbf{v}(x, y)} D(I_R(x, y), I_L(x + v_x, y + v_y)) \quad (2.2)$$

Most often, the sum of squared differences (SSD), sum of absolute differences (SAD) or normalized cross-correlation (NCC) is taken as the distance metric. For epipolar constraint, we may set $v_y = 0$ and thus disparity exists only in horizontal direction. However, in practice, the matching point of any current point of *right image* I_R is not always rigorously on the epipolar line because of sensor noise, discretization errors and the deviation from pinhole camera model. As a consequence, a strip along the epipolar line is considered and all the points falling within this strip are considered as potential matching candidates to be paired with the current point [Kaaniche 2009b]. Hence, all disparity vectors will form the Disparity Map (DM).

After the *DM* is obtained and the *left image* I_L is compensated, the *residual error image* I_e is formed by subtracting the *compensated left image* $\tilde{I}_L$ from the right image I_R as:

$$I_e(x, y) = I_R(x, y) - I_L(x + v_x, y + v_y) \quad (2.3)$$

In the following I_e may also be referred to as *Disparity Compensated Difference* (DCD).

Generally, the disparity vectors are losslessly encoded using Differential Pulse Code Modulation (DPCM) followed by arithmetic coding, whereas the reference and residual images can be coded using different transforms.

In disparity compensated coding, the target image in the stereo pair is replaced with the *DM* field and the *DCD* image, and thus the given bits are distributed among the reference image, the *DM* field and the *DCD* frame [Woo 2000]. Figure 2.3 illustrates the canonical predictive (dependent) stereo image encoder, where the *DE/DC* is done either using the original left image I_L (open loop) or it is done with the quantized version of I_L .

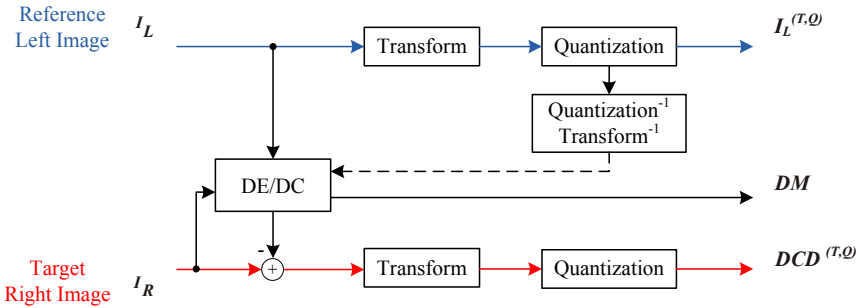


Figure 2.3: Canonical dependent stereo image encoder

N.B: In this work, our main concern is the design of an efficient transform for stereo image coding, independently from the *DE/DC* process. However, and

while tackling the prediction issue, we have run some experiments related to Illumination Change Compensation (ICC) and *DE/DC* process and hence we came out with some contributions that are presented respectively in chapter 3 and chapter 4

2.4 Predictive coding scheme

In dependent coding schemes, much of the literature focuses on improving the *DE/DC* process [McNerney 2007]. Much less work has been done on the design of an "efficient transform" to generate the residue from the disparity compensated image.

In [Moellenhoff 1998], Moellenhoff investigated specialized coding methods for the residual image. The observed characteristics of the disparity compensated stereo residual are used to select the appropriate transforms and quantization methods. In [Perkins 1992], Perkins has proposed a technique called disparity compensated transform domain predictive coding. He has also discussed the adaptive quantization of a disparity-compensated residual image. One of the recent papers proposed by Kaaniche et al. [Kaaniche 2009a] discusses the generation of the residual during and after compensation by using the lifting scheme. This method proposes to code the left image separately, and then introduces the disparity compensation of the left image details to compute the right image details. Nevertheless, the existing redundancy in the stereo pair is not fully exploited, because of the fact that the correlation in high frequencies is much smaller than in low frequencies. This is illustrated in Figure 2.4.

2.4.1 Idea of Disparity Compensated Difference Transform

In this work, our objective is to improve the exploitation of the high redundancy in the stereo pair. This work addresses an efficient transform specialized to generate the disparity compensated residual image. Our goal is to improve the overall performance in terms of entropy and bitrate regardless of the disparity estimation/compensation approach that is used. We propose a new scheme by designing a transform that reduces the intra and inter-image redundancy.

In 2D compression, especially with high resolution images, the Discrete Wavelet Transform (DWT) has shown better performance and flexibility than the Discrete Cosine Transform (DCT) [Xu 2002]. JPEG2000, a DWT-based coder not only outperform the DCT-based JPEG coder but also offers scalability in reconstruction quality and spatial resolution for many applications. Moreover, JPEG2000 is the adopted standard by the Digital Cinema Initiatives (DCI). The proposed scheme is inspired by the Lifting Scheme (LS) which correspond to classical wavelet transform (WT)[Sweldens 1996a].

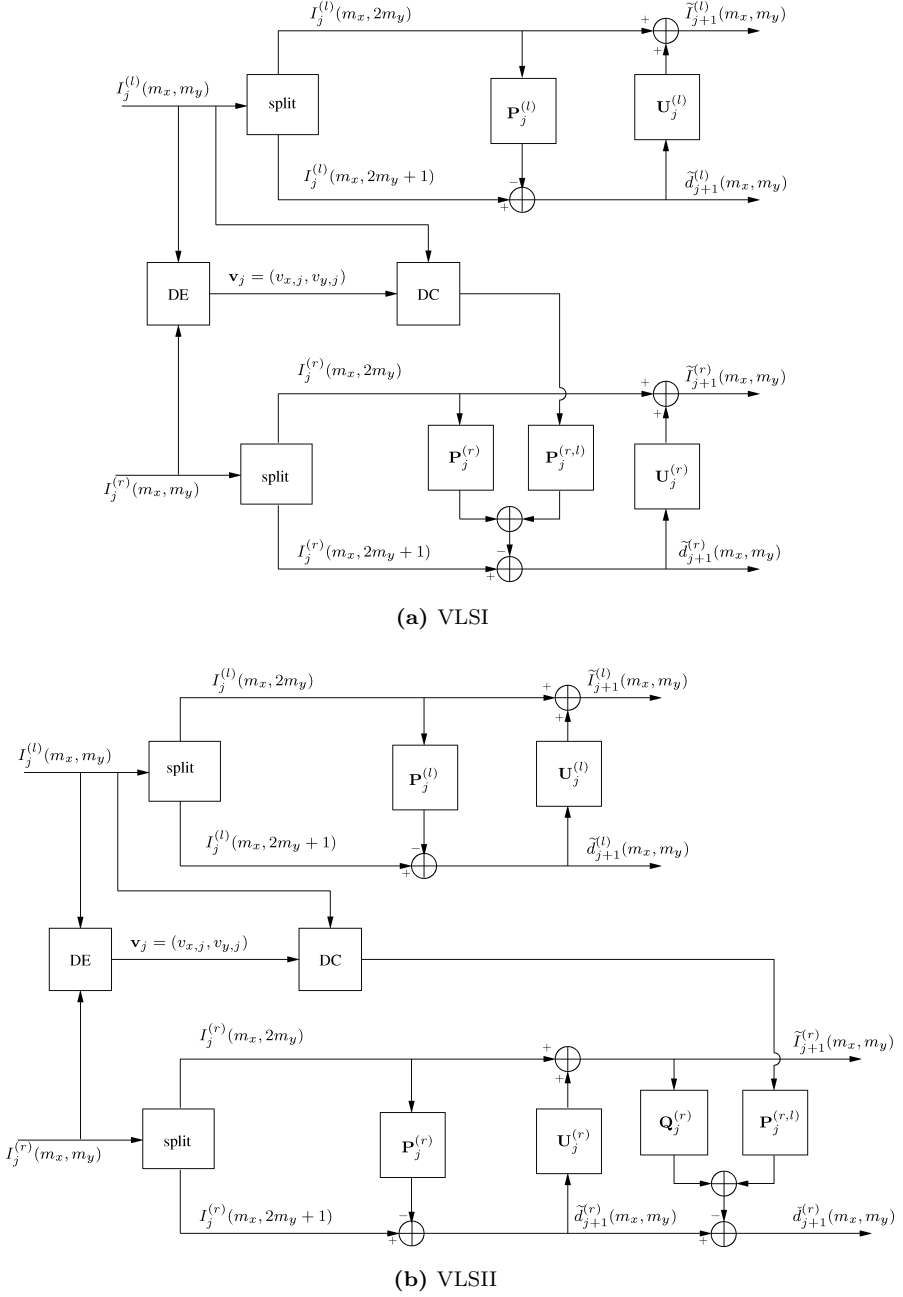


Figure 2.4: Joint Stereo Lifting Scheme proposed by Kaaniche et al.

2.4.2 Directional Disparity Compensated Transform, 2D Directional Transforms

How to fully exploit the directional correlation in the either spatial or frequency domains has been a research interest for many years in 2D image compression. In [Ding 2007], Ding et al. reviewed direction-aware wavelet-based image coding methods into two major categories in their ways of using and coding directional information: the first category is Directional Filter and Transform, and the second one is Directional Prediction. In Directional Filter and Transform, many wavelet transforms have been proposed to preserve fine directional information in the wavelet domain, such as ridgelet [Candès 1999], curvelet [Do 2003], directional wavelet transform [Piella 2002, Piella 2005, Kitanovski 2008, Gerek 2006], contourlet [Do 2005], complex wavelet, brushlet, bandlet [Maalouf 2010].

These directional filters and transforms provide good presentations of directional data in the frequency domain. They are extensively used for feature extraction, image enhancement, denoising, classification, and even retrieval. However, they are not suited for compression because it does not exist yet or because it is due to the nature of these transforms. In addition, these methods require a considerable amount of side information. On the other hand, in Directional Prediction: The key issue on applying directional prediction in wavelet decomposition is the conflict of global transform and local features. The prediction is locally adapted by exploiting local information orientation of the signal. Major work in this line of thinking was also reported in [Ding 2007]. A new technique of seamlessly integrating directional prediction for stereo image coding is presented in chapter 5.

2.5 Background on lifting scheme

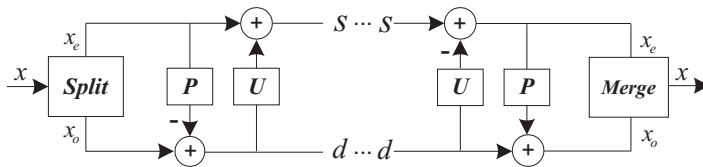


Figure 2.5: Canonical Lifting Scheme

The fundamental principle of the LS is to exploit efficiently the correlation in the signal, leading to a more compact information set [Sweldens 1996b]. LS corresponds to classical WT (2^{nd} generation Wavelet decomposition). In [Daubechies 1998] Daubechies has demonstrated that any wavelet transform implemented by transversal finite impulse response filters can also be implemented

by equivalent lifting scheme with a finite number of stages.

The lifting scheme implementation is simple, fast, efficient and easily invertible [Sweldens 1995a]. The canonical scheme consists of three main steps [Sweldens 1996a, Calderbank 1998] as illustrated in Figure 2.5.

First, the original signal, denoted x , is split into two disjoint subsets, x_e (the set of even samples of x), and x_o (the set of odd samples of x). Second, a prediction operator, denoted P , is applied on the subset x_e , in order to predict the subset x_o :

$$d = x_o - P * x_e \quad (2.4)$$

where the prediction error d represents the details of x . This stage represents a lifting step. Third, an update operator, denoted U , is applied on d :

$$s = x_e + U * d \quad (2.5)$$

This stage represents a dual lifting step. The lifting is built by alternating different prediction and update steps. The inverse LS consists of making the inverse operation and a merge (inverse of the split) as shown in Figure 2.5. The above implementation of LS is presented in 1-D. Similar lifting structure steps operating along the lines and columns of an image are carried out simultaneously in the horizontal and the vertical direction. Something to be mentioned is that the most decisive advantage of LS for temporal filtering of a image sequence is the ability to change operators prediction and update even by non-linear operations such as motion compensation, without compromising its invertibility.

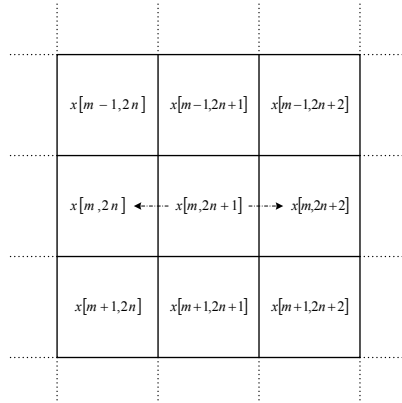


Figure 2.6: Involved samples in canonical 1D - Lifting Scheme

To illustrate the potential of LS to produce a compact representation, and for the sake of comparison in the following, we consider the example of the

well-known 5/3 wavelet filters. The 5/3 filters generally outperforms other filters for natural and smooth images [Daubechies 1998]. They are represented by the transversal filters with $h_1 = [-1/2, 1, -1/2]$ and $h_2 = [-1/8, 1/4, 3/4, 1/4, -1/8]$. Where h_1 and h_2 are respectively the coefficients of filter p and u . These aforementioned filters are equivalent to the LS(2,2) with the following equations [Sweldens 1995b]:

$$d[m, n] = x[m, 2n + 1] - \sum_{k \in \mathcal{P}} p_k \cdot x[m, 2n - 2k] \quad (2.6)$$

and

$$s[m, n] = x[m, 2n] + \sum_{k \in \mathcal{U}} u_k \cdot d[m, n - k] \quad (2.7)$$

Where $\mathcal{P} = \{-1; 0\}$ and $\mathcal{U} = \{0; 1\}$ are respectively the support index of the filter p and the filter u . And $p = [1/2; 1/2]$ and $u = [1/4; 1/4]$ their related weights. Rounding operation $\lfloor \cdot \rfloor$ for the second term in Equation 2.6 and Equation 2.7 results to a integer-to-integer transform.

Note that, as illustrated in Figure 2.6, the prediction of even sample $x[m, 2n + 1]$ involves equally neighboring odd samples $x[m, 2n]$ and $x[m, 2n + 2]$ only in the horizontal direction here.

Illumination Change Compensation

Contents

3.1	Introduction	27
3.2	Global illumination compensation	28
3.3	Block-based and local ICC	30
3.4	ICC performance and transmission cost, Global vs. Local approach	31
3.5	Implementation of local ICC	31
3.6	Conclusion	34

3.1 Introduction

¹ The Human Visual System is far more sensitive to the brightness information than to color information. Most of the image and video standards consider more the luminance component and reduce the chrominance components. One of basic constraints, that stereo correspondence requires, is *constancy*, which means that corresponding pixel on right image should have same intensity as left image. But, sometimes the left and right images show significant illumination difference leading to poor correspondence. When multiview video/image data are recorded, significant variations are often observed between the luminance and chrominance components of the different views [Fecker 2008]. These discrepancies mainly originate from calibration differences between cameras, leading to global changes between the views. In addition to global changes, local illumination variations occur because the same object surface can appear differently from different angles.

Applying an Illumination compensation step can have a big impact on the prediction results.

¹Part of this work is jointly done with Amit Kumar KC.

As previously mentioned, our main concern is not to develop a new DE/DC algorithm, however the *ICC* step is closely dependent on the used compensation method. In order to tackle this problem, we have considered two class of DE/DC algorithms: Block-based and Energy-based. We have adapted two strategies for ICC: Global compensation and Local compensation.

In the chapter we summarize those techniques, and we present the pros and cons for each approach.

3.2 Global illumination compensation

Fecker *et al.* have used a cumulative histogram mapping technique [Fecker 2008] for multiview video. For perfect reconstruction, the whole lookup table has to be sent to the decoder. Moreover, the histogram exhibits sudden drops to null for some intensity values. The difference in illumination level can be easily observed from their corresponding histograms and can be numerically computed with Kullback-Leibler (KL) distance between normalized histograms. Having low value for KL distance implies that the left image is quite closer to right image in terms of intensity distribution. According to illumination-reflectance model, an image is obtained as a result of multiplicative interaction between object (which is modeled by reflectance factor), and illumination (modeled by illuminance factor). Assuming that the reflectance factor is the same for both left and right images (because both correspond to the same object), we can adjust the illuminance parameter so that their histograms are as close as possible.

3.2.1 Illumination-reflectance model and homomorphic filtering

An image is generated from a physical process, its values are proportional to energy radiated by a physical source [Delac 2006]. In other words, an image is an array of measured light intensities and is a function of the amount of light reflected of the objects in the scene. The intensity is a product of *illuminance* (the amount of source illumination incident on the scene being viewed) and *reflectance* (the amount of illumination reflected by the objects in the scene). Thus the observed image $c[m, n]$ is the result of multiplicative interaction of illumination factor with the object. Thus,

$$c[m, n] = L[m, n].R[m, n] \quad (3.1)$$

where $L[m, n]$ is the illumination factor and $R[m, n]$ represents the reflectance factor of *illumination-reflectance model*. This model has been well used to address the problem of improving the quality of an image that has been acquired under poor illumination conditions.

It is worth noting that the illumination factor $L[m, n]$ is usually slow-varying whereas $R[m, n]$ is presumably not. Thus, if we take the natural logarithm of $c[m, n]$, we will obtain two additive terms one of which is low-pass and another is high-pass. Applying low-pass filter to the logarithm, thus, gives illumination component. Similarly, we can obtain the reflectance component by high pass filtering the logarithm of $c[m, n]$. In formula:

$$\begin{aligned} c[m, n] &= L[m, n]R[m, n] \\ \ln(c[m, n]) &= \ln(L[m, n]) + \ln(R[m, n]) \\ L'[m, n] &= \text{LowPass}\{\ln(c[m, n])\} \\ R'[m, n] &= \text{HighPass}\{\ln(c[m, n])\} \end{aligned} \quad (3.2)$$

After the lowpass and highpass components are separated, they are processed separately and then merged.

3.2.2 Histogram equalization

After the illumination component has been separated from left image, we can multiply it with some factor α (which depends on right image), we can adjust the illumination level of left according to right image. The optimum value of α is chosen such that the histogram of left image h_l is aligned with the right one h_r , or in other words minimizing the Kullback-Leibler (KL) distance between normalized histograms.

The KL distance measures the expected number of extra bits required to code samples from one distribution when using a code based on another distribution [KL-]. Hence, minimizing KL distance between h_r and h_l implies that less bits are required to encode right image from left which leads to better correlation between stereo image pairs and thus, improved disparity estimation/compensation. By sweeping α , we can modify the histogram of left image horizontally so as to best align it with that of right image. The particular value of α for which two histograms are best aligned is designated as α_{opt} . Mathematically,

$$KL(h_l, h_r) = \frac{1}{2} \left[\sum h_l \log_2 \left(\frac{h_l}{h_r} \right) + \sum h_r \log_2 \left(\frac{h_r}{h_l} \right) \right] \quad (3.3)$$

$$\alpha_{opt} = \underset{\alpha}{\operatorname{argmin}} KL(h_l, h_r) \quad (3.4)$$

where h_l and h_r represent the normalized histograms of left and right images respectively.

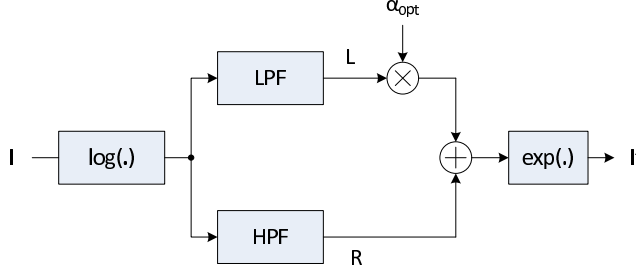


Figure 3.1: Homomorphic filtering based illumination compensation

3.2.3 Implementation

The algorithm takes left and right images as inputs and provides compensated left, compensated right, α_{opt} , *Send* flag and I_1-I_2 flag as outputs. *Send* flag is used to indicate whether α_{opt} needs to be sent or not. Often, the illumination levels are same in left and right images. In such cases, it is worthless to transmit α_{opt} . For numerical implementation, α has been swept from 0.9 to 1.1. The flag I_1-I_2 indicates whether right image or the left image has been modified. This flag is introduced because it is always wise to lower the illumination level of one image regarding the other, so that the dynamic range of both images does not change. Gaussian filter has been used as the low pass filter. The σ of filter should be taken so that it is larger than the object size in the image [tu-]. For implementation purpose, σ is taken as 10 and filter size is 20×20 .

3.3 Block-based and local ICC

An alternative to the global compensation approach is the local illumination compensation. This approach can fit with block-based DE algorithms where the compensation is done jointly with the DE. Luminance correction is done for each block after the best matching. This highlights the existing spatial redundancy and consequently improves our results in the prediction step. We considered as in [Hur 2007] that illumination variation in the stereo pair can be approximated to a luminance shift. For this end, one additive value per block needs to be sent to the decoder. For each matching block where NCC has been used as a similarity measure, ICC is done as follows:

$$Block(\tilde{I}_L) = Block(\tilde{I}_L) + Mean(Block(I_R)) - Mean(Block(\tilde{I}_L)) \quad (3.5)$$

Where $\tilde{I}_L$ is the disparity compensated left image block obtained as:

$$\tilde{I}_L(x, y) = I_L(x + u(x, y), y + v(x, y)) \quad (3.6)$$

While $(u(x, y), v(x, y))$ represents the disparity vector.

3.4 ICC performance and transmission cost, Global vs. Local approach

This first technique is primarily based on the assumption that illumination change between left and right images is global. So, it tries to compensate illumination discrepancy globally (and not locally) just by adjusting the illumination level. On the other hand, the local method is quite insensitive to illumination change but at the cost of increased overhead. This method requires *one component per block to be sent*.

As for the transmission cost of the global compensation technique, it essentially requires two flag bits, viz. *Send* and I_1-I_2 , to be sent. Occasionally, when the illumination correction needs to be performed, α_{opt} is also required. So, it can be said that the transmission overhead is not much. If *Send* flag is zero, then α_{opt} is not required to be transmitted.

3.5 Implementation of local ICC

This section provides results of homomorphic filtering ICC procedure and local block-based ICC. Figures below depict the qualitative improvement after global compensation. We can see that the algorithm performs well in case of image where the illumination level is significantly different. Figures 3.2 and 3.3 depict histograms of images and that of residues before and after compensation. It can be noted that no compensation is needed for images like *Tsukuba* (as $\alpha_{opt} = 1$), whereas, images like *Shrub*, *Fruit* etc. show improved performance in terms of bitrate (and KL distance).

Table 3.1 provides numerical results of illumination compensation for different stereo pairs. Images like *Apple* and *Tsukuba* do not have significant illumination difference. The left and right histograms for these stereo pairs are almost aligned. For such images, *Send* flag is asserted LOW, which means that α_{opt} does not need to be sent. However, stereo pairs like *J1*², *Shrub*³, *Walker1*⁴ and *Walker9*⁵ have significant illumination difference between left and right images. Sometimes left image seems to be brighter than the right and vice versa. Depending upon which image is brighter, the algorithm automatically scales its illumination level down to a value so that its histogram aligns with that of another image. If left histogram is modified, then I_1-I_2 flag is set ONE, otherwise it is asserted ZERO. We can observe that even such simple compensation process is quite effective. For example, in case of *Shrub* image, the bpp rate for disparity reduces to 0.10

²Right eye is approximately 13% brighter than the left

³Right eye is approximately 5% brighter than the left

⁴Right eye is approximately 11% brighter than the left

⁵Right eye is approximately 13% brighter than the left

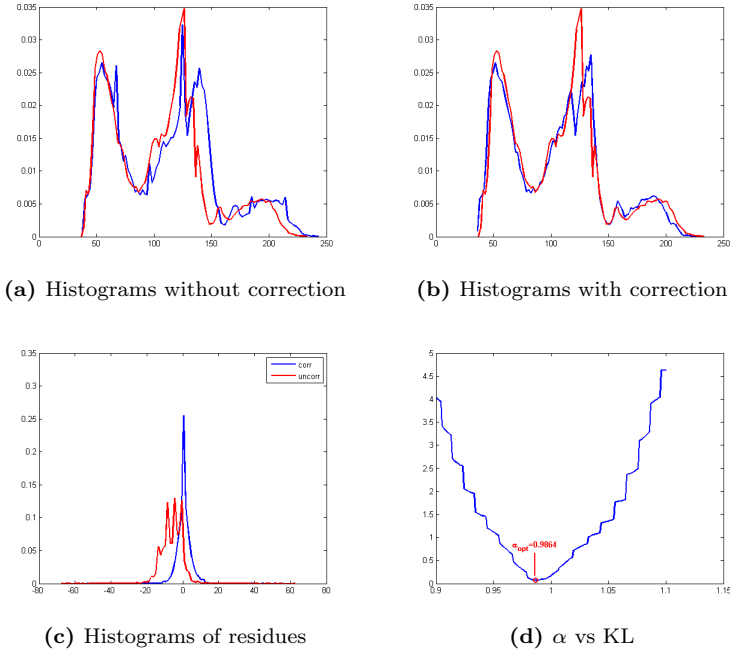


Figure 3.2: Results of illumination compensation for *Shrub* image

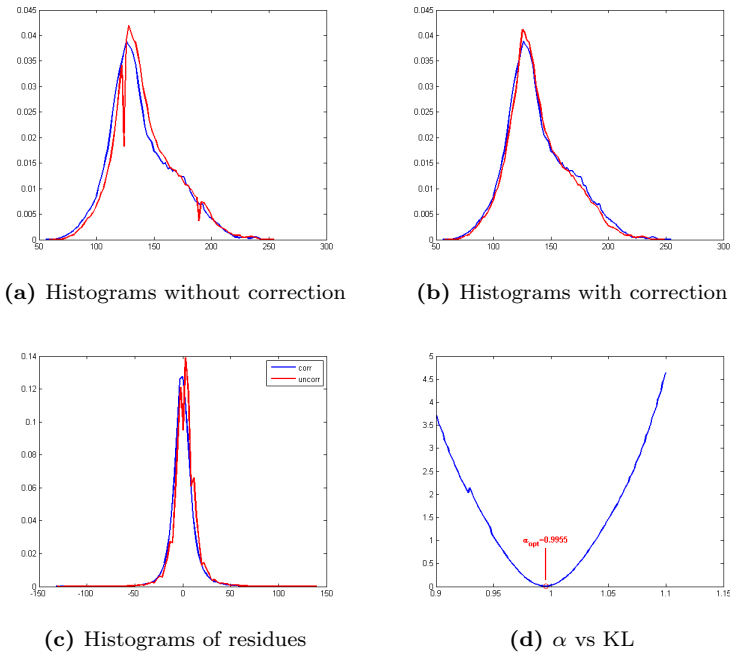


Figure 3.3: Results of illumination compensation for *HouseOf* image

from 0.13 and that for DCD is decreased from 3.45 to 3.27, thereby reducing the total bit rate by around 0.2 bpp at the cost of two bits (for flags) and one coefficient.

Table 3.1: Rate [bpp] comparison for global illumination compensation

Image	Before Compensation			After Compensation			Flags		α_{opt}
	Bitrate		KL Distance	Bitrate		KL Distance	I_1 - I_2	$Send$	
	DV	DCD		DV	DCD				
Apple	0.85	4.54	6.09	0.87	4.54	5.35	1	0	0.9997
Shrub	0.13	3.45	1.12	0.10	3.27	0.07	1	1	0.9867
J1	1.28	4.56	0.62	0.89	4.24	0.38	0	1	0.9857
Walker1	0.95	4.98	0.30	0.91	4.88	0.20	0	1	0.9692
Walker9	0.97	5.06	0.27	0.95	4.94	0.19	0	1	0.9687
Fruit	0.54	4.08	1.31	0.49	4.05	0.10	1	1	0.9892
Tsukuba	0.35	3.51	0.44	0.35	3.51	0.11	0	0	1.0003
HouseOf	0.72	5.37	1.09	0.71	5.36	0.01	0	1	0.9952

The final simulations on global ICC have been carried out using two different datasets. The first dataset denoted $\mathcal{D}_1$, contains 38 stereo pairs from the well-known Middlebury stereo dataset. The resolution of the images is about 370×435 pixels, they represent scenes with complex objects and textures. These pairs are all rectified so that only horizontal disparity appears. The second dataset denoted $\mathcal{D}_2$ contains, 28 pairs which represent a number of well known stereo pairs (e.g: Pentagon, Fruit, Tsukuba ...). They are about 512×512 pixels resolution, they can be considered as more natural images compared to dataset $\mathcal{D}_1$. These stereo pairs are not rectified and posses significant illumination difference. The hall stereo images can be downloaded at [car b, car a, mid].

To simplify the presentation of results, we define the following notations:

- S_I denotes the independent coding scheme.
- S_B corresponds to the baseline scheme using only disparity estimation/compensation.
- S_{LC} is the extension of S_B when adding the illumination compensation process.

dataset	S_I	S_B	S_{LC}
$\mathcal{D}_1$	3.97	4.12	4.12
$\mathcal{D}_2$	3.90	4.35	4.31

Table 3.2: Average entropy and bitrate of residual image (in bpp) over dataset $\mathcal{D}_1$ and $\mathcal{D}_2$

In Table 3.3, numerical results for local block-based *ICC* are presented. In order to highlight the impact of *ICC* on the prediction of the residual image, the

original left image I_L and the residual image I_e are coded using the lossless 5/3 transform over 3 levels. I'_e refer to I_e subject to the *ICC* process. Note that the cost related to the *DM* and the *ICC* parameters can be encoded in the same way for the two cases using DPCM and are not added to the total bitrate.

Table 3.3: Comparision for local ICC in terms of entropies (in bpp).

	Original	5/3 for (I_L, I_e) without ICC	5/3 for (I_L, I'_e) after ICC
Fruit	6.36	4.10	3.88
Pentagon	6.66	5.27	5.15
Houseof	6.75	5.37	5.24

3.6 Conclusion

In this chapter, the illumination compensation procedure using global and local approach is discussed. The global approach is based on the assumption that the reflectance factor for both left and right images remains invariant. This approach can be used with energy based DE/DC algorithm like optical flow. As for the local approach, the illumination change is approximated to a luminance shift and each block is modified. This approach is more suitable for block-based DE/DC algorithm where the illumination compensation is done in a parallel way to the DE/DC process. The global compensation process is not as efficient as block-based approach. Nonetheless, its transmission overhead is almost negligible as compared to block based approach. The results show that on average while applying Global ICC for non rectified stereo images (dataset $\mathcal{D}_2$), the gain is better than rectified stereo images (dataset $\mathcal{D}_1$) where there is no gain.

