



**THESE DE DOCTORAT CONJOINT TELECOM SUDPARIS
et L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité : Informatique et Télécommunication

Ecole doctorale : Informatique, Télécommunications et Electronique de Paris

**Méthodes de tatouage robuste
pour la protection de l'imagerie numérique 3D**

**Présentée par
Afef Chammem**

Soutenu le 27 Mai 2013 à paris

Devant le jury composé de :

Prof. Touradj Ebrahimi	Head of the MMSPG, EPFL	Rapporteur
Prof. Laurent Lucas	Head of the SIC Group, URCA	Rapporteur
Prof. Patrick Gallinari	Head of the LIP6, UPMC	Président
Dr. Frédéric Dufaux	CNRS Research Director, TelecomParisTech	Examineur
Prof. Faouzi Ghorbel	Head of the CRISTAL Lab, ENSI	Examineur
Mme. Maryline Clare	Head of the New Formats and New Coding Project, Orange Labs	Examineur
Prof. Françoise Preteux	Deputy Scientific Director, Mines ParisTech	Directeur de thèse
HDR. Mihai Mitrea	Associate Professor, Telecom SudParis	Co-Directeur de thèse

Table of Contents

Abstract.....	1
Chapter 1 Context.....	5
1.1. Stereoscopic content	7
1.1.1. Content proliferation	8
1.1.2. Device explosion	14
1.1.3. Potential protection techniques.....	17
1.2. Watermarking properties.....	18
1.2.1. Watermarking applicative panorama.....	18
1.2.2. Main watermarking constraints	20
1.2.3. From 2D to stereoscopic video watermarking	26
1.2.4. Theoretical framework for stereo watermarking	28
1.3. Thesis structure.....	30
Chapter 2 Objectives.....	31
2.1. Disparity maps computation.....	33
2.1.1. Basic concepts	33
2.1.2. Block matching disparity map computation in the literature	33
2.1.3. Conclusion	38
2.2. Watermarking embedding techniques.....	38
2.2.1. Principles.....	38
2.2.2. State-of-the-art.....	39
2.2.3. Discussion	43
2.3. Conclusion and thesis objectives.....	48
Chapter 3 A new insertion domain: 3D Video-New Three Step Search	51
3.1. Method presentation	53
3.1.1. Principles.....	53
3.1.2. Pseudo-code and flowchart.....	53
3.2. Experimental validation.....	58
3.2.1. Test bed	58
3.2.2. Horizontal and vertical displacements tendency	59
3.2.3. Visual investigation of the disparity maps.....	60
3.2.4. Reconstructed image quality assessment.....	62
3.2.5. Computational cost	67
3.3. Discussion	70
3.4. Conclusion.....	70

Chapter 4 A reliable watermarking technique: Fast-IProtect	71
4.1. IProtect: hybrid embedding technique.....	73
4.1.1. IProtect embedding scheme	73
4.1.2. IProtect detection.....	78
4.2. Towards Fast-IProtect	78
4.2.1. Statistical models for watermarking attacks	80
4.2.2. Monte Carlo simulation for watermarking attacks.....	83
4.2.3. Experimental validation.....	86
4.3. Conclusion.....	87
Chapter 5 Stereoscopic watermarking benchmark	89
5.1. Embedding procedure	92
5.2. Embedding domains	94
5.3. Transparency evaluation	96
5.3.1. Subjective protocol	96
5.3.2. Objective assessment	104
5.4. Robustness evaluation.....	107
5.5. Computational cost	110
5.5.1. Processing time	110
5.5.2. Algorithmic complexity.....	110
5.6. Discussion	111
Chapter 6 Conclusion and future work.....	113
6.1. Conclusion.....	115
6.2. Future work.....	116
Appendix.....	117
A.1 Processed corpora.....	119
A.2 The considered embedding techniques	124
A.3 Transparency evaluation.....	126
A.4 Robustness evaluation	132
References.....	135
List of acronyms	143
List of publications.....	147

List of figures

Figure 1.1 Binocular human vision vs. stereoscopic content acquisition.....	7
Figure 1.2 Stereoscopic content display.....	8
Figure 1.3 Content market value in the stereo 3D technology segment.....	9
Figure 1.4 Timeline of 3D movies releases [Kun11].....	10
Figure 1.5 Overall TV and 3D TV sales 2009-2014 [Kin11].....	12
Figure 1.6 3D TV content need [Teu10].....	12
Figure 1.7 3D TV channel in the world [Teu10] [Kun11].....	13
Figure 1.8 Professional technologies for 3D TV [3DF13].....	14
Figure 1.9 Digital personal stereo vision systems [Koc10].....	15
Figure 1.10 Global 3D-enabled devices sales forecast from 2010 to 2014.....	15
Figure 1.11 Samples of passive and active glasses for 3D viewer technologies [Kaw02] [Kun11].....	16
Figure 1.12 Lenticular lens and parallax barrier for autostereoscopic displays [Dod05].....	17
Figure 1.13 Samples of autostereoscopic displays [Kun11].....	17
Figure 1.14 The watermarking workflow.....	19
Figure 1.15 “Big Buck Bunny” [Ble08], original frame with samples of its attacked versions.....	24
Figure 1.16 Test image with samples of its corresponding attacked versions.....	25
Figure 1.17 Watermarking as a noisy channel.....	29
Figure 1.18 The theoretical model for stereo noisy channel.....	29
Figure 2.1 Stereo matching.....	33
Figure 2.2 Block searching strategies: (a) ES, (b) NTSS, and (c) FS –MPEG.....	35
Figure 2.3. Synoptic representation of the state-of-the-art studies.....	37
Figure 2.4 The watermarking embedder scheme.....	39
Figure 2.5 Synoptic overview of the state of the art of the stereoscopic watermarking approaches.....	40
Figure 2.6 Synoptic representation of the state-of-the-art studies.....	45
Figure 3.1 Block searching strategies: (a) NTSS, and (b) 3DV-NTSS.....	53
Figure 3.2 The 3DV-NTSS algorithm flowchart.....	56
Figure 3.3 The main three steps of the 3DV-NTSS algorithm.....	57
Figure 3.4 Left and right views sampled from the 3DLive corpus.....	58
Figure 3.5 Left and right views sampled from the MPEG corpus.....	58
Figure 3.6 Search area weights variation (α and β) for 3DLive and MPEG corpora.....	59
Figure 3.7 NTSS (a), FS-MPEG (b) and 3DV-NTSS (c) disparity maps.....	60

Figure 3.8 NTSS (a), FS-MPEG (b), and 3DV-NTSS (c) disparity maps.....	61
Figure 3.9 NTSS (a), FS-MPEG (b), and 3DV-NTSS (c) disparity maps.....	61
Figure 3.10 NTSS (a), FS-MPEG (b), and 3DV-NTSS (c) disparity maps.....	62
Figure 3.11 Objective evaluation of the reconstructed image quality.	66
Figure 4.1 The watermarking method synopsis.	73
Figure 4.2 Treillis representation. Each row corresponds to one of the K states.....	74
Figure 4.3 The selected sub-bands.....	75
Figure 4.4 The embedded ARTEMIS logo.....	75
Figure 4.5 The IProtect embedding algorithm synopsis.	77
Figure 4.6 IProtect time processing.....	79
Figure 4.7 The five steps undergone when applying an attack procedure.	79
Figure 4.8 Fast-IProtect embedding algorithm synopsis.	80
Figure 4.9 The Monte Carlo attack simulation algorithm flowchart.....	85
Figure 4.10 The quality of the Monte Carlo generators for the StirMark attack.	86
Figure 4.11 The attack procedure during the embedding process.....	87
Figure 4.12 Time processing for different watermarking operations.	87
Figure 5.1 The watermarking assessment procedure.....	91
Figure 5.2 View-based watermarking scheme.	95
Figure 5.3 Disparity-based watermarking scheme.....	96
Figure 5.4 Presentation structure of the assessment session.....	99
Figure 5.5 Side-by-side spread spectrum watermarked samples.....	99
Figure 5.6 Side-by-side 2-QIM watermarked samples.....	100
Figure 5.7 Side-by-side 5-QIM watermarked samples.....	101
Figure 5.8 Side-by-side Fast-IProtect watermarked samples.....	101
Figure 5.9 Subjective evaluations for image quality, depth perception and visual comfort.....	104
Figure 5.10 Objective evaluation of the watermarked content visual quality.....	106
Figure 5.11 Watermark robustness against five different attacks.	108
Figure 5.12. Watermarking chain computational cost.....	110
Figure A.1.1 Left and right views sampled from the 3DLive corpus.....	119
Figure A.1.2 Left and right views sampled from the MPEG corpus.	120
Figure A.1.3 Experimental HD3D ² database.....	121
Figure A.1.4 Experimental EPFL database.....	121
Figure A.1.5 Experimental ARTEMIS database.....	122

List of tables

Table 1.1 The highest-grossing 3D films: Costs vs. incomes (in USD million).....	11
Table 1.2 Data payload requirements for some examples of watermarking applications.	26
Table 2.1 Performance evaluation between the state of the art algorithms.	36
Table 2.2 A comparative view on the state of the art of stereoscopic image/video watermarking methods.	46
Table 2.3 The thesis main challenges, the underlying current limitations and the thesis objectives.....	49
Table 3.1 Reconstructed image visual quality, expressed by PSNR, IF, NCC, SC and SSIM.	64
Table 3.2 95% error limits in reconstructed image quality assessment.	65
Table 3.3 Computational cost expressed in number of search points per 16x16 block.	68
Table 3.4 Gain factor on computational cost 3DV-NTSS vs. NTSS and 3DV-NTSS vs. FS-MPEG.....	69
Table 4.1 Attacks statistical models in the (9,7) <i>DWT</i> hierarchy.....	82
Table 5.1 The considered insertion methods and their main parameters.....	93
Table 5.2 General viewing conditions for subjective assessments at the ARTEMIS laboratory environment. ...	97
Table 5.3 Preferred Viewing Distance-PVD for moving image.....	98
Table 5.4 A sample of a transparency: subjective evaluation sheet.	98
Table 5.5 Computational complexity of the watermarking modules involved in Fast-IProtect/3DV-NTSS.....	111
Table A.3.1 The considered attacks and their main parameters.	127
Table A.3.2 The Mean Opinion Score values for a valid number of 25 subjects.	128
Table A.3.3 95% error of the MOS for a valid number of 25 subjects.	129
Table A.3.4 Watermarking transparency: 3DV-NTSS vs. NTSS vs. FS-MPEG, see Figure 5.10.	130
Table A.3.5 95% error in watermarked transparency: 3DV-NTSS vs. NTSS vs. FS-MPEG, see Figure 5.10.	131

Abstract

Context

Coined in the mid' 50s for audio signals, the stereoscopic world is nowadays the challenge in image/video processing. On the one hand, medical diagnosis, fault detection in manufactory industry, army, arts all of them consider multi-view imaging as a key enabler for professional added value services. On the other hand, the worldwide success of movie releases (Avatar, Alice in Wonderland) and the deployment of 3D TV chains made the non-professional user aware about a new type of multimedia entertainment experience.

This explosion in stereoscopic video distribution increases the concerns over its copyright protection. Watermarking can be considered as the most flexible property right protection technology, since it potentially supports all the requirements set by real life applications without imposing any constraints for a legitimate user. Actually, instead of restricting the media distribution, watermarking provides means for tracking the source of the illegitimate usage.



Watermarking and its properties (transparency, robustness and data payload), from the embedding to the detection.