Optimal DM quantization using optical flow method

Contents

4.1	Introduction	35
4.2	Disparity map estimation and optimal quantization .	37
4.3	Experimental Results	38
4.4	Conclusion	40

4.1 Introduction

Stereo images are associated with the same scene, observed from two view points. The left and right views represent the stereo pair. In dependent stereo image coding, one image is considered as the reference image and other as the target image. The disparity between these two images is estimated and the residual image is obtained by computing the difference between the target image and the disparity compensated reference image. There exist a variety of methods to estimate the disparity map. Classically, block matching technique has been used to estimate the disparity. Another class of disparity map estimation is optical flow for which disparity is estimated at each point of the image assuming that the pixel intensity does not change on corresponding points.

According to Quentin *et al.* and Yu *et al.* [Davis 1995, Chi 2007], the two classes, namely *optical flow* and *block matching*, are equivalent for subpixel displacements. Specifically, block matching that uses bilinear interpolation can be recast into equivalent optical flow formulations.

Since block matching and optical flow methods generate numerically equivalent estimates for subpixel displacements, other features may suggest choosing one over another. For example, the implementation of the optical flow algorithm may require significantly fewer operations than an equivalent block matching algorithm (BMA). The computational effort of a full search BMA is directly related to the size of the search window. In case of optical flow, there is no

specific search window and the effective range is determined from characteristics of the image. In addition, the optical flow algorithm does not require any overhead to obtain subpixel disparity vectors as the interpolation is amortized within the gradient filters [Chi 2007].

Research in stereo image coding has focused on the disparity estimation and compensation process to exploit the inter-ocular (or cross-view) redundancies. The goal of stereo image coding is not to estimate the true disparity but rather to achieve a high compression ratio. Therefore, it may not be worthwhile to compute a dense disparity map if the cost of handling (transmitting or storing) the disparity vector field is too high. That is why, fixed size block matching has been widely used, even though the true disparity maps are blockwise constant [Woo 1996]. Most of the reported methods use a classical block-based technique in order to estimate the disparity map. However, this estimation technique does not always provide an accurate disparity map, which may affect the compensation step. Specifically, the block-based technique fails at object edges and boundaries. Besides, if the block size becomes smaller, the overhead required to transmit the disparity information becomes large with respect to overall bit rate.

Regarding disparity compensation, dense disparity map is superior but at the price of huge transmission/storage cost. However, the dense disparity map provides a great flexibility to tradeoff between the disparity map and the resulting residual image bitrates. The disparity map is computed only once. Then, one can set the bitrate for disparity map and compute the corresponding residual image. Since our aim is to minimize the sum of bitrates of disparity map and residual image, it allows an optimization on the accuracy of the disparity map with respect to required bitrate for the residual image. This is not possible for block-matching as the disparity map has to be computed every time for different block-size.

Morvan *et al.* [Morvan 2007] have used rate-distortion (RD) framework based quadtree decomposition for depth map. In order to reduce bpp cost for disparity map, the previous can be segmented using quadtree decomposition followed by a median filtering as proposed by Kaaniche *et al.* [Kaaniche 2009b]. The idea behind [Kaaniche 2009b] is to divide the estimated field into macro blocks. If the block is homogeneous, then it is kept intact, else it is recursively divided into four sub blocks. Thus, homogeneous areas will be represented by larger blocks whereas boundaries will have smaller blocks.

However, above-mentioned works consider either the residual image or the disparity map but not both for quantization. Thus, these techniques are difficult to adopt for lossless encoding which requires both the disparity map and the residual image to be considered for encoding. In this paper, the goal is to compute the optimal disparity map (in terms of bitrates) by jointly exploiting the trade-off between the disparity map and the residual image. The novelty

of this work is that both the disparity map and the residual image are jointly considered and consequently the optimum quantization of the disparity map is done such that the overall bitrate is minimized.

The organization of this chapter is as follows. In section 4.2, the proposed optimization scheme is presented. Experimental results are shown in section 4.3, followed by conclusion in section 4.4.

4.2 Disparity map estimation and optimal quantization

4.2.1 Disparity map estimation

In the optical flow method, the disparity map estimation problem can be stated as a minimization problem which minimizes the following energy function [Zach 2007].

$$J = \int |\nabla u| d\underline{x} + \lambda \int |I_L(\underline{x} + \underline{u}(\underline{x})) - I_R(\underline{x})| d\underline{x} \quad (4.1)$$

where ∇u is the gradient of the disparity map u , the data term is represented by the L_1 norm of the difference between the left image I_L and the right image I_R , the regularization term is represented by total variation (TV) norm of u .

The Lagrangian parameter λ is used to determine which term (data or regularization) should be given more weight during optimization process. In other words, it balances between the data fidelity and the regularization terms. Larger value of λ exhibits more priority to the data term and hence the disparity compensated difference (DCD) is improved. Thus, we expect that the entropy of the DCD decreases with increasing λ . On the other hand, smaller value of λ exhibits more priority to the disparity map and hence the entropy of the disparity map is reduced. Therefore, the free parameter λ shows a tradeoff between the accuracy of the DCD and the disparity map.

Note that the detailed resolution of the aforementioned energy function is presented in Appendix A.

4.2.2 Optimal quantization of disparity map

Once the dense disparity map is obtained, an optimal quantization is performed on the dense disparity map such that the total bitrate of resulting disparity map and the residual image is minimized. The block diagram is presented in Figure 4.1:

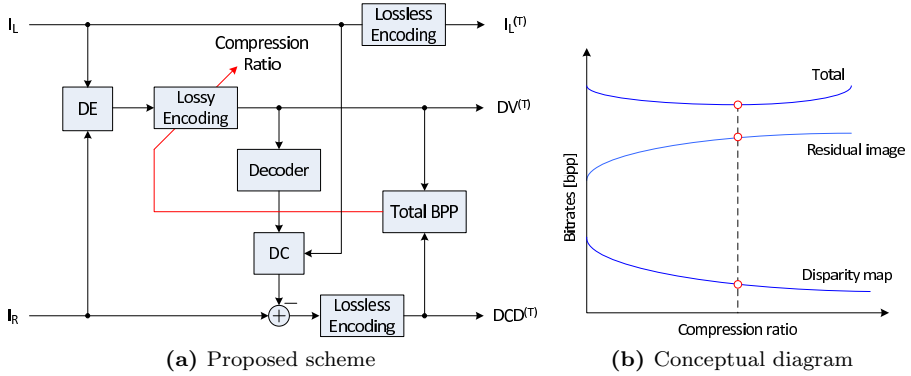


Figure 4.1: Proposed scheme for optimal quantization of disparity map

We quantize disparity map with different compression ratios using JPEG2000 codec as discussed earlier. It is obvious that as we quantize the disparity map, the residual image quality gets worsened i.e., the bitrate for residual image $R^{(e)}$ increases, whereas the bitrate for the disparity map $R^{(v)}$ decreases. It is worth noting the rate at which $R^{(v)}$ changes is different than the rate at which $R^{(e)}$ changes. Specifically, $R^{(v)}$ decreases much rapidly and $R^{(e)}$ increases slowly until some quantization level. Beyond this level, $R^{(v)}$ decreases slowly whereas $R^{(e)}$ starts increasing sharply. Since we are interested at the total bpp (i.e., $R^{(v)} + R^{(e)}$), the total bitrate exhibits a minimum. This can be illustrated in Figure 4.1. It can be easily observed that our proposed scheme allows to have a direct tradeoff between the bitrate of disparity map and that of residual image. The advantages of the proposed method as compared to quadtree decomposition method(QTD) [Kaaniche 2009b] are:

- QTD technique requires side information about the block size and segmented disparity map to be sent. This is not required in the proposed scheme.
- QTD is not directly compatible to JPEG2000 as it incorporates differential pulse code modulation (DPCM) encoding for transmission of encoded bits. Our method uses the JPEG2000 encoder to encode the disparity map as normal image, thereby alleviating the need of using DPCM. The proposed method is an extension of our previous work which deals with compression of stereo images for digital cinema using directional and optimized prediction techniques.[Darazi 2009b].

4.3 Experimental Results

For the simulations, stereo images are downloaded from Carnegie Mellon and Middlebury websites[car b, car a, mid]. Figure 4.2 depicts the qualitative behaviour of the disparity map and the residual image for different compression ratios in the case of *Tsukuba* image.

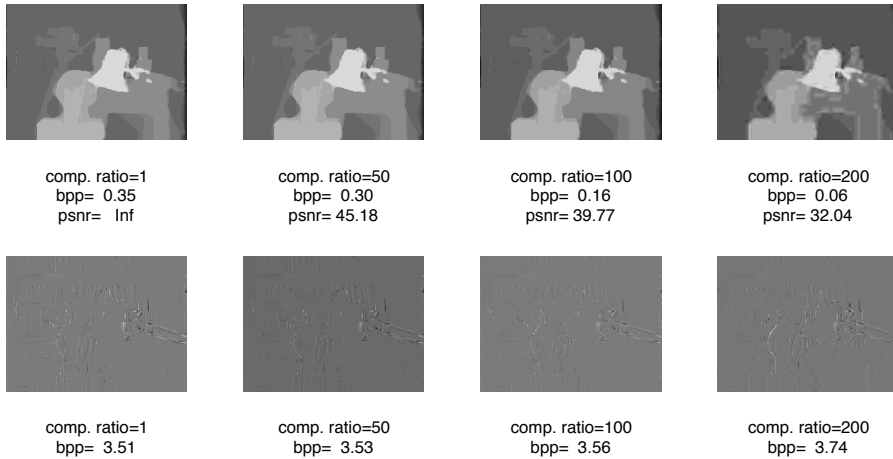


Figure 4.2: Quality of disparity map and residual image

From the above figure, it can be noted that the quality of the disparity map does not change much even for high compression ratio. The reason behind it is that the disparity map (obtained by using total variation as regularization) has cartoon shape, i.e., it has almost smooth regions separated by edges. One possible drawback of compressing disparity map at very high compression ratio is appearance of spurious edges in the disparity map or disappearance of some weak edges.

Figure 4.3 verifies our assumption about the resulting bitrates (bpp curves) for the residual image and the disparity map. It depicts the behavior of the disparity map and the residual image for various compression ratios (and hence various bitrates). Moreover, the reconstruction quality of the disparity map is also presented in terms of peak-signal-to-noise-ratio (PSNR) between the unquantized disparity map u , and the reconstructed disparity map after quantization, $\tilde{u}$. Table 4.1 gives numerical results of above-mentioned optimization procedure. It depicts the bitrates of the disparity map, the residual image and the total rate before and after optimization for different images. It can be easily concluded that the optimization procedure gives an advantage of about 0.3 bpp on total bitrate. We can observe from the Table 4.1 that the bitrate for the disparity map drops drastically from an average of 0.56 bpp to 0.07 bpp. On

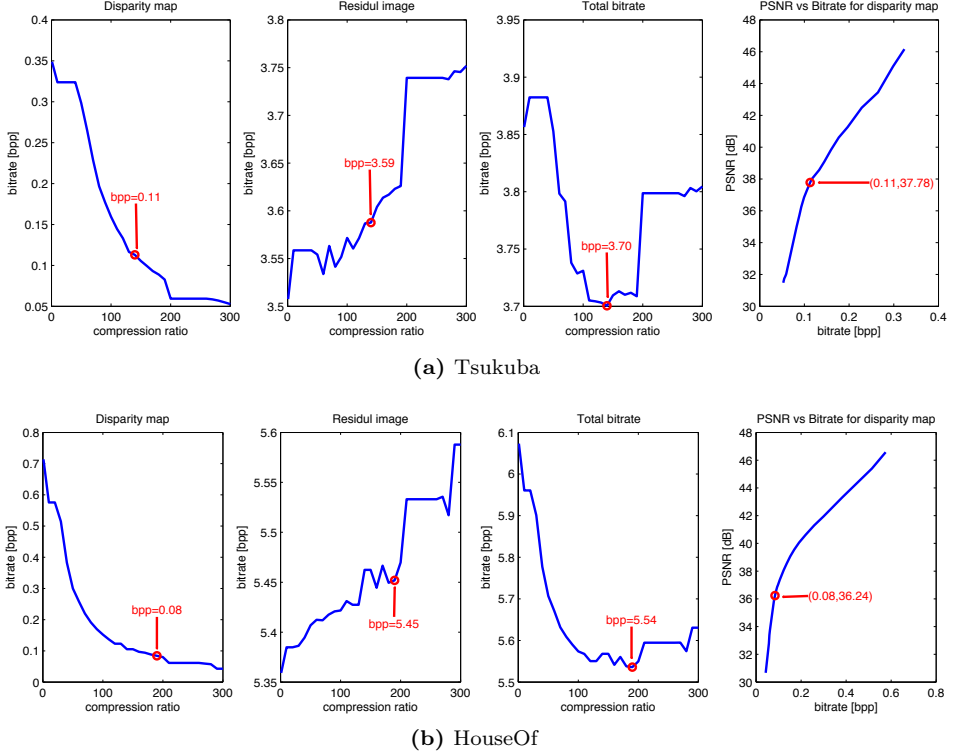


Figure 4.3: Bitrates and PSNR for (a) *Tsukuba* and (b) *HouseOf* images

the other hand, the increase of the bitrate of the residual image bpp is not very significant. It increases from 4.04 bpp to 4.13 bpp on average. Figure 4.4 shows the quality of the reconstructed right image $\hat{I}_R$ using the original dense disparity map and the optimal quantized version. The performance in both cases is presented in terms of PSNR having different bitrates. It can be noticed that the two curves follow almost the same trajectory, and the price to pay with the optimal quantized disparity map is about ± 0.5 dB even for low bitrates.

4.4 Conclusion

In this chapter, an optical flow based disparity estimation method for stereo images has been presented. For estimation of the disparity map, $TV - L_1$ based approach has been used for the reason that it preserves the edges in disparity map and is quite robust to outliers and noise. Although the dense disparity map is not suitable for direct coding, it offers flexibility to be encoded such that the total bitrate of the disparity map and the residual image is

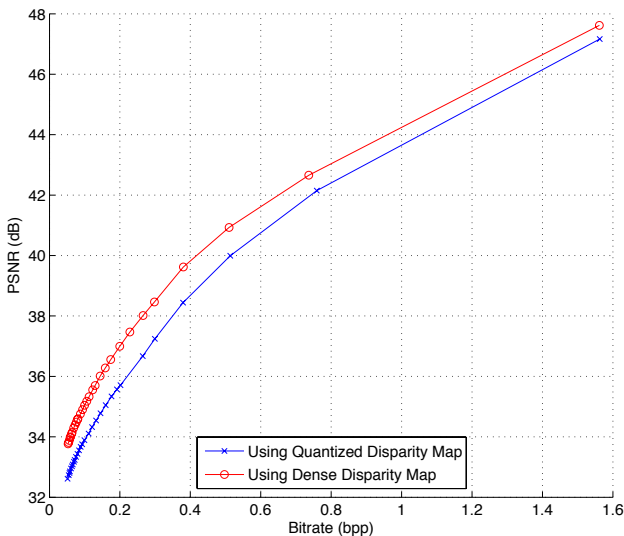


Figure 4.4: PSNR vs. Bitrate of the reconstructed right image using the *dense* and the *quantized* disparity map

Table 4.1: Results of optimization of dense disparity map

Image	Bitrates before optimization			Bitrates after optimization		
	Disparity	Residue	Total	Disparity	Residue	Total
Tsukuba	0.35	3.51	3.86	0.11	3.59	3.70
Fruit	0.49	4.05	4.54	0.06	4.10	4.16
Pentagon	0.64	5.11	5.75	0.06	5.21	5.27
Apple	0.87	4.54	5.41	0.10	4.72	4.82
HouseOf	0.71	5.36	6.07	0.08	5.45	5.53
Corridor	0.28	2.05	2.33	0.06	2.09	2.15
Pm	0.10	3.27	3.37	0.05	3.27	3.32
Birch	0.97	4.56	5.53	0.05	4.72	4.77
Shrub	0.10	3.27	3.37	0.05	3.27	3.32
Average	0.56	4.04	4.59	0.07	4.13	4.20

minimized. In addition, the proposed method offers an empirical framework for optimal quantization of the disparity map without any prior segmentation. So, the disparity image is compressed as a normal image and is compatible with JPEG2000 environment. Results in terms of bpp show an improvement of about 0.4 bpp on average. One of the perspectives is to design an analytical model in order to find the optimal quantization rather than doing in empirically.

Stereo Lifting Scheme

Contents

5.1	Introduction	43
5.2	Proposed Lifting-based structure	44
5.3	Joint Lifting Structure with Optimized Prediction	45
5.4	Experimental results	54
5.5	Conclusion	61

5.1 Introduction

In the first part of this chapter, a lifting structure is used to predict the left compensated image $\tilde{I}_L$ ¹ from the right image I_R by using an optimized filter. Aiming to minimize the predicted error variance, as done in [Boulgouris 2001] and [Gouze 2004] for intra prediction, we design an optimized 1D and 2D filter for inter prediction. A local spatial prediction is incorporated after the disparity compensation. The prediction is locally adapted and is based on local orientation information of disparity compensated left image $\tilde{I}_L$. Preliminary results of this work were first reported in [Darazi 2009c] [Darazi 2009b]. An other novelty in this work is the use of the directional information provided by the disparity map to reduce the prediction error d which represents the transformed residue. The decision of the prediction can be recovered at the reconstruction, without the need to transmit any overhead information.

In the second part of this chapter, the transformed residue d is decomposed by applying an adaptive scenario using LS. Two transform schemes are proposed. The first consists in a *Directional LS* where in each lifting stage, the prediction operation are carried out in the direction of image edges and textures. Whereas in the second scheme, a *Hybrid LS* is performed. Depending on each pixel, a choice of applying either Haar or 5/3 transform is carried out. The proposed schemes are not far to be compatible with the JPEG2000 framework. It is a lossy-to-lossless scheme, which can be used for digital cinema where we can

¹We recall that $\tilde{I}_L$ refers to the disparity compensated left image with ICC

tolerate lossy coding, or it can be adopted in medical imaging with lossless coding where the reconstruction of minor information is required.

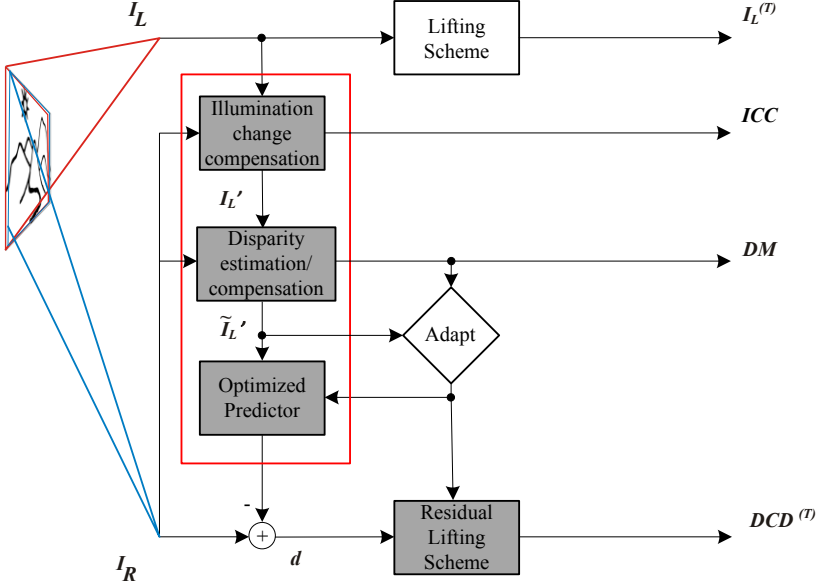


Figure 5.1: Lifting structure in a block diagram

The block diagram of the proposed scheme is presented in Figure 5.1 and consists of four main blocks: *Illumination Change Compensation*, *Disparity Estimation and Compensation*, *Optimized Filter*, *Residual Lifting Scheme*. As the first two blocks were presented respectively in chapter 3 and chapter 4, this chapter is devoted to the last two blocks.

5.2 Proposed Lifting-based structure

The main idea of this work is to efficiently exploit the existing correlation between the two images. Since the two images in the stereo pair are slightly different, exploiting joint information enables high efficiency in compression. Regardless of the DE/DC process, a lifting structure is used to predict the compensated left image $\tilde{I}_L$ from the right image I_R by using an optimized filter. Aiming to minimize the predicted residual image, as done in [Boulgouris 2001] and [Gouze 2004] for intra prediction, we design an optimized 1D and 2D filter for inter prediction. A local spatial prediction is incorporated after the disparity

compensation and is not only limited to high frequency correlation between the two images.

In the first step of the basic *LS*, the input image is split into two parts, the even and odd samples. In the proposed scheme, we omit this step and assume that the stereo pair is assimilated to two sub images as they represent the same scene. Thus, joint information from I_L and I_R are involved to compute the detail coefficients d . This is illustrated in Figure 5.2.

The second step corresponds to the prediction lifting step. Note that now the predictor P incorporates: Disparity Estimation and Compensation, and Optimized Filter as presented in Figure 5.1. In this paper we focus on the filtering part.

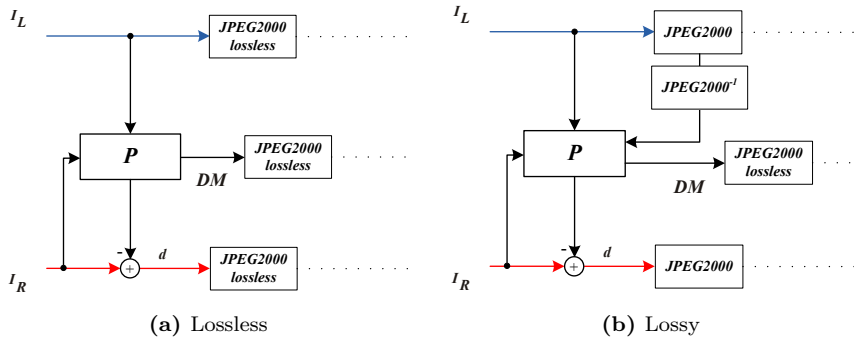


Figure 5.2: Stereo Lifting Scheme for (a) Lossless, and (b) Lossy coding

5.3 Joint Lifting Structure with Optimized Prediction

In dependent stereo image coding, the correlation in the stereo pair has been exploited in order to reduce the redundancies between the two views. Early work has been focusing mostly on Disparity Estimation and Disparity Compensation (DE/DC). The disparity compensated difference *DCD* referred also as the residual image is computed as:

$$I_e = I_R - \tilde{I}_L \quad (5.1)$$

The residual image contains impulsive components, this is due to occluded regions and mismatched disparity pixels near edges. As a consequence, two special properties can be assigned to the residual image. First, the majority

of the content has intensity values near zero. Second, the remaining contents are narrow vertical edges. These are the consequence of strictly horizontal apparent disparity between the two views. Accordingly, an optimized predictor is designed and applied to the compensated $\tilde{I}_L$ in order to predict the right image I_R . Joint information from the left and right image are involved to compute the detail coefficients d in analogy to I_e as shown in equation 5.2.

$$d = I_R - P * \tilde{I}_L \quad (5.2)$$

In the following section, we describe the optimization criterion and the procedure to define the coefficients of the predictor. Equation 5.3 shows that, the prediction of a pixel of I_R involves corresponding pixel in $\tilde{I}_L$ and its neighboring pixels in the horizontal direction.

$$d[m, n] = I_R[m, n] - \sum_{k \in \mathcal{P}} p_k \cdot \tilde{I}_L[m, n - k] \quad (5.3)$$

$$\text{with } \mathcal{P} = \left\{ k \mid -\frac{M-1}{2} \leq k \leq \frac{M-1}{2} \right\}$$

The choice of the horizontal direction relies on the fact that only horizontal disparity compensation is applied on I_L . Intuitively, using such a predictor allows a good tradeoff between prediction of well-matched pixel (during *DC*) and bad ones. Other directions for prediction are also investigated and used for comparison. These predictors are defined by

$$P_v * \tilde{I}_L = \sum_{k \in \mathcal{P}} p_k \cdot \tilde{I}_L[m - k, n] \quad (5.4)$$

$$P_{d_1} * \tilde{I}_L = \sum_{k \in \mathcal{P}} p_k \cdot \tilde{I}_L[m - k, n + k] \quad (5.5)$$

$$P_{d_2} * \tilde{I}_L = \sum_{k \in \mathcal{P}} p_k \cdot \tilde{I}_L[m - k, n - k] \quad (5.6)$$

where v , d_1 and d_2 stands for 90° , 45° and 135° respectively. In sufficiently smooth region, filters with larger support produce better prection but fail near edges or in textured areas. A good way to increase the support size without being discriminated by edges, is to extends the filter to two dimensions leading to the following predictor:

$$P_{2D} * \tilde{I}_L = \sum_{k \in \mathcal{P}} \sum_{l \in \mathcal{P}} p_{k,l} \cdot \tilde{I}_L[m - k, n - l]$$

Unlike one dimensional filters, this filter aims to exploit the correlation in all the neighboring directions.

5.3.1 Linear Optimized Inter Predictor

In predictive stereo coding frameworks, the total bitrate is given by:

$$R_{total} = R_L + R_e + R_{DM} \quad (5.7)$$

Since R_L depends only on the performance of the intra coder and is coded independently, it is not concerned by the choice of the predictor. Likewise, the optimal coefficients are independent of the term R_{DM} . Thus, given a disparity compensated image using a DE/DC algorithm, the optimal predictor should be provided such that:

$$P^* = \arg \min_P R_e \quad (5.8)$$

We have to find a way to solve problem in equation 5.8. We know that an image can be considered as a discrete source of symbols with a probability distribution given by its normalized histogram. The entropy $\mathcal{H}$ of such a source characterizes the number of bits required to encode it. Assuming that the residual image having the lowest entropy $\mathcal{H}_e$ produces the lowest bitrate R_e after being encoded with a given intra coder, we can reformulate the problem defined in 5.8 as:

$$P^* = \arg \min_P \mathcal{H}_e \quad (5.9)$$

Even if this problem is independent of the intra coder, there is still no exact analytical relationship between coefficient values and entropy of residual image. However, when predictive coding is used in image compression, the residual image histogram is generally well modeled by a generalized Gaussian distribution. The entropy of such a distribution is a monotonic function of its variance. Hence, a good criterion is to choose the predictor coefficient such as to minimize the variance of the residual image.

According to the coding theory, a signal with minimum variance is coded with the highest efficiency. A good prediction filter is the one that generates a signal of coefficients with suitable properties for compression. Hence, a reliable criterion is to minimize the variance of the residual image [Gouze 2004].

Theoretical analysis:

Our aim is to minimize the variance of the residual signal d . In the following, $y[n]$ denotes each pixel of I_R . $x[n - k]$ is the corresponding pixel in $\tilde{I}_L$ and its neighbors. And h_k is the k^{th} coefficient of the defined filter.

$$d[n] = y[n] - \sum_{k=1}^{M-1} h_k x[n - k] \quad (5.10)$$

$$\sigma_d^2 = E[d^2] - (E[d])^2 \quad (5.11)$$

$$\begin{aligned} \sigma_d^2 &= E[d^2] - (E[d])^2 \\ &= E \left[\left(y[n] - \sum_{k=1}^{M-1} h_k x[n-k] \right)^2 \right] - \left(E \left[y[n] - \sum_{k=1}^{M-1} h_k x[n-k] \right] \right)^2 \end{aligned}$$

Consider the first term as $f(\underline{h})$:

$$f(\underline{h}) = E \left[\left(y[n] - \sum_{k=1}^{M-1} h_k x[n-k] \right)^2 \right] \quad (5.12)$$

and the second term as $g(\underline{h})$:

$$g(\underline{h}) = - \left(E \left[y[n] - \sum_{k=1}^{M-1} h_k x[n-k] \right] \right)^2 \quad (5.13)$$

$$\sigma_d^2 = f(\underline{h}) + g(\underline{h}) \quad (5.14)$$

$$\begin{aligned} g(\underline{h}) &= - \left[(E[y[n]]) - E \left(\sum_{k=1}^{M-1} h_k x[n-k] \right) \right]^2 \\ &= - \left[\mu_y - \sum_{k=1}^{M-1} h_k E(x[n-k]) \right]^2 \\ &= - \left[\mu_y[n] - \sum_{k=1}^{M-1} h_k \mu_x[k] \right]^2 \end{aligned} \quad (5.15)$$

where,

$$E(y[n]) = \mu_y = \frac{1}{N} \sum_{n=0}^{N-1} y[n] \quad \text{and} \quad E(x_n[k]) = \mu_x[k] = \frac{1}{N} \sum_{n=0}^{N-1} x_k[n]$$

$$g(\underline{h}) = - \left[\mu_y - \sum_{k=1}^{M-1} h_k \mu_x[k] \right]^2 = -\mu_d = 0 \quad (5.16)$$

Then the optimization problem is:

$$\begin{aligned}
 \underline{h}_{opt} &= \arg \min_{\underline{h}} \sigma_d^2 \text{ such that } \mu_d = 0 \\
 &= \arg \min_{\underline{h}} (f(\underline{h}) + g(\underline{h})) \\
 &= \arg \min_{\underline{h}} (f(\underline{h})) \text{ s.t. } g(\underline{h}) = 0
 \end{aligned} \tag{5.17}$$

Then, we redefine:

$$\underline{h}_{opt} = \arg \min_{\underline{h}} E \left[\underbrace{\left(y[n] - \sum_{k=1}^{M-1} h_k x[n-k] \right)^2}_{\text{}} \right] \tag{5.18}$$

$$\text{s.t. } \mu_y = \underbrace{\sum_{k=1}^{M-1} h_k \mu_x[k]}_{\text{}} \tag{5.19}$$

This is equivalent to

$$\underline{h}_{opt} = \arg \min_{\underline{h}} \frac{1}{N} \sum_{n=0}^{N-1} \left(y[n] - \sum_{k=1}^{M-1} h_k x[n-k] \right)^2 \tag{5.20}$$

$$\text{s.t. } \mu_y = \sum_{k=1}^{M-1} h_k \mu_x[k] \tag{5.21}$$

This is equivalent of using the Least-Squared Method with the constraint $\mu_y = \sum_{k=1}^{M-1} h_k \mu_x[k]$.

As a consequence:

$$\underline{h}_{opt} = \arg \min_{\underline{h}} \frac{1}{N} \|Y - X \cdot H\|^2 \text{ s.t. } \mu_y = \underline{\mu}_x^T \cdot H \tag{5.22}$$

with $Y = [y_1, y_2, \dots, y_N]^T$, and $X = [X_1, X_2, \dots, X_N]^T$ where $X_i = [x_{ij}, j \in N(i)]$, and $H = [h_1, h_2, \dots, h_k]^T$.

Since the objective function is convex linear and the constraint is also linear w.r.t h_k , it can be solved by linear programming with constraints.

5.3.2 Non Linear Directional Inter Predictor

In the proposed lifting-based method, directional inter prediction has been investigated. The signal is adapted by exploiting local orientation information.

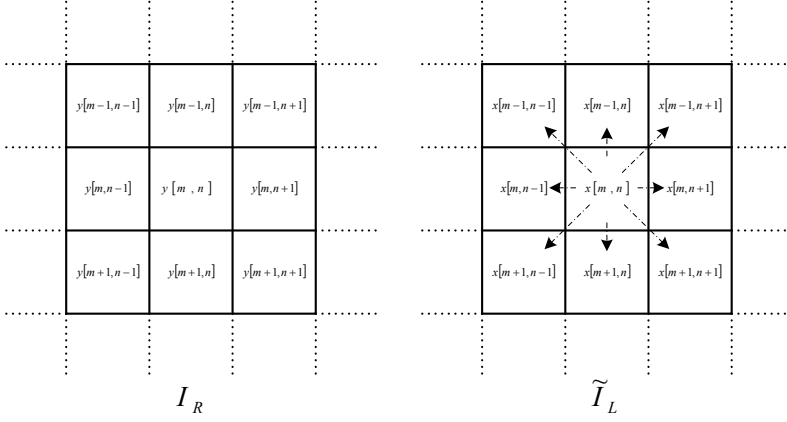


Figure 5.3: Directional Inter Prediction

In order to discriminate between different geometrical information, the system makes use of a criteria, denoting to the optimal way of prediction for each pixel. This is illustrated in Figure 5.3.

Optimal 2D filter can be seen as globally directional. In the sense that, for a given pixel, corresponding coefficients to the best correlated direction are given more weight than others. However, these coefficients can see their weight decreasing at other pixel location where best correlated neighbors occur in different direction. Directional predictors are non-linear filters that try to use the best direction of prediction at each pixel location. So they locally exploit only the best direction. Pixels in other directions may still have some good correlation but are not considered.

The problem of building such a predictor can be split into two subproblems. First, the need of defining a good criterion for the selection of direction. Second, the actual predictor have to be designed in order to behave according to the selected direction. This problem separation allows to focus on one or both part and to group many directional filters under a common framework.

We set up the condition under which the decision can be recovered at the reconstruction, without the need to transmit any overhead information. The non-linear predictor is defined as follows:

$$P_{hv} * \tilde{I}_L = \begin{cases} \sum_{k \in \mathcal{P}} p_k \tilde{I}_L(m-k, n) & \text{if } \nabla_x \tilde{I}_L(m, n) \leq \nabla_y \tilde{I}_L(m, n) \\ \sum_{k \in \mathcal{P}} p_k \tilde{I}_L(m, n-k) & \text{otherwise} \end{cases}$$

where $\nabla_x \tilde{I}_L(m, n)$ and $\nabla_y \tilde{I}_L(m, n)$ denote the partial derivatives of image $\tilde{I}_L$ at pixel location (m, n) respectively in the horizontal and vertical directions.

In other word, if the luminance variation rate change in horizontal direction is greater than the rate of change in the vertical direction, then prediction is done vertically. Otherwise it is done horizontally. This rely on the fact that smaller rate of change implies higher correlation. Since decision is based only on the value of $\tilde{I}_L$, it can be recovered at the decoder without any overhead transmission.

It is possible that for some pixels, the best correlated direction is neither horizontal nor vertical. Thus, these filters are extended so as they can choose between four main directions. The direction of prediction filtering is triggered by a decision obtained by computing the partial derivative of each pixel in the horizontal, vertical and diagonal directions respectively $\nabla_h I(x, y)$, $\nabla_{d_1} I(x, y)$, $\nabla_v I(x, y)$ and $\nabla_{d_2} I(x, y)$ and corresponding to $0^\circ, 90^\circ, 45^\circ$ and 135° . The extended predictor is defined as:

$$P_{4d} * \tilde{I}_L = \begin{cases} \sum_{k \in \mathcal{P}} p_k I(m-k, n) & \text{if } \nabla_* \tilde{I}_L(m, n) = \nabla_h \tilde{I}_L(m, n) \\ \sum_{k \in \mathcal{P}} p_k I(m-k, n+k) & \text{if } \nabla_* \tilde{I}_L(m, n) = \nabla_{d_1} \tilde{I}_L(m, n) \\ \sum_{k \in \mathcal{P}} p_k I(m, n-k) & \text{if } \nabla_* \tilde{I}_L(m, n) = \nabla_v \tilde{I}_L(m, n) \\ \sum_{k \in \mathcal{P}} p_k I(m-k, n-k) & \text{otherwise} \end{cases}$$

where

$$\nabla_* \tilde{I}_L(m, n) = \min \left\{ \nabla_h \tilde{I}_L(m, n), \nabla_{d_1} \tilde{I}_L(m, n), \nabla_v \tilde{I}_L(m, n), \nabla_{d_2} \tilde{I}_L(m, n) \right\}$$

Moreover, and in order to highlight the improvement that can be induced due to the Directional information, we introduce four additional directions corresponding to $30^\circ, 60^\circ, 120^\circ$ and 150° . Instead of providing formal definition for the 8-directional extension of predictors P_{4d} , we represent in Figure 5.4 the convolution mask of linear filter, for $M = 3$, corresponding to these four new directions. Formal definition can be directly deduced from previously defined 4-directional predictors. Note that these convolution masks are defined by five coefficients. Other directions only require three coefficients for the same value of $M = 3$. Thus, if we want to use a single set of coefficients common for all directions, we have to impose two relations between coefficient in the masks. To be coherent with the angle definition, we decide to impose $p_{-2} = p_{-1} = 0.5p_{-1}^*$ and $p_2 = p_1 = 0.5p_1^*$ so that we can use the set $\mathbf{p}^*$ for all directions. This section discusses the adaptation of the LS for the obtained residual image by using the disparity map. Two adaptation methods are described: *Hybrid Lifting Scheme* and *Directional Lifting Scheme*. In Hybrid LS, the adaptation is related to the length of the filter used in the prediction and update steps. This choice is tuned by a edge map generated from the disparity map. As for the Directional LS, the adaptation is related to the choice of prediction direction or neighborhood. This is done with the help of the orientation map, that is also computed from the disparity map.

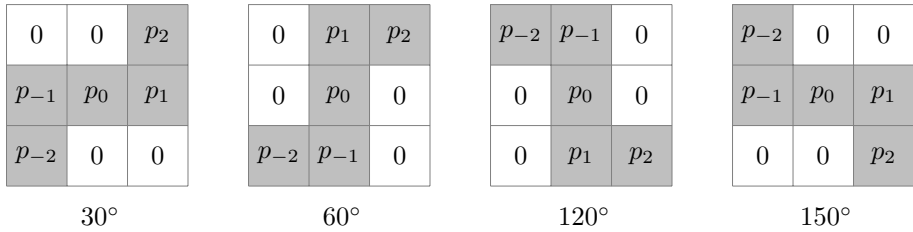


Figure 5.4: Convolution mask for linear filtering directions of 30°, 60°, 120° and 150°. Illustration for filter length of $M = 3$.

5.3.3 Hybrid LS and Edge Map

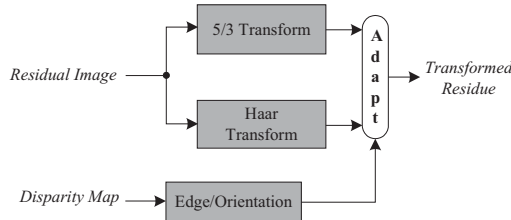


Figure 5.5: Residual Hybrid Lifting Scheme

As previously mentioned in the literature by Daubechies et al. [Daubechies 1998], there is no filter that performs better than all the others on all type of images. Wavelet filters with higher order perform well with natural images and smooth images, besides lower order filters perform better in presence of texture and edges content.

In this section, our aim is to locate high-energy edges in the residue that result from unmatched edges or occluded regions. Typically, the problem of imperfect matching occurs because of the existing textures in the image. Moreover, occluded regions are also reflected in residual image. Hence, the edge locations in residual image are relatively approximated by edges in disparity map. Since the Haar filter has to be applied on a neighborhood of the edge, the edge map is dilated. Otherwise, 5/3 filter is applied. This process is illustrated in Figure 5.5. One novelty of this idea is that the edge locations are determined without considering the residual image. The obvious question in mind is if the same edge map can be reconstructed at the decoder side or not. It should be noted that edge map has been constructed by logical ORing edges of disparity field and edges of disparity compensated image (after filtering). Since we already know the disparity map, we can exactly recover the edge locations for lossless scenario. Even for lossy circumstances, where the left image is altered we can obtain the exact edge locations. Thanks to the feedback mechanism that has been

integrated to the disparity map, the encoder has an *a priori* knowledge of the disparity map. Since we have performed the disparity compensation on the left image which is decoded after compression, this can be precisely accomplished at the decoder too.

5.3.4 Directional LS and Orientation map

One of the special properties of the residual image is the narrow vertical edges. These are the consequence of strictly horizontal apparent disparity between the two views. This assumption is true when we have perfect matching in the stereo pair. Besides, in some stereo pairs, other directional edges can be present. This may be related to the intrinsic content of the stereo image itself, consequently exploiting those particular characteristics in the obtained residual image by applying a special transform will reduce significantly the entropy and the bits-per-pixel measure.

In order to discriminate between different geometrical information, the system makes use of a criteria, denoting to the optimal way of prediction. The method computes the orientation map based on the partial derivatives of each pixel of the disparity map in the horizontal, vertical and diagonal directions (∇_{0° , ∇_{90° , ∇_{45° and ∇_{135°). The condition under which the decision is made can be recovered at the reconstruction, without the need to transmit any overhead information. The proposed method consists of selecting the direction that minimizes the partial derivative among 4 filtering directions for each pixel of the residual image. At the reconstruction, the prediction filter is applied in the same way, and this is done by inverting the operation.

Let us consider an horizontal processing of the residual image around a pixel at position $x[m, 2n + 1]$ as shown in Figure 5.6. The prediction filter inherently assumes that pixel $x[m, 2n + 1]$, the right neighboring pixel $x[m, 2n + 2]$, and left neighboring pixel $x[m, 2n]$ are closely related with pixel $x[m, 2n + 1]$. As a result, $\sum_{k \in \mathcal{P}} p_k \cdot x[m, 2n - 2k]$ with $\mathcal{P} = \{-1; 0\}$ will be an estimation of $x[m, 2n + 1]$. Hence, by subtracting this prediction value from the true value of $x[m, 2n + 1]$, a small residue is obtained. In the case of an object edge very close to the central pixel, the correlation between $x[m, 2n + 1]$ and $x[m, 2n]$ or $x[m, 2n + 2]$ may be reduced due to the sharp change. On the other hand, some of the four other immediate diagonal neighbors and $x[m + 1, 2n]$, $x[m - 1, 2n + 2]$, or $x[m + 1, 2n + 2]$ and $x[m - 1, 2n]$ may be more correlated with the pixel $x[m, 2n + 1]$. Therefore, it may be better to use two of these six neighbors in the prediction stage of the lifting structure in a judicious manner. Our directional adaptive predictor is obtained by relaxing the condition where the predictor should use samples from the current row in the horizontal direction that is being processed. Something to be mentioned is while predicting the even sample $x[m, 2n + 1]$ (after a column-wise sampling), we cannot use the vertical neighboring pixels (90° prediction).

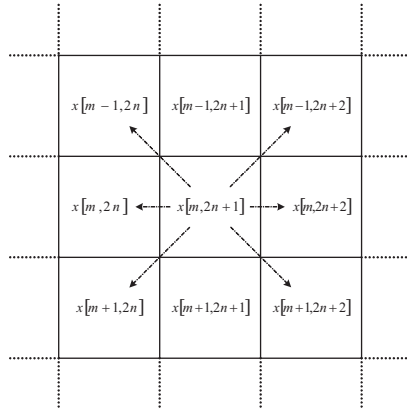


Figure 5.6: Residual Directional Prediction using Lifting Scheme

Because as shown in equation 2.4 the odd samples are predicted from the even samples. That is why, in case of vertical prediction, one way is to apply the haar transform, otherwise keep the horizontal prediction. The same process is done in a row-wise sampling for the horizontal direction (0° prediction).

5.4 Experimental results

Simulations have been carried out using two different stereo image datasets. The first dataset denoted $D1$, contains 38 stereo pairs from the well-known Middlebury stereo dataset. The resolution of the images is about 370×435 pixels, they represent scenes with complex objects and textures. The second dataset denoted $D2$ contains, 17 pairs which represent a number of well known stereo pairs (e.g: Pentagon, Fruit, Tsukuba ...). They are about 512×512 pixels resolution, they can be considered as more natural images compared to dataset $D1$. The hall stereo images can be downloaded at [car b, car a, mid].

For all the experiment, the filter's length parameter is fixed to $M = 3$ and the predefined coefficient values for non optimal one dimensional filters are given as $\mathbf{p} = (1/4 \ 1/2 \ 1/4)^T$. These coefficients are obtained simply by considering that p_0 should be given more weight than all others and that there is no reason, a priori, to give more weight to p_{-1} than to p_1 (and vice-versa).

We recall that in the following results, DCD represents the traditional residual image I_e . P_h, P_v, P_{d1}, P_{d2} represent the obtained transformed residue when applying respectively non-optimal linear filter in horizontal, vertical and diagonal directions. $*$ is used to recognize the optimized predictors.

5.4.1 Linear Optimized Predictor

Average results of image residual entropy $\mathcal{H}_e$ and bitrate $\mathcal{R}_e$ over all stereo pairs of datasets $\mathcal{D}_1$ and $\mathcal{D}_2$ are shown in table 5.1 and 5.2 respectively. Performance of some stereo pairs from both dataset are illustrated in table 5.3 (entropy) and table 5.4 (bitrate). Values are rounded to the second decimal place because more precision doesn't lead to significant difference. In those tables, the first row describes the predictor used for residual image computation. Optimal filters are denoted by the asterisk symbol. The performance of the proposed method are also evaluated using a lossy coder. Different predictors are compared in terms of PSNR in Figure 5.8, where the PSNR of I_R and reconstructed I_R is computed. Figure 5.9 represents in terms of SSIM metric [Wang 2004], the reconstructed I_R using coded I_e at different birates and DM .

	P_{DCD}	P_h	P_v	P_{d_1}	P_{d_2}	P_h^*	P_v^*	$P_{d_1}^*$	$P_{d_2}^*$	P_{2D}^*
$\mathcal{H}_e$	4.55	4.63	4.73	4.81	4.82	4.56	4.58	4.59	4.59	4.57
$\mathcal{R}_e$	3.68	3.66	3.75	3.86	3.86	3.59	3.64	3.68	3.67	3.59

Table 5.1: Average entropy and bitrate of residual image (in bpp) over dataset $\mathcal{D}_1$.

	P_{DCD}	P_h	P_v	P_{d_1}	P_{d_2}	P_h^*	P_v^*	$P_{d_1}^*$	$P_{d_2}^*$	P_{2D}^*
$\mathcal{H}_e$	5.10	5.07	5.07	5.10	5.10	5.06	5.03	5.04	5.04	5.00
$\mathcal{R}_e$	4.21	3.94	3.94	4.02	4.02	3.97	3.97	4.03	4.02	3.87

Table 5.2: Average entropy and bitrate of residual image (in bpp) over dataset $\mathcal{D}_2$

5.4.2 Non Linear Directional Predictor

Given a directional predictor with a predefined set of coefficient, it is easy to build a reference angle map $\mathcal{A}_0(m, n)$ using both right and disparity compensated left images. In order to highlight the performances while using directional predictors with perfect choice of direction, we consider the case of non optimal predictors. In this section we first define what is the optimal angle map in the particular case of non optimal filter. Then we find $\mathcal{A}_0(m, n)$ and try to find a way to compute an angle map $\mathcal{A}(m, n)$ as close as possible to $\mathcal{A}_0$ that can be recovered without side information at the decoder.

We propose to define the optimal angle map as

$$\mathcal{A}_0(m, n) = \underset{dir}{\operatorname{argmin}} |I_e(m, n)| \quad (5.23)$$

Stereo pair	DCD	P_h	P_v	P_{d_1}	P_{d_2}	P_h^*	P_v^*	$P_{d_1}^*$	$P_{d_2}^*$	P_{2D}^*
cones	4.93	5.11	5.21	5.36	5.36	4.95	4.99	4.99	4.99	4.97
sawtooth	4.32	4.48	4.67	4.80	4.81	4.31	4.24	4.30	4.31	4.22
teddy	4.51	4.59	4.71	4.83	4.80	4.51	4.56	4.56	4.56	4.53
mid2	4.06	4.08	4.23	4.27	4.29	4.06	4.07	4.08	4.07	4.05
moebius	4.63	4.71	4.71	4.82	4.83	4.62	4.64	4.65	4.64	4.63
lampshade2	4.32	4.31	4.34	4.35	4.35	4.34	4.36	4.36	4.36	4.34
plastic	3.93	3.94	3.97	3.99	3.98	3.98	3.99	3.99	3.98	3.97
reindeer	5.14	5.16	5.17	5.24	5.26	5.28	5.28	5.34	5.34	5.33
pentagon	5.45	5.31	5.30	5.36	5.38	5.31	5.30	5.33	5.34	5.26
houseof	5.55	5.44	5.45	5.49	5.50	5.45	5.45	5.44	5.46	5.39
fruit	5.07	4.98	4.98	4.97	4.97	4.96	5.11	5.12	5.10	5.09
book	5.66	5.64	5.64	5.64	5.63	5.70	5.70	5.70	5.70	5.69
schrub	4.43	4.38	4.48	4.50	4.47	4.03	3.95	3.96	3.96	3.91

Table 5.3: Entropy of residual image (in bpp) before intra coding.

Stereo pair	DCD	P_h	P_v	P_{d_1}	P_{d_2}	P_h^*	P_v^*	$P_{d_1}^*$	$P_{d_2}^*$	P_{2D}^*
cones	4.54	4.63	4.72	4.87	4.86	4.50	4.54	4.58	4.57	4.52
sawtooth	4.15	4.26	4.35	4.46	4.46	4.10	4.05	4.12	4.12	4.03
teddy	4.12	4.10	4.20	4.30	4.29	4.03	4.08	4.10	4.10	4.04
midd2	2.87	2.77	2.84	2.88	2.90	2.80	2.84	2.85	2.86	2.79
moebius	3.63	3.65	3.68	3.80	3.83	3.58	3.61	3.64	3.63	3.60
lampshade2	2.89	2.72	2.76	2.79	2.80	2.70	2.77	2.80	2.81	2.66
plastic	2.21	2.10	2.13	2.17	2.15	2.11	2.11	2.17	2.14	2.08
reindeer	4.01	3.77	3.82	3.92	3.94	3.77	3.83	3.92	3.92	3.74
pentagon	5.15	4.69	4.65	4.85	4.86	4.69	4.65	4.85	4.87	4.54
houseof	5.35	5.02	5.04	5.16	5.17	5.04	5.05	5.15	5.17	4.95
fruit	4.07	3.65	3.60	3.73	3.72	3.66	4.13	3.96	4.03	3.90
book	3.52	3.23	3.23	3.30	3.27	3.19	3.19	3.26	3.22	3.10
schrub	3.42	3.21	3.29	3.34	3.33	3.19	3.28	3.29	3.28	3.13

Table 5.4: Bitrate of residual image (in bpp) after JPEG2000 lossless encoding.

(a) DCD (b) P_h^* (c) P_{2D}^*

Figure 5.7: Absolute value of residual images obtained using (a) DCD and (b) optimal horizontal, and (c) optimal two-dimensional predictors for Schrub stereo pair.

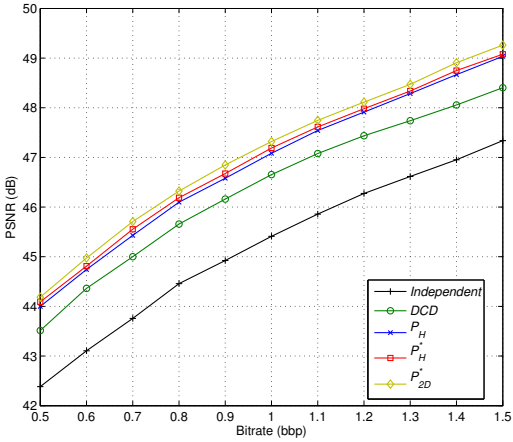


Figure 5.8: PSNR versus bitrate using lossy JPEG2000 encoder for Book stereo pair.

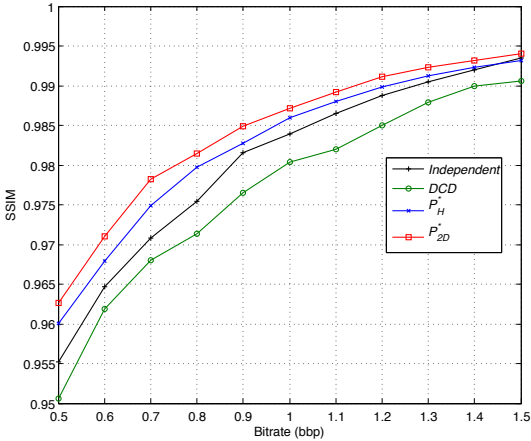


Figure 5.9: SSIM versus bitrate using lossy JPEG2000 encoder for I_e and DM of Pentagon stereo pair.

where we choose the direction dir that lead to the smallest residue $I_e(m, n)$. Having $\mathcal{A}_0$, we can compute the best achievable performances of the corresponding non optimal directional predictor. These performances are computed for predictors P_{hv} , P_{4d} and P_{8d} described in section 5.3.2 with $M = 3$ and $\mathbf{p} = (1/4 \quad 1/2 \quad 1/4)^T$.

Table 5.7 and Table 5.8 show the performance of non linear directional predictors using the criterion defined in 5.3.2 to select the direction of prediction. Whereas in Table 5.5, and Table 5.6 the performance of no linear directional predictors are presented using the optimal angle map. It can noticed that when directions are appropriately selected the bitrate is significantly reduced by 0.31 bpp and 0.46 bpp respectively for dataset $\mathcal{D}_1$ and dataset $\mathcal{D}_2$. This highlights the importance of using Directional prediction. How to improve the angle map to approach the optimal case is of great interest but is not part of the scope of this paper.

	P_{DCD}	P_h	P_h^*	P_{hv}	P_{4d}	P_{8d}	P_{2D}
$\mathcal{H}_e$	4.55	4.63	4.56	4.33	4.12	4.03	4.57
$\mathcal{R}_e$	3.68	3.66	3.59	3.53	3.41	3.37	3.59

Table 5.5: Average entropy and bitrate of residual image (in bpp) using a non linear directional predictor with optimal direction selection over dataset $\mathcal{D}_1$.

	P_{DCD}	P_h	P_h^*	P_{hv}	P_{4d}	P_{8d}	P_{2D}
$\mathcal{H}_e$	5.10	5.07	5.06	4.86	4.70	4.65	5.00
$\mathcal{R}_e$	4.21	3.94	3.97	3.84	3.77	3.75	3.87

Table 5.6: Average entropy and bitrate of residual image (in bpp) using a non linear directional predictor with optimal direction selection over dataset $\mathcal{D}_2$.

	P_h	P_h^*	P_{hv}	P_{hv}^*	P_{4d}	P_{4d}^*	P_{8d}	P_{8d}^*
$\mathcal{H}_e$	4.63	4.56	4.56	4.55	4.55	4.55	4.54	4.55
$\mathcal{R}_e$	3.66	3.59	3.66	3.64	3.66	3.65	3.64	3.65

Table 5.7: Average entropy and bitrate of residual image (in bpp) using a non linear directional predictor over dataset $\mathcal{D}_1$.

	P_h	P_h^*	P_{hv}	P_{hv}^*	P_{4d}	P_{4d}^*	P_{8d}	P_{8d}^*
$\mathcal{H}_e$	5.07	5.06	5.06	5.06	5.07	5.06	5.06	5.05
$\mathcal{R}_e$	3.94	3.97	4.04	4.07	4.08	4.10	4.07	4.08

Table 5.8: Average entropy and bitrate of residual image (in bpp) using a non linear directional predictor over dataset $\mathcal{D}_2$

5.4.3 Residual LS

Table 5.9 depicts entropies of 3-level transformed residual image in case of usual 5/3 transform, Haar transform, Directional LS, 5/3 transform with only 45° prediction, 5/3 transform with only 135° prediction and Hybrid transform. We can easily notice that Directional and Hybrid transforms outperform both the 5/3 transform and the Haar transform. This is expected because the Directional and Hybrid transform gets the best of both other transforms. The Hybrid transform shows better results than the Haar and 5/3 transforms. Whereas, the proposed Directional transform shows improvement in terms of entropy compared to traditional 5/3 and Haar transform and outperforms the hybrid transform. As the disparity map is coded in the same way for all the tested methods, the computation of the average entropy of the images is a significant evaluation criterion. The entropy is computed and is independent of the performance of any coder. Figure 5.10 shows a 3-level decomposition of Pentagon residual image using the 5/3 transform and the Directional LS. We can see in the magnified figure that using directional transform exhibit less details coefficients in high frequency subbands.

Table 5.9: Results of adaptive Directional vs. Hybrid LS for transformed residue

Image	Entropy[bpp]					
	5/3 transform	Haar	Directional (45°, 135°)	5/3 transform Directional 45°	5/3 transform Directional 135°	Hybrid transform
Fruit	3.09	3.36	2.75	2.75	3.13	3.15
Pentagon	4.22	4.19	3.99	4.04	4.04	4.07
Tsukuba	2.60	2.53	2.53	2.53	2.61	2.52
Ball	3.14	3.16	3.20	3.17	3.15	3.18
Apple	3.55	3.51	3.51	3.48	3.54	3.46
HouseOf	4.21	4.31	4.08	4.12	4.18	4.24
Average	3.47	3.52	3.34	3.35	3.44	3.44

5.5 Conclusion

In this work, an adaptive scheme for stereo image coding is presented. A efficient transform aiming to minimize the stereo residue is progressively designed. An optimized filter has been incorporated in the predictive framework after

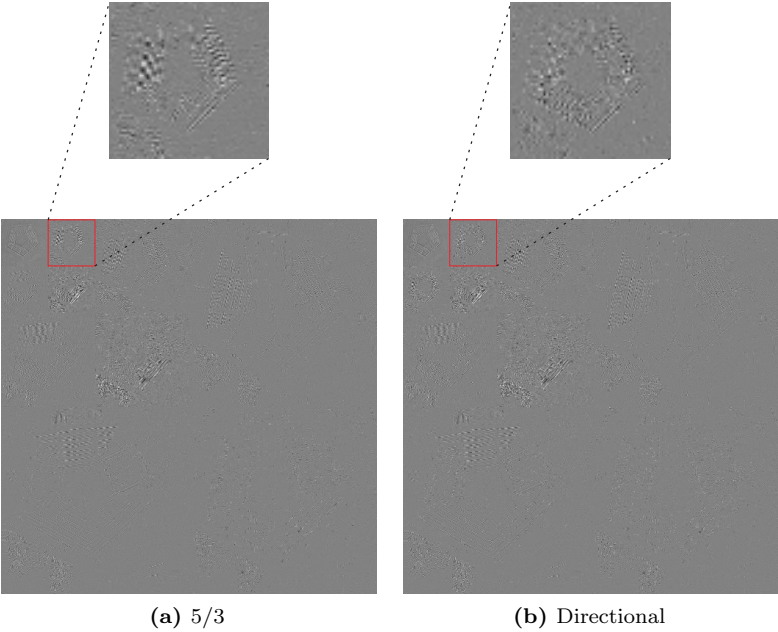


Figure 5.10: 3-level decomposition of Pentagon residual image: (a) $5/3$ -(b) *Directional Residual LS*

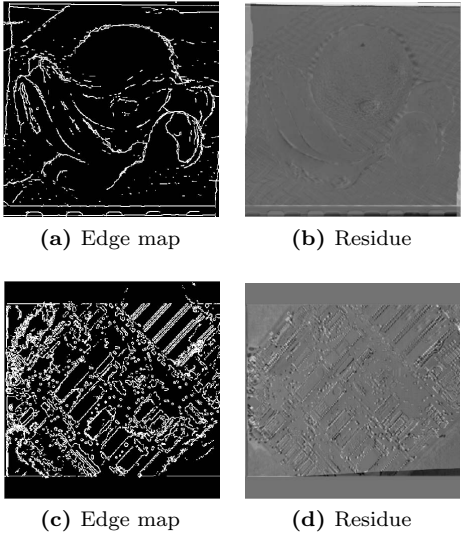


Figure 5.11: Approximation of edges of residue by disparity map and disparity compensated image: (a)-(b) *Fruit*, (c)-(d) *Apple*

disparity compensation. Results show the improvements in lossless and lossy cases compared to traditional predictive scheme and non-optimal filters. An adaptation of the prediction has been applied using directional filtering without requiring additional side information. Our simulations have shown that the selection of the direction of prediction for each pixel based on the compensated left image does not perform better than the optimized filters. However, we have proved that when the directions are more appropriately selected, directional predictors outperform all the other predictors. Finally, we have adapted the lifting scheme to the particularities of the residual image by considering filter length and the angle of prediction. The residual LS is performed because of the feedback of the disparity map which is also provided at the decoder.

Part II

**From 2D to 3D Stereo image
Watermarking**

Digital Watermarking

Contents

6.1	Origins and History: Information Hiding	68
6.2	Applications and Requirements	70
6.3	Watermarking Systems	71
6.4	Survey on QIM methods	76

The wide distribution of Digital Media is more and more suffering from the lack of effective content protection. Digital Watermarking is a widely proposed solution to identify and prove the ownership of digital data. It is mostly used to detect malicious tampering of a digital data and to track illegal copies of the copyrighted material. In this chapter, we give an overview on digital watermarking, the main properties, and the potential applications. We also present the related work to our contributions in the next chapter.

6.1 Origins and History: Information Hiding

The story of Watermarking is referred to the time of the ancient Greeks around 480 BC, when Histiaeus wanted to secretly notify the regent of the Greek city Miletus to start a revolt against the Persian occupier Darius I, king of Persia at that time. Histaeus chose to shave the head of a slave, tattoo the message on his skull, allow the hair to grow back, and finally dispatch the slave to Miletus. There the slave was shaved again to reveal the secret message.

Information Hiding is a general term gathering **Watermarking** and **Steganography**. As presented in the book of [Ingemar J. Cox 2008], the term *Hiding* can refer to either making the embedded information imperceptible (as in Watermarking), or keeping the embedded information secret (as in Steganography). The word *Watermark* has been introduced near the end of the eighteenth century. It may have been derived from the German term *wassermarke*, or the German word could be derived from English. The most important in this nomination, is that the term water is not important in the creation of the mark. It was affected to this nomination because the marks looks like the effects of water on paper.

As for the word *Steganography*, it was invented by Trithemius who is the author of the earliest publication on Cryptography. The term is derived from the Greek words, *Steganos* which means "covered", and *Graphia* which means "writing". Hence, Steganography is considered as the art of concealed communication.