Applicative constraints

The watermarking applicative issue is to reach the trade-off between the properties of transparency, robustness and data payload. The transparency refers to the imperceptibility of the embedded additional information in the watermarked media. The robustness is the ability of detecting the watermark after applying some mundane or malicious attacks on the marked document (spatial filtering, lossy compression, recording, etc). The data payload represents the quantity of information that is inserted into the host document. In a real time context, the embedding and detecting modules must not slow down the production chain.

A specific study on the optimal (with respect to the above-mentioned properties) insertion domain is also required. While the capturing and displaying of the 3D content are solely based on the two left/right views, some alternative representations, like the disparity maps should also be considered during transmission/storage.

In order to pave the way from methodological development towards real life applications, the watermarking benchmarking should be application independent, reinforced by statistical relevance and fostered by standard recommendations.

Scientific and technical challenges

For 2D video watermarking applications, several classes of insertion techniques already proved their efficiency. For instance, SS (Spread Spectrum) techniques are generally connected to excellent transparency properties while ensuring good robustness for a quite small size of the mark. Conversely, SI (Side Information) techniques result in large sizes of the inserted marks, while decreasing the transparency/robustness properties. These two directions have been also extended for stereoscopic watermarking system. However, they still lack in achieving good transparency, they deal only with a restricted class of attacks and fail in embedding the required amount of data payload. Moreover, the transparency evaluation was solely guided by the PSNR values; no other objective transparency metrics or subjective evaluation protocols have been deployed. Consequently, specifying a watermarking method able to reach this three-folded trade-off is still an open issue. Such a study should also be accompanied by an evaluation of the computational cost.

Nowadays, stereoscopic video watermarking studies let the insertion domain selection to the experimenter's choice rather than identifying it by objective studies. This way, the disparity maps are directly inherited from 2D video and simply neglect the stereoscopic video peculiarities. For such a content, the disparities between the left and the right views are predominant on the horizontal direction (where the very depth information is conveyed) while the vertical disparities, although non-zero, are mainly connected to the video shooting errors. Moreover, in the block matching stage, basic visual quality metrics like mean squared error – MSE or sum of absolute differences - SAD are considered. As such metrics are unrelated to the human visual system, the obtained results are sub-optimal from a quality of experience perspective. Consequently, specifying a disparity map, jointly exploiting the horizontal/vertical peculiarities of the stereoscopic content and a visual quality metric related to the HVS remains a challenging research topic.

Concerning the experimental validation, the processed data sets are restricted in their content heterogeneity and their size. Consequently, constructing a large and representative corpus to ensure statistical relevance for the results and performing the evaluation protocol according to some referenced standards are still desiderata.

Methodological contributions and achievements

The present thesis tackles the three above-mentioned challenges.

First, by reconsidering some 2D video inherited approaches and by adapting them to the stereoscopic video content and to the human visual system peculiarities, a new disparity map (3D video-New Three Step Search - 3DV-NTSS) is designed. The inner relation between the left and the right views is modeled by some weights discriminating between the horizontal and vertical disparities. The block matching operation is achieved by considering a visual related measure, namely the normalized cross correlation - NCC. The performances of the 3DV-NTSS were evaluated in terms of visual quality of the reconstructed image and computational cost. When compared with state of the art methods (New three step search NTSS and FS-MPEG) average gains of 2dB in PSNR and 0.10 in SSIM are obtained. The computational cost is reduced by average factors between 1.3 and 13.

Second, a comparative study on the main classes of 2D inherited watermarking methods and on their related optimal stereoscopic insertion domains is carried out. Four insertion methods are considered; they belong to the SS, SI (binary QIM and 5-symbols QIM) and hybrid (Fast-IProtect) families. The fast-IProtect establishes synergies between SS and SI in order to achieve the transparency/robustness/data payload trade-off and relays on Monte Carlo generators, following the attack theoretical models, in order to meet time constraints. Each of these four methods is successively applied on the left view of the video sequences as well as on three disparity maps, computed according to the NTSS, FS-MPEG and 3DV-NTSS disparity map algorithms. The experiments brought to light that the Fast-IProtect performed in the new disparity map domain (3DV-NTSS) would be generic enough so as to serve a large variety of applications:

- it ensures the imperceptibility according to subjective tests performed based on three different criteria: image quality, depth perception and visual comfort;
- it offers $PSNR > 35\text{dB}$ and IF , NCC ; SC and $SSIM$ values larger than 0.95^1 ;
- it ensures robustness expressed by a BER lower than 0.05 after filtering and JPEG compression and lower than 0.1 after the geometric random bending attacks;
- it features a non-prohibitive computational cost (e.g. insertion time lower than the frame rate in video, $T_{\text{insertion}}=38\text{ms}$, obtained on a Core2 PC, CPU@2.13GHz, RAM 2 Go).

Finally, concerning the performance evaluation, all the quantitative results are obtained by processing two corpora (3DLive and MPEG) of stereoscopic visual content, organized according to three general criteria: significance, acceptability and exploitability. Each of these two corpora combines indoor/outdoor, unstable and arbitrary lighting, still and high motion scenes. The 3DLive corpus sums up about 2 hours of HD 3D TV content captured by French professionals. The MPEG 3D video reference corpus is composed of 17 minutes of video sequences provided by both academic/industry and encoded at different resolutions (from 320×192 to 640×480 pixels).

The statistical relevance of the results is given by the 95% confidence limits computed for all the values reported in our study (for both transparency and robustness), and by their underlying the relative errors which were lower than $\varepsilon_r < 0.1$.

Two standards have been considered in our study. The transparency of the watermarked content is subjectively assessed according to the ITU-R BT 500 – 12, 710 – 4 and BT 1438 recommendations. The robustness and data payload were considered so as to comply with the Digital Cinema Initiatives (DCI) prescriptions.

¹ Ideal transparency is given by $PSNR \rightarrow \infty$, image fidelity $IF = 1$, normalized cross correlation $NCC = 1$, structural content $SC = 1$ and structural similarity $SSIM = 1$.

Stereoscopic video watermarking: constraints, challenges, current limitations and thesis contributions.

Constraints	Challenges	Current limitations	Thesis contributions
Embedding domain	Disparity map for watermarking application	2D video inherited disparity map <ul style="list-style-type: none"> ignoring the vertical disparities block matching metrics unrelated to the human visual system prohibitive computational cost for HD 3D content 	New disparity map: 3DV-NTSS Principle <ul style="list-style-type: none"> content adaptive (discriminating weights between the horizontal and vertical disparities) human visual system metrics Experimental validation <ul style="list-style-type: none"> reconstructed image quality: gains of 2dB in PSNR and 0.10 in SSIM computational cost: reduction by factor between 1.3 and 13
Embedding technique	Transparency/robustness /data payload trade-off Low computational cost	Transparency <ul style="list-style-type: none"> no subjective evaluation a single objective metric (PSNR) Robustness <ul style="list-style-type: none"> fragility to geometric random bending transformations Computational cost <ul style="list-style-type: none"> Never investigated 	Fast-IProtect/3DV-NTSS watermarking Principle <ul style="list-style-type: none"> Joint SS and SI watermarking Monte Carlo generators following the attacks theoretical models Transparency <ul style="list-style-type: none"> no perceptual difference between the original and watermarked sequences difference and correlation based measures: PSNR> 35dB and IF, NCC, SC and SSIM >0.95 Robustness <ul style="list-style-type: none"> filtering and compression: BER<0.05 geometric random bending transformations: BER<0.1 Computational cost <ul style="list-style-type: none"> real time compatibility
Performance evaluations	Benchmarking <ul style="list-style-type: none"> statistical relevance of the results standard recommendations 	No statistical background <ul style="list-style-type: none"> limited structure and small size in processed corpus no statistical relevance for the results Application-driven benchmarking <ul style="list-style-type: none"> no referenced standards 	Corpora (significance, acceptability and exploitability) <ul style="list-style-type: none"> 3Dlive, HD full encoded (2h11min24sec) MPEG, multiple resolutions and frame rates (17min29sec) Statistical relevance for all results <ul style="list-style-type: none"> 95% confidence limit with relative errors $\varepsilon_r < 0.1$ Transparency <ul style="list-style-type: none"> ITU-R recommendations (ITU-R BT 500-12, 710-4 and BT 1438) Robustness and data payload <ul style="list-style-type: none"> According to DCI

Chapter 1 Context

Abstract

This Chapter browses the thesis basis, from the advent of the video stereoscopic content to the watermarking potentiality to solve issues related to copyright protection. Finally, the thesis structure is identified by facing the stereoscopic video protection to the main watermarking properties.

1.1. Stereoscopic content

From the etymological point of view, the word *stereopsis* derives from the Greek words “stereo” meaning solid and “opsis” meaning appearance or sight [May81] [Rog82] [Pog84].

Under the image processing framework, the word *stereoscopy* encompasses the research and technological efforts related to the design and deployment of tools able to create/enhance the illusion of depth in image [Rog82] [Pog84].

From the anatomic point of view, the human brain calculates the depth in a visual scene mainly by processing the information brought by the images seen by the left and the right eyes. These left/right images are slightly different because the eyes have biologically different emplacements², Figure 1.1 a. Consequently, the straightforward way of achieving stereoscopic digital imaging is to emulate the Human Visual System (HSV) by setting-up (under controlled geometric positions), two traditional 2D cameras, see Figure 1.1.b. Such a representation is commonly referred to as two-view stereo [Sch02] [Fra02].

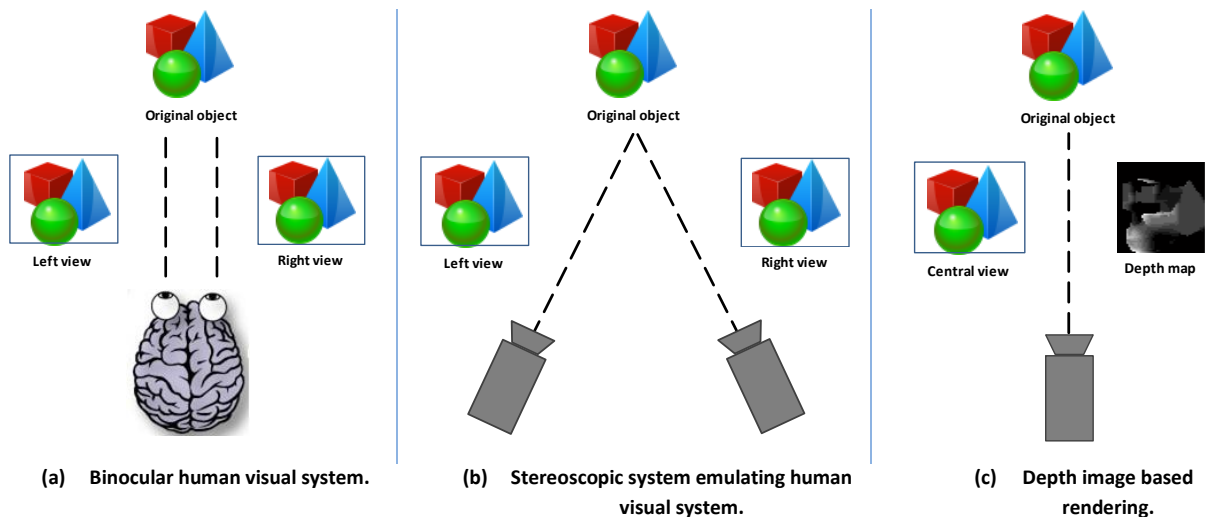


Figure 1.1 Binocular human vision vs. stereoscopic content acquisition.