The main difference between Watermarking and Steganography lies in the embedded information (message). If the embedded information is related to the cover work, then this corresponds to a watermarking system. Otherwise if the embedded information is unrelated to the cover work, and the very existence of the message is kept secret, consequently it is considered as a Steganographic system. We can consider the example of Histaeus to highlight the difference between Watermarking and Steganography. The skull of the slave is the cover work (CW). If the tattooed message on the slave's skull is something related to the slave e.g. "Personal property of Histaeus", then this is considered as a watermark. If someone else claims the possession of the slave, Histaeus can shave the slave's head and prove his ownership.

Information Hiding systems are summarized into four categories as shown in Table 6.1. The history of watermarking is including Invention progress and related people and events are illustrated in Figure 6.1.

Digital Watermarking is modernly defined as: *"the practice of imperceptibly altering a digital media in order to embed a message about that media."*

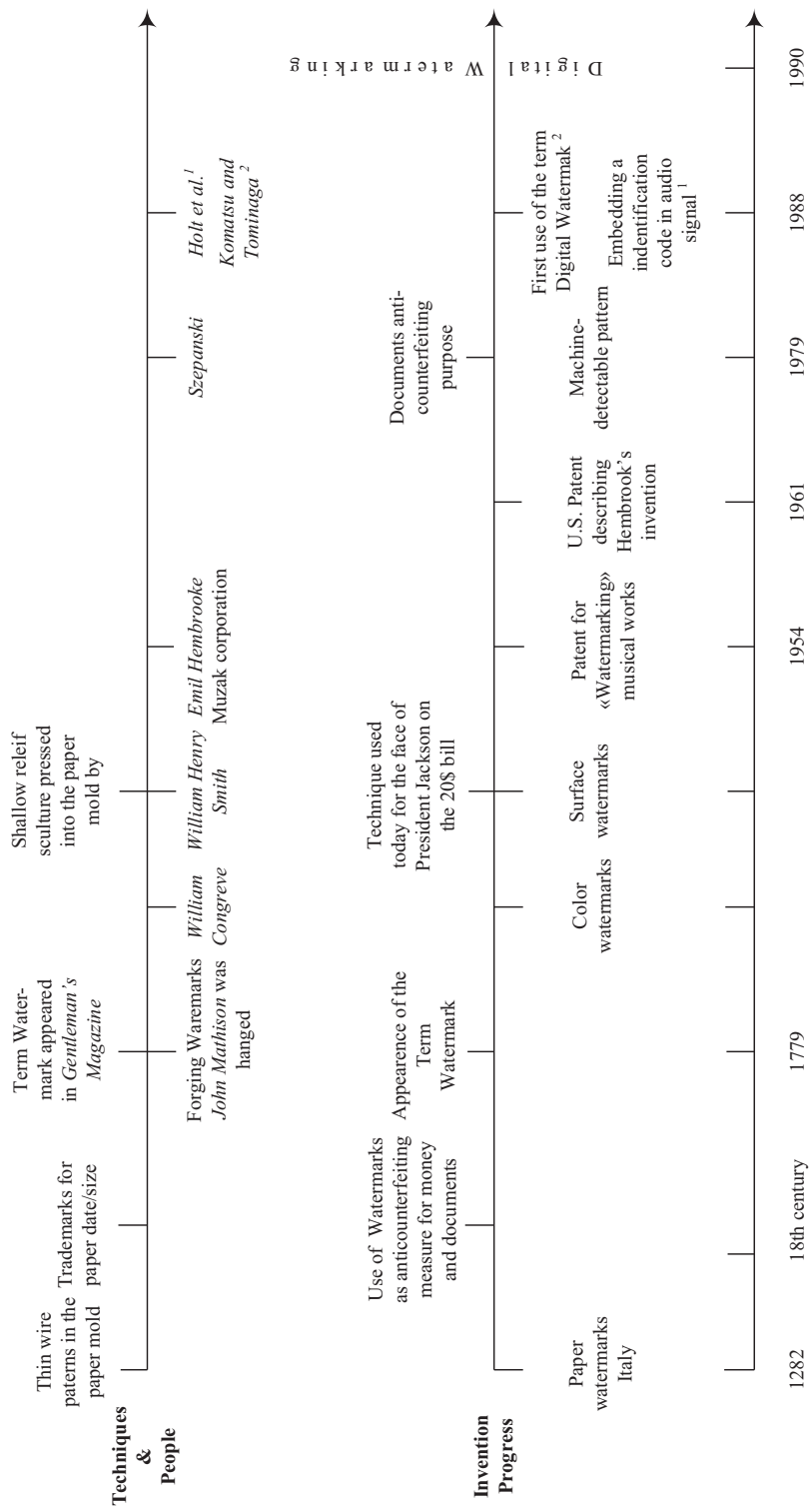


Figure 6.1: Watermarking history in a variable timeline.

Table 6.1: Categories of Information Hiding as presented in [Ingemar J. Cox 2008]

CW Dependent Message		CW Independent Message
Existence Hidden	Covert Watermarking	Steganography (Covert Steganography)
Existence Known	Overt Watermarking	Overt Embedded Communications

6.2 Applications and Requirements

6.2.1 Modern Applications

A number of applications have been derived in order to benefit from the advantages of Digital Watermarking. The include but are not limited to: Copy Control, Broadcast Monitoring, Copyright Protection/Owner Identification, Transaction Tracking/Fingerprinting and Content Authentication.

6.2.2 Requirements/Properties

Generally, a watermarking scheme must satisfy a number of constraints, most often having different aims. These constraints are interdependent and most often contradictory. This makes it difficult to study all aspects simultaneously or consider each one separately. Depending on the application type, a trade-off should be made regarding a number of constraints. In the following we present these constraints:

- **Fidelity/Imperceptibility/Transparency** This property refers to how much the original work is perceptually similar to the watermarked one. This is the most important property of a watermarking system.
- **Robustness** refers to the ability of the watermark to be retrieved even after modifications of the content. These modifications, also known as watermarking attacks, can be either malicious (performed with the intention of erasing the watermark) either not. There is no international consensus on which attacks should be taken into account for a given watermarking application and a given media type. Frequent attacks are the compression attack (JPEG, JPEG2000), the noise attack (AWNG), valuemetric scaling attack and also geometric distortion.
- **Security** Security refers to the fact that non authorized people (those who do not know the necessary secret key) should not detect nor read the

watermark. We distinguish between passive attacks when the adversary tries to read the message, while active attacks are when the attackers is able to make unauthorized embedding or removal.

- **Capacity/Payload** is referred as the number of bits inserted the cover work. Capacity also referred to as payload, is expressed in bits. The capacity needed for a watermarking scheme depends on the targeted application.
- **Complexity** Lower implementation complexity and cost can ba an added value to the watermarking scheme.
- **Real time** is often a requirement for digital video watermarking.

6.2.3 Evaluation- Effectiveness watermarking systems

The evaluation of any watermarking system mainly depends on the specified application. Before any evaluation, we need first to specify the important properties for the concerned application.

6.3 Watermarking Systems

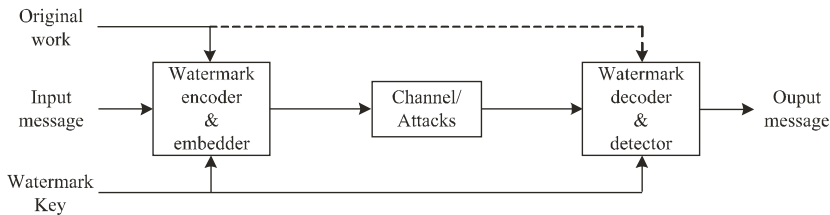


Figure 6.2: General Watermarking System.

Generally, a watermarking system consists of embedding a watermark m (input message), into a host signal x (original work) using an embedding function $e(x, m)$. As illustrated in Figure 6.2, the information to be inserted in the host signal x is encoded using a watermarking key k , in order to form the input binary message m , hence the embedding function is now $e(x, m, k)$. The watermarking key k is shared by the embedder and the detector of the watermarking system. $\hat{m}$ is the decoded output message after passing the channel where the signal is subject to different kind of attacks.

6.3.1 Watermarking Models: Informed/Blind

In this section, we distinguish between different watermarking models. This is mainly referred to the practice of using or not the properties of the host signal (cover work) at the watermark embedder or/and at the watermark detector. This is considered regarding the type of the application and the constraints related to it. Note that, in the literature, the term Informed/Blind watermarking is referred to Informed/Blind Detection. It is more appropriate to refer to Informed/Blind detection in order to prevent the confusion with Informed/Blind embedding.

A. Informed/Blind Embedding

We refer to a watermarking scheme with a blind embedder, when the embedding of the watermark is done independently from the original work. On the other hand, an informed embedding is referred to a watermarking system where the embedding is done by using a prior knowledge about the original work. In Figure 6.3, the dashed line is omitted in the case of a blind embedder.

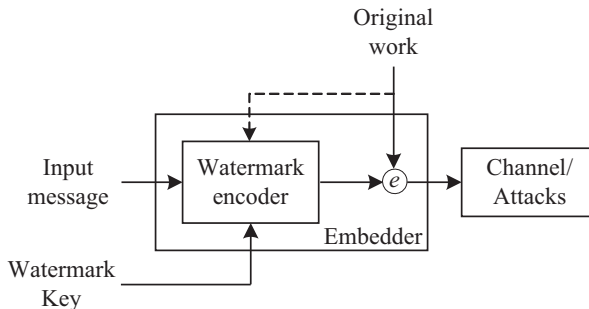


Figure 6.3: Watermarking system with an Informed/Blind Embedder.

B. Informed/Blind Detection

In analogy to the embedder, a watermarking system with a informed detector necessitate the use of the original work or part of it in order to detect the output message. Alternatively, in the case of a Blind detector, the detection process of the output does not require the presence of the original work. Note that whether a watermark system uses an Informed/Blind detector can be determining if it can be used for a given application.

Watermarking systems with Informed detection improves the performance of the detector and are good for registration to remedy against geometrical

and temporal distortions. However, for public systems like Fingerprinting/Transaction tracking for Digital Cinema, distributing the original work to all consumers is not only impractical but also loses the main purpose of the watermarking system. A watermarking system with an informed embedder and a blind detector is illustrated in Figure 6.4.

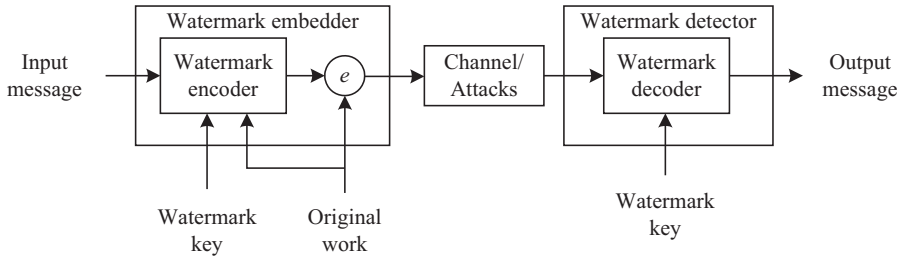


Figure 6.4: Watermarking system with an Informed embedder and a Blind detector.

6.3.2 Watermarking Media Space/ Marking Space

In a watermarking system, the embedding and the detection may occur in the original work, which is referred in the literature as the Media space. However, in advanced systems, the embedding/detection occurs in a transformation or projection of the original work, referred to as the Marking space. Note that the choice of the watermarking space can significantly affect the performance of the watermarking system not only in terms of robustness but also in terms of cost and complexity.

Focusing on image watermarking, information embedding can be done in spatial or in some transform domain. Spatial-domain methods apply the watermark directly to visual data of the content. The watermark is embedded by modifying directly the pixel values or the intensity values of the luminance of the original image [Nikolaidis 1998]. Transform-domain methods embed the watermark into image coefficients of a properly chosen linear transform or from some non-linear transform of the host signal. Usually these methods are called frequency-domain methods, where watermarking techniques are applied to coefficients obtained as the result of a frequency transform as the Discrete Cosine Transform (DCT) [Barni 1998], the Discrete Fourier Transform (DFT) [Pereira 2000] or the Discrete Wavelet Transform (DWT) [Xia 1998, Dugad 1998], and other types of transforms.

Table 6.2 highlights the difference between Spatial and Frequency marking performance *w.r.t.* some properties. The Fourier transform is the less used

Table 6.2: Spatial vs. Frequency domain watermarking

	Spatial Domain Marking	Frequency Domain Marking
Complexity/Cost	Low✓	High
Capacity	Low	High✓
Robustness	Low	High✓
Visual Models/Fidelity	Minor	More studied ✓
Security	Lower	Higher ✓

frequency domain method because of its modest results in watermarking applications due to its complexity to preserve complex conjugate symmetry of the DFT coefficients and the lack of Human Visual System (HVS) models . The DCT domain is far the most popular one because of the familiarity with it in image processing and the fact that the major compression techniques were developed in this domain (JPEG, MPEG, H26x). Furthermore, much research was carried out in developing perceptual models for the DCT domain [Watson 1993] that could be easily applied to watermarking.

6.3.3 Watermarking Classes and techniques: SS/QIM

Evaluating watermarking techniques as previously evoked in subsection 6.2.3, is a very difficult task. This is related to the large variety of image watermarking methods in the literature. The same issue persists with the classification. However, a common way to classify watermarking algorithms is based on the embedding strategy: how host data is modified in order to embed the message. This classification leads us to distinguish between linear additive techniques and non-linear quantization techniques.

Linear additive techniques (additive class) are designed based on the premise that the host signal is like a source of noise or interference. These methods are considered as host-interference non-rejecting methods: the host signal x acts as an additive interference that inhibits the decoder’s ability to estimate m . Thus, these methods are useful mainly when either the host signal is available at the decoder (non-blind/informed detection schemes) or when the host signal interference is much smaller than the channel interference. These methods have purely additive embedding functions and are often referred to as spread spectrum (SS) methods.

Non-linear quantization techniques (substitutive class) are referred as host-interference rejecting methods. These methods exploit knowledge of the host

signal at the encoder, so they reject host interference. These methods use quantization embedding functions and are referred to Quantization Index Modulation (QIM) methods. In the following we present the canonical form of the two classes and therefore highlight the difference between the two embedding methods.

6.3.3.1 Host interference non-rejecting methods

The SS methods were introduced by Cox et al. [Cox 1997]. The additive embedding function is of the form:

$$s(x, m) = x + w(m) \quad (6.1)$$

Where $w(m)$ is typically a pseudo-noise sequence.

$$w(m) = a(m)v \quad (6.2)$$

Where v is a unit-energy spreading vector and $a(m)$ is a scalar function of the message. As presented in [Chen 2001], additive SS can be viewed as a perturbation of a projection. Substituting Equation 6.2 into Equation 6.1 and having v as a unit energy vector yields to:

$$s = x + a(m)v \quad (6.3)$$

While projecting s into v , we can obtain:

$$\tilde{s} = s^T v = \tilde{x} + a(m) \quad (6.4)$$

where $\tilde{x}$ is the projection of the host signal x onto v

$$\tilde{x} = x^T v \quad (6.5)$$

At the end, substituting Equation 6.4 into Equation 6.3 yields to the reconstructed signal from projection:

$$s = x + (\tilde{s} - \tilde{x})v \quad (6.6)$$

As shown in Equation 6.1, for this class of methods, the host signal acts as a source of interference to the message m and consequently inhibits the decoder to estimate $\hat{m}$. As assumed in [Cox 1997] this class of methods is useful when the host signal is present at the decoder or when the host signal interference is smaller than channel interference.

6.3.3.2 Host interference rejecting methods

Substitutive QIM methods were introduced by Chen and Wornell [Chen 2001]. The embedding function is of the form:

$$s(x, m) = q(x) + d(m) \quad (6.7)$$

where $q(\cdot)$ represents a coarse quantizer that determines the most significant bits, and $d(m)$ a perturbation value indexed by the message m determined only to modify the least significant bit. Thus the embedding is of the same form as in Equation 6.4, but with $\tilde{s}$:

$$\tilde{s} = s^T v = q(\tilde{x}) + d(m) \quad (6.8)$$

where $q(\cdot)$ is a uniform scalar quantization function of step size Δ and $d(m)$ is considered as a perturbation value.

A general characteristic of these systems is that the embedding never alters the most significant bits of the host signal which is expressed in terms of constraint

$$q(s) = q(x) \quad (6.9)$$

6.4 Survey on QIM methods

Known also as dither modulation codes, quantization methods have been initially derived from work in Information Theory done by Costa [Costa 1983], such as lattice structured codes and other quantization-based schemes. Later, simplest techniques proposed embedding by modifying the least significant bit(s), known as low-bit modulation (LBM) [Barton 1997, Swanson 1997]. In LBM systems, the least significant bit(s) in the binary representation of a prior quantized host sample is (are) simply replaced with message bits (quantize and replace strategy). A more advanced quantization-based technique introduced by [Chen 2001] allows more robust watermarking is the general class of host-interference rejecting methods. It was shown that these advanced quantization-based embedding functions result in better rate-distortion-robustness performance than that obtainable by LBM. In the following, we give an insight on QIM methods.

6.4.1 Quantization Index Modulation

As prior quantization schemes, QIM methods perform non-linear modifications, and detect the embedded message by quantizing the received samples, mapping them to the nearest reconstruction point.

The design of a QIM model starts by defining a number of embedding functions

$s(x, m)$ that are indexed by a message m . QIM concept is characterized by two canonical properties. The first property is to have a small embedding induced distortion (least perceptibility). Hence, each embedding function must be approximately equal to identity function as shown in Equation 6.10:

$$s(x, m) \approx x, \quad \forall m \quad (6.10)$$

This is referred to the approximate identity property. And the second one is the non intersection property (robustness) which means that each point in the range of one function should be far away from a point in the range of any other function. It is mainly due the non intersection property that QIM systems are qualified as host-signal interference rejection.

In order to introduce QIM, we consider the example of a host signal x and a message $m \in \{S_1, S_2\}$ to be embedded. Thus, two different quantizers are required. Hence, to embed $m = S_1$, we use the quantizers Q_1 , and to embed $m = S_2$, we quantize the host signal x using Q_2 . This is illustrated in Figure 6.5 where $\circ$ corresponds to the samples quantized by Q_1 and $\times$ corresponds to the samples quantized using Q_2 .

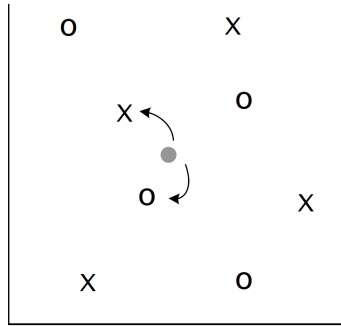


Figure 6.5: Quantization Index Modulation.

The approximation property of QIM requires that the host signal x and the watermarked signal s should be perceptually close. The embedding distortion of the host signal due to the watermark is measured quantitatively by the means of a distortion function between the host signal and the watermarked one. The most common distortion measure is the squared error / mean square error:

$$D(s, x) = \|s - x\|^2 \quad (6.11)$$

or

$$D(s, x) = \frac{1}{L} \|s - x\|^2 \quad (6.12)$$

where L is the number of the host signal samples.

On the other hand, the second canonical property is non intersection property. It is represented by the minimum distance between the sets of reconstruction points of different quantizers in the ensemble. This determines effectively the robustness of the embedding, measuring the size of the perturbation vectors that can be tolerated by the system. Thus, the minimum distance is:

$$d_{min} \triangleq \min_{(i,j): i \neq j} \min_{x_i, x_j} \|s(x_i; i) - s(x_j; j)\| \quad (6.13)$$

Hence, a significant measure relating the embedding-induced distortion in Equation 6.12 and the minimum distance between quantizers in Equation 6.13 will characterize the achievable performance of a particular QIM realization. This measure is defined as the distortion-normalized squared minimum distance:

$$d_{norm}^2 = \frac{d_{min}^2}{D_s} \quad (6.14)$$

In the following, we consider the signal y as the resulting watermarked signal, subject to various kind of attacks, it is represented as:

$$y = s + n \quad (6.15)$$

where n is a perturbation due an attacker.

Minimum distance decoder

Detection in QIM can be performed without access to the original data or the original watermark. The QIM decoder is a minimum-distance decoder. It finds the quantizer point closest to y and outputs the estimated message $\hat{m}$

$$\hat{m} = \arg \min_{0,1} |y - s(y, m)| \quad (6.16)$$

Note that error correction coding of the embedded message m is usually performed to achieve better decoding rates. In that case, the coded bit sequence is embedded as it is m , even though the coding rate and the coding gain must be considered in the measures used.

6.4.2 Scalar-QIM

A simple way to understand the basic idea of QIM is to consider the embedding of one bit message $m \in \{0, 1\}$ in a real valued sample x . A scalar, uniform

quantizer $Q_m(\cdot)$ is indexed by the message m , and is used to quantize the signal sample x . Hence the marked sample s is defined as:

$$s = Q_m(x) = Q(x, \Delta_m) = \lfloor (\frac{x}{\Delta_m}) \rfloor \Delta_m \quad m \in \{0, 1\} \quad (6.17)$$

Where Δ represents the quantization factor (also called quantization step size), and where $\lfloor \cdot \rfloor$ denotes a rounding operation.

6.4.3 Basic Dither Modulation with Scalar-QIM

Dither Modulation is used as a practical implementation of QIM. The advantage of such an implementation is that the quantization and reconstruction point of a given quantizer is a shifted version of the quantization and the reconstruction point of another quantizer.

In the same context of previous section subsection 6.4.2, to embed one bit message $m \in \{0, 1\}$ in a host signal x , we use the dithering to generate two quantizers based on the embedding information m . Thus, satisfying Equation 6.17.

$$s(x, m) = Q_m(x) = Q(x - d_m) + d_m, \quad m = \{0, 1\}$$

The dither value and, for instance, the quantizer used is modulated by m , so we have two dither values, d_0 and d_1 chosen usually

$$d_0 = -\frac{\Delta}{4}, \quad d_1 = \frac{\Delta}{4} \quad (6.18)$$

The condition to be satisfied is to choose dither signals such as the distance defined in Equation 6.13 between the two corresponding sets of reconstruction points of the quantizers is the maximum possible.

Another option of having a deterministic dither signal, is to choose d_0 pseudo-randomly from a uniform distribution over $[-\Delta/2, \Delta/2]$. Depending on the sign of d_0 , $\Delta/2$ is either added or subtracted from d_0 to form d_1 . This pseudo-random dither choose tends to give a perceptually better quantized output and makes the quantization noise produced independent of the quantizer input.

$$\text{If } d_0 < 0 \quad d_1 = d_0 + \Delta/2 \quad (6.19)$$

$$\text{If } d_0 \geq 0 \quad d_1 = d_0 - \Delta/2 \quad (6.20)$$

As shown in Figure 6.6 and in Figure 6.10, the maximum induced error due to embedding is $d_{min} = \Delta/2$.

If the quantization error $(Q_m(x) - x)$ is uniformly distributed over $[-\Delta/2, \Delta/2]$,

then the mean-squared distortion defined in Equation 6.11 and Equation 6.12 is of the form:

$$D(s, x) = \frac{\Delta^2}{12} \quad (6.21)$$

The minimum-distance decoder defined in Equation 6.16 estimates the message embedded finding the reconstruction point within the two sets of reconstruction points closest to the received signal y . Therefore, no decoding error is assured if $|n| < \Delta/4$.

To summarize, having two quantizers allows us to embed one bit of information taking into account that the bigger the step size used, the more robust is the embedded message against attacks whereas more distortion is induced by the embedding process.

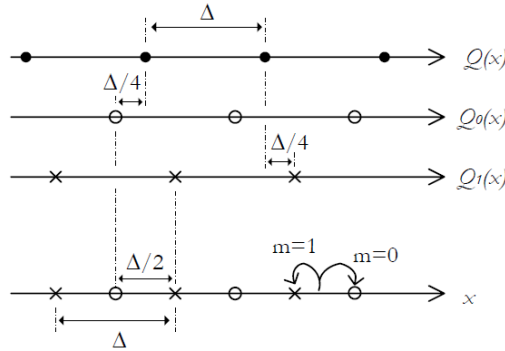


Figure 6.6: Quantization Index Modulation.

Generalized Dither Modulation with Scalar-QIM

From the simple example of embedding one bit message in a host signal sample x , more generalized case is presented, and where it is considered that the high dimensional base quantizer $Q(\cdot)$ is the Cartesian product of scalar quantizers. The host signal x is divided into N/L non-overlapping blocks/vectors of length L . Each i_{th} information bit is embedded into the i_{th} non-overlapping block/vector of length L . The rate can be therefore defined as:

$$R_m = 1/L \quad (6.22)$$

For this end, as done previously in subsection 6.4.3, two dither sequence $d[k,0]$ and $d[k,1]$ of length L and a sequence of uniform scalar quantizers $\Delta_{1,...,L}$ of

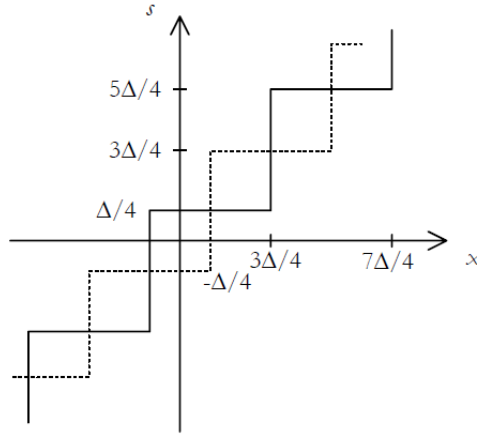


Figure 6.7: Marked sample s given sample x with $m \in \{0, 1\}$. The solid curve shows the embedding function for $m = 0$. The dashed curve shows the embedding function for $m = 1$

length L are created. such that the following constraint is respected:

$$d[k, 1] = \begin{cases} d[k, 0] + \Delta_k/2, & d[k, 0] < 0 \\ d[k, 0] - \Delta_k/2, & d[k, 0] \geq 0 \end{cases}, \quad k = 1, \dots, L \quad (6.23)$$

The aim of this constraint is to assure that the dithered quantizers are maximally distant from each other. This constraint ensures that the two corresponding L -dimensional dithered quantizers are the maximum possible distance from each other.

Usually, $d[k, 0]$ is chosen either from a pseudorandom sequence of $\pm\Delta_k/4$, or pseudorandomly with a uniform distribution over $[-\Delta_k/2, \Delta_k/2]$. Note that both ways satisfy the previous constraint in Equation 6.23. Also, the two dither sequences need not be the same for each length- L block. The i th block/vector is quantized with the corresponding dithered quantizer $d[k, i]$.

Hence, the minimum-distance decoder in Equation 6.16 used in detection can be seen, for a received altered marked-vector y :

$$\hat{m} = \arg \min_{0,1} \sum_{k=1}^L |y[k] - Q_m(y[k])| \quad (6.24)$$

The squared-minimum distance Equation 6.13 over all N dimensions is:

$$d_{min}^2 = \sum_{k=1}^L \left(\frac{\Delta_k}{2}\right)^2 = \frac{1}{4} \sum_k \Delta_k^2 \quad (6.25)$$

The expected quantization error or embedding distortion of a uniform, scalar quantizer with step size Δ_k is $\Delta_k^2/12$ as presented in Equation 6.21. Hence, the overall average embedding distortion Equation 6.11 is:

$$D_s = \frac{1}{12L} \sum_k \Delta_k^2 \quad (6.26)$$

6.4.4 Distortion-Compensated QIM

Distortion Compensation-QIM (DC-QIM) is sve the achievable rate-distortion-robustness tradeoffs of QIM methods. If we assume that for a given ensemble of quantizers, we apply a scaling factor $\alpha \leq 1$. A consequence of the previous is the increase of d_{min}^2 by a factor of $1/\alpha^2$, hence increasing the embedding robustness. However, the embedding distortion is increased by a factor of $1/\alpha_2$. Note that if a reconstruction point is at q , it is scaled by a factor α by moving it to q/α . In order to eliminate or compensate the increased distortion DC-QIM adds a fraction $(1 - \alpha)$ of the quantization error $x - Q(x, \Delta/\alpha)$ to the quantization value $Q(x, \Delta/\alpha)$. The resulting embedding function is:

$$s(x, m) = Q(x, \Delta/\alpha) + (1 - \alpha)(x - Q(x, \Delta/\alpha)) \quad (6.27)$$

The DC-compensation was rewritten by [Moulin 2005] as follow:

$$s(x, m) = Q(\alpha x) + (1 - \alpha)(x - Q(\alpha x)) \quad (6.28)$$

$$s(x, m) = Q(\alpha x) + (1 - \alpha)x \quad (6.29)$$

$$s(x, m) = x + (Q(\alpha x) - \alpha x) \quad (6.30)$$

However, we think that there is a slight difference between Equation 6.27 and Equation 6.28, in the compensation part. Moulin's expression is more suited because the compensation should be related to x not $(x - Q(\alpha x))$.

Reconstruction points of these quantizers are now separated by a distance Δ/α . However, adding the distortion compensation term expands these reconstruction points to reconstruction intervals of α -dependent width $(1 - \alpha)\Delta/\alpha$.

Observe in Figure 6.8 the reconstruction points resulted of the new embedding functions in Equation 6.27, the quantization error added back leads to a source of interference to the decoder. In most of the cases the probability density functions of the quantization error for all quantizers in the QIM ensemble are similar, so the distortion compensation term is nearly statistically independent of m and can be treated as independent noise in decoding. However, there is a trade-off between minimum distance and distortion-compensation interference at the decoder. Decreasing α leads to greater minimum distance, but for a

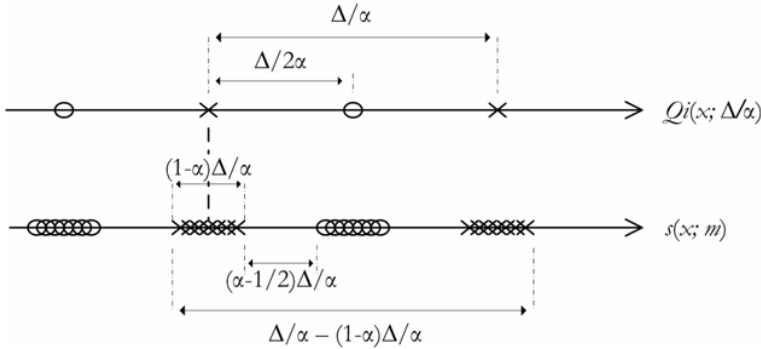


Figure 6.8: DC-QIM

fixed embedding-induced distortion, it increases the distortion-compensation interference.

Post-quantization distortion compensation may be combined with each of the QIM realizations.

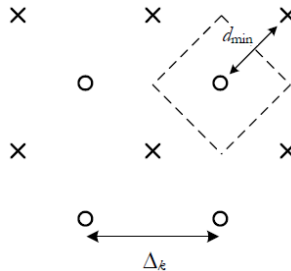


Figure 6.9: Binary dither modulation with uniform, scalar quantization. Reconstruction points of quantizers Q_1 and Q_0 are marked with $\circ$ and $\times$ respectively. The quantization step size Δ_k is the same for both dimensions.

QIM methods are generally fragile to valumetric scaling (e.g. brightness changes), and re-quantization (e.g. numerical rounding in lossy compression). Rational dither modulation was recently proposed by Perez-Gonzalez et al. [Perez-Gonzalez 2005] to overcome valuemetric scaling. On the other hand, spread transform dither modulation is the QIM variant which exhibits most robustness to noise addition re-quantization. It is described as follows and chosen for the proposed algorithm chapter 7. We investigate the possibility of improving STDm to withstand valumetric scaling. We refer to Unspread Dither Modulation (UDM) as the case of binary dither modulation with uniform, scalar

quantization where all the host signal components are quantized with the same step size. The term unspread refers to not projecting the host signal onto a spreading vector before quantization, as it is done in next method outlined in subsection 6.4.5.

6.4.5 Spread Transform Dither Modulation

STDM is a special case of QIM. It exhibits robustness to re-quantization and meanwhile maintains low complexity. Therefore, instead of quantizing the host signal x itself, the quantization occurs entirely in the projection of the host signal x onto a randomly generated vector p (unit length vector, unit energy vector). Consequently, the embedding of each information bit occurs in the projection of the host signal x by quantizing it with a uniform, scalar, dithered quantizer [Chen 2001].

Note that in this method, each bit of the watermark message is spread into n samples instead of one sample.

We can illustrate that by considering a host signal $x = (x_1, x_2, \dots, x_N)$ of length N . The whole signal x is divided into segments with equal length n , and one bit is embed into each segment. The embedded bit is spread in the projection of the segment onto a random vector p .

Thus, the watermarked signal is given by:

$$y = x + (Q_m(x^T p) - x^T p)p \quad m \in \{0, 1\} \quad (6.31)$$

The STDM decoder makes a decision based on the projection of the channel output y onto the spreading vector p . The detection can be performed with a minimum distance decoder as of the form:

$$\hat{m} = \arg \min_{\hat{m} \in \{0, 1\}} |y^T p - Q_{\hat{m}}(y^T p)| \quad (6.32)$$

Note that an additional specification of STDM compared to other QIM methods is the ability of mixing the advantage of an additive SS and a LBM to a STDM:

i.e. by replacing Equation 6.8 (LBM):

$$\tilde{s} = s^T v = q(\tilde{x}) + d(m)$$

into Equation 6.6 (SS):

$$s = x + (\tilde{s} - \tilde{x})v$$

gives equation of Equation 6.31 (STDM),

$$s = x + ([q(\tilde{x}) + d(m)] - \tilde{x})v \quad (6.33)$$

The demonstration performed by Chen et al. shows that for the same embedding-induced distortion, and for a reasonable output channel output. At the decoder for reconstruction, the squared difference, hence the SNR for additive spread spectrum and Generalized LBM are worst than STDM.

Figure 6.10 shows the sets of reconstruction points of two quantizers for embedding one bit in a vector of two host signal samples, where the projection of the host signal vector is done before quantization. The host signal is quantized to the nearest point on a $\circ$ -line to embed a 0-bit and on a $\times$ -line to embed a 1-bit.

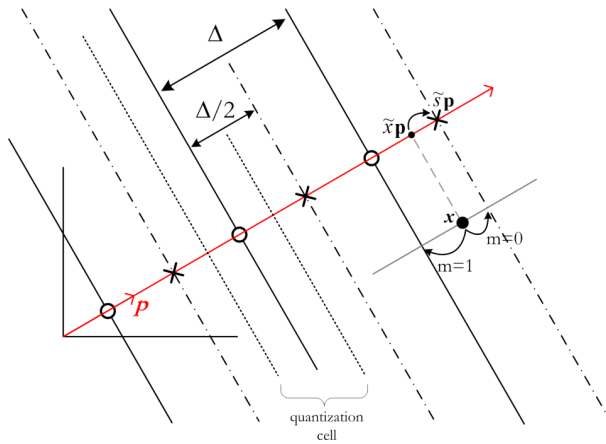


Figure 6.10: Geometrical representation of Spread Transform Dither Modulation with uniform, scalar quantizer of step size Δ . Points on solid-lines represent embedding for $m = 1$, whereas dashed-lines are for $m = 0$

As in these systems quantization occurs entirely in the direction of p , the minimum distance defined in Equation 6.25 between quantizers in the ensemble is $d_{min} = \Delta/2$. Note that distortion due to embedding, $(Q_m(\tilde{x}) - \tilde{x})p = (\tilde{s} - \tilde{x})p$, takes place in direction p only, though the embedder can allocate the entire distortion budget in direction p . This enables the use of a larger quantizer step size Δ , as the maximum quantization error incurred in each host signal component, $s_k - x_k$, is only a fraction of the quantization step, $p_k \Delta/2$. Then, the mean distortion D_s defined in Equation 6.12 per embedded-bit (per host vector), assuming the quantization error to be uniformly distributed over $[-\Delta/2, \Delta/2]$ is:

$$D_s = \frac{\Delta^2}{12L} \quad (6.34)$$

If the projection vector p is chosen pseudo-randomly, the embedding-induced distortion will also be spread. Thus, that is why this type of dither modulation is

called spread transform dither modulation. Thereby, the distortion-normalized squared minimum distance Equation 6.14 of binary dither modulation and STDM does not depend on the distribution of the distortion across samples within the length- L vector,

$$d_{norm}^2 = 3/R_m \quad (6.35)$$

Thus, the choice of the quantization step sizes Δ_k in subsection 6.4.3 or the projection vector p , can turn around the optimization of other embedding characteristics without sacrificing the performance of these systems against bounded perturbations.

Good spread spectrum systems can be designed, for example, by designing a projection vector p with good perceptual distortion properties.

Decision regions for Basic QIM are marked in Figure 6.7 with solid, dashed lines respectively. These regions specify where a watermarked signal can reside, after being disturbed by a perturbation vector, without having an error at the decoding. As this region is bigger in spread transform dither modulation, we can conclude that the probability of a correct decision in STDM is higher than in basic binary dither modulation.

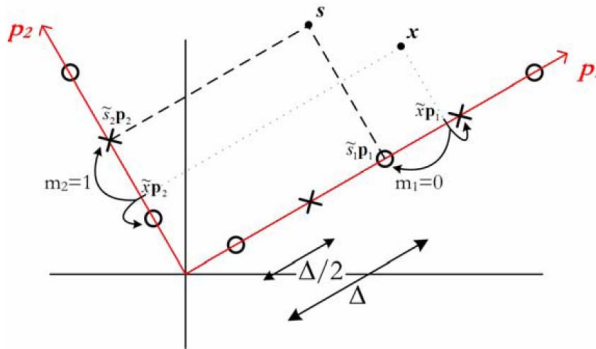


Figure 6.11: Embedding 2 bits, $m \in \{0, 1\}$ using STDM with a uniform, scalar quantizer of step size Δ , using two unit-length orthogonal projection vectors p_1 and p_2 .

One important characteristic of spread transform dither modulation methods is that various information bits can be embedded in a length- L host vector increasing the embedding rate R_m . In that case, each of the J bits to be embedded in a host signal vector is embedded in a different projection of the host signal vector $\tilde{x}_j$ onto a set of length- L orthogonal vectors to form the composite signal projections $\{\tilde{s}_1, \dots, \tilde{s}_J\}$. Then, the quantization error occurred in each direction, $(\tilde{s}_j - \tilde{x}_j)p_j$, is added to the unquantized signal to form the

overall composite vector:

$$s = x + \sum_{j=1}^J (\tilde{x}_j - \tilde{s}_j) \tilde{p}_j \quad (6.36)$$

6.4.6 Distortion Measures

Other distortions measures are based on the characteristics of the human visual system HVS or mathematical tractability [Voloshynovskiy 2001, Wolfgang 1996], as far as the expectations of these distortions take over a probability distribution of the host signal and/or the embedded information. These measures are more close to subjective measures. After embedding the watermarked signal can be corrupted intentionally or unintentionally by common signal processing such as lossy compression, random noise addition, filtering, re-sampling etc...

Exhibition DWT-based Watermarking for D-Cinema

Contents

7.1	Introduction	90
7.2	Digital Cinema Watermarking requirements	90
7.3	Exhibition Watermarking	91
7.4	Flexible DWT marking space	92
7.5	QIM-based watermarking using perceptual models .	93
7.6	Valuometric scaling attacks	99
7.7	Implementation and Results	100
7.8	Conclusion	108

¹In this chapter, a watermarking method for Digital Cinema is presented. In the first part of this chapter, the context of the Digital Cinema application is introduced. The requirements of the watermarking method are defined based on the DCI related specifications. The motivation of the proposed approach is presented. In the second part of this chapter, the fixed gain attack or scaling attack is investigated. The proposed watermarking method is described with a particular attention to the improved fidelity, robustness. A perceptual mask to modulate the watermarking technique is defined and applied. The evaluation of the effectiveness of the algorithm is then analyzed. The performance of the algorithm after various attacks is presented in section 7.7.

¹The major part of this work has been published in [Callau 2009] and [Darazi 2009a].

7.1 Introduction

In a Digital Cinema system, piracy can occur at various points in content distribution, in particular at the end of a given distribution, decoding pathway, where the content is decrypted, decoded, and displayed. The privacy holes are illustrated in Figure 7.1. The final stage of the distribution chain, the exhibition, has been identified by members of the Motion Picture Association (MPA) as the most susceptible and damaging source of theft and unauthorized distribution of motion picture data [Lubin 2003].

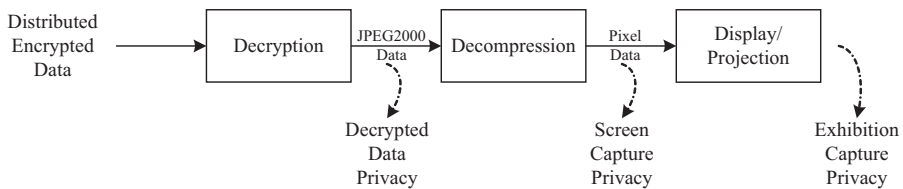


Figure 7.1: Privacy holes in the decoding flow in Digital Cinema.

7.2 Digital Cinema Watermarking requirements

Although several digital watermarking solutions for DCinema have been proposed and implemented in [Lubin 2003, Vural 2005, Haitsma 2001], there is still room for improvements on invisibility and robustness, including retrieval from compressed camcorder captures of exhibited data; as well as for increasing the implementation security and reducing the complexity cost. Prior to the DCI specifications, the SMPTE DC28.4 have stated properties for "Download Watermark" and "Exhibition Watermark". Most of the SMPTE requirements are similar to the DCI specifications, however an additional requirement related to the detection of the watermark is assigned for DCinema.

Relatively to the SMPTE and the DCI specifications, a forensic marking process used for identifying any of the projected films should be compliant to the following:

- **Fidelity:** Satisfy the strict imperceptibility criterions imposed by content owners.
- **Robustness:** The watermark should survive the numerous degradations that occur when a movie is copied from the screen using a DV-camera. It should robust to both compression and exhibition capture.

- Complexity Cost: The embedding must be part of the decoding chain and should be performed in real-time without adding a delay. Being able to exploit perceptual properties in real time.
- Payload: The watermarking data payload is required to be at least 35 bits, out of which, 16 bits are reserved for a time stamp to be embedded every 15 minutes. The remaining 19 bits are used to identify the location of the exhibition. All 35 bits are required to be included in each five minute segment. The specifications allow the use of up to 30 minutes of content for recovery of the watermark.
- Security: The exhibition watermark must be secure against unauthorized reading and unauthorized embedding.
- Detection: The detection of the watermark should be done in a Blind way without the use of any reference. The detection is not required to be performed in real-time.

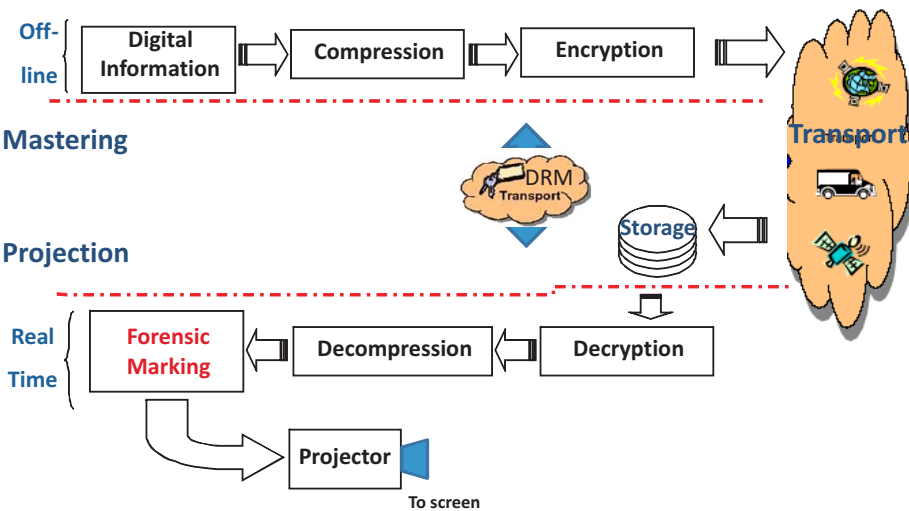


Figure 7.2: Functional encode and decode flow in the digital cinema part of DCI specifications.

7.3 Exhibition Watermarking

In a Digital Cinema environment, the playback system should be able to insert the watermark identifying the time and location of the projection at the same

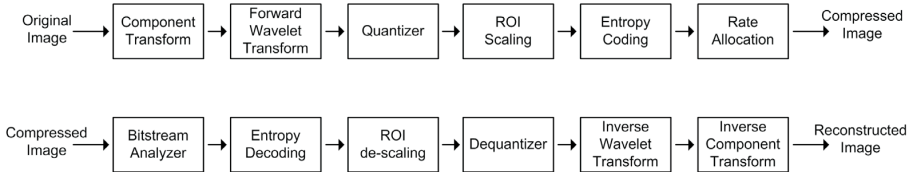


Figure 7.3: The watermark is embedded in the DWT-domain, after the de-quantization step, in the JPEG2000 decoding pipeline

rate when the motion pictures are decoded. The practice of embedding the watermark before the projection or the display of the content is referred to "Exhibition Watermarking". Such digital solution can be used to fill the privacy gap for digital cinema and can be similarly employed to video download.

Since the decoding flow in digital cinema is executed in real time, the exhibition watermarking system should meet the real-time constraint, and hence the complexity of the watermarking algorithm should be as low as possible. The functional encode and decode flow in the digital cinema is illustrated in Figure 7.2 [DCI].

7.4 Flexible DWT marking space

Due to the excellent space-frequency features and the matching with the human visual system (HVS) characteristics, DWT has been used for digital watermarking, especially after becoming the basic technology in JPEG2000 standard. One of the main advantages of watermarking in the wavelet domain is the hierarchical representation, which offers the possibility of analyzing a signal at different resolution levels allowing more degree of freedom compared to DCT. The main disadvantage of frequency-domain techniques are their computational cost. These techniques apply some invertible transform to the host image before embedding the watermark. Then, the transform domain coefficients are modified to embed the watermark and finally the inverse transform is applied to obtain the watermarked image.

In the DCinema decoding flow, precisely during JPEG2000 decompression, the inverse-DWT is performed after the de-quantization process as illustrated in Figure 7.3. Integrating the watermarking system at this stage has the additional advantage of direct wavelet coefficients marking, without performing the DWT before the embedding process and inverting the transform after it. Therefore, the computational cost to derive the signal transform in the watermarking embedding function is saved, even though during the detection stage, it is required to perform DWT before watermark extraction.

Another major challenge consists in being able to exploit the perceptual properties in real-time. The multi-resolution decomposition and reconstruction property of DWT, offers the flexibility to exploit frequency and spatial information of wavelet coefficients in order to adapt the embedding watermark strength w.r.t perceptual characteristics.

Based on this, an adaptive exhibition watermarking method is applied in the DWT-domain during the decoding phase. This method is designed to fit in the JPEG2000 decoding pipeline as shown in Figure 7.4, where the embedding is done after entropy decoding and de-quantization and prior to inverse-DWT. The watermarking technique is based on QIM-based technique called STDm that was introduced in chapter 6. Each watermarking bit is embedded in a segment of wavelet coefficients such that minimum perceptual impact is induced to the image quality.

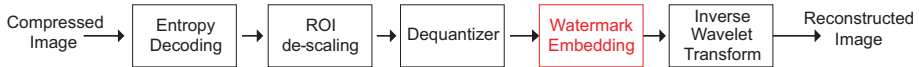


Figure 7.4: Watermarking in the JPEG2000 decoding pipeline

7.5 QIM-based watermarking using perceptual models

In [Barni 2001], Barni et al. presented a SS watermarking method, where the watermark is embedded in the DWT-domain through pixel-wise masking. In this approach, the image is decomposed into sub-levels and the low-frequency coefficients are watermarked by using a controlled quantization process. More recently, Li and Cox presented a STDm watermarking algorithm in the DCT-domain using Watson’s perceptual models [Li 2007]. In this method, for each DCT coefficient, they define a “Slack” which refers to the quantity by which the DCT coefficient can be altered. The perceptual weighting consists on three main considerations: noise sensitivity, local brightness and neighborhood texture activity. In [Yu 2008], Yu et al. proposed the same algorithm with an improved luminance-masked threshold.

In chapter 6, we have introduced the QIM methods and highlighted the performance/improvement compared to additive spread spectrum methods. However a lot of effort has already been dedicated for optimizing additive spread spectrum methods, especially related to perceptual aspect of the human visual system. Nevertheless, there is still room to work on optimizing such aspects with QIM methods.

In the proposed method, our aim is to adapt the STDm method in the

DWT-domain, taking into account the fidelity vs.robustness trade-off while preserving the security of the watermark.

7.5.1 DWT perceptual modeling

The DWT properties linked to HVS are generally exploited in three fields and are described as follows:

- Human eyes are less sensitive to the noise in high resolution bands, and especially in the diagonal subbands of the DWT decomposition.
- The changes in high or low brightness areas of the image are less perceptible.
- Human eyes are less sensitive to noise in highly textured areas but, among these, more sensitive near the edges.

In the proposed method, only the luminance component is watermarked. This choice is mainly motivated by robustness reasons. Note that, JPEG compression attack preserves better the luminance component than the chrominance components. In addition, human eyes are more sensitive to slight color changes than to slight luminance changes.

In STDM, the induced distortion to each DWT coefficient is initially controlled by the quantization factor(see subsection 6.4.5). From Equation 6.31 and Figure 6.10, we notice that changes occur entirely in the direction of the spreading vector p . Thus, we propose the inclusion of a visual model in this framework by designing the vector p in the direction of the least perceptual distortion.

Given an image I and its k -level discrete wavelet transformation coefficients $h(i, j)$, the local brightness of each (i, j) coefficient in each resolution level $l = 0 \dots k - 1$ is computed based on the grey-level values of the low pass subband I_k^{LL} of the image. The local brightness is written as follow:

$$L(l, i, j) = (\frac{1}{I_{kmax}^{LL}}) I_k^{LL} (1 + \lfloor \frac{i}{2^{k-l}} \rfloor, 1 + \lfloor \frac{j}{2^{k-l}} \rfloor) \quad (7.1)$$

As human eyes are less sensitive in very dark and bright regions, we intend to introduce more weights during the embedding process to the corresponding points. Consequently, the local luminance factor is modified as follows:

$$L'(3, i, j) = \begin{cases} 1 - L(3, i, j) & \text{if } L(3, i, j) < 0.5 \\ L(3, i, j) & \text{otherwise} \end{cases} \quad (7.2)$$

The larger the value of $L'(l, i, j)$, the more we may change the corresponding DWT coefficient before the change becomes noticeable.

The proposed adaptive STDm assigns to the corresponding projection vector p , the values computed by Equation 7.1 and Equation 7.2 for each wavelet coefficient in the host vector. Note that the projection vector p is now a function of the host vector; thus, it is variable for each embedded bit.

Discussion : A 2D digital cinema frame from "Magic Hour" film is shown in Figure 7.5 where (a) represents the original frame, (b) the watermarked frame, and (c) the induced distortion due to the watermarking process. This illustrates the performance of the modified projection vector p . Note that the fidelity is improved, the watermark is more adapted to the content and is not spread in the same way every where in the image. Whereas the fact of having a vector p that is totally dependent on some properties of the image can have a negative impact on the security property STDm.

Note that the fidelity is improved, whereas the fact of having a vector p totally dependent on some properties of the image can have a negative impact on the security property STDm. In order to overcome this drawback, we investigate the modulation of the vector p by using a *pixel-wise masking vector*. And instead of defining the vector p as a function from the local luminance of the image, we generate a vector p that holds two major properties: security by preserving the randomness of p , and fidelity by exploiting perceptual properties related to HVS.

7.5.2 Perceptual Mask

Based on the properties of the DWT related to the HVS, we define a pixel-wise masking vector v , that modulates the random vector p . Consequently, the modeled vector denoted as p' is obtained as follows:

$$[p'] = [v][p] \quad (7.3)$$

This provides us with a kind of flexibility to overcome the security drawback of choosing a vector p that is a function of the host signal. Hence, we preserve the security of the vector p by holding its randomness and at the same time modulating it by a pixel-wise vector that considers the properties of the HVS. This issue is discussed later, where we give some insight on a practical scenario.

Therefore, each entry of the vector v evaluates the amount by which each DWT coefficient can be modified. It corresponds to a weighting function that is the product of three main constraints: Resolution, Luminance, Texture and Edges.



(a) Original frame



(b) Watermarked frame



(c) Difference

Figure 7.5: From Wedding film - DCI

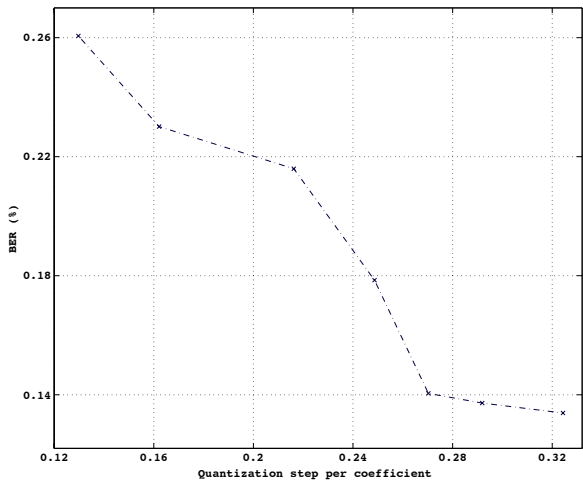


Figure 7.6: Bit error rate (BER) vs. JPEG compression using an embedding rate of 1/120 and a PSNR of around 60dB

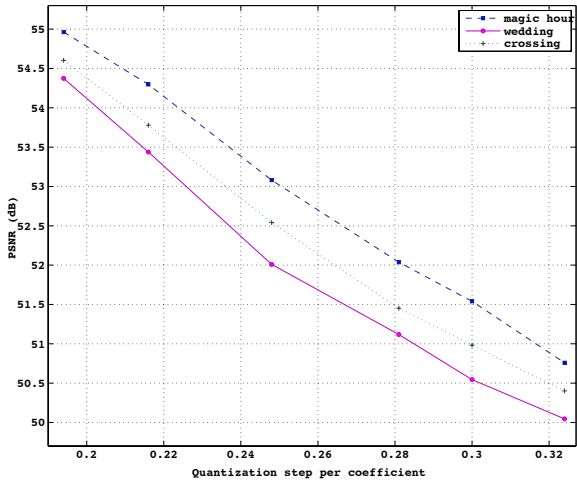


Figure 7.7: Bit error rate (BER) vs. JPEG compression using an embedding rate of 1/120 and a PSNR of around 60dB

- **Resolution** The first constraint is used to define in which decomposition level the watermark embedding occurs. As DWT has poor directional selectivity for diagonal features (there is only one filter for diagonal features), no embedding is done in diagonal subbands. Initially, we decided to embed information bits in the third-level horizontal and vertical subbands ($LH3, HL3$) of a $k = 5$ -level wavelet decomposition. We assume that in 3^{rd} level decomposition $l = 3$, more detail coefficients are available than in previous decomposition levels and are less vulnerable than those in the following levels.
- **Luminance** To indicate the capacity of each DWT coefficient to support changes based on Luminance perceptual properties, we compute the local luminance in the embedding resolution based on the gray level values of the low pass version of the image as presented in Equation 7.1 and in Equation 7.2.
- **Texture and Edges** The DWT performs good de-correlation of the signal and distributes the energy of the image on a few significant coefficients. It also provides a good tool for edge and texture detection. However, high activity regions are essential for both compression and watermarking: compression needs to preserve their integrity and the HVS is less sensitive to modifications in these regions, which is a useful property for watermarking. In the masking model proposed in [Barni 2001], edge and texture activities need the analysis of several subbands, which may increase computational complexity of the watermarking method. Instead of that and for computational efficiency, we use the diagonal subband $HH3$ as a significant index for edges and textures details.

Hence, each entry of the vector v will be generated as follows:

$$\begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(n) \end{bmatrix} = \begin{bmatrix} L'_3(1) * [\varepsilon + HH_3(1)] \\ L'_3(2) * [\varepsilon + HH_3(2)] \\ \vdots \\ L'_3(n) * [\varepsilon + HH_3(n)] \end{bmatrix} \quad (7.4)$$

ε is a constant parameter introduced to avoid the neutralization of the luminance constraint in case of insignificance texture and edge details in the $HH3$ subband. ε should be small and will be defined in the implementation.

Another alternative to exploit edge and texture sensitivity, is to apply a non-linear scaling function before the embedding. Thus, applying the corresponding inverse scaling after embedding. And this to put more watermark energy in edges and textured areas. Hence, the host vectors after scaling the coefficients can be written as follows:

$$x(i, j) = \text{sign}(h(i, j)) \cdot |h(i, j)|^\beta, \quad \beta \leq 1 \quad (7.5)$$

We will analyze the performance of the proposed scheme with the non-linear scaling function or without it. Note that, this step can introduce some distortions. The scaling factor β is controlled in a way to avoid any perceptual distortions. To embed the watermark, a non-overlapping window of size $(n \times 1)$ is slid over the *HL3* and *LH3* coefficients. The size of the window is determined by the coding rate and the size of watermark to be embedded. At each window position one bit of the watermark is embedded by using the adaptive Spread-Transform Dither Modulation scheme introduced in section 6.4.5.

7.6 Valuemetric scaling attacks

We recall the embedding function of *STDM*:

$$y = x + (Q_m(\tilde{x}) - \tilde{x}) \cdot p \quad (7.6)$$

with $\tilde{x}$ the projection of x onto p :

$$\tilde{x} = x^T \cdot p \quad (7.7)$$

The estimated watermark message $\hat{m}$ is decoded as follow:

$$\hat{m} = \arg \min_{0,1} |\tilde{y} - Q_m(\tilde{y})| \quad (7.8)$$

$$\hat{m} = \arg \min_{0,1} |y^T \cdot p - Q_m(y^T \cdot p)| \quad (7.9)$$

Problem formulation:

When the watermarked signal y is subject to a Fixed Gain Attack (FGA) or Scaling attack, it becomes z and it can be written as:

$$z = \rho \cdot y \quad (7.10)$$

Then $\hat{m}$:

$$\hat{m} = \arg \min_{0,1} |\tilde{z} - Q_m(\tilde{z})| \quad (7.11)$$

$$\hat{m} = \arg \min_{0,1} |\rho \cdot \tilde{y} - Q_m(\rho \cdot \tilde{y})| \quad (7.12)$$

As it can be seen, the problem is that:

$$Q_m(\rho \cdot \tilde{y}) \neq \rho \cdot Q_m(\tilde{y}) \quad (7.13)$$

Proposed solution:

Hence, we need to make the quantization step dependent in a causal way on the watermarked samples. We propose to modify the embedding function *s.t.*:

$$y = x + (\|x\|Q_m(\frac{x^T \cdot p}{\|x\|}) - x^T \cdot p)p \quad m \in \{0, 1\} \quad (7.14)$$

with

$$\|x\| = \sqrt{x_1^2 + x_2^2 \dots x_n^2} \quad (7.15)$$

Then if we have a FGA *s.t.* $z = \rho y$, then the embedding function in Equation 7.6 is modified as in Equation 7.14.

This modification is also applied to the minimum distance decoder in Equation 7.8 and is written as:

$$\hat{m} = \arg \min_{0,1} |\tilde{z} - \|z\|Q_m(\frac{\tilde{z}}{\|z\|})| \quad (7.16)$$

$$\hat{m} = \arg \min_{0,1} |\rho \cdot \tilde{y} - \rho\|y\|Q_m(\frac{\rho \cdot \tilde{y}}{\rho\|y\|})| \quad (7.17)$$

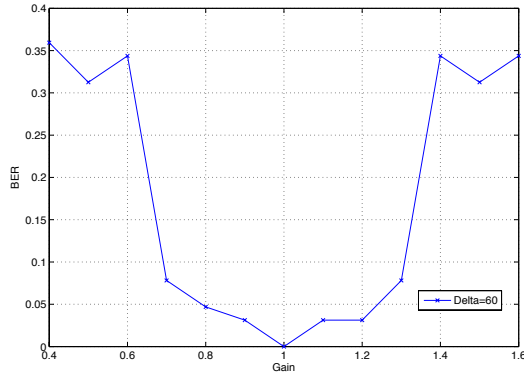


Figure 7.8: Performance of the STDM under Valuemetric Scaling Attack, Gain Attack for $\Delta = 60$

As it can be noticed in Figure 7.8, Figure 7.9 the performance of the STDM under Valuemetric Scaling Attack, Gain Attack. On the contrary, in Figure 7.10, ??, Figure 7.11 the impact of Gain Attack is somehow neutralized. Thus the non-linear impact of the ρ factor in the Quantization is now linear and consequently will not affect the process of decoding the message.