An alternative way of achieving depth in images is by exploiting the Depth Image Based Rendering (DIBR) principle [Feh03] [Fra11]: this time, the stereoscopic left/right views are created from an image captured with a traditional 2D camera centered in-between the eyes and from the depth estimated by the camera itself, see Figure 1.1 c.

² The difference between the left and right eye images is called binocular disparity.

Regardless of the capturing method, the stereoscopic video content is displayed in the same way, see Figure 1.2.

The viewer brain is provided with two different images, representing two perspectives of the same object. For basic stereoscopic imaging, these two images are the left/right views captured by two cameras. For DIBR, the two images are rendered from the central view and the depth information³.

An additional mechanism, a filter (as illustrated in Figure 1.2) should be considered in order to merge (for visual perception purposes) these images. This filter can be on the user side (i.e. a headgear or glasses) or on the screen side (i.e. a lenticular lens or a parallax barrier). A discussion in this respect is presented in Chapter 1.1.2 (3D displays).

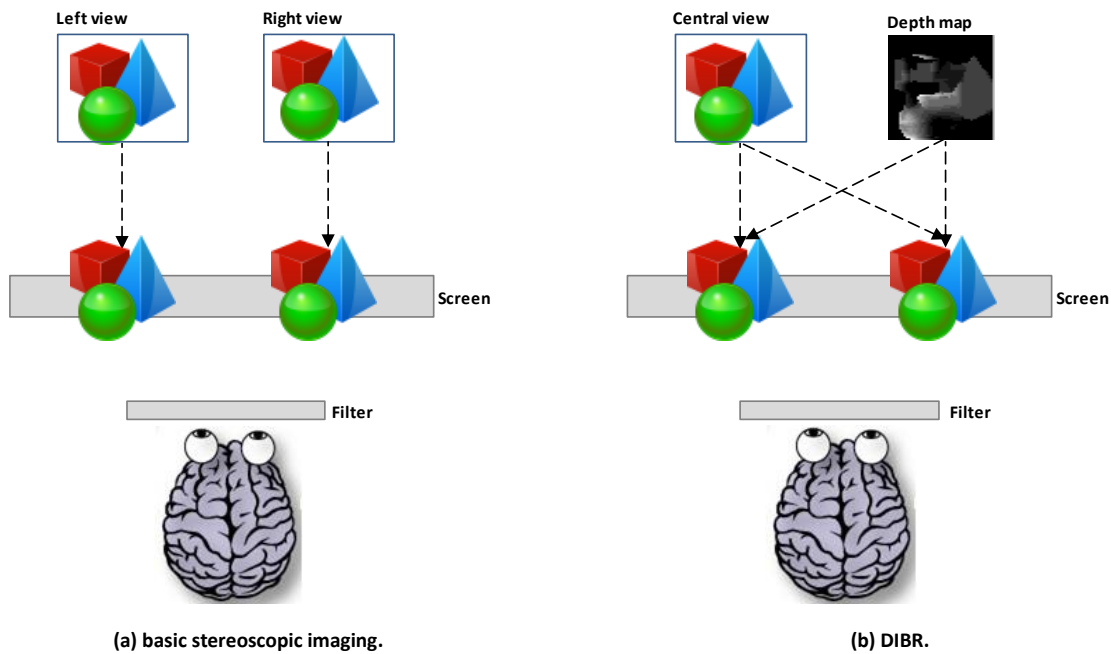


Figure 1.2 Stereoscopic content display.

1.1.1. Content proliferation

To the best of our knowledge, the first experiments devoted to the creation of stereoscopic images were reported at the beginning of the 19th century [Wad87] [Lat38]. They considered a stereoscope which is an optical device with two eyepieces, through which two slightly different images of the same scene are presented individually to each eye, thus giving the illusion of depth.

Today, the stereoscopic content is preponderantly digital and its market is forecast to reach more than 4 billion USD in 2014, see Figure 1.3, [Ped13].

³ The intrinsic parameters describing the camera and its geometric set up are out of the scope of our work.

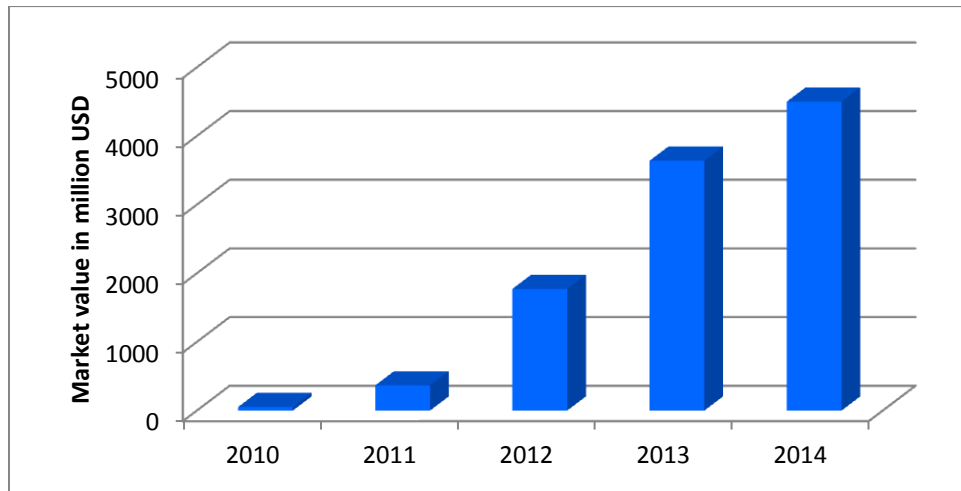


Figure 1.3 Content market value in the stereo 3D technology segment.
(from 2010 to 2014 in million U.S. dollars [Ped13]).

3D Cinema

From the final user's point of view, the 3D cinema with its 120 years history, represents the most appealing instantiation of the stereoscopic imaging [Kun11], Figure 1.4.

Experiments and first patent

The first important stereoscopic era of motion pictures began in the late 1890s when British film pioneer William Friese-Greene filed a patent for a 3D movie process. In his patent, two films were projected side by side on a screen. The viewer looked through a stereoscope to converge the two images. Later on, some movie excerpts were played and presented for tests to the audience.

Early stereoscopic production

In 1922, the American silent movie "The power of love", produced by Harry K. Fairall, was widely reported as the first full length 3D movie presented in anaglyph (see Chapter 1.1.2, 3D displays). The late 1920s to early 1930s witnessed no interest in stereoscopic pictures. The 1940s saw a mini-boom for 3D releases especially in the USSR. The first talking 3D movie was Alexander Andreyevsky's Soyuzdet film production "Robinson Crusoe" which was shown at the Vostok cinema in 1947. A few other 3D films followed to the end of the 1940s.

The golden era (1952–1955)

The early 1950s were a bad time for the movie industry. The popularity of television kept people out of the theaters. The movie industry used 3D as a way of increasing the audience, a strategy which made 3D enter its booming years. Hence the "golden era" of 3D began in 1952 with the release of the first color stereoscopic film, Bwana Devil, produced, written and directed by Arch Oboler. Starting from this moment, most 3D movies used the polarizing process (see Chapter 1.1.2, 3D displays).

Rebirth of 3D (1970–1986)

The huge number of movie releases during the golden era allowed the audience to experience headaches and 3D gained a bad reputation. Consequently, the audience dropped down and 3D entered a hibernation period. The late 1970s witnessed a revival of the 3D with the design of new stereovision caption and projection technologies.

In the mid-1980s, IMAX theaters provided a perfect environment for 3D movie projection. They emphasized mathematical correctness of the 3D rendering and thus largely eliminated the eye fatigue and pain of previous 3D presentations. Unfortunately, at the end of the 1980s, the 3D movie production collapsed.

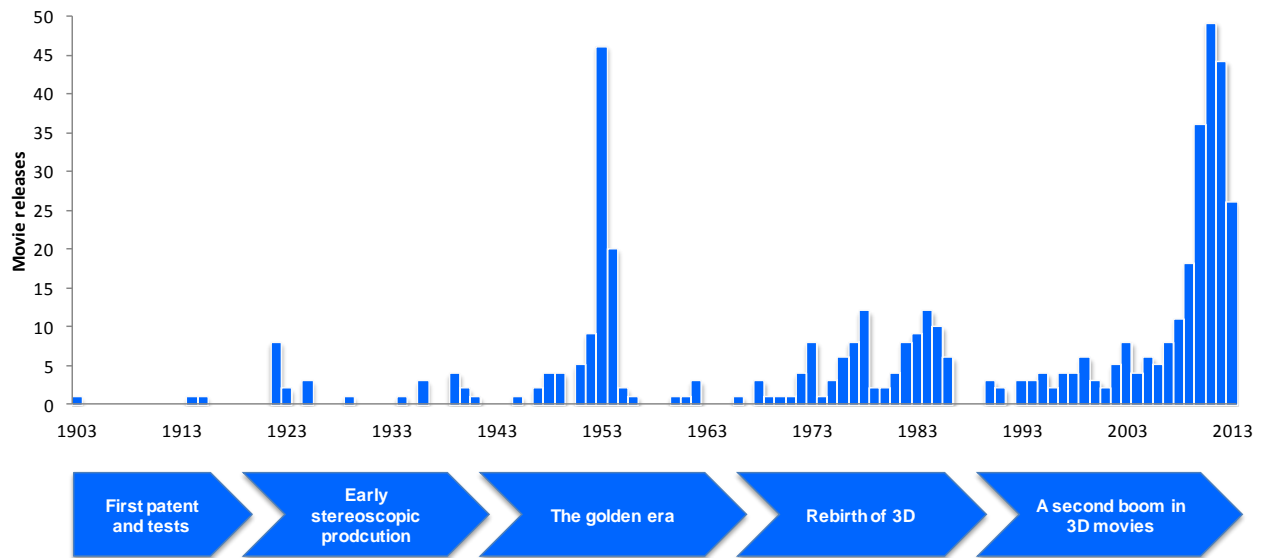


Figure 1.4 Timeline of 3D movies releases [Kun11].

Second boom in 3D movies (1993-present)

From 1993, numerous films were produced announcing the comeback of 3D. In the early 2000s, first full-length 3D IMAX feature movies were released using the new technologies (Real D 3D, Dolby 3D, Xpand 3D, MasterImage 3D, and IMAX 3D)

In 2009, *Avatar* was one of the most influential film with about 285 millions viewers worldwide. Three major 3D releases followed in 2010: *Alice in Wonderland* hitting US theaters on March 5, *How to Train Your Dragon* on March 26 and *Clash of the Titans* on April 2.

Today, 3D movies are extremely expensive to make, and only big-budget studio production can afford them, see Table 1.1. Comparing to the highest-grossing 2D movies, 3D movie production is holding the record with acknowledged costs larger than a nominal value of USD 150 million and going usually over USD 300 million. The white paper [Saf11] looked at the actual costs of a 2D and 3D production of the same movie and studied the cost increase of a 3D film over its 2D counterpart under various trade-offs between cost and staff expertise. The research showed that going from 2D to 3D production entails

average increase cost of 18%. The vast majority of cost increases comes from the increase in 3D cameras rental costs and from the addition of staff members.

3D movies are not only on the top of cost production but also on the top of the highest grossing films, see Table 1.1. For instance, with a worldwide box-office gross of about USD 2.8 billion, *Avatar* is often proclaimed to be the highest-grossing film ever. The *Transformer* comes in the second place with about USD 1.1 billion. *Alice in wonderland* gross reached USD 1.024 billion. The incomes above include revenue coming from theatrical exhibition, home video, television broadcast rights and merchandising.