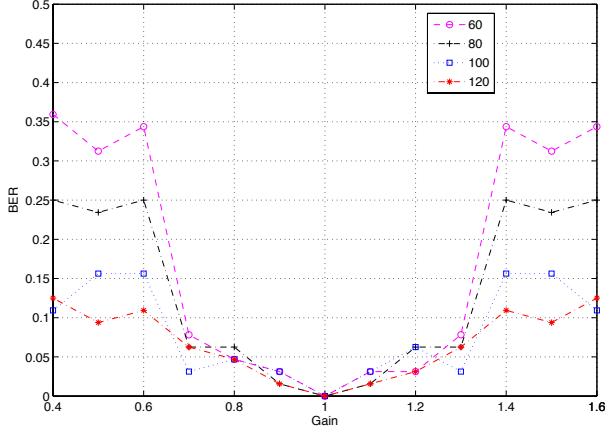


Figure 7.9: Performance of the STDM under Scaling Attack, FGA

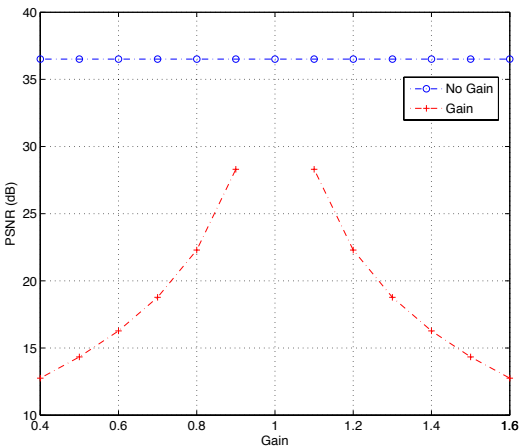
7.7 Implementation and Results

A set of eight 2K-row frames of digital cinema content were obtained and are subjected to the watermark insertion process. Test material was obtained from StEM DCI material [DCI], RED cameras sample shots and Big Buck Bunny open movie project. All images have a 1920x1080 pixel spatial resolution and are 8-bit depth for each color component channel.

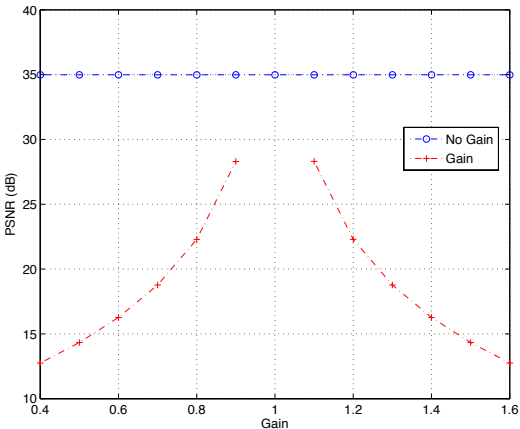
As specified for JPEG2000 compression of 2K resolution content in [DCI], a 5-level 9/7 wavelet decomposition is performed firstly. The total information payload is 35bits, embedded in the level-3 of *LH3* and *HL3* entire subbands. The watermark embedding strength is controlled by the quantization step size Δ used in STDM. We take $\varepsilon = 1$ and the scaling factor parameter β is fixed to 0.85. Experiments were performed to determine both fidelity and robustness of the watermarking algorithm.

7.7.1 Perceptual Quality

Subjective and quantitative evaluation of the distortion introduced due to the watermarking process has been analyzed. As mentioned previously, the watermark robustness depends directly on the embedding strength, which in turn influences the visual degradation on the image. In the proposed scheme, we fix the embedding intensity Δ , however this quantization step is applied differently at each coefficient due to the masking vector. The number of the embedded data is 70 bits/frame and PSNR stands for the average value for



(a) STDM norm $\Delta = 4$



(b) STDM $\Delta = 570$

Figure 7.10: Performance of the (a) improved STDM with $\Delta = 4$ under FGA and the (b) STDM with $\Delta = 570$ both with PSNR=35dB

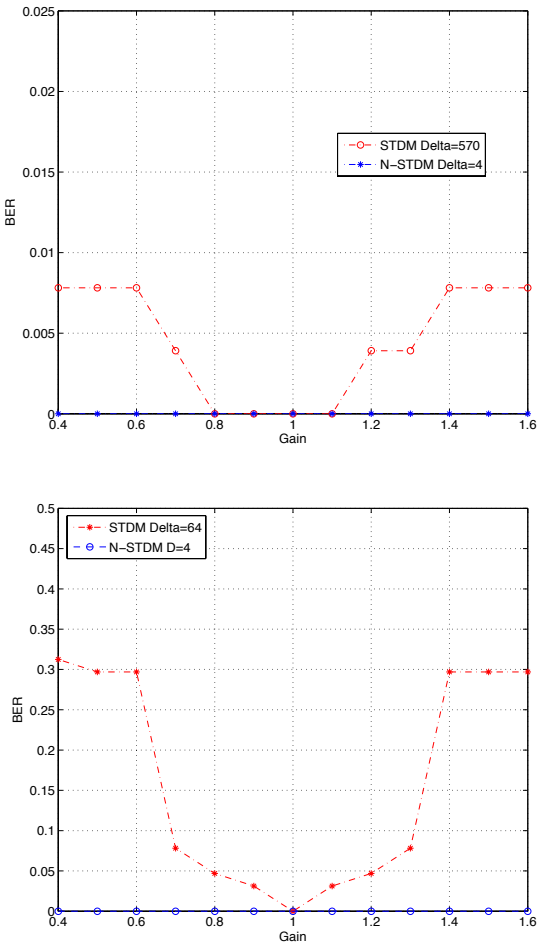


Figure 7.11: Performance of Norm STDM vs. STDM under gain attack

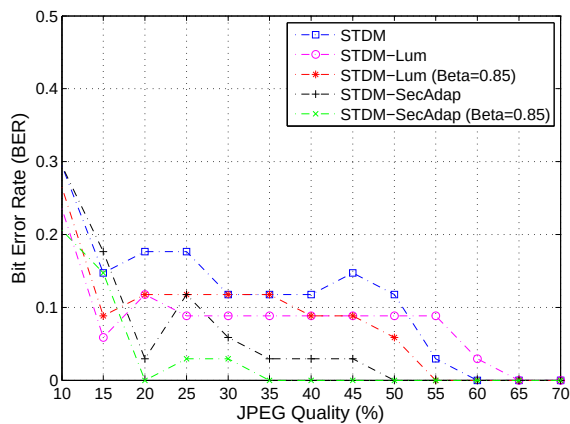


Figure 7.12: Bit error rate (BER) vs. JPEG compression using an embedding rate of 1/120 and a PSNR of around 60dB



(a) Original



(b) Watermarked

Figure 7.13: From Big Buck Bunny film - DCI

several repetitions on each image.

Butterfly tests on large-screen projections were performed with five critical observers to evaluate image fidelity. Each of the eight test frames were watermarked with three different embedding intensities, keeping always the PSNR above 55 dB. Watermarked frames were presented in couple respectively with the original one, and observers were required to detect the watermarked frame in each couple. The results show that no perceptible distortion is produced in the watermarked images for PSNRs above 55 dB. In any case, none of the viewers was able to reliably identify the watermarked frame. A sample watermarked frame is shown in Figure 7.13.

7.7.2 Robustness

We have considered two possible attack scenarios: Perfect Capture and Camcorder Capture. In Perfect Capture, we assume that the decoded bit-stream can be captured from some digital output port after embedding without any distortions. Experimental results show that the proposed method is robust against image processing attacks. A 5% quality image compression rate degrades severely the image so to consider the algorithm robust to compression. Figure 7.12 shows that the proposed method produces lower BER compared to traditional STDM and other STDM improved methods. The STDM-Lum curve illustrates the results when the projection vector is defined from the local luminance of each pixel. And STDM-Lum Beta=0.85 when we apply a non-linear scaling function with a scaling factor $\beta = 0.85$. And finally, STDM-SecAdap illustrates the results for the proposed method with or without a non-linear scaling.

7.7.3 Anti-camcording

As can be seen in the sample frame in figure Figure 7.14, the frame capture introduce arbitrary scaling(size/resolution change), rotation, some perspective projection, blurring and compression. Each captured frame has been processed such that all black-bars or boundary pixels are removed, resulting on approximately (500x375) resolution images. Detection is performed after individual image registration using the morphon method [Knutsson 2005]. Registered images present a PSNR of about 27-28dB.

7.7.4 Intended implementation scenario and perspectives

Discussion: Security

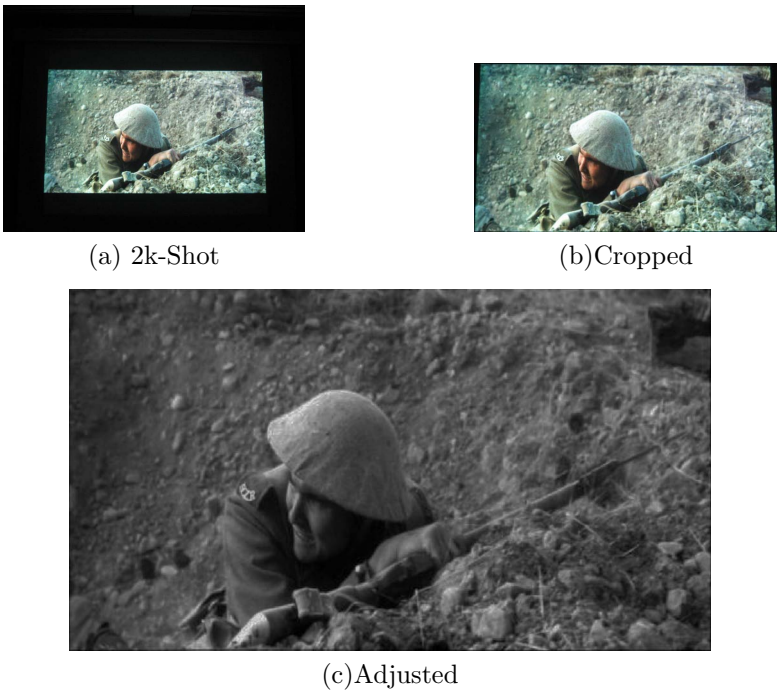


Figure 7.14: Pre-processing for Watermark detection using an advanced informed decoder

In this paragraph, we discuss the security in the watermarking scheme and hence we give some recommendations of how the scheme can be practically implemented in order to preserve the security property.

- A Unauthorized reading can occur when the watermarking scheme is known by the attacker, and also when all the parameters related to that technique are public, e.g the quantization step size. This kind of attacks is referred to as passive attacks.

Ciphering m : The message m should be ciphered, this is done to overcome unauthorized reading by the adversary. This can happen also in case the secrecy of the vector p is compromised.

- B Unauthorized embedding or removal is referred to Active attacks.

Securing p

- Same p : the fact of using the same key p modulated by a perceptual mask which can be computed by the adversary at the decoder is vulnerable to be detected. The possibility of parametrizing the computation of the mask may reduce the risk at the cost of increased complexity.
- Different p : For each frame we have different p . This is very secure and do not alter the property of basic STDM scheme.

Drawbacks:

- * Vulnerable to synchronization attack in case of frame removal or frame rate change.
- * A considerable amount of header information is required to be available at the decoder. Moreover, this is not compliant to DCI.
- Changing p each n frame.

Ideal case: At the encoder side, provided by one secret key, a particular key generator generates N different keys. At the decoder side, having the same secret key and the same key generator, the same N keys are reproduced. A new p is generated each n frames.

Advantages:

- * The security of p should hold since a new p is available only each n frames, so the adversary won't be able by any chance to detect p .
- * Robustness to de-synchronization: if one frame is omitted the decoder can still reconstruct the embedded bit from the following frames, the same bit is embedded in n frame using same p
- * There is no considerable extra information to be provided to the decoder. Thanks to the pattern generator based on one single key? Note that in this case, even that there is not such key generator. it can be also possible to have a different p each

n frames. Hence, the embedder should communicate all the corresponding $totalframes/n$ to the decoder.

Detector

The detection of the Watermark is subject to various kind of attacks that mostly are not related to each other. To deal with that, the idea of having two kind of detectors can be a key solution. At a first level, a *Normal Blind Detector* is used. It should resist for the largest number of attacks. On the other hand for cases where the content have encountered serious distortions, an *Advanced Informed Detector* can be used. In the later, part of the original content is available to perform advanced processing for detection.

7.8 Conclusion

We have proposed a robust exhibition watermarking algorithm. The embedding occurs in the decoding pipeline before the projection of a Digital Cinema movie. The goal is to deter capture and unauthorized redistribution of electronically distributed movies. The algorithm has a very low complexity and is suitable for implementation inside the JPEG2000 decoder. The tests show that the proposed method highly preserves the image quality. The watermark survives low bit-rate compression and capture with a video camera, allowing us to reach the watermark recovering constraint of the DCI without compromising image quality.

Image Content Authentication Using Robust Hashing and Watermarking

Contents

8.1	Introduction	110
8.2	System requirement and overview	112
8.3	Image authentication by hashing and watermarking	116
8.4	Experiment results and analysis	118
8.5	Conclusion	122

¹Robust hashing is a promising tool for multimedia content authentication. Digital watermarking is a convenient way of data hiding. By combining the two, we get a more efficient and versatile solution. In a typical scenario, multimedia data is sent from a server to a client. The corresponding hash value is embedded in the data. The data might undergo incidental distortion and malicious modification. In order to verify the authenticity of the received content, the client can compute a hash value from the received data, and compare it with the hash value extracted from the data. The advantage is that no extra communication is required – the original hash value is always available and synchronized. However, on the other hand, image quality can be degraded due to watermark embedding. There is interesting interaction between hashing and watermarking. We investigate this issue by proposing a content authentication system. The hash algorithm and the watermarking algorithm are designed to have minimal interference. This is achieved by doing hashing and watermarking in different wavelet subbands. Through extensive experiments we show that the parameters of the watermarking algorithm have significant influence on the authentication performance. This work gives useful insights into the research and practice in this field.

¹This chapter is part of a joint work with Li Weng and was submitted to [Weng 2011].

8.1 Introduction

Multimedia has become an important element of digital life. Recent technology development has made it much easier for everyone to produce and distribute multimedia content. For example, people often take photos and share them via social network websites. When there is abundant media content being distributed in a large scale, an important issue is content protection. Typically, we not only need to ensure the correct delivery of the content, but also need to ensure the content is used for the intended purpose.

The requirements of content protection vary w.r.t the application. A commonly studied issue is copyright protection [Wolfgang 1996]. Recently, another important problem is gaining more and more attention. That is – trustworthiness protection. Nowadays, software for editing and modifying digital content is easily accessible. Consequently, modified or synthesized content often appears. Unless one knows the original content, it is difficult to find out whether the content has been professionally modified.

In this work, we focus on the trustworthiness protection of images and video. We assume a general application scenario, as shown in Figure 8.1. In a heterogeneous network, there are servers, clients, and intermediate nodes with different computing capabilities. Clients receive multimedia data from servers through intermediate nodes that form a distribution chain. The distribution chain is not perfectly reliable, due to the following issues:

- Incidental distortion – the content may undergo re-encoding. It can be a format change or re-compression, since it is necessary to adjust the data stream according to the client’s capability and the network condition. Properties, such as resolution, contrast, brightness, sharpness, etc., may change. Data packets may arrive at destinations through different paths and losses can happen.
- Malicious modification – moreover, there might be malicious nodes that modify or replace the content.

In such a circumstance, an important question of the client is *whether the received content is authentic*. Therefore, it is necessary to have active means that can guarantee the authenticity of the content.

The above problem is easy when the original content is available for comparison, but in practice it is usually not the case. When the original content is not available, a possible solution is to generate a *hash value* on the server side and send it securely to the client side. The hash value is a compact abstract of the content. A client can re-generate a hash value from the received content, and compare it with the original hash value. If they match, the content is considered

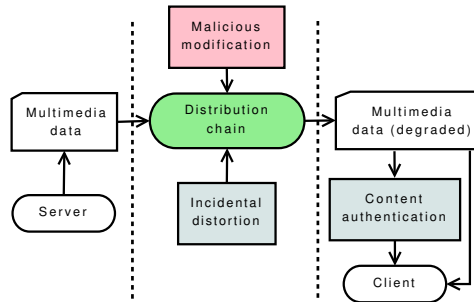


Figure 8.1: The application scenario.

as authentic. Conventionally, there exist cryptographic hash algorithms for data authentication, such as MD5, SHA-1 [Menezes 1996, Schneier 1996]. However, they are not suitable for multimedia data, because they are extremely sensitive and do not tolerate any change of the data. In order to successfully authenticate the content even when it has undergone some incidental distortion, the hash value must possess some *robustness*. Therefore, a new generation of hash algorithms has emerged in the multimedia domain, called robust or perceptual hash algorithms [Schneider 1996, Venkatesan 2000, Arnold 2005, Swaminathan 2006, Moulin 2010].

Perceptual hash algorithms enable robust content authentication, whereas sometimes it is not easy to setup a secure channel for transmitting the hash values. Therefore, a more convenient approach comes fourth – instead of sending through a secure channel, we can deliver the hash value by embedding it into the content itself using robust digital watermarking techniques [Cox 1997, Chen 2001]. This kind of techniques can embed a message in multimedia data in an imperceptible way. The new approach avoids extra communication, and ensures that the original hash value is always available and synchronized. The authentication procedure is simple: a client extracts the original hash value from the watermarked data, and compares it with the re-computed hash value. Such a system was envisioned by Fridrich and Goljan in [Fridrich 1999, Fridrich 2000]. They motivated the usefulness of content-dependent watermarks and designed a robust hash algorithm. A similar concept was exploited by Cannons and Moulin in [Cannons 2004], where they proposed a hash-aided watermarking system. Their aim was to improve the performance of watermark detection. Since previous work focuses on either the hashing, or the watermarking, so far there has been no in-depth study on a complete content authentication system.

In this work, we revisit the concept of content authentication by robust hashing and watermarking from a different angle. In our view, the hash algorithm is the essential component that achieves content authentication, watermarking is a tool to deliver the hash value, and there is interaction between the two.

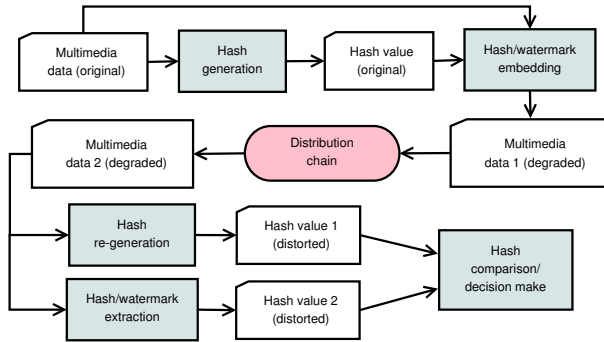


Figure 8.2: Diagram of image content authentication.

This work provides a comprehensive introduction to both robust hashing and watermarking, with a focus on the hash performance under the influence of watermarking. We define the overall system performance as the sum of the authentication performance and the image quality. We design a new system and evaluate the performance using these metrics by extensive experiments. The results serve as a useful reference for research and practice in this field.

The rest of the work is organized as follows. In Section 2, the roles of hashing and watermarking are clearly established. We describe the system requirement and the properties of each component, including some security definitions. Section 3 describes the details of the proposed image hash algorithm and watermarking algorithm. Section 4 shows the performance of the proposed system. The receiver operating characteristics of the hash are derived using different watermarking parameters. Section 5 concludes the work.

8.2 System requirement and overview

The content authentication system includes a watermarking algorithm and an image hash algorithm. They have their own requirements, but also interact. The schematic diagram of content authentication is shown in Figure 8.2.

8.2.1 The image hash system

The purpose of the hash value is to fairly represent the corresponding content. This is the essential requirement. Specifically, the hash algorithm should possess the following properties:

- Compactness – the hash value is compact;

- Robustness – the hash computation is insensitive to a certain level of distortion to the input image;
- Discriminability – perceptually different content results in significantly different hash values;
- Key dependence – the hash generation depends on a key.

It is difficult to define “perceptually different”. However, we could define perceptually similar images as images that are generated by certain incidental distortion. If two images are not perceptually similar, they are considered as perceptually different.

In the content authentication scenario, the distance d between two hash values is compared with a threshold T . The decision is made from two hypotheses:

- $\mathbb{H}_0$ – the images correspond to different content;
- $\mathbb{H}_1$ – the images correspond to similar content.

If $d \leq T$, we choose $\mathbb{H}_1$; otherwise we choose $\mathbb{H}_0$.

The overall performance of the hash algorithm can be characterized by the true positive rate P_d and the false positive rate P_f . They are defined as:

- $P_d = \text{Probability } \{d \leq T | \mathbb{H}_1\}$;
- $P_f = \text{Probability } \{d \leq T | \mathbb{H}_0\}$.

When the threshold T is increased, the hash is likely to tolerate more distortion, but that also increases the chance of false positive. A good algorithm should suppress the false positive rate while maintaining a high true positive rate. The relationship between P_d and P_f is known as the receiver operating characteristic (ROC) curve. In order to choose the most suitable threshold value, we also need to take into account of the false negative rate P_m , which is defined as

- $P_m = \text{Probability } \{d > T | \mathbb{H}_1\}$.

By definition, $P_m = 1 - P_d$. We can see that P_f and P_m are contradicting requirements. Depending on the severity of false positive and false negative, different applications give their own bias towards P_f or P_m . By default, we can choose the equal error rate point (EERP), where $P_f = P_m$, as the working point.

8.2.2 The watermarking system

The watermarking scheme, as a method for data hiding, generally concerns the following issues:

- Payload capacity– the scheme is able to carry a certain amount of information;
- Robustness – the scheme should withstand certain distortion to the image;
- Perceptual quality – the watermark should have little impact on the image quality;
- Key dependence – the algorithm depends on a key.

The performance of a watermarking scheme typically depends on the embedding strength and the size of the payload. Generally, increasing the embedding strength offers stronger robustness but worse perceptual quality; increasing the payload size reduces both the robustness and the perceptual quality.

We can distinguish two main classes of image watermarking techniques: the additive class known as spread spectrum (SS) methods introduced by Cox et al. [Cox 1997], and the substitutive class known as Quantization Index Modulation (QIM) methods introduced by Chen and Wornell [Chen 2001]. In this work, we adopt QIM. It is very robust to noise. In addition to the high capacity-robustness trade-offs, QIM methods are simple and have a small computational cost.

The basic idea of QIM is the quantization of a signal sample x using a quantizer Δ . A scalar, uniform quantizer $Q(x)$ with a step size Δ is defined by $Q(x) = Q(x, \Delta) = \lfloor (\frac{x}{\Delta}) \rfloor \Delta$, where $\lfloor \cdot \rfloor$ denotes a rounding operation. For instance, to embed a message $m \in \{0, 1\}$ in a host signal x , we use dithering to generate two quantizers based on the embedding information m .

$$Q_m(x) = Q(x - d_i) + d_i, \quad i = 0, 1$$

where

$$d_0 = -\frac{\Delta}{4}, \quad d_1 = \frac{\Delta}{4} \tag{8.1}$$

Detection in QIM is performed without access to the original data or the original watermark.

8.2.3 The attacker

As an attacker, his goal is to compromise the content authentication. He could target at either the hash algorithm or the watermarking algorithm and try to generate false positive or false negative cases:

- False positive attack – the content is significantly modified, but the extracted hash value is similar to the re-computed one;
- False negative attack – the content is not significantly modified, but the extracted hash value is drastically different from the re-computed one.

In a false positive attack, after the attacker modifies the content, he has to embed the hash value of the modified content. Therefore, the watermark must be dependent on a secret key K_w . The key is only shared between the server and the client, so that unauthorized embedding is prohibited.

In a false negative attack, the attacker can possibly try two approaches:

1. Gradually modify the content, re-compute the hash value, and check if it is significantly different;
2. Destroy or modify the watermark while maintaining the content.

They intend to modify the hash value 1 and the hash value 2 respectively in Figure 8.2. The first approach implies that the hash generation should depend on a secret key K_h , so that the attacker cannot verify whether the attack is successful. However, this approach inevitably modifies the hash value 2 as well.

In order to thwart the above attacks, the watermarking should be robust and secure, and the hashing should be discriminative and secure. Additionally, there is also incidental distortion that must be tolerated, so the hashing should be robust too.

The attacker can also try to derive the keys K_w and K_h from existing knowledge. If he knows both keys, he could forge any authentic content. There is some study on the security of the key dependence of hashing or watermarking, e.g., [Radhakrishnan 2005, Mao 2007] and [Cayre 2005]. However, for a hybrid system like our proposal, the keys' security is still an open problem. We do not focus on this issue here.

8.2.4 Performance evaluation

The purpose of the proposed system is to authenticate the content under incidental distortion. The effect is the authentication performance. The cost is the quality loss due to watermarking. The overall performance is a balance between the two.

A typical issue in such a hybrid system is that the watermark embedding also brings distortion to the original image, so it might affect the re-computation

of the hash value. This interference, denoted as I_{wh} , should be kept as small as possible.

Besides, the parameters of the two algorithms also influence each other's performance. For example, increasing the hash size typically increases the discrimination performance, but also increases the payload and decreases the robustness of the watermark. Increasing the watermark embedding strength may increase I_{wh} , but on the other hand achieves better robustness of the watermark. Different parameters also implicitly influence the key dependence. In some cases it is easier for an attacker to derive the key.

Therefore, the design of such a system involves the interaction between the two components. We need to optimize the authentication performance and the perceptual quality. In practice, we can fix one of them and try to get the best performance for the other.

8.3 Image authentication by hashing and watermarking

8.3.1 The watermarking algorithm

This method has been presented in chapter 6 and used in chapter 7. Spread transform dither modulation (STDm) is a variant of QIM. It exhibits robustness to re-quantization and meanwhile maintains low complexity. Instead of quantizing the host signal x itself, the quantization occurs entirely in the projection of the host signal x onto a randomly generated vector p , which depends on a secret key K_w . Consequently, the embedding of each information bit occurs in the projection of the host signal x by quantizing it with a uniform, scalar, dithered quantizer [Chen 2001]. Note that in this method, each bit of the watermark message is spread into n samples instead of one sample.

Let us consider a host signal $x = (x_1, x_2, \dots, x_N)$ of length N . We break the whole signal into segments with equal length n , and embed one bit into each segment. This bit is spread in the projection of the segment onto a random vector p . Thus, the watermarked signal is given by:

$$y = x + (Q_m(x^T p) - x^T p)p \quad m \in \{0, 1\} \quad (8.2)$$

The STDm decoder makes a decision based on the projection of the channel output y onto the spreading vector p . The detection can be performed with a minimum distance decoder.

$$\hat{m} = \arg \min_{\hat{m} \in \{0, 1\}} |y^T p - Q_{\hat{m}}(y^T p)| \quad (8.3)$$

In the proposed scheme, the considered marking space is the two-dimensional discrete wavelet transform (DWT) space, where we exploit the flexibility of multi-resolution decomposition. This flexibility allows us to have a non-interfering

hashing and watermarking scheme. We use the biorthogonal 4.4 wavelet. The watermark embedding happens on the third decomposition level of luminance component using the horizontal and vertical subbands.

8.3.2 The image hash algorithm

In this work, we focus on the interaction between the watermarking and hashing algorithms. We do not necessarily need an image hash algorithm with the best performance. Instead, we need an algorithm which is typical, light-weight, and has moderate performance. In addition, this algorithm should be easily configurable and compatible with the watermarking algorithm.

Based on these considerations, we design an image hash algorithm which also works in the wavelet domain. The advantage of using the same domain as the watermarking algorithm is that the influence of watermark embedding to the hash re-computation I_{wh} can be minimized by doing watermarking and hashing in different wavelet subbands.

The algorithm has the following steps:

1. Convert the input image to gray-scale and proportionally resize it to make the maximum dimension equal to 512 pixels;
2. Compute X so that level X has more than L_h wavelet coefficients;
3. Two-dimensional wavelet transform until level X ;
4. Apply a two-dimensional discrete cosine transform (DCT) to the approximate subband;
5. Extract the sign bits of the first L_h DCT coefficients using a zigzag scan, excluding the DC;
6. Concatenate all the extracted bits and perform a random permutation according to K_h .

When two hash values are compared, the bit error rate (BER), i.e., the normalized Hamming distance, is used as the distance metric. Since the hash generation happens in the approximation subband and the watermark embedding happens in the detail subbands, hash re-computation is almost not affected by watermark insertion.

The DCT step is actually optional, but in practice we find that it helps to slightly increase the hash performance. Note that the permutation step only achieves key-dependence with a moderate security level. It is not highly secure

according to the criteria described in [Mao 2007]. However, the hash key K_h is not the only security factor in the system – there is also the watermark key K_w . In order to know the embedded hash value, the attacker has to break the watermark key first. This complicates the security analysis. Future investigation is required for this issue. We do not cover it in this work.

8.4 Experiment results and analysis

The system performance depends on the parameters of both algorithms. In this work, we investigate how the watermarking algorithm affects the content authentication performance. We fix the hash size to 128 bits, and vary the watermarking parameters n and Δ , i.e., the length of the spreading vector and the quantization step. The following values are used:

- n – 16, 32, 64, 128, 256;
- Δ – 40, 60, 80, 100, 120, 160, 200, 240.

For each combination of n and Δ , we evaluate the content authentication performance by three tests:

- Robustness test – hash values of original images and their distorted versions are compared;
- Discrimination test – hash values of different images are compared;
- Hypothesis test - the false positive rate and false negative rate are computed for all thresholds.

A database of 100 images is used. They are natural scene photos of different genres such as architecture, art, humanoids, landscape, objects, vehicles, etc. In the robustness test, distorted versions are generated for each original image. Two kinds of incidental distortion are simulated:

- Additive white Gaussian noise (AWGN) – PSNR from 20 dB to 50 dB with step 5 ;
- JPEG compression – quality factor (QF) from 10 to 70 with step 10.

There are 100 original images and 1400 distorted images. Hash values are generated for all of them. Note that a hash value of an original image is generated by:

Table 8.1: EERP values for different $\{n, \Delta\}$ combinations.

$n \setminus \Delta$	40	60	80	100	120	160	200	240
16	.1316	.0395	.0071	.0065	.0050	.0032	.0021	.0005
32	.1788	.1046	.0237	.0199	.0129	.0175	.0117	.0020
64	.2080	.1636	.1076	.0342	.0130	.0039	.0019	.0013
128	.2872	.1895	.1307	.1144	.1057	.0230	.0125	.0043
256	.3947	.2827	.2026	.1379	.1269	.1097	.0503	.0279

Table 8.2: Average PSNR (dB) values for different $\{n, \Delta\}$ combinations.

$n \setminus \Delta$	40	60	80	100	120	160	200	240
16	53.10	49.94	47.68	45.90	44.43	42.10	40.28	38.78
32	56.28	52.79	50.62	48.92	47.50	45.23	43.44	41.97
64	60.89	56.07	53.55	51.82	50.45	48.25	46.53	45.10
128	69.00	61.16	57.62	55.47	53.95	51.71	50.03	48.64
256	84.71	70.26	64.46	61.23	59.11	56.36	54.51	53.11

1. Embedding the original hash value into the image;
2. Distort the image;
3. Extract the the watermark.

The hypothesis test is done by pair-wise comparing all the hash values. There are 1500 runs of hash generation and $\binom{1500}{2} = 1124250$ hash comparisons for each $\{n, \Delta\}$ pair.