The releases in multiple formats impacts the number of people purchasing movie tickets. Almost two-thirds of the tickets sold were for 3D showings with an average price of USD 10.

Table 1.1 The highest-grossing 3D films: Costs vs. incomes (in USD million).

Movie	Avatar	Alice in Wonderland	Toy Story 3	Transformers: Dark of the Moon
Year	2009	2010	2010	2011
Cost	280	236	200	195
Worldwide gross	2782	1024	1063	1123

3D Television

The awareness about the 3D movie experience is currently extended and promoted from theater to homes.

The report, "Global 3D-Enabled Device Sales Forecast" [Kin11] predicts that 34% of US homes will own a 3D-ready TV by 2014. In Europe, 3DTV ownership will grow a bit faster, with 42% of homes owning a 3DTV by 2014. Moreover, the study affirms that the TV market is proliferating and most major TV producers are expected to launch new 3D TV technologies. Consequently, 3D TV is expected to surpass 59 million units in 2013 (hence to account 23% of the overall market) and to exceed 83 million units in 2014 (hence to account about 31% of overall TV market), see Figure 1.5.

According to the report made by the market intelligence firm NPD DisplaySearch [Whe12], the 3D television industry is forecast to almost quadruple over the next seven years. In this respect, analysts predict that the tri-dimensional display market will see worldwide revenue growth from USD 13.2 billion/year to more than USD 67 billion/year by 2019. In the same line, the demand for 3D TV products is expected to grow to 180 million units per year by 2019 [Whe12].

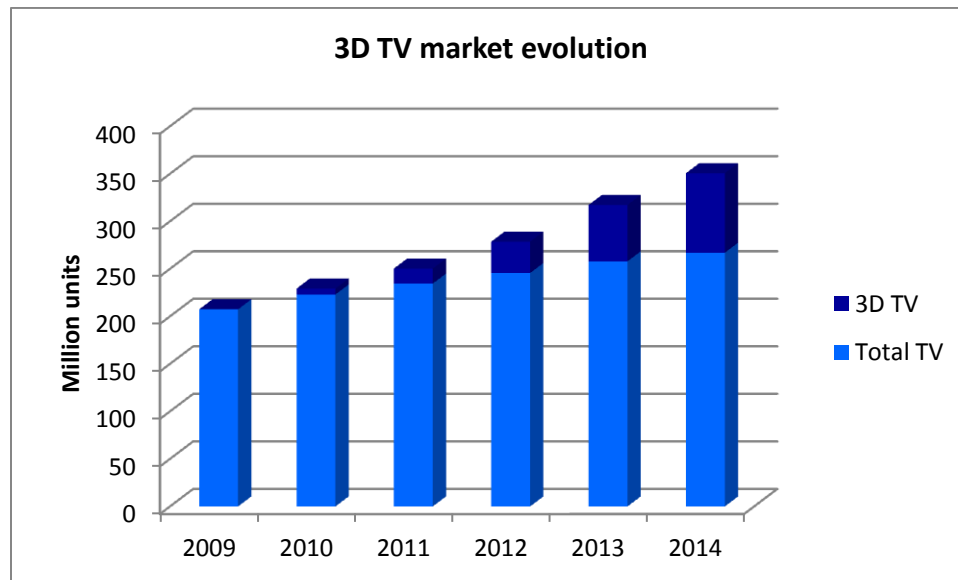


Figure 1.5 Overall TV and 3D TV sales 2009-2014 [Kin11].

3D television must not only face the issue of TV devices but also the need for high-quality content. Although we could not find financial estimation concerning the 3DTV content trends, the increasing number of 3D channels in the world (see Figure 1.7), along with the increasing number of viewers interested in such programs (see Figure 1.6) imposes it as a high added value market [Teu10].

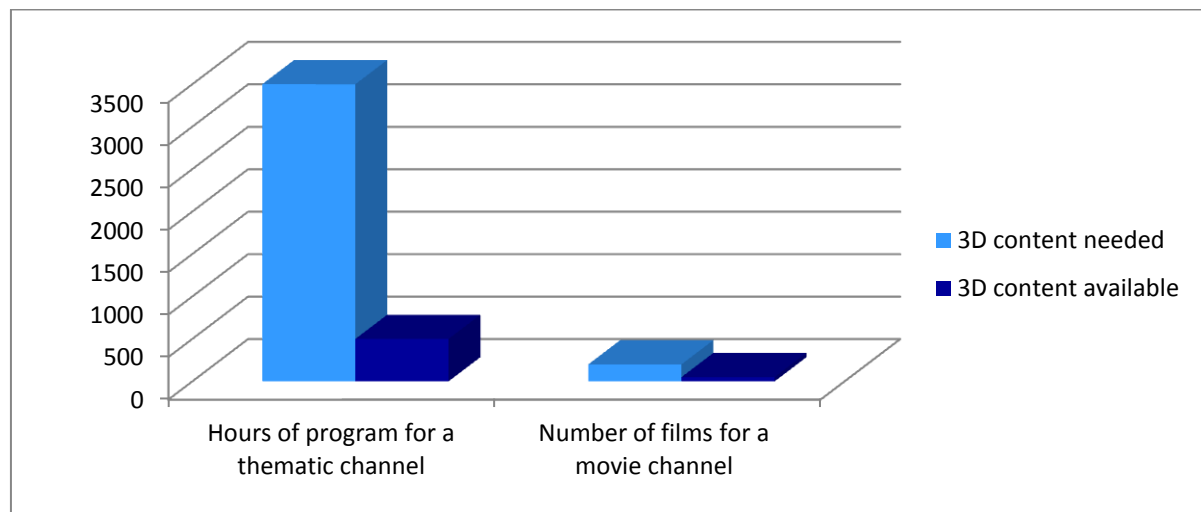


Figure 1.6 3D TV content need [Teu10].

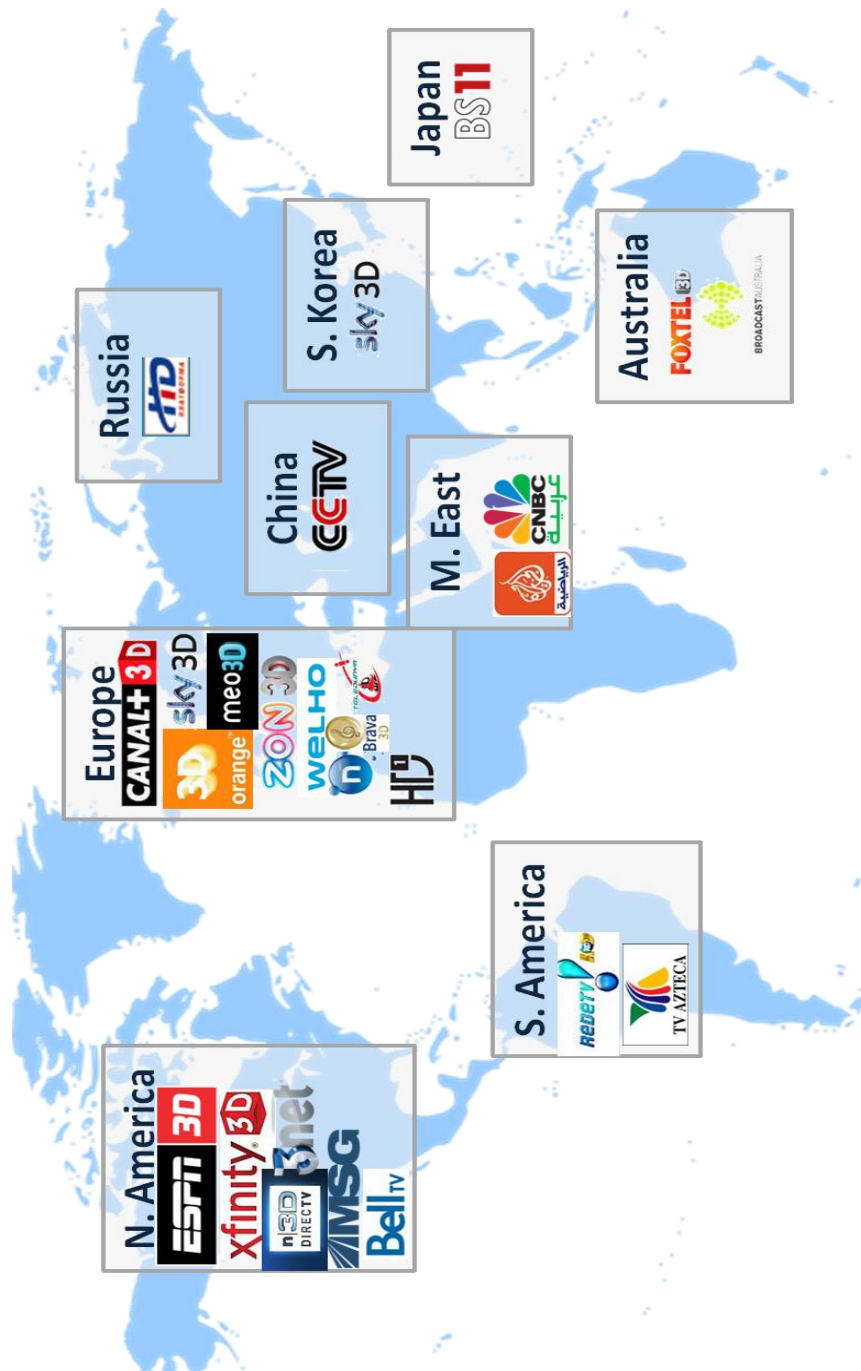


Figure 1.7 3D TV channel in the world [Teu10] [Kun11].

1.1.2. Device explosion

The content proliferation is supported by explosive effort in 3D devices manufacturing:

3D capturing devices

For side-by-side stereoscopic shooting, two synchronized cameras must be used. The distance between the center of the lenses of the two cameras is called the *interaxial*, and the cameras' convergence, is called the *angulation*. These two parameters can be modified according to the expected content peculiarities.

The two cameras must be correctly aligned, identically calibrated (i.e. brightness, color, etc...) and perfectly synchronized (frame-rate and scan-wise).

To hold and align the cameras, a 3D-rig is used (see Figure 1.8). The rigs can be of two main types:

- the *side-by-side rig*, where the cameras are placed side by side (Figure 1.8.a). This kind of 3D-rig is mostly useful for large landscape shots since it allows large interaxials; however, it doesn't allow small interaxials because of the physical size of the cameras.
- the *mirror rig*, also called the *beam splitter rig* (Figure 1.8.b), where one camera films through a semi-transparent mirror, and the other films the reflection in the mirror. These rigs allow small and medium interaxials, useful for most shots, but not the very large interaxials (because the equipment would be too large and heavy).

Monoblock cameras have been designed as well, where the two cameras are presented in a fixed block and are perfectly aligned, which avoids cameras desynchronization (Figure 1.8.c).

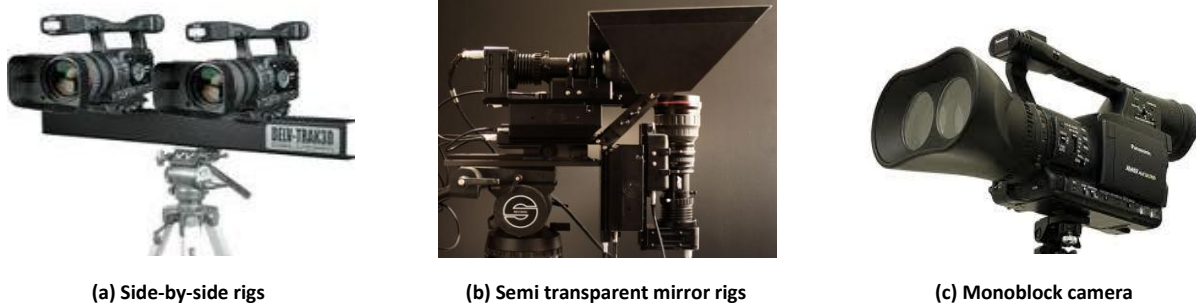


Figure 1.8 Professional technologies for 3D TV [3DF13].

A second category of 3D shooting devices is presented in Figure 1.9. These electronic devices are less expensive and are targeting the user-created stereoscopic picture/movie distribution.



Figure 1.9 Digital personal stereo vision systems [Koc10].

3D displays

Improvements in camera, post-production and projector technology lead to the increase of 3D production's quality. Although these new technologies have high costs, film production companies are competing to release high quality 3D movies and to grant the audience a pleasant experience of 3D. Moreover the 3D devices ownership is growing steadily and will accelerate rapidly over the next three years, according to the latest research from Strategy Analytics [Kin11]. Global sales of 3D-enabled devices are expected to grow with 40% in 2013, to reach 555 million units and 900 million units in 2014, see Figure 1.10. This forecast includes 3D-ready TVs, 3D TV set-top boxes, 3D Blu-ray players, 3D media players, 3D phones and 3D fixed and portable games consoles.

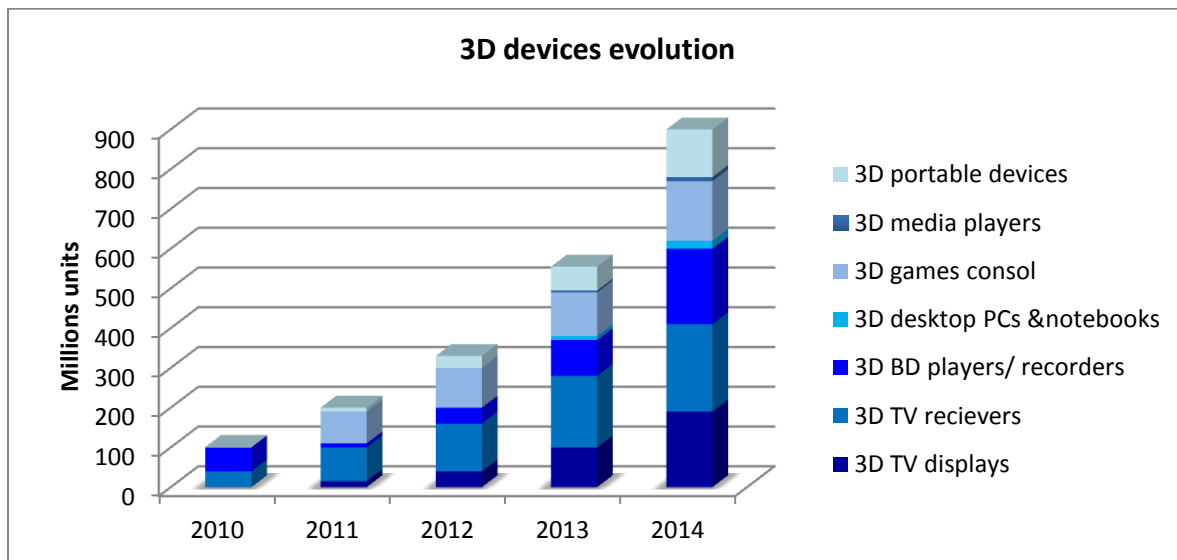


Figure 1.10 Global 3D-enabled devices sales forecast from 2010 to 2014.

(source: Strategy Analytics, "connected home devices services", March 2011). The portable devices include cameras, camcorders, media players and phones [Kin11].

From 1900 onward, the 3D movie releases and television broadcasts were based on two types of 3D displays [Kaw02]: stereoscopic 3D and auto-stereoscopic.

For the first category (stereoscopic 3D displays), the viewer needs to wear special glasses which separate the views of the stereoscopic image for the left and the right eye. These 3D glasses can be active or passive.

On the one hand, active glasses are controlled by a timing signal that allows to alternatively darken one eye glass, and then the other, in synchronization with the refresh rate of the screen. Hence presenting the image intended for the left eye while blocking the right eye's view, then presenting the right-eye image while blocking the left eye, and repeating the process at a high speed which gives the perception of a single 3D image. This technology generally uses liquid crystal shutter glasses, see Figure 1.11.a.

On the other hand, passive glasses are polarization-based systems [Kaw02] and contain a pair of opposite polarizing filters; each of them passes light with similar polarization and blocks the opposite polarized light (Figure 1.11.b). The polarized glasses are associated with a display with polarized filters, providing each eye with the appropriate image.

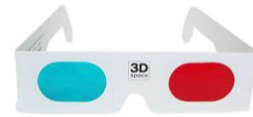
The color anaglyph-based systems are a particular case of the passive glasses and use a color filter for each eye, typically red and cyan, Figure 1.11.c. The anaglyph 3D image contains two images encoded using the same color filter, thus ensuring that each image reaches only one eye.



(a) LCD shutter glasses



(b) Polarized glasses



(c) Anaglyph glasses

Figure 1.11 Samples of passive and active glasses for 3D viewer technologies [Kaw02] [Kun11].

For the second category (auto-stereoscopic), the viewer doesn't need to use any headgear or glasses. This category called also *glasses-free 3D* includes lenticular lens, parallax barrier, holography and volumetric displays [Dod05] [Hal05] [Kun11], Figure 1.12.

The lenticular lens and the parallax barrier are devices placed in front of an image source, such as a projector or an LCD monitor, to allow it to discriminate between the content to be presented to the left/right eyes.

The lenticular lens is a layer of lenses, designed so that when viewed from slightly different angles, different images are magnified, Figure 1.12.a. However, the parallax barrier consists of a layer of material with a series of precision slits, allowing each eye to see a different set of pixels, therefore creating a sense of depth through the parallax, Figure 1.12.b. This technology is used in some auto-stereoscopic phones and video game consoles such as the Nintendo 3DS. Their major limitation lies in the so-called *sweet-spot* which represents the region where the viewer must be placed to perceive the 3D effect. This region can be very narrow depending on the technology, and if the viewer changes position, he/she will not be able to see the stereoscopic image any more.

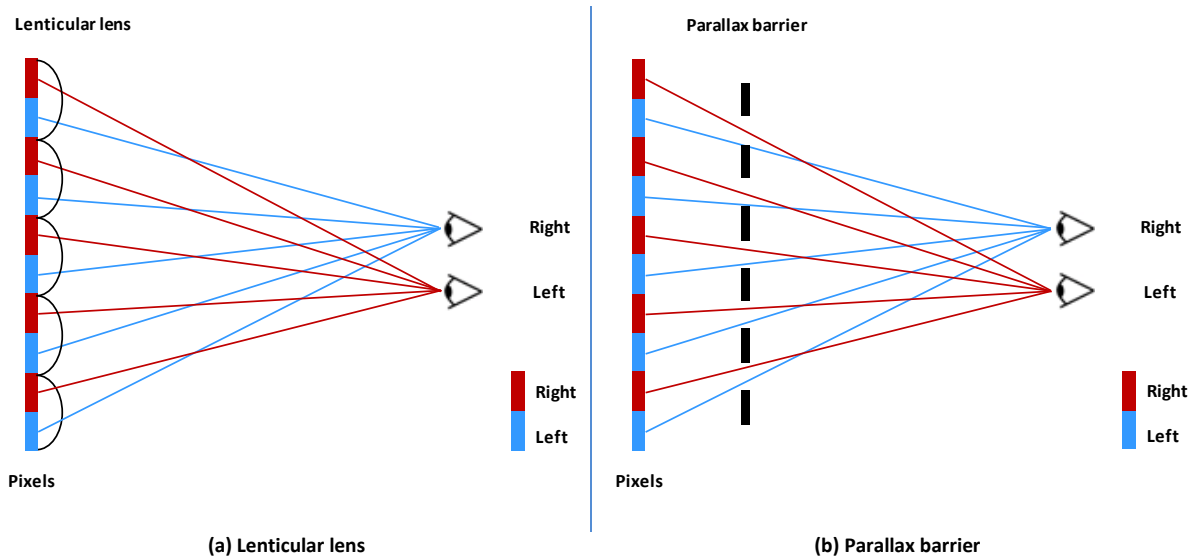
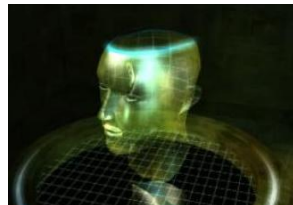


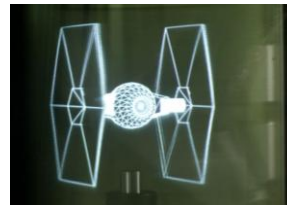
Figure 1.12 Lenticular lens and parallax barrier for autostereoscopic displays [Dod05].

Holography-based devices create a light field with both horizontal and vertical parallax across a large range of viewing angles. They use some optical phenomena (i.e. interference, diffraction) and light intensity to reproduce 3D images from the original scene, see Figure 1.13.a.

Volumetric devices (e.g. multi-planar displays), have stacked up displaying planes, and as well as rotating panel displays. These technologies display points of light within a volume giving 3D depth to the objects, see Figure 1.13.b.



(a) Holography



(b) Volumetric displays

Figure 1.13 Samples of autostereoscopic displays [Kun11].

1.1.3. Potential protection techniques

The proliferation of content production and the explosion of its devices turn the issues related to the stereoscopic content traceability and copyright protection into a hot research topic.

Digital Rights Management

The Digital Rights Management (DRM) refers to a set of policies that define the digital content owner rights and specify the appropriate content consumption and distribution modes. Thus, it enables the secure exchange of high-value content over the Internet or any electronic devices (i.e. DVD, mobile

network, etc). Unfortunately, the restrictions imposed by DRM have made people uncomfortable when using digital medias [Sub06] [Har00].

Cryptography

Cryptography is a term derived from the Greek words *cryptos*, which means hiding or secret, and *graphei*, which means writting. Cryptography is the most common method of protecting digital content, and consists in encrypting the content with an encryption key that is later provided only for legitimate users for decryption. The content can be uploaded on the Internet without any risk of piracy as long as the key is not provided. Unfortunately, once the pirate has access to the decryption key after a legitimate purchase of the product, he/she is able to distribute illicit copies. Hence, cryptography is a good mean of data protection during its transmission, but once the content is decrypted, no further protection is ensured [Cox02].

Steganography

Steganography is also of Greek origin, derived from *steganos*, meaning covered or protected, and *graphei* which means writing. It is the art of writing hidden messages in a way that no one but the sender and the receiver know about the existence of the message. Hence, a secret message is encoded in a manner that its very existence is covered. The main goal of steganography is to communicate securely in a completely undetectable manner. The security of steganography systems relies on the security of the data encoding system. Once this later is known, the steganogrpahy system is defeated. Moreover, the hidden message can be easily removed after applying some common multimedia operations on the host-media content [Cox02].

Digital watermarking

Watermarking can be considered as the most flexible property right protection technology, since it potentially supports all the requirements set by real life applications without imposing any constraints for a legitimate user. This technology adds some information (a mark, i.e. copyright information) in the original content without altering its visual quality. Such a marked content can be further distributed/consumed by another user without any restriction. Still, the legitimate/illegitimate usage can be determined at any moment by detecting the mark. Actually, the watermarking protection mechanism, instead of restricting the media copy/distribution/consumption, provides means for tracking the source of the content illegitimate usage [Cox02] [Mit07].