The content authentication performance is represented by the EERP, i.e., P_f or P_m when they are equal. Different EERP values w.r.t different $\{n, \Delta\}$ combinations are listed in Table Table 8.1. A lower value means better performance.

The image quality loss can be measured by comparing the images before and after watermark embedding. Three quality metrics are considered here: the peak signal-to-noise ratio (PSNR), the signal-to-noise ratio (SNR), and the structural similarity (SSIM) [Wang 2004]. The average results are listed in Table 8.2, Table 8.3, and Table 8.4 respectively. A higher value means better perceptual quality.

Based on the above results, there are two observations:

1. The authentication performance increases with smaller n and larger Δ ;
2. The image quality loss increases with smaller n and larger Δ .

Table 8.3: Average SNR (dB) values for different $\{n, \Delta\}$ combinations.

$n \backslash \Delta$	40	60	80	100	120	160	200	240
16	46.34	43.18	40.92	39.14	37.67	35.34	33.51	32.02
32	49.51	46.03	43.86	42.15	40.74	38.47	36.68	35.21
64	54.13	49.31	46.79	45.06	43.68	41.49	39.77	38.34
128	62.24	54.40	50.86	48.71	47.19	44.95	43.27	41.88
256	77.95	63.50	57.70	54.47	52.35	49.60	47.75	46.35

Table 8.4: Average SSIM values for different $\{n, \Delta\}$ combinations.

$n \backslash \Delta$	40	60	80	100	120	160	200	240
16	.9976	.9953	.9923	.9888	.9847	.9750	.9635	.9507
32	.9988	.9974	.9959	.9941	.9920	.9869	.9808	.9739
64	.9996	.9988	.9978	.9968	.9958	.9932	.9901	.9866
128	.9999	.9996	.9991	.9986	.9980	.9968	.9954	.9937
256	1.000	.9999	.9998	.9996	.9994	.9988	.9982	.9976

Implicitly, there is a third relationship on watermarking:

- 3) The security of watermarking decreases with smaller n and larger Δ .

This is because a longer spreading vector is in general more difficult to discover; on the other hand, when the quantization is stronger, the watermark pattern becomes more noticeable to an attacker.

In practice, a user can first choose an acceptable image quality level and a watermarking security level, then he can find the corresponding authentication performance, or vice versa. For example, we can choose $n = 64$, $\Delta = 120$ to get a PSNR around 50 dB in Table 8.2. The corresponding EERP is $P_f = P_m = 0.0130$ in Table 8.1. The detailed authentication performance results are presented below. Table 8.5 shows the robustness test results. Except for JPEG QF=10, the average hash distances are all quite small (less than 0.1). They indicate strong resistance to JPEG and AWGN. Figure 8.3 shows the discrimination test results. They are derived from 1113750 hash comparisons between different images. Most hash distances for different images are close to 0.5, which is similar to the ideal situation. Figure 8.4 shows the hypothesis test results, where we see efficient trade-offs between the false positive rate and the false negative rate. These results prove that the proposed system works properly.

Table 8.5: Robustness test for $\{n = 64, \Delta = 120\}$.

JPEG QF	10	20	30	40	50	60	70
Mean BER	.2901	.0132	.0126	.0223	.0732	.0030	.0025
AWGN PSNR	20	25	30	35	40	45	50
Mean BER	.0818	.0089	.0055	.0030	.0022	.0016	.0012

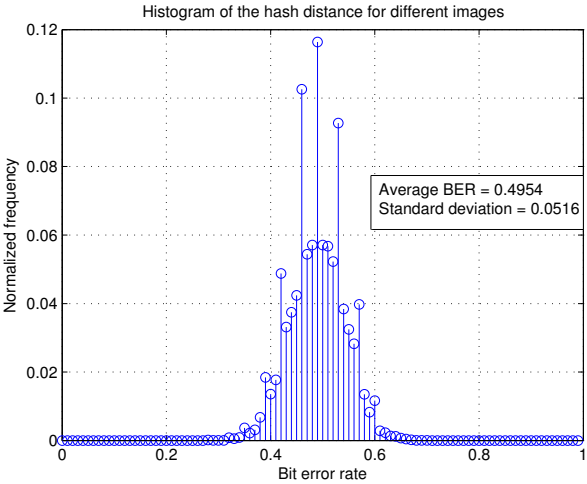


Figure 8.3: BER Histogram for different images using $\{n = 64, \Delta = 120\}$.

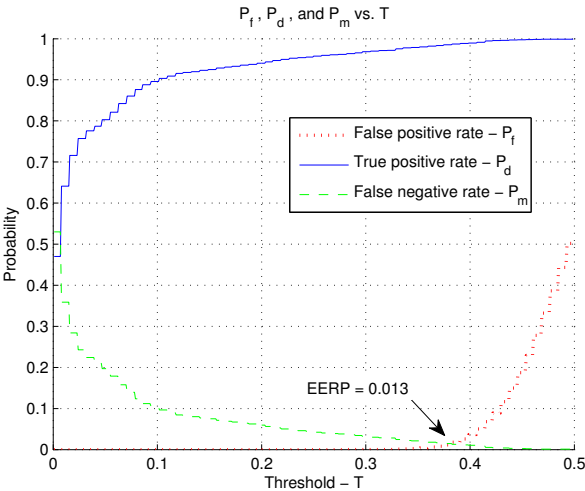


Figure 8.4: Hypothesis test results for $\{n = 64, \Delta = 120\}$.

8.5 Conclusion

Robust hashing is a promising solution to the multimedia content authentication problem. In a typical scenario, multimedia data and the corresponding hash value are sent from a server to a client. The data might undergo incidental distortion and malicious modification. In order to verify the authenticity of the received content, the client can compute a hash value from the received data, and compare it with the hash value from the server. Since the original hash value must be sent securely, there is additional cost which is not always affordable. This limits the application scope of robust hashing.

The above limit can be overcome by digital watermarking techniques. Instead of being transmitted separately, hash values can be embedded in the corresponding multimedia data. The advantage is that no extra communication is required – the original hash value is always available and synchronized. This makes robust hashing a more versatile solution. However, on the other hand, image quality can be degraded due to watermark embedding.

There is interesting interaction between hashing and watermarking. We investigate this issue by proposing a content authentication system. The hash algorithm and the watermarking algorithm are designed to have minimal interference. This is achieved by doing hashing and watermarking in different wavelet subbands. We define the authentication performance as the receiver operating characteristics of the hash algorithm. Through extensive experiments we show that the parameters of the watermarking algorithm have significant influence to the authentication performance. There is a trade-off between the authentication performance and the perceptual quality. In practice, we can fix one of them and try to get the best performance for the other. This work gives useful insights into the research and practice in this field.

Part III

Towards a combining scheme for 3D Stereo image Coding and Watermarking

Combining Scheme

Contents

9.1 Stereo Image Particularity	126
9.2 Stereo Watermarking Scenario	126
9.3 Results and Evaluation	129
9.4 Conclusions and Perspectives	131

¹Nowadays, 3D stereo image technologies are getting more and more mature. They are used in a wide range of applications such as digital cinema, geographical information systems, and video games etc... Digital watermarking has been used as a potential tool for copyright protection of digital media.

Unlike the 2D mono images, the 3D stereo images offer the perception of 3D experience. This phenomenon presents a number of particularities that need to be taken into consideration while applying any kind of processing to the stereo pair. The 3D experience mainly depends on the depth of a point in the scene.

¹The major part of this chapter has been published in [Darazi 2011].

9.1 Stereo Image Particularity

Generally, the human eye is more sensitive to modifications in the focus zone which normally corresponds to either the zero disparity or the foreground regions where the object is closer to the eyes [Lipton 1982]. In 3D digital cinema production, a practice of blurring all the distracting zones is used in order to have more attention to the object in the focus region [Lipton 1982].

In 3D stereo watermarking, altering the images without considering these particularities can be very inconvenient. This can compromise the 3D matching in the stereo pair and also can make the transparent watermark in 2D, visible in 3D [Desmedt 1998a, Desmedt 1998b]. Another requirement of a consistent 3D stereo watermarking is the analysis of the dependency in the stereo pair. Exploiting the disparity information between the two images is essential to prevent the flickering effect. Besides the previous requirements, watermarking one of the images in a stereo pair is not optimal, an attacker by some means can benefit from the other image stream. All these aforementioned issues motivate us to find a kind of complementarity between the coding and the watermarking scheme.

9.1.1 Constraints

In 2D watermarking, a lot of interest has been paid to approaches that use the Human Visual System (HVS) properties to adapt the watermark. These approaches differ in the way they fix the watermarking strength. The maximum amount of changes that can be tolerated before human eyes detect a difference or Just Noticeable Difference (JND), is determined in those methods by the exploitation of HVS perceptual characteristics [Barni 2001, Li 2007].

The evolution of watermarking for 3D stereo images follow the progress of 2D watermarking. Most of the recent methods usually use already existing techniques for 2D images. In [Bae 2008], the right image is watermarked without any specific adaptation to the visual constraints of the 3D stereo case. In [Bhatnagar 2009, Rawat 2010], stereo and disparity information are jointly used to adapt a 2D watermarking method to a specific scenario in the stereo case. In the same context of the proposed work, Campisi et al. uses the depth map to adapt the Quantization Index Modulation (QIM) watermarking scheme in the wavelet subbands [Campisi 2008].

9.2 Stereo Watermarking Scenario

The novelty of our work is the integration of the watermarking in a predictive stereo image decoding framework. The embedding process occurs in the

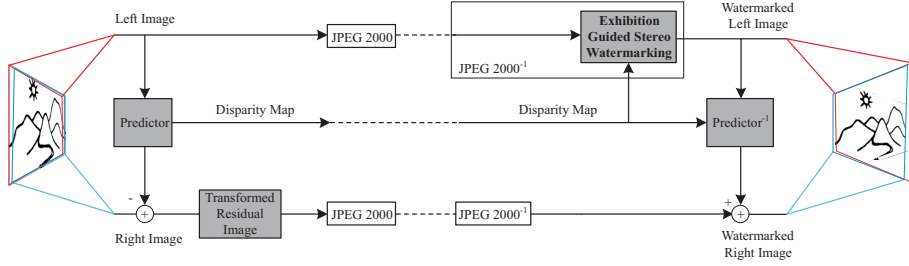


Figure 9.1: Block diagram of the global framework where the exhibition 3D watermarking is integrated

JPEG2000 decoding pipeline of the reference (left) image. We design a 3D-visual-mask with respect of the HVS properties related to the DWT of 2D images, and the depth information related to the 3D experience of the stereo images.

9.2.1 Feedback/Content Analysis from the compression process

The novelty of our work is the integration of the watermarking in a predictive stereo image decoding framework. The embedding process occurs in the JPEG2000 decoding pipeline of the reference (left) image. We design a 3D-visual-mask with respect of the HVS properties related to the DWT of 2D images, and the depth information related to the 3D experience of the stereo images.

9.2.2 Depth map guided watermarking for Stereo Images

9.2.3 3D-Visual Mask

In the proposed method, our aim is to adapt the STDM method in the DWT-domain, taking into account the particularity of the stereo image constraints. Based on our previous work, we extend the 2D-pixel-wise masking vector proposed in [Darazi 2009a] to the 3D scenario. The perceptual mask is updated by considering a new constraint related to the depth map. Consequently, the 3D-visual mask is used to modulate the vector p in the STDM watermarking method. This way the vector p is adapted to the 3D properties of the stereo pair. The random vector p is multiplied by the 3D-visual mask, in order to take into account the perceptual properties related to the human visual system for 3D stereo images.

9.2.4 Disparity guided STDM

In STDM, the quantity by which each DWT coefficient can be altered, is initially controlled by the quantization factor Δ . From (6.31) we can see that changes occur entirely in random spreading vector p . We propose the inclusion of a perceptual model in the STDM framework by modulating the vector p towards the least 3D perceptual distortion. Consequently, the vector denoted as p' is obtained as follows:

$$[p'] = [v_{3D}][p] \quad (9.1)$$

where v_{3D} is the designed 3D mask. We recall the 2D-pixel-wise masking vector designed previously in [Darazi 2009a]:

$$\begin{bmatrix} v_{2D}(1) \\ v_{2D}(2) \\ \vdots \\ v_{2D}(n) \end{bmatrix} = \begin{bmatrix} L'_l(1) * [\varepsilon + HH_l(1)] \\ L'_l(2) * [\varepsilon + HH_l(2)] \\ \vdots \\ L'_l(n) * [\varepsilon + HH_l(n)] \end{bmatrix} \quad (9.2)$$

where L' represents the modified local luminance in the embedding resolution based on the gray level values of the image's low pass version. This component indicates the potential of each DWT coefficient to support changes based on luminance perceptual properties. HH_l represents the diagonal subband of the DWT as a significant index for edges and textures details. Parameter l denotes the level of DWT decomposition and is chosen to be 3, where more detail coefficients are available than in previous decomposition levels and are less vulnerable than those in the following levels. Parameter ε is a constant parameter introduced to avoid the neutralization of the luminance constraint in case of insignificance texture and edge details in the HH_l subband. Hence, the designed 3D mask is obtained by multiplying the 2D mask by a function of the disparity map and is written as follow:

$$\begin{bmatrix} v_{3D}(1) \\ v_{3D}(2) \\ \vdots \\ v_{3D}(L) \end{bmatrix} = \begin{bmatrix} v_{2D}(1) \\ v_{2D}(2) \\ \vdots \\ v_{2D}(L) \end{bmatrix} * \begin{bmatrix} \mathcal{F}_{DM}(1) \\ \mathcal{F}_{DM}(2) \\ \vdots \\ \mathcal{F}_{DM}(L) \end{bmatrix} \quad (9.3)$$

We first approximate the depth map to a gaussian distribution $f(DM)$ with a mean μ equal to 0 corresponding the zero disparity plane and a variance σ that we define as $\sigma = \sqrt{\mu_{DM - \min(DM)} / k}$. The watermark is intensified around the zero disparity region and is reduced elsewhere. This offers the advantage of a smooth transition between the watermarked regions depending on their location in the depth plane.

Another alternative is to approximate the disparity map to the cumulative distribution function of the normal distribution that converges to zero for

negative disparity

$\lim_{x \rightarrow \infty} f(DM) = 0$ and to one for positive disparity

$\lim_{x \rightarrow -\infty} f(DM) = 1$. A consequence of this function is the presence of more watermark in the background plane where the attention of the viewer is less requested.

In order to preserve both of the previous properties. We propose a function having a main lobe in zero and which converges to 1 when the disparity tends to infinity.

$\lim_{x \rightarrow -\infty} \mathcal{F}(DM) = 1$. It can be presented as follows:

$$\mathcal{F}(DM) = \max\left(a \cdot \frac{f(DM)}{\max(f(DM))}, b \cdot DM^3\right) \quad (9.4)$$

where

$$f(DM) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{\frac{1}{2}\left(\frac{DM-\mu}{\sigma}\right)^2} \quad (9.5)$$

Parameters a and b allow to more accentuate the lobes. In our implementation we take $a = 0.3$, $b = 1$, and we choose $\sigma = 1/8$ in order to have more cavity between the zero disparity and the background region.

9.3 Results and Evaluation

In the proposed scheme, stereo image tests of different sizes were used, 512*512 pixels for pentagon and fruit, 2k for digital cinema stereo images like "monstre" and 720p for knight stereo sequence. We fix the embedding intensity Δ , however this quantization step is applied differently at each coefficient due to the masking vector. The number of the embedded data is 512 bits/frame and the PSNR stands for the average value for several repetitions on each image. To embed the watermark, a non-overlapping window of size $(n \times 1)$ is slid over the HL_3 and LH_3 coefficients. The size of the window is determined by the size of watermark to be embedded and hence the size of the stereo image. At each window position one bit of the watermark is embedded by using the Spread-Transform Dither Modulation scheme introduced in Section 6.4.5. Parameter ε is chosen to be 1.

Subjective and quantitative evaluation of the distortion introduced due to the watermarking process has been analyzed. The watermark robustness depends directly on the embedding strength, which influences the visual degradation on the image. The quality comparison metric used in this procedure is the structural We present the performance of the proposed watermarking scheme, where we show in Figure 9.2 the quality gain in terms of SSIM under JPEG attack compared to the 2D mask or without any mask. We can notice also in Figure 9.3 the improvement compared to the 2D mask in terms of BER under the Additive Gaussian Noise attack. Note that the price to pay in this

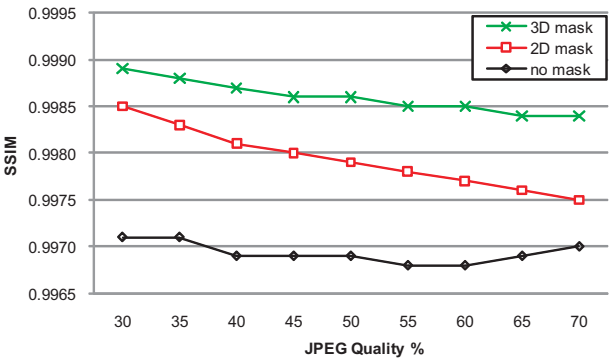


Figure 9.2: Performance of the proposed method in terms of Structural Similarity metric under JPEG attack

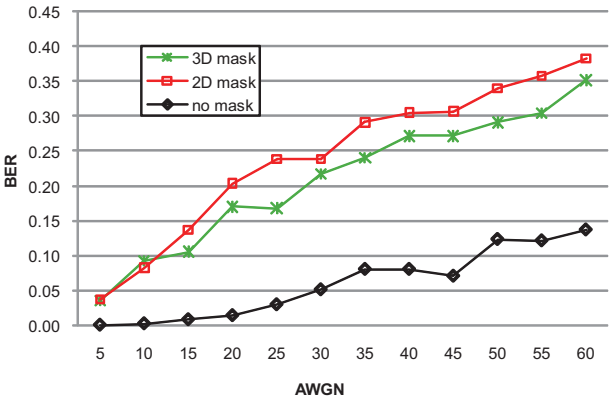


Figure 9.3: Performance in terms of Bit error rate under noise attack

scheme is in terms of robustness as represented in Figure 9.3 compared to the no mask watermarking. In Figure 9.4, we show the left image of "monstre" stereo pair, and respectively the different watermarks where no mask is used, the 2D mask and finally the 3D mask. Butterfly tests using Barco stereo projection system were performed with five critical observers to evaluate image fidelity. Watermarked images were presented in couple respectively with the original one, and observers were required to detect the watermarked frame in each couple. A part of the "knight" stereo test sequence was used. A 20s stereo sequence is watermarked and projected. The results show that no perceptible distortion is produced in the watermarked images where we have used the 3D mask for PSNRs above 55 dB. None of the viewers was able to reliably identify the watermarked sequence.

9.4 Conclusions and Perspectives

A watermarking scheme for 3D stereo Digital Cinema images is presented. The watermarking is done in a predictive stereo image decoding framework. We design a 3D-visual-mask that guides the watermarking method with respect to the HVS properties related to the DWT of 2D images, and the depth information related to the 3D experience of the stereo images. The results show improvements in terms of objective and subjective quality tests. One of our perspectives is to have a deeper study of the optimization related to the quantization factor and the parametrization of the 3D masking vector.

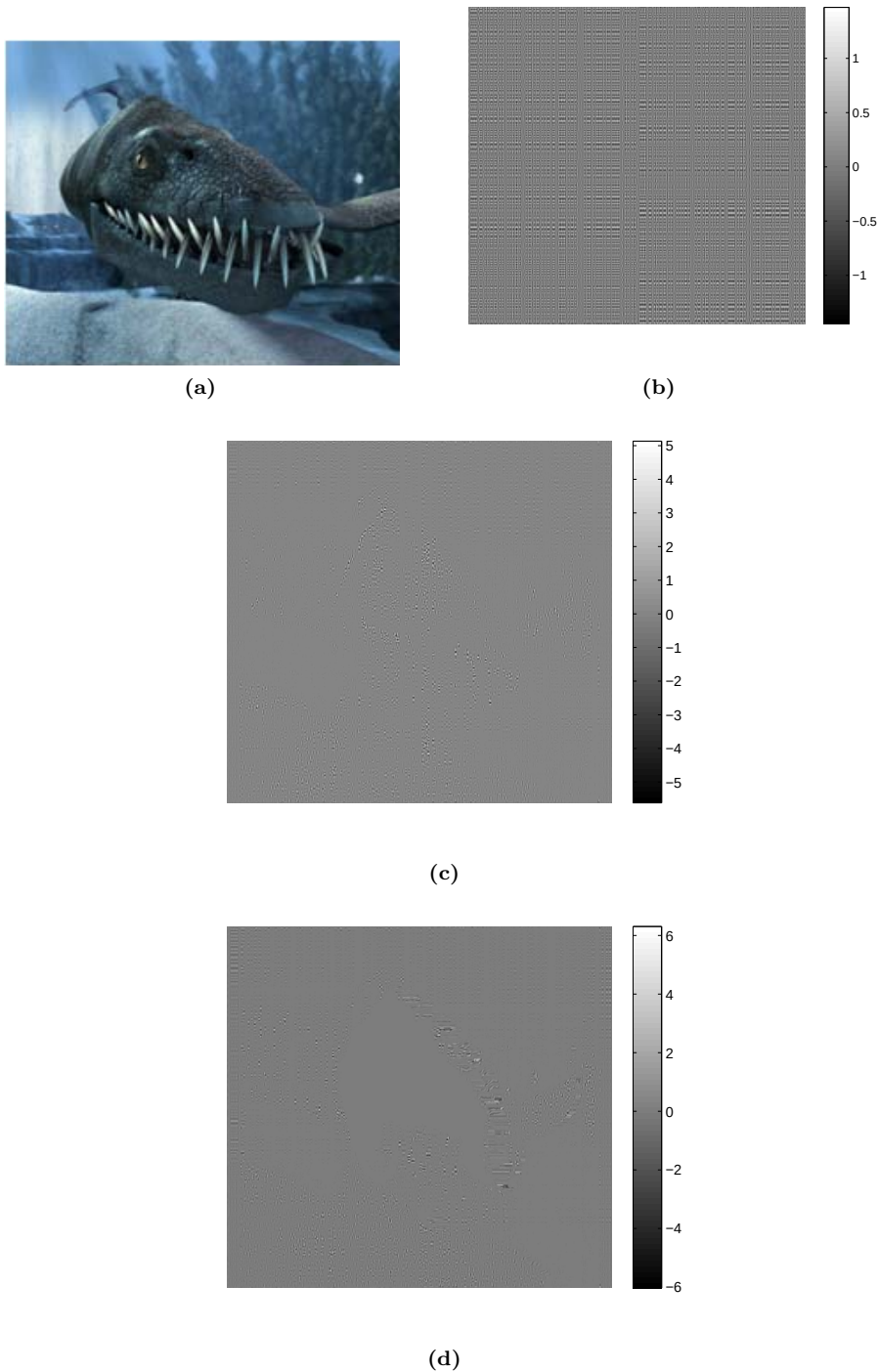


Figure 9.4: The watermark for the (a) *monstre stereo image* using respectively (b) *no mask*, (c) *2D mask*, and (d) *3D mask*

Conclusions

10.1 Discussion, Back to the main contributions

In this thesis, the main focus had been on efficient compression and intellectual property rights protection associated with 3D stereo images.

Provided that more information is displayed within stereo images, supporting technologies are required such as digital compression in order to ensure the storage and transmission of these sequences. Such compression techniques cannot be achieved without efficiently exploiting the correlation in the stereo pair. Within a predictive framework, we have proposed the modification of the traditional disparity compensated coding scheme in order to optimize the system at various stages.

Working within a predictive framework was a key issue not only for designing a transform that exploits the redundancy in the stereo pair, but also for proposing an exhibition watermarking technique with good perceptual distortion properties. The disparity map has been practically used and exploited during all the parts of this research.

In chapter 1, we have presented a global insight on the problem and specified the challenges related to each subject. In chapter 2, we had a brief introduction on stereo modeling and coding techniques. The fundamental predictive coding framework has been introduced. Lifting Scheme corresponding to 2nd generation wavelet transform was presented. We have described in Chapter 3 illumination compensation problem. In chapter 4, an optimal quantization of the Disparity Map was proposed with respect to the global bitrate.

chapter 5 was devoted for the development and implementation of optimized prediction filter that minimizes the variance of residual image so that coding is efficient. In addition, we have proposed a modified Lifting Scheme for the residual image.

In chapter 6, we have introduced digital watermarking, the main properties, and the potential applications. We have surveyed the Quantization Index Modulation class of methods.

In the first part of chapter 7, the context of the 2D Digital Cinema application

was presented, and the proposed watermarking method was described with a particular attention to the improved fidelity, robustness. A perceptual mask to modulate the watermarking technique had been designed and applied. In chapter 8, we revisited the concept of content authentication by robust hashing and watermarking from a different angle. In our view, the hash algorithm is the essential component that achieves content authentication, watermarking is a tool to deliver the hash value, and there is interaction between the two.

Finally in chapter 9, an extension of the 2D exhibition watermarking scheme was applied to stereo images. We have proposed a disparity guided watermarking technique where the perceptual mask is extended to fit with the particularities of stereo. This was inserted in the predictive coding scheme that was investigated in part I of the thesis.

Publications

Journal papers

R. Darazi, A. K. K.C., A. Gouze and B. Macq. "Lifting-based Residual Transform for Stereo Image Coding" *Image Processing, in IEEE Transactions on*, submitted 2011.

International conference papers

R. Darazi, A. Gouze and B. Macq, "Lifting Scheme-based Method for Joint Coding 3D-Stereo Digital Cinema with Luminance Correction and Optimized Prediction." *Visual Communication and Image Processing VCIP, in Proc. Electronic Imaging 2009, IS&T/SPIE 2009*.

R. Darazi, P. Callau and B. Macq. "Secure and HVS-adaptive exhibition Spread Transform Dither Modulation watermarking for Digital Cinema." *In Information Forensics and Security, 2009. WIFS 2009. First IEEE International Workshop on*, pages 1-5, 2009.

R. Darazi, A. Gouze and B. Macq. Adaptive Lifting Scheme- based method for joint coding 3D-stereo images with luminance correction and optimized prediction." *In Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 917-920, 2009.

R. Darazi, M. Montanola, L. Weng, B. Preneel and B. Macq. "Disparity Guided Exhibition Watermarking for 3D Stereo Images." *In Security and Privacy in e-Societies, 2011. SECES 2011. ACM First International Workshop on*, pages 1-8, 2011.

L. Weng, R. Darazi, B. Preneel, B. Macq and A. Dooms. "Image Content Authentication Using Robust Hashing and Watermarking." *In Acoustics, Speech and Signal Processing, 2012. ICASSP 2012. IEEE International Conference on*, pages 1-4, submitted, 2011.

A. K. K.C., R. Darazi and B. Macq. "Optimal optical flow based disparity map estimation for lossless stereo image coding." *In Visual Information Processing and Communication II. in Proc. Electronic Imaging 2011, IS&T/SPIE 2011*.

P. Callau, R. Darazi and B. Macq "Exhibition QIM-based Watermarking for Digital Cinema. " *In Media Forensics and Security, in Proc. Electronic Imaging 2011, IS&T/SPIE 2009*.

R. Darazi, R. Hu, B. Macq, "Applying Spread Transform Dither Modulation for 3D-mesh watermarking by using perceptual models," *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, vol., no., pp.1742-1745, 14-19 March 2010

M. Montanola, R. Darazi, J. Giard, P. Rondao and B. Macq "Using feature point-based extraction for STDM 3D-mesh watermarking that withstands the cropping attack," *In Media Forensics and Security, in Proc. Electronic Imaging 2011, IS&T/SPIE 2011.*

Appendices

Optical Flow-based Disparity map estimation

A.0.1 Disparity map estimation

In the optical flow method, the disparity map estimation problem can be stated as a minimization problem which minimizes the following energy function [Zach 2007].

$$J = \int |\nabla \underline{u}| d\underline{x} + \lambda \int |I_L(\underline{x} + \underline{u}(\underline{x})) - I_R(\underline{x})| d\underline{x} \quad (\text{A.1})$$

where ∇u is the gradient of the disparity map u , the data term is represented by the L_1 norm of the difference between the left image I_L and the right image I_R , the regularization term is represented by total variation (TV) norm of u . Assuming only the horizontal disparity such that $\underline{u}(\underline{x}) = u(\underline{x})$ and linearizing $I_L(\underline{x} + u(\underline{x}))$ around initial point $x + u_0(\underline{x})$, we can write the data term of Equation A.1 as:

$$I_L(x + u(\underline{x})) = I_L(x + u_0(\underline{x})) + (u(\underline{x}) - u_0(\underline{x}))I_L^x(x + u_0(\underline{x})) \quad (\text{A.2})$$

where u_0 is the given disparity map and I_L^x is the horizontal derivative of I_L . For simplicity, we write $u \equiv u(\underline{x})$ and $x \equiv \underline{x}$.

The above problem is reduced to the following equation:

$$\begin{aligned} J &= \int |\nabla u| dx + \lambda \int |u I_L^x(x + u_0) + I_L(x + u_0) - u_0 I_L^x(x + u_0) - I_R(x)| dx \\ &= \int |\nabla u| dx + \lambda \int |\rho(u)| dx \end{aligned} \quad (\text{A.3})$$

where $\rho(u) = u \cdot I_L^x(x + u_0) - u_0 \cdot I_L^x(x + u_0) + I_L(x + u_0) - I_R(x)$.

To solve the above optimization problem, we introduce an auxiliary variable v and rewrite the above Equation A.3 as:

$$J_\theta = \int \left[\lambda |\rho(u)| + \frac{1}{2\theta} (u - v)^2 + |\nabla u| \right] dx \quad (\text{A.4})$$

where θ is a small constant such that v is close approximation of u .

Thus, we can decouple the above minimization problem of Equation A.4 into two problems:

- For fixed v , solve

$$\min_u \int \left[|\nabla u| + \frac{1}{2\theta} (u - v)^2 \right] dx \quad (\text{A.5})$$

This is solved by using *Chambolle's approach* [Chambolle 2004] and the solution is given by $u = v - \theta \text{div}(\underline{p})$, where $\text{div}(\cdot)$ is the divergence and $\underline{p}$ is obtained iteratively as

$$\underline{p}^{k+1} = \frac{\underline{p}^k + \tau \nabla \left(\text{div}(\underline{p}^k) - \frac{v}{\theta} \right)}{1 + \tau \left| \nabla \left(\text{div}(\underline{p}^k) - \frac{v}{\theta} \right) \right|} \quad (\text{A.6})$$

with $\underline{p}^0 = \underline{0}$ and time step $\tau \leq 1/8$

- For fixed u , solve

$$\min_v \int \left[\lambda |\rho(v)| + \frac{1}{2\theta} (u - v)^2 \right] dx \quad (\text{A.7})$$

The solution is

$$v = u + \begin{cases} \lambda \theta I_L^x & \text{if } \rho(u) \leq -\lambda \theta (I_L^x)^2 \\ -\lambda \theta I_L^x & \text{if } \rho(u) > -\lambda \theta (I_L^x)^2 \\ -\frac{\rho(u)}{I_L^x} & \text{otherwise} \end{cases} \quad (\text{A.8})$$

Equation A.8 can be equivalently considered as soft-thresholding of u to obtain v .

Since the original problem is a non-convex function and is transformed into convex problem by linearization, the algorithm is valid only for small displacements and hence the energy minimization procedure is embedded into a coarse-to-fine approach. The pyramid is constructed with a down-sampling factor of 2. Beginning with coarsest level, we solve the algorithm and the results are propagated downwards the pyramid.

Since the dimensions of sub-bands at resolution level j are half of those at level $j - 1$, the disparity vectors are multiplied by a factor of 2 while propagating

downwards the pyramid. The coarse-to-fine approach not only accelerates the convergence, but also avoids local minima. It is important to note that the above equation may yield non integer values of the disparity vectors. In that case, the corresponding right image pixel is obtained by using bilinear interpolation.

$$u_{j-1}(x, y) = 2u_j(2x, 2y) \quad \text{for } j = N \text{ to } 1 \quad (\text{A.9})$$

Structural Similarity Metric

Structural Similarity Metric - SSIM

Subjective and quantitative evaluation of the distortion introduced due to the watermarking process has been analyzed. The watermark robustness depends directly on the embedding strength, which influences the visual degradation on the image. The quality comparison metric used in this procedure is the structural similarity (SSIM), proposed by Wang *et al.* [Wang 2004]. This metric is a block-based algorithm. It measures image quality by luminance, contrast, and structure information. The input is two gray images of the same size. The output is a real number between zero and one, where a larger number means higher similarity. It is defined as

$$\text{SSIM}(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \quad (\text{B.1})$$

where x and y represent two test images; functions $l()$, $c()$, and $s()$ correspond to luminance, contrast, and structure similarity, respectively; α , β , and γ are weighting factors. In the simulation, a simplified form is used

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (\text{B.2})$$

where μ represents the mean; σ represents the (co)variance; C_1 and C_2 are numerical constants for stability (by default 6.5025 and 58.5225 are used). This metric has the following properties:

- Symmetry: $\text{SSIM}(x, y) = \text{SSIM}(y, x)$;
- Boundedness: $\text{SSIM}(x, y) \leq 1$;
- Unique maximum: $\text{SSIM}(x, y) = 1$ iff $x = y$.

These properties make it easy to interpret an SSIM score. The advantage of this metric is that it is less sensitive to noise than conventional quality metrics such as the peak signal to noise ratio or the mean square error. It has been shown to be more suitable for image quality assessment [Wang 2004]. Another reason to use structural similarity is that most information contained in a natural scene image is about structures.

JPEG2000 in DCI

High quality digital cinema exhibition is actually viable due to recent developments in areas such as high-resolution film scanners, digital projection, and image compression. To establish broad-based commercially viable digital cinema exhibition, some kind of industry-wide standards would need to be created. Digital Cinema Initiatives (DCI) was created in March 2002 [1] as a joint venture of seven major Hollywood studios (Disney, Fox, Paramount, MGM, Sony Pictures Entertainment, Universal and Warner Bros Studios) to establish a specification for an open architecture for digital cinema. Such a standard would ensure the members of the movie industry that their products and services will be compatible and interoperable with the products and services of other industry members. The final version of the DCI Digital Cinema Specification was published online in July 2005 [1] selecting JPEG2000 as the compression format for future digital distribution of motion pictures. The image compression standard JPEG2000 proposed a large set of features although, unfortunately, its complexity is greater than older standards. This complexity can be a problem for real-time applications as Digital Cinema. However, a video format named Motion JPEG2000 has been designed, which encapsulates JPEG2000 frames and enables synchronization with audio data [2][13]. In this section, we provide a brief overview of the DCI specification and discuss how JPEG2000 is utilized within this specification.

C.1 Digital Cinema system overview

We present a brief overview of the Digital Cinema System as defined by the DCI specification. As depicted in Fig. A.1, a digital cinema system can be divided into four stages: Mastering, Transport, Storage and Playback, and Projection. The DCI specification discusses each of these stages. Figure A.1. Digital Cinema System The output of the digital cinema post-production process is referred to as the Digital Cinema Distribution Master (DCDM). At the mastering stage, the image data in the DCDM is compressed using JPEG2000, then encrypted and packaged for delivery to the theatres. There are two image structures defined in DCDM: 2K resolution (up to 2160x1080 pixels) and 4K resolution (up to 4096x2160 pixels). A device-independent $X'Y'Z'$ colour space is used.

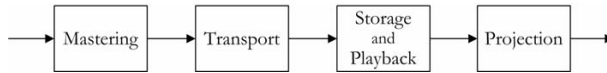


Figure C.1: Digital Cinema system

The bit depth of each colour component is 12 bits. The frame rate is set to be 24Hz. In addition, a frame rate of 48Hz is also allowed for 2K content. It is worth discussing the reason behind allowing two different image resolutions. It is anticipated that there will be both 2K and 4K projectors installed in theatres. The studios can choose to deliver a movie in either 2K or 4K format. The multi-resolution scalability properties of JPEG2000 may be utilized for this purpose. The data is then transported to the exhibition site. The DCI specification does not specify a particular mode of transport but no loss is allowed during transmission. At the exhibition site, the data is unpackaged, unencrypted, and uncompressed. The DCI specification also contemplates the insertion of forensic watermarks¹ into both audio and images. It is required the forensic watermarking data payload to be at least 35 bits. Out of these stages. A forensic watermark uniquely identifies the owner and the origin of a digital document. It is usually repeated at random locations within the content, making it difficult to detect or remove.

35 bits, 16 bits are allotted for a time stamp that can identify the time of the year within 15 minutes. The remaining 19 bits are used to identify the location. All 35-bits are required to be included in each five minute segment. The insertion of the forensic watermarks is required to be performed in real-time. The specifications allow the use of up to 30 minutes of content for recovery of the forensic watermarks. DCI specifications define also several aspects of the projection system, whose function is to convert the digital image data into light appearing on the screen.

C.2 DCI specifications for JPEG2000

The DCI specification defines the size of each movie frame to be as large as 4096x2160 pixels. With three colour components, 12 bits/pixel/colour component, and 24 frames/second, the total size of a three hour feature film exceeds 9 terabytes. Such large sizes make the distribution of uncompressed digital movies impractical. The DCI specification requires frames to be compressed individually via JPEG2000 which is required to support high bit depth (12 bits per colour component), the $X'Y'Z'$ colour space without chroma subsampling and both 2K and 4K resolution projectors from the same file. The JPEG2000 standard is published in multiple parts. Part I describes the minimal compliant

decoder and the code-stream syntax. Other parts of the standard describe other features and added-value technologies such as the Motion JPEG2000 file format. The DCI specification is based on Part-1 of the JPEG2000 standard and the particular set of parameters that will be used in digital cinema applications are defined in JPEG2000 profiles. A JPEG2000 profile is a set of parameters that are designed to best serve the needs of a particular application. Current JPEG2000 profiles to support digital cinema applications are the 2K digital cinema profile and the 4K digital cinema profile. Following, we present a general operation overview of JPEG2000 standard and, finally, some of the interesting restrictions over JPEG2000 of the DCI specification to be used in digital cinema applications.

C.2.1 JPEG2000 overview

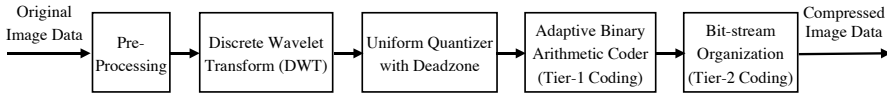


Figure C.2: Fundamental blocks in JPEG2000 standard

Coding

After the possible application of a colour transform (ICT or RCT), the image is split into rectangular blocks called tiles which will be compressed individually. Discrete wavelet transform is carried out on each tile. Then, for lossy JPEG2000 compression, the wavelet coefficients are quantized with different quantization step sizes depending on the resolution level and the sub-band they belong to. Every sub-band is then split into rectangular entities called code-blocks. Each code-block will be quantized and compressed independently from the others using a context-based adaptive entropy coder. It reduces the amount of data without losing information by removing redundancy present in the binary sequence. Practically, entropy coding consists of first encoding the bits which contribute to the largest distortion reduction for the smallest increase in file size as in the Embedded Block Coding with Optimized Truncation (EBCOT) algorithm. The coefficients in the code-block are bit-plane encoded, starting with the most significant bit-plane. Each bit-plane is encoded (arithmetic coding) in three passes with the possibility of truncation the bit-stream at the end of each coding pass. During the rate allocation step, segment from each code-block are scanned in order to find the optimal truncation points to achieve various targeted bit-rates. Quality layers are then created using the incremental contributions from each code-block. Compressed data corresponding to the same component, resolution, spatial region and quality layer is then inserted in a packed. Packets, along with additional headers, form the final JPEG2000

code-stream.

Decoding

First, the bit-stream analyzer analyses the headers, unpacks the compressed data and sends them to the entropy decoding unit. Entropy decoder gets compressed coding passes from the parser and converts them to quantized code blocks. Thus, code-blocks are reconstructed bit plane by bit plane from most significant to least significant, and relevant bits are placed in zigzag order in each bit plane. The arithmetic decoder generates, then, the decoded bits placed at the right location. Sub-bands are reconstructed based on the code blocks coming from the entropy decoding channels. The inverse quantizer applies inverse quantization steps as specified in the stream header. A different inverse quantization step is available for each sub-band resulting in differently weighted frequency sub-bands. Then, inverse 2D wavelet transform is performed.

C.2.2 DCI restrictions for JPEG2000

DCI Digital Cinema Specification [1] defines also some restrictions to ensure compatibility. A general overview of some of these restrictions concerning the watermarking context used in this document is presented as follows: . DCI specification requires all decoders to decode each colour component at 12 bits/sample. And, as stated above, chroma subsampling is disallowed. . Both profiles require the use of the 9/7 irreversible wavelet transform. . The use of the irreversible colour transform (ICT) is required. Although the ICT is the well-known RGB to YCbCr transform, in this case, the input color space is $X'Y'Z'$. Thus, the transformed components do not correspond exactly to Y, Cb, and Cr. . Full-frame coding and decoding (tiling is disallowed): the entire image should be encoded as a single tile. . The maximum number of wavelet transform levels is 5 for 2K content and 6 for 4K content. . Code-block sizes are restricted to be 32x32. . Precinct sizes at all resolutions are set to be 256x256, except for the lowest frequency sub-band, where a precinct size of 128x128 is used. . Each compressed code-stream of a 2K distribution has 3 tile parts where each tile part contains one colour component. Similarly, each compressed code-stream of a 4K distribution has 6 tile parts. . Only a single quality layer is allowed. . There are constraints on the maximum sizes of code-stream segments as well as the total size of the code-stream that correspond to a maximum total rate of 250 Mbits per second.

Bibliography

- [Arnold 2005] M. Arnold, J. Dittmann, W. Funk, R. Norcen, W. Spinnler, M. Steinebach and S. Voloshynovski. *First Summary Report on Forensic Tracking*. Rapport technique, European Network of Excellence for Cryptology (ECRYPT), 2005. 111
- [Bae 2008] K.-H. Bae, D.-C. Hwang and E.-S. Kim. *Stereo Vision Based Digital Watermarking Using Regularized Feature-based Disparity Estimation*. Journal of Imaging Science and Technology, vol. 52, no. 5, page 050503, 2008. 126
- [Barni 1998] M. Barni, F. Bartolini, V. Cappellini and A. Piva. *A DCT-domain system for robust image watermarking*. Signal Process., vol. 66, pages 357–372, May 1998. 73
- [Barni 2001] M. Barni, F. Bartolini and A. Piva. *Improved wavelet-based watermarking through pixel-wise masking*. Image Processing, IEEE Transactions on, vol. 10, no. 5, pages 783–791, May 2001. 93, 98, 126
- [Barton 1997] J. M. Barton. *Method and apparatus for embedding authentication information within digital data*, July 1997. United States Patent 5646997. 76
- [Bhatnagar 2009] G. Bhatnagar, S. Kumar, B. Raman and N. Sukavanam. *Stereo image coding via digital watermarking*. J. Electronic Imaging, vol. 18, no. 3, page 033012, 2009. 126
- [Boulgouris 2001] N. Boulgouris, D. Tzovaras and M. Strintzis. *Lossless image compression based on optimal prediction, adaptive lifting, and conditional arithmetic coding*. Image Processing, IEEE Transactions on, vol. 10, no. 1, pages 1–14, Jan 2001. 43, 44
- [Calderbank 1998] R. Calderbank, I. Daubechies, W. Sweldens and B.-L. Yeo. *Wavelet transforms that map integers to integers*. Appl. Comput. Harmon. Anal., vol. 5, no. 3, pages 332–369, 1998. 24
- [Callau 2009] P. Callau, R. Darazi and B. Macq. *Exhibition QIM-based Watermarking for Digital Cinema*. In Media Forensics and Security. SPIE, in Proc. 2009. 89
- [Campisi 2008] P. Campisi. *Object-oriented stereo-image digital watermarking*. J. Electronic Imaging, vol. 17, no. 4, page 043024, 2008. 126
- [Candès 1999] E. J. Candès and E. J. C. *Monoscale Ridgelets for the Representation of Images with Edges*, 1999. 23

- [Cannons 2004] J. Cannons and P. Moulin. *Design and statistical analysis of a hash-aided image watermarking system*. IEEE Transactions on Image Processing, vol. 13, no. 10, pages 1393–1408, October 2004. 111
- [car a] <http://vasc.ri.cmu.edu/idb/html/jisct/index.html>. 33, 38, 54
- [car b] <http://vasc.ri.cmu.edu/idb/html/stereo/index.html>. 33, 38, 54
- [Cayre 2005] F. Cayre, C. Fontaine and T. Furon. *Watermarking security: theory and practice*. Signal Processing, IEEE Transactions on, vol. 53, no. 10, pages 3976 – 3987, oct. 2005. 115
- [Chambolle 2004] A. Chambolle. *An algorithm for total variation minimization and applications*. Journal of Mathematical Imaging and Vision, pages 89–97, 2004. 140
- [Chen 2001] B. Chen and G. Wornell. *Quantization index modulation: a class of provably good methods for digital watermarking and information embedding*. Information Theory, IEEE Transactions on, vol. 47, no. 4, pages 1423–1443, May 2001. 75, 76, 84, 111, 114, 116
- [Chi 2007] Y. Chi, T. Tran and R. Etienne-Cummings. *Optical Flow Approximation of Sub-Pixel Accurate Block Matching for Video Coding*. In Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, volume 1, pages I–1017 –I–1020, 15-20 2007. 35, 36
- [Costa 1983] M. Costa. *Writing on dirty paper (Corresp.)*. Information Theory, IEEE Transactions on, vol. 29, no. 3, pages 439 – 441, may 1983. 76
- [Cox 1997] I. Cox, J. Kilian, F. Leighton and T. Shamoon. *Secure spread spectrum watermarking for multimedia*. Image Processing, IEEE Transactions on, vol. 6, no. 12, pages 1673–1687, Dec 1997. 75, 111, 114
- [Darazi 2009a] R. Darazi, P. Callau and B. Macq. *Secure and HVS-adaptive exhibition Spread Transform Dither Modulation watermarking for Digital Cinema*. In Information Forensics and Security, 2009. WIFS 2009. First IEEE International Workshop on, pages 1 –5, 2009. 89, 127, 128
- [Darazi 2009b] R. Darazi, A. Gouze and B. Macq. *Adaptive Lifting Scheme-based method for joint coding 3D-stereo images with luminance correction and optimized prediction*. In Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, pages 917 –920, 2009. 38, 43
- [Darazi 2009c] R. Darazi, A. Gouze and B. Macq. *Lifting Scheme-based Method for Joint Coding 3D-Stereo Digital Cinema with Luminance Correction and Optimized Prediction*. In Visual Communication and Image Processing VCIP. SPIE, in Proc. 2009. 43

- [Darazi 2011] R. Darazi, M. Montanola, L. Weng, B. Macq and B. Preneel. *Disparity Guided Exhibition Watermarking for 3D Stereo Images*. In Security and Privacy in e-Societies, 2011. SECES 2011. ACM First International Workshop on, pages 1–8, 2011. 125
- [Daubechies 1998] I. Daubechies and W. Sweldens. *Factoring wavelet transforms into lifting steps*. Journal of Fourier Analysis and Applications, vol. 4, no. 3, pages 247–269, 1998. 23, 24, 52
- [Davis 1995] C. Davis, Z. Karul and D. Freeman. *Equivalence of subpixel motion estimators based on optical flow and block matching*. In Computer Vision, 1995. Proceedings., International Symposium on, pages 7–12, 21-23 1995. 35
- [DCI] *Digital Cinema Initiatives, LLC*. 5, 92, 100, 101
- [Delac 2006] K. Delac, M. Gregic and T. Kos. *Sub-Image Homomorphic Filtering Technique for Improving Facial Identification under Difficult Illumination Conditions*. 2006. 28
- [Desmedt 1998a] Y. Desmedt, S. Hou and J.-J. Quisquater. *Audio and Optical Cryptography*. In ASIACRYPT, pages 392–404, 1998. 126
- [Desmedt 1998b] Y. Desmedt, S. Hou and J.-J. Quisquater. *Cerebral Cryptography*. In Information Hiding, pages 62–72, 1998. 126
- [Ding 2007] W. Ding, F. Wu, X. Wu, S. Li and H. Li. *Adaptive Directional Lifting-Based Wavelet Transform for Image Coding*. Image Processing, IEEE Transactions on, vol. 16, no. 2, pages 416–427, feb. 2007. 23
- [Do 2003] M. N. Do and M. Vetterli. *The Finite Ridgelet Transform for Image Representation*. IEEE Transactions on Image Processing, vol. 12, pages 16–28, 2003. 23
- [Do 2005] M. N. Do and M. Vetterli. *The Contourlet Transform: An Efficient Directional Multiresolution Image Representation*. IEEE Transactions on Image Processing, vol. 14, no. 12, pages 2091–2106, 2005. 23
- [Dugad 1998] R. Dugad, K. Ratakonda and N. Ahuja. *A new wavelet-based scheme for watermarking images*. In Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on, volume 2, pages 419–423 vol.2, oct 1998. 73
- [Fecker 2008] U. Fecker, M. Barkowsky and A. Kaup. *Histogram-Based Pre-filtering for Luminance and Chrominance Compensation of Multiview Video*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 18, no. 9, pages 1258–1267, sept. 2008. 27, 28

- [Frajka 2002] T. Frajka and K. Zeger. *Residual Image Coding for Stereo Image Compression*. ICIP International Conference on Image Processing, IEEE, 2002. 3, 19
- [Fridrich 1999] J. Fridrich. *Robust bit extraction from images*. In Proc. of IEEE International Conference on Multimedia Computing and Systems, volume 2, pages 536–540, 1999. 111
- [Fridrich 2000] J. Fridrich and M. Goljan. *Robust hash functions for digital watermarking*. In Proc. of International Conference on Information Technology: Coding and Computing, 2000. 111
- [Gerek 2006] O. Gerek and A. Cetin. *A 2-D orientation-adaptive prediction filter in lifting structures for image coding*. Image Processing, IEEE Transactions on, vol. 15, no. 1, pages 106–111, jan. 2006. 23
- [Gouze 2004] A. Gouze, M. Antonini, M. Barlaud and B. Macq. *Design of signal-adapted multidimensional lifting scheme for lossy coding*. Image Processing, IEEE Transactions on, vol. 13, no. 12, pages 1589–1603, 2004. 43, 44, 47
- [Haitsma 2001] J. Haitsma and T. Kalker. *A watermarking scheme for digital cinema*. In Image Processing, 2001. Proceedings. 2001 International Conference on, volume 2, pages 487–489 vol.2, oct 2001. 90
- [Hur 2007] J. Hur, S. Cho and Y. Lee. *Adaptive Local Illumination Change Compensation Method for H.264/AVC-Based Multiview Video Coding*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 17, no. 11, pages 1496–1505, Nov. 2007. 30
- [Ingemar J. Cox 2008] J. A. B. J. F. Ingemar J. Cox Matthew L. Miller and T. Kalker. Digital watermarking and steganography. Morgan Kaufmann, second edition édition, 2008. 68, 70
- [Kaaniche 2009a] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu and J.-C. Pesquet. *Vector Lifting Schemes for Stereo Image Coding*. Image Processing, IEEE Transactions on, vol. 18, no. 11, pages 2463–2475, nov. 2009. 21
- [Kaaniche 2009b] M. Kaaniche, W. Miled, B. Pesquet-Popescu, A. Benazza-Benyahia and J.-C. Pesquet. *Dense disparity map representations for stereo image coding*. In Image Processing (ICIP), 2009 16th IEEE International Conference on, pages 725–728, 7-10 2009. 20, 36, 38
- [Kitanovski 2008] V. Kitanovski, M. Kseneman, D. Gleich and D. Taskovski. *Adaptive lifting integer wavelet transform for lossless image compression*. Systems, Signals and Image Processing, 2008. IWSSIP 2008. 15th International Conference on, pages 105–108, June 2008. 23

- [KL-] http://en.wikipedia.org/wiki/Kullback-Leibler_divergence. 29
- [Knutsson 2005] H. Knutsson and M. Andersson. *Morphons: segmentation using elastic canvas and paint on priors*. volume 2, pages II-1226-9, Sept. 2005. 105
- [Li 2007] Q. Li and I. Cox. *Using Perceptual Models to Improve Fidelity and Provide Resistance to Volumetric Scaling for Quantization Index Modulation Watermarking*. Information Forensics and Security, IEEE Transactions on, vol. 2, no. 2, pages 127-139, June 2007. 93, 126
- [Lipton 1982] L. Lipton. *Foundation of the stereoscopic cinema*. Van Nostrand Reinhold Company Inc., 1982. 16, 126
- [Lubin 2003] J. Lubin, J. A. Bloom and H. Cheng. *Robust content-dependent high-fidelity watermark for tracking in digital cinema*. 2003. 5, 90
- [Maalouf 2010] A. Maalouf and M.-C. Larabi. *Bandelet-based stereo image coding*. In Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, pages 698-701, march 2010. 23
- [Mao 2007] Y. Mao and M. Wu. *Unicity distance of robust image hashing*. IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pages 462-467, September 2007. 115, 118
- [McNerney 2007] P. McNerney, J. Konrad and M. Betke. *Block-Based MAP Disparity Estimation Under Alpha-Channel Constraints*. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, vol. 17, no. 6, page 785, 2007. 3, 19, 21
- [Menezes 1996] A. Menezes, P. van Oorschot and S. Vanstone. *Handbook of applied cryptography*. CRC Press, 1996. 111
- [mid] <http://vision.middlebury.edu/stereo/data/scenes2001/>. 33, 38, 54
- [Moellenhoff 1998] M. Moellenhoff and M. Maier. *Transform coding of stereo image residuals*. Image Processing, IEEE Transactions on, vol. 7, no. 6, pages 804-812, jun 1998. 21
- [Morvan 2007] Y. Morvan, D. Farin and P. de With. *Depth-Image Compression Based on an R-D Optimized Quadtree Decomposition for the Transmission of Multiview Images*. In Image Processing, 2007. ICIP 2007. IEEE International Conference on, volume 5, pages V-105-V-108, sept. 2007. 36
- [Moulin 2005] P. Moulin and R. Koetter. *Data-Hiding Codes*. Proceedings of the IEEE, vol. 93, no. 12, pages 2083-2126, dec. 2005. 82

- [Moulin 2010] P. Moulin. *Statistical modeling and analysis of content identification*. In Information Theory and Applications Workshop (ITA), pages 1–5, feb. 2010. 111
- [Nikolaidis 1998] N. Nikolaidis and I. Pitas. *Robust Image Watermarking in the Spatial Domain*. In Signal Processing, pages 385–403, 1998. 73
- [Pereira 2000] S. Pereira and T. Pun. *Robust template matching for affine resistant image watermarks*. Image Processing, IEEE Transactions on, vol. 9, no. 6, pages 1123 –1129, jun 2000. 73
- [Perez-Gonzalez 2005] F. Perez-Gonzalez, C. Mosquera, M. Barni and A. Abrardo. *Rational dither modulation: a high-rate data-hiding method invariant to gain attacks*. Signal Processing, IEEE Transactions on, vol. 53, no. 10, pages 3960 – 3975, oct. 2005. 83
- [Perkins 1992] M. Perkins. *Data compression of stereopairs*. Communications, IEEE Transactions on, vol. 40, no. 4, pages 684 –696, apr 1992. 21
- [Piella 2002] G. Piella and H. Heijmans. *Adaptive lifting schemes with perfect reconstruction*. Signal Processing, IEEE Transactions on, vol. 50, no. 7, pages 1620–1630, Jul 2002. 23
- [Piella 2005] G. Piella, G. Pau and B. Pesquet-Popescu. *Adaptive lifting schemes combining seminorms for lossless image compression*. Image Processing, 2005. ICIP 2005. IEEE International Conference on, vol. 1, pages I–753–6, Sept. 2005. 23
- [Rabbani 2002] M. Rabbani and R. Joshi. *An overview of the JPEG2000 still image compression standard*. In Signal Processing: Image Communication, pages 3–48, 2002. 4, 19
- [Radhakrishnan 2005] R. Radhakrishnan, Z. Xiong and N. Memon. *On the security of the visual hash function*. Journal of Electronic Imaging, vol. 14, page 10, 2005. 115
- [Rawat 2010] S. Rawat, G. Gupta, R. Balasubramanian and M. S. Rawat. *Digital Watermarking Based Stereo Image Coding*. In IC3 (1), pages 435–445, 2010. 126
- [Schneider 1996] M. Schneider and S.-F. Chang. *A robust content based digital signature for image authentication*. In Proc. of International Conference on Image Processing (ICIP96), volume 3, pages 227–230, 1996. 111
- [Schneier 1996] B. Schneier. Applied cryptography: Protocols, algorithms, and source code in c. John Wiley & Sons, 2nd édition, 1996. 111

- [Swaminathan 2006] A. Swaminathan, Y. Mao and M. Wu. *Robust and secure image hashing*. IEEE Transactions on Information Forensics and Security, vol. 1, no. 2, pages 215–230, June 2006. 111
- [Swanson 1997] M. Swanson, B. Zhu and A. Tewfik. *Data hiding for video-in-video*. In Image Processing, 1997. Proceedings., International Conference on, volume 2, pages 676–679 vol.2, oct 1997. 76
- [Sweldens 1995a] W. Sweldens. *The lifting scheme: A new philosophy in biorthogonal wavelet constructions*. Wavelet Applications in Signal and Image Processing III, SPIE, pages 68–79, 1995. 24
- [Sweldens 1995b] W. Sweldens. *The Lifting Scheme: A New Philosophy in Biorthogonal Wavelet Constructions*. In A. F. Laine and M. Unser, editors, Wavelet Applications in Signal and Image Processing III, pages 68–79. Proc. SPIE 2569, 1995. 25
- [Sweldens 1996a] W. Sweldens. *The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets*. Applied and Computational Harmonic Analysis, vol. 3, no. 2, pages 186–200, 1996. 21, 24
- [Sweldens 1996b] W. Sweldens and P. Schröder. *Building Your Own Wavelets at Home*. In Wavelets in Computer Graphics, pages 15–87. ACM SIGGRAPH Course Notes, 1996. 23
- [tu-] <http://www.ph.tn.tudelft.nl/Courses/FIP/noframes/fip-Shading-2.html>. 30
- [Venkatesan 2000] R. Venkatesan, S.-M. Koon, M. Jakubowski and P. Moulin. *Robust image hashing*. In Proc. of IEEE International Conference on Image Processing, volume 3, pages 664–666, Vancouver, CA, 2000. 111
- [Voloshynovskiy 2001] S. Voloshynovskiy, S. Pereira, V. Iquise and T. Pun. *Attack Modelling: Towards a Second Generation Watermarking Benchmark*. In Signal Processing, Special Issue on Information Theoretic Issues in Digital Watermarking, pages 1177–1214, 2001. 87
- [Vural 2005] S. Vural, H. Tomii and H. Yamauchi. *Robust Digital Cinema Watermarking*. International Journal of Information Technology, vol. 2, no. 2, pages 255–259, 2005. 90
- [Wang 2004] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli. *Image quality assessment: from error visibility to structural similarity*. Image Processing, IEEE Transactions on, vol. 13, no. 4, pages 600–612, april 2004. 55, 119, 141
- [Watson 1993] A. B. Watson. *DCT quantization matrices visually optimized for individual images*. 1993. 74

- [Weng 2011] L. Weng, R. Darazi, B. Preneel, B. Macq and A. Dooms. *Image Content Authentication Using Robust Hashing and Watermarking*. In Information Forensics and Security, 2011. WIFS 2011. IEEE International Workshop on, pages 1–6, submitted, 2011. 109
- [Wolfgang 1996] R. Wolfgang and E. Delp. *A watermark for digital images*. In Image Processing, 1996. Proceedings., International Conference on, volume 3, pages 219–222 vol.3, sep 1996. 87, 110
- [Woo 1996] W. Woo and A. Ortega. *Stereo Image Compression with Disparity Compensation Using the MRF Model*. In Proc. SPIE VCIP, pages 28–41, 1996. 36
- [Woo 2000] W. Woo and A. Ortega. *Overlapped block disparity compensation with adaptive windows for stereo image coding*. IEEE Trans. Circuits Syst. Video Techn., vol. 10, no. 2, pages 194–200, 2000. 19, 20
- [Xia 1998] X. G. Xia, C. Boncelet and G. Arce. *Wavelet transform based watermark for digital images*. Opt. Express, vol. 3, no. 12, pages 497–511, Dec 1998. 73
- [Xu 2002] J. Xu, Z. Xiong and S. Li. *High Performance Wavelet-Based Stereo Image Coding*. ISCAS International Symposium on Circuits and Systems, IEEE, 2002. 21
- [Yu 2008] D. Yu, L. Ma, G. Wang and H. Lu. *Adaptive spread-transform dither modulation using an improved luminance-masked threshold*. pages 449–452, Oct. 2008. 93
- [Zach 2007] C. Zach, T. Pock and H. Bischof. *A Duality Based Approach for Realtime TV-L1 Optical Flow*. In Pattern Recognition (Proc. DAGM), pages 214–223, Heidelberg, Germany, 2007. 37, 139