1.2. Watermarking properties

A watermarking technique [Cox02] [Mit07] consists in imperceptibly (transparently) and persistently (robustly) associating some extra information (a watermark) with some original content. A synopsis of the watermarking process is presented in Figure 1.14.

1.2.1. Watermarking applicative panorama

Initially devoted to fighting piracy, nowadays watermarking can be the core of a large variety of applications for each type of media content, and for both analogous and digital formats: secure

documents, photos, audio, video, etc [Dig03]. Stereoscopic imaging comes with an extensive range of applications, such as 3D television, 3D video applications, robot vision, virtual machines, medical surgery and so on, which emphasize the importance of the watermarking solution.

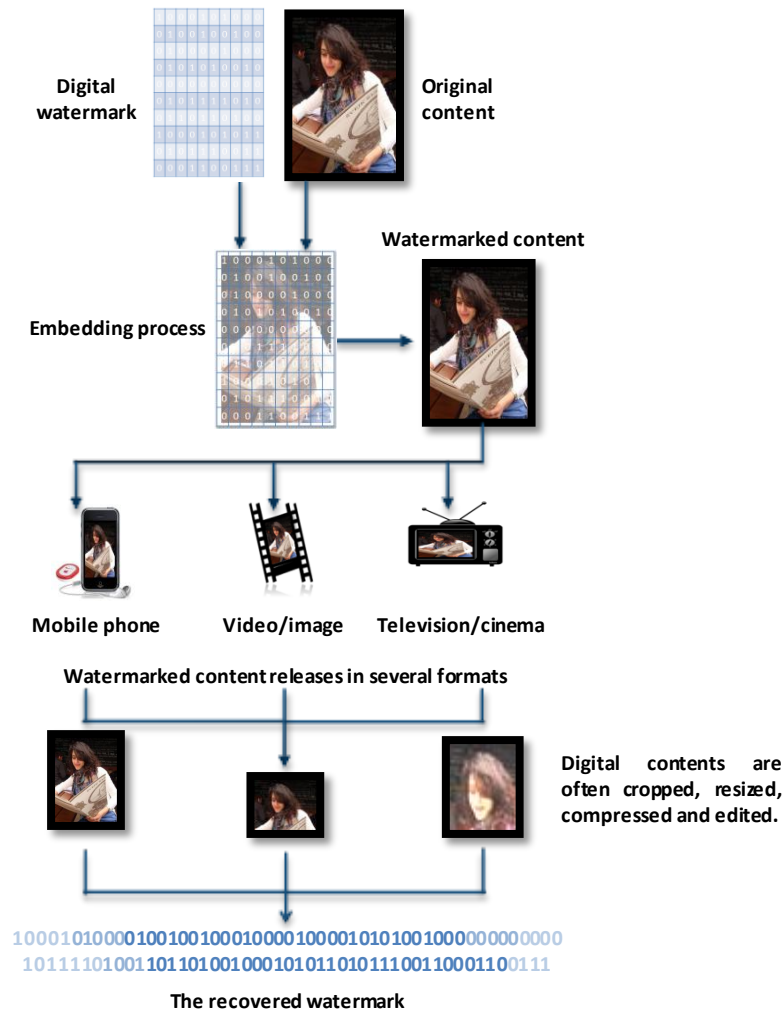


Figure 1.14 The watermarking workflow.

Digital Right Management

Copyright protection is the main application of watermarking techniques aiming to prevent or deter unauthorized copying of digital media. Digital watermarks contain a set of copy control instructions, which tell the copy devices if copies are allowed, or not.

Forensics and piracy tracking

Forensic watermarking applications enhance the content owner's ability to detect and respond to the misuse of his/her assets. Digital watermarking is used not only to gather evidence for criminal acts, but

also to enforce contractual usage agreements between a content owner and the customers. It provides positive, irrefutable evidence of misuse for leaked content assets [Dig03].

Authentication and integrity

Digital watermarks are imperceptibly embedded into all forms of media content, be they individually or aggregately distributed/stored. The watermark can uniquely identify each specific item or instance of content and carry information about its consumption chain and intended destinations. Watermarks can be encrypted and secured so that only authorized reading devices can detect and access the data. Altering a watermark is virtually impossible and the carried data can immediately indicate if the content is genuine or a counterfeit [Sam09].

Broadcast and Internet monitoring

Over the last few years, the amount of content flowing through television/radio channels continues to grow exponentially. Watermarking techniques offer the best solution to automate the monitoring of digital content. The original content, which has been embedded conveys a unique identifier (owner, distributor, data/time information) allowing the content owners and distributors to track their media [Sam09].

Asset and content management

Watermarking techniques enable effective content identification by giving a unique digital identity to any type of media content and by embedding this identity as additional hidden information [Sam09]. The watermarks have to be imperceptible and to have minimal or no impact on the visual quality of the original content. Hence they can be used as a persistent tag, acting as keys into a digital asset management system (DAM). Tagged content can lead back to the original content stored in the DAM system; it can also be linked to metadata in the DAM, such as keywords, rights and permissions.

Filtering/classification

Digital watermarks offer new opportunities for content owners, advertisers and more generally marketers searching for new ways to engage consumers in richer media experiences. In fact, the embedded information enables the identification, classification and filtering of multimedia content. The watermarking systems are able to selectively filter potential inappropriate content (e.g. parental control).

1.2.2. Main watermarking constraints

To assess the performances of a watermarking technique, the following main properties are generally considered: transparency, robustness, data payload, false positive rate and computational cost. According to the targeted application, different constraints are set on these properties.

1.2.2.1. Transparency

Transparency can be defined as the imperceptibility of the embedded additional information in the watermarked media. This may signify either that the user is not disturbed by the artefacts induced by the watermark in the host document or that the user cannot identify any difference between the marked and the unmarked document.

Subjective evaluation

In order to assess the impact of these artefacts from the human visual point of view, subjective protocol and objective quality measures can be considered.

The International Telecommunication Union (ITU) defined some recommendations to be followed for the subjective evaluation [BT98] [BT00] [BT02]. First, it states the required material for the test, such as the monitor and its calibration parameters and defines the environment conditions such as the testing room luminance in order to ensure the reproducibility of results. Second, the panel size is designed depending on the sensitivity and reliability of the test procedure adopted and upon the anticipated size of the effect sought. The minimum number of observers is fixed at 15. The observer's visual acuity and color vision have to be tested according to Snellen chart and Ishihara test, respectively. Other tests are required for different types of content (i.e. fine stereopsis and dynamic stereopsis tests for stereoscopic content assessment).

The ITU-R BT.500-12 also describes methods for subjective quality assessment of the quality of television pictures. Among these, the research community considers the Double-Stimulus Continuous Quality-Scale (DSCQS), the Double-Stimulus Impairment Scale (DSIS) and the single Stimulus for Continuous Quality Evaluation (SSCQE).

Finally, rules for score computation and results analysis were defined.

Objective quality metrics

The visual quality of the watermarked content can also be objectively evaluated. The objective metric is a function that takes as input some video/image information (e.g. pixel intensity), calculates the distance to some reference information and outputs a value giving the distortion level. In our study we consider five objective metrics [Esk95] [Chi11] [Wan04], of two types, namely:

- **Pixel-based measures:** the Peak Signal to Noise Ratio-PSNR and the Image Fidelity-IF;
- **Correlation based measures:** the Normalized Cross Correlation-NCC, Structural Content-SC and Structural SIMilarity- SSIM.

These values are individually computed at the frame level, and then averaged over all the frames in the video sequence.

Consider two frames, S and \hat{S} ; then:

$$PSNR(S, \hat{S}) = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \text{ where } MSE(S, \hat{S}) = \frac{1}{wh} \sum_{i=1}^w \sum_{j=1}^h (S_{i,j} - \hat{S}_{i,j})^2$$

$$\begin{aligned}
IF(S, \hat{S}) &= 1 - \frac{\sum_{i=1}^w \sum_{j=1}^h (S_{i,j} - \hat{S}_{i,j})^2}{\sum_{i=1}^w \sum_{j=1}^h S_{i,j}^2}, & NCC(S, \hat{S}) &= \frac{\left(\sum_{i=1}^w \sum_{j=1}^h (S_{i,j} - \bar{S}) (\hat{S}_{i+u, j+v} - \bar{\hat{S}}) \right)}{\sigma_S \sigma_{\hat{S}}}, \\
SC(S, \hat{S}) &= \frac{\sum_{i=1}^w \sum_{j=1}^h S_{i,j}^2}{\sum_{i=1}^w \sum_{j=1}^h \hat{S}_{i,j}^2}, & SSIM(S, \hat{S}) &= [l(S, \hat{S})]^\alpha \cdot [c(S, \hat{S})]^\beta \cdot [s(S, \hat{S})]^\gamma
\end{aligned}$$

where $S_{i,j}$ and $\hat{S}_{i,j}$ are the pixel intensity at (i, j) location in S and \hat{S} , respectively. h and w are the height and width of the frames while μ and σ are the corresponding mean values and standard deviations, $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are parameters adjusting the relative importance of the three components [Wan04]. In order to simplify the expression, the study in [Wan04] suggests to consider $\alpha = \beta = \sigma$. The luminance $l(S, \hat{S})$, the contrast $c(S, \hat{S})$ and structure $s(S, \hat{S})$ of the two frames S and \hat{S} are defined by

$$\begin{aligned}
l(S, \hat{S}) &= \frac{2\mu_S \mu_{\hat{S}} + C_1}{\mu_S^2 + \mu_{\hat{S}}^2 + C_1} & c(S, \hat{S}) &= \frac{2\sigma_S \sigma_{\hat{S}} + C_2}{\sigma_S^2 + \sigma_{\hat{S}}^2 + C_2} & s(S, \hat{S}) &= \frac{2\sigma_{S\hat{S}} + C_3}{\sigma_S \sigma_{\hat{S}} + C_3}
\end{aligned}$$

where C_1 , C_2 et C_3 are small constants given by $C_1 = (K_1 \cdot L)^2$, $C_2 = (K_2 \cdot L)^2$, and $C_3 = C_2/2$, L is the dynamic range of pixel values, $K_1 = 0.01$ and $K_2 = 0.03$.

1.2.2.2. Robustness

The robustness refers to the ability of detecting the watermark after applying some signal modifications operations and malicious attacks on the marked document, such as spatial filtering, lossy compression, recording, etc.

The watermarking techniques can be divided in three main classes according to their robustness level set by the targeted application: robust, fragile and semi-fragile watermarking [Lin00] [Bar01a] [Wol99]. Robust watermarking is designed to withstand attacks. The presence of the embedded watermark must be detected after common data manipulations and malicious transforms. Fragile watermarking is designed to detect any changes in the original content. Thus, it refers to the case where the inserted watermark is lost or altered as soon as the host content suffers modifications. This way, the watermark loss or alteration is taken as evidence that data has been tampered with. The watermark is said semi-fragile if it survives only a particular kind of manipulations (e.g. moderate compression or video format change) thus allowing that particular content modification to be identified.

The distortions a watermarked content can suffer can be classified in three classes: valumetric, temporal and geometric [Cox02].

The valumetric distortions include the additive noise, linear filtering, lossy compression and quantization. The additive noise can be added to the content when applying some usual processing or when transmitting the signal over a communication channel during the broadcast. Linear filtering (such as blurring) and non-linear filtering (such as sharpening) are included in some image processing software. Lossy compression can affect the embedded watermark, as it removes the redundancy exploited in

watermarking schemes. Finally, a good watermarking system has to survive the quantization which generally occurs during a Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) used for compression purposes.

The temporal distortions include delay and temporal scaling. This type of attacks often occurs during the conversions between different television standards, for example the frame-rate changes.

The geometric distortions include rotations, translations, spatial scaling, cropping and changes in aspect ratio.

All these types of distortions are considered as malicious attacks and can occur due to user manipulations. For images, such manipulations can be printing and scanning. However, for video content, geometric distortions commonly occur during format changes.

Attacks are very likely to appear during the stereoscopic content processing workflow and it is hard to design a watermarking system able to survive all types of distortions while preserving the watermarked content's fidelity.

StirMark is a generic tool developed for robustness assessment of image watermarking algorithms [Pet98] [Pet00]. Among a large class of attacks, it also includes StirMark random bending, a simulator of the random errors induced when printing the image on a high quality printer and then scanning it with a low quality scanner. It also applies minor geometric distortions: the image is slightly stretched, sheared shifted and rotated by a random amount and the re-sampled by using Nyquist interpolation, see Figure 1.15.c (right) and Figure 1.16.c (right)⁴.

Figures 1.15 and 1.16 show some of the most common attack effects on one image taken from the corpus processed in the thesis, see Appendix A.1 and on a test image respectively. The distorted versions of the original image are represented in the second and the third row according to the type of the applied attack.

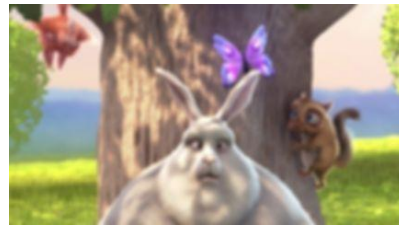
⁴ All through this thesis, the StirMark random bending attack is considered.



a) Original image.



Sharpening (11×11).



Gaussian filtering (7×7).

b) Valumetric distortions.



Rotation $+2^\circ$.



StirMark random bending.

c) Geometric distortions.

Figure 1.15 "Big Buck Bunny" [Ble08], original frame with samples of its attacked versions.

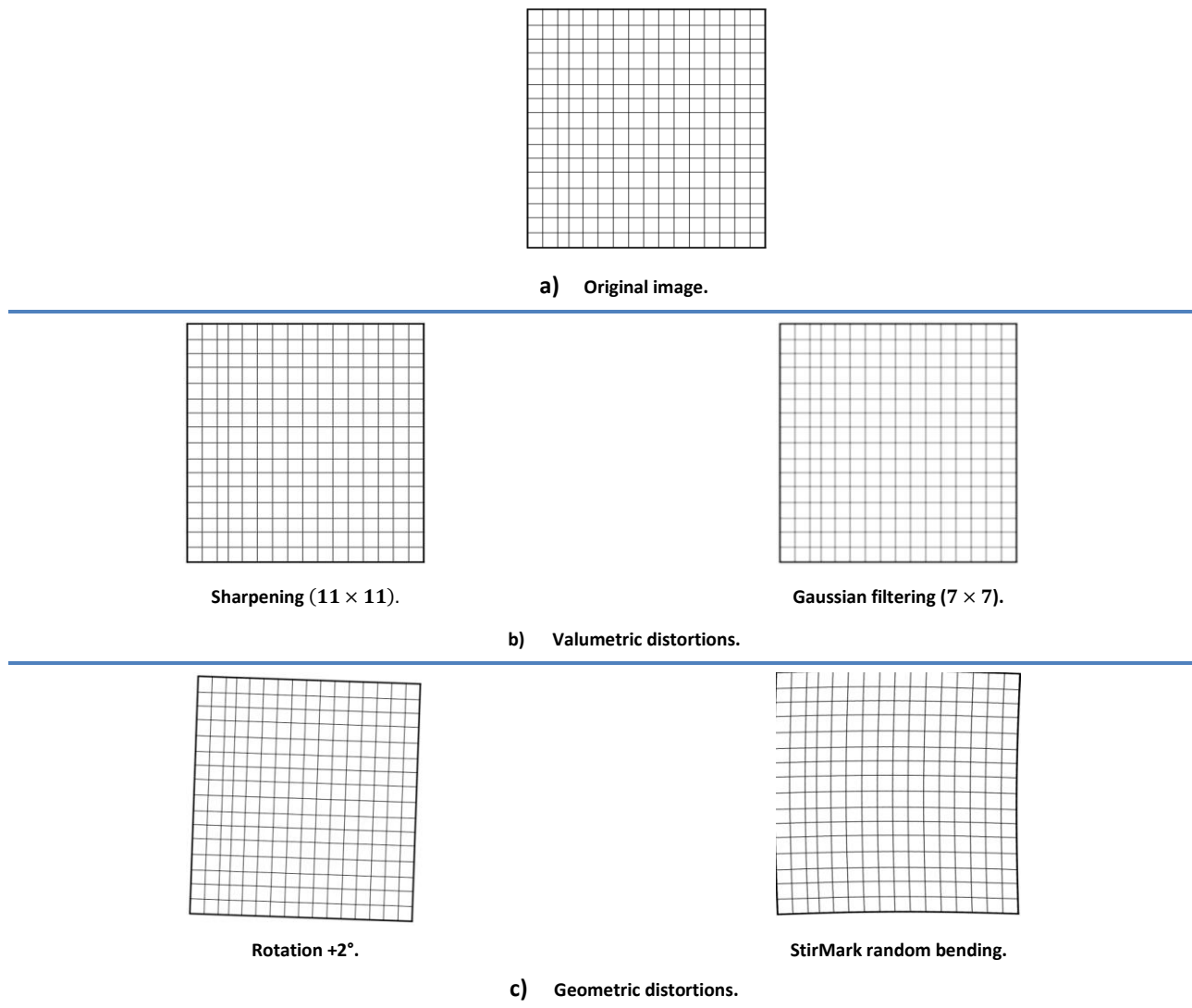


Figure 1.16 Test image with samples of its corresponding attacked versions.

1.2.2.3. Data payload

The data payload is the quantity of information that is inserted into the host document. For copyright application, this amount of information should be high enough so as to allow the owner and/or his/her rights to be identified. The data payload can be defined in number of bits within a document or within a unit of time (e.g. 64 bits would correspond to an ISBN number). Different applications require different data payloads, see Table 1.2. For instance, for e-commerce applications, the additional data (the watermark) could bring information about the document buyer, vendor, date and time of purchase. In a right management context, the embedded watermark has to identify the content and specify the usage rules as well as the billing information. For authentication and integrity applications, the watermark details the content's main modification and the corresponding date of modification.

Table 1.2 Data payload requirements for some examples of watermarking applications.

Applications	Data payload
Digital Right Management	4 to 8 bit in 5 min of video [Ala03]
Forensics and piracy tracking	30 bit in 2 min of video [Cox07]
Authentication and integrity	200 bit per video sequence [Xu05]
Broadcast and internet monitoring	24bit/s [Cox07]
Asset & content management	31 bit per video sequence [Sha02]
Filtering/classification	1 bit/frame [lev01]

1.2.2.4. False positive rate

The false positive rate is the number of false positives expected to occur within N runs of watermark detection. A false positive is the detection of a watermark in an un-watermarked document. This probability should be as low as possible, but its upper limit is application dependent.

For example, for ownership proof purposes, where the detector is used only when disputing the ownership, the highest accepted value can be considered of 10^{-6} . However, for copy control applications the false positive rate should be smaller, e.g. 10^{-12} .

1.2.2.5. Computational cost

The computational cost can be determined by calculating the processing time or by defining the algorithmic complexity.

The processing time is calculated in milliseconds (ms) and, in some applications must meet the real time requirement: the embedding and detecting modules must not slow down the production schedule. Examples of real time applications include live TV, videoconference, online gaming, community storage solutions and some e-commerce transactions.

The algorithmic complexity is defined as a numerical function $O(n)$ – number of operations versus the input size n and gives the *a priori* estimation for the time taken by an algorithm independent of the PC performances.

1.2.3. From 2D to stereoscopic video watermarking

In its 20 years history, watermarking theory has been already put into practice, for audio, still images, video and 3D objects. While, from the applicative point of view the same properties are searched for, each type of content comes with different methodological frameworks.

Some illustrative examples will be further discussed.

Image watermarking

Digital images are often printed and scanned again, downloaded and reused for illustrations. Consequently, many watermarking algorithms have been advanced to protect them from illicit use [Nik98] [Gro01].

The visual quality is obviously crucial in image watermarking which leads to a limited data payload. When used as illustrations, images are often compressed by lossy compression algorithms like JPEG and undergo many other common processing operations including softening, sharpening, denoising, scaling, cropping and colour corrections.

Currently, the most challenging attacks on the robustness of image watermarks are nonlinear transformations, which make the watermark undetectable for the detection algorithm. But even common image processing operations, like scaling and rotations, can be serious challenges for the re-synchronisation of an image watermark.

Another important aspect of image watermarking is the broad variety of image types. There are photos, figures based on line drawings, rendered synthetic images, bitmap representations of textual information, etc... Challenges with respect to transparency and robustness often depend on the characteristics of these image types.

Video watermarking

In the literature, video watermarking has been initially approached as a direct extension of still image watermarking, i.e. by considering a video as a set of still images which are individually protected [Kal99]. However, such straightforward application does not consider the peculiarities of the video content. First, with respect to still images, a larger variety of both hostile and non-hostile processing, such as video editing, lossy compression, filtering, chrominance re-sampling, change of the frame rate during transmission, and formats interchanging, have to be considered. Secondly, in digital video the content is usually dynamic and the human attention cannot focus on all the areas of the video frame simultaneously.

Consequently, for video watermarking, it is of paramount importance to select proper frame regions that can guarantee the data masking, and it is crucial to decide the embedding technique (frame by frame [Bar00], or multiple frames [Kal99]). Given that some particular frame-based attacks (frame dropping, frame exchanging, and frame rate variation) can occur, it seems that frame by frame techniques are preferable, since, in this case, each frame contains the entire watermark, and time synchronization is not needed. Nevertheless, the watermark embedding and recovery can greatly benefit from exploiting the information which is contained in a sequence of frames, i.e. watermark recovery should be performed on a video sequence basis.

Another important issue regarding the video content is related to the possibility of embedding the same watermark in every frame, thus obtaining a system which is sensible to statistical attacks. The other option is to change the watermark from frame to frame with the risk of producing visible temporal artefacts [Hol00].

3D watermarking

3D data is mainly represented by meshes and Non-Uniform Rational Basis Spline (NURBS); these two representations also define the classes of 3D watermarking techniques.

The largest part (more than 90%) of the studies on 3D data watermarking are mesh-based [Yeo99], no matter if they are devoted to fragile watermarking or to robust watermarking [Lin00] [Bar01a] [Wol99].

Generally, the mesh-based techniques embed the watermark into the vertex positions. Their main problem remains always the robustness. When dealing with NURBS representations, the main advantage is the availability of a natural 2D surface parameterization which makes it possible to extend the 2D watermarking techniques to 3D data. As far as we know, [Lee02] is the first study which exploits the NURBS parametric definition. In this respect the authors advanced a method where the mark is not embedded directly into the 3D model but into some virtual images derived from the 3D model. They propose two methods, the first for steganography and the second for watermarking. For the latter method, three virtual images are obtained by uniformly sampling the 3D model and by recording the x , y , and z co-ordinates of the sampled images.

Streoscopic visual content

In the watermarking context, the stereo data can be represented in two modalities. The first one simply considers the right and the left views as two independent images. This way, the stereo data can be straightforward exploited with basic 2D methods. However, such an approach has no guaranteed success in the stereoscopic image processing, since it neglects the very intimate mechanisms governing their interdependence. The second modality considers derived representations from the stereo pair, as a disparity map, for instance. Consequently, the stereo watermarking scheme can be classified according to these two directions: view based stereo approaches [Don03a] [Don03b] [Bha09] [Ell09] and disparity based stereo approaches [Zen07].

1.2.4. Theoretical framework for stereo watermarking

From the theoretical point of view, and independently on the protected data peculiarities, a watermarking technique can be identified within the communication theory as a noisy channel, Figure 1.17: the mark is sampled from the information source and should be sent through a channel where a strong noise (the original content and the malicious attacks) acts. The watermarking challenge consists of optimizing the transmission through this channel, i.e. specifying a method achieving prescribed transparency and robustness while maximizing the data payload. This model shows that transparency, robustness and data payload properties are three-folded antagonistic constraints in the design of a watermarking scheme. First, for a fixed data payload, the more transparent the mark, the lower its power at the transmitter, hence the lower its robustness. Second, for a fixed robustness, the larger embedded amount of data, the lower watermark transparency. Third, for a fixed transparency, the more robust the mark, the lower the data payload.

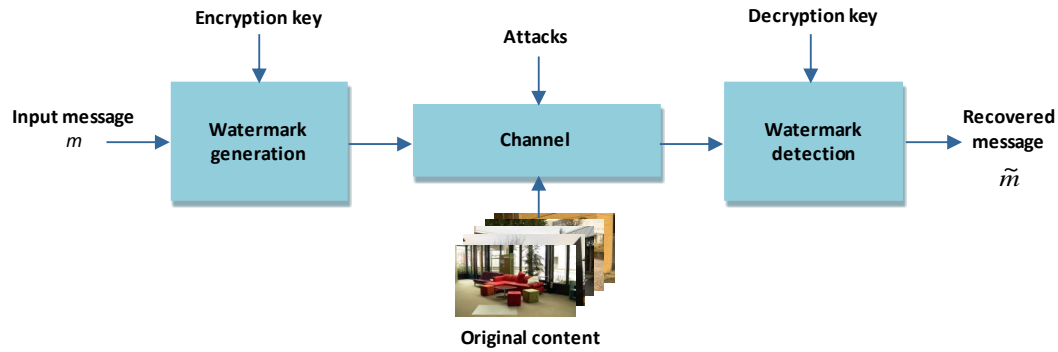


Figure 1.17 Watermarking as a noisy channel.

In the stereoscopic context, the original content is defined by the right and the left images composing the same scene, Figure 1.18. Hence, the corresponding noisy channel is affected by an extra noise source comparing to the 2D-video model. The main issue is still to maximize the data payload on such a channel, under an additional power constraint set by the human visual system. However, when extending this model to the stereoscopic video case, three more issues should be dealt with. First, the theoretical probability density functions modeling the various transforms the stereoscopic video suffers during its distribution are not yet investigated. Secondly, the two stereo views represent correlated side information noise sources; no theoretical result is nowadays available for handling such a situation. Finally, the human stereoscopic vision is not yet modeled with precision, at least not so as to be directly integrated as a power limit constraint for the inserted mark.

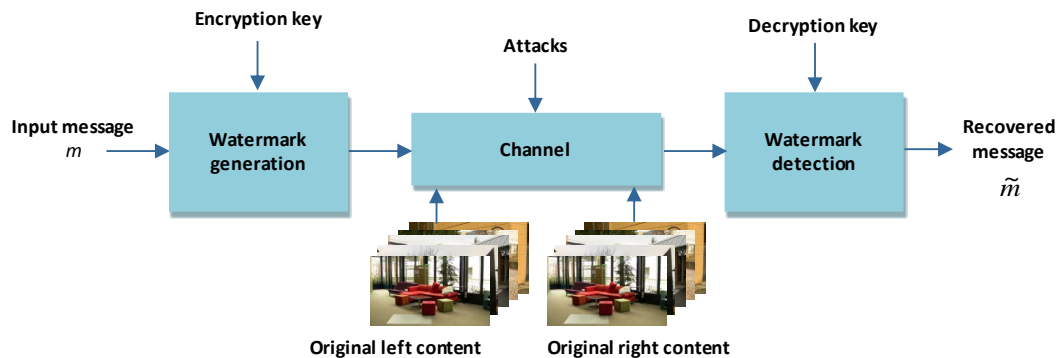


Figure 1.18 The theoretical model for stereo noisy channel.

1.3. Thesis structure

In order to investigate the practical impact the watermarking technique can have for stereoscopic video content protection, this thesis is structured as follow.

The second Chapter is composed of two main parts. The first stands for a state-of-the-art of the disparity map computation algorithms and the second presents a critical overview of stereoscopic watermarking approaches.

Chapter 3 introduces a new disparity map for stereo watermarking scheme. First, the new disparity map algorithm (further referred to as the 3DV-NTSS) is presented and the experimental validation is described.

Chapter 4 explicitly describes the Fast-IProtect stereo embedding technique. First, it introduces the hybrid watermarking technique IProtect and presents its advantage and main drawbacks. Second, it presents the advanced solution to enhance the IProtect performances in terms of computational cost.

In Chapter 5, a benchmark of four class of embedding technique applied on four insertion domains is presented.

The last Chapter is devoted to concluding remarks and perspectives.

The thesis has four appendixes. Appendix 1 is devoted to the processed corpora. Appendix 2 presents the considered embedding techniques. The results on the subjective and objective evaluation of the watermarked videos transparency and their corresponding 95% error limits are detailed in Appendix 3. Appendix 4 details the numerical values of the robustness evaluation and their corresponding 95% error limits.

Chapter 2 Objectives

Abstract

This Chapter contains two main parts. The first part presents the state of the art for the disparity map computation methods and compares their performances. The second part introduces the two main directions in watermarking (Spread Spectrum and Side Information) and presents a critical overview of the underlying stereo watermarking approaches. This makes it possible for the thesis main objectives to be identified and positioned with respect to the current day methodological limitations.

2.1. Disparity maps computation

2.1.1. Basic concepts

Under the digital image processing framework, the concept of “disparity” is preponderantly associated with 2D video compression: it covers the issue related to the specifications of algorithms able to identify visually similar image blocks inside the same frame and/or in neighbor frames. Due to the 2D video visual content homogeneity, such visually similar blocks are expected to be equally likely distributed on the vertical/horizontal directions.

Under the stereoscopic video framework, the term disparity was first introduced to describe the difference in the position of the same features, seen by the left and right eyes [Mar82], see Figure 2.1.a. The horizontal disparity is the most commonly studied phenomenon, while the vertical disparities are generally neglected and *a priori* supposed to be induced by camera synchronization errors.

Such an approach results in the so-called rectified stereos image. The rectification process projects both the left and right images onto a common image plane in a way that the corresponding points have the same row coordinate, see Figure 2.1.b.

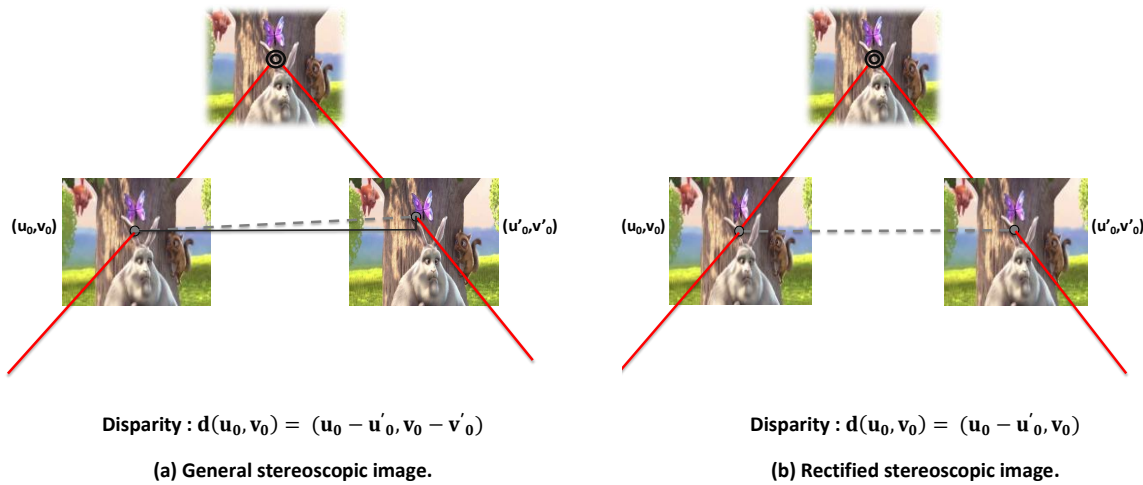


Figure 2.1 Stereo matching.

2.1.2. Block matching disparity map computation

In its widest acceptance, a disparity map provides information about the coordinates at which similar blocks are located in two images (the reference and the target images). Consequently, computing a disparity map requires the design of a rule specifying how the reference blocks are searched for in a given area of the target image and to define a similarity metric establishing whether or not a reference block matches a target block.

By exploiting the spatio/temporal correlation between successive frames in 2D video content, several disparity maps have already been advanced and proved their efficiency in various research fields, like compression, indexing or segmentation. They generally assumed that the differences between the target and reference frames are homogeneous on the two directions and they performed the block matching by measures based on the differences between the values in the two blocks (e.g. Mean Squared Error – MSE or Sum of Absolute Differences – SAD).

The Exhaustive Search (ES) algorithm [Kua99] is widely used for block motion estimation in video coding in order to determine effective similarity while providing minimal error estimation for a fixed search window size, see Figure 2.2.a. The estimated MSE provided by the ES algorithm for content of medium to high motion is of 168.85; this value decreases to 16.11 for content with small motion. The main disadvantage of ES is the massive computation required for running the full search window. The average search points is estimated to be 225 for a block of 16×16 pixels and the algorithm complexity is of $(2p + 1)^2$, where p is the searching distance.

Several fast algorithms were developed to reduce the computation cost and to remedy the huge time consumption issue. The Three Step Search (TSS) was advanced by Koga *et al.* [Kog81] and consists of three evaluation steps, each of which composed of nine uniformly spaced search points. The best candidate search point in a step becomes the center of the following step. In the first step, a nine-point search pattern on a 9×9 grid is considered. TSS requires a fixed $(9 + 8 + 8) = 25$ search points per block, which leads to a speedup ratio of 9 over the ES in a block of 16×16 pixels. The algorithm complexity is logarithmic: $1 + 8 \times \log_2(p + 1)$. However the visual quality is degraded comparing to the ES algorithm; the estimated MSE values for medium/large motion content and for small motion content being of 192.77 and 16.89, respectively. The main drawback of TSS is the relatively large search pattern considered in the first step, which makes it inefficient for finding blocks with small motions.

The Four Step Search (4SS) [Jai81] was introduced to reduce the average computational cost and to give better motion estimation. 4SS requires four search steps for the same 16×16 search window. It reduces the initial step size comparing to the TSS algorithm by considering a nine-point search pattern on a 5×5 grid in the first step. Hence, it increased efficiency for small motions, but it is more complex than TSS. 4SS gave less compensation errors than TSS: the estimated MSE values being of 190.77 and 16.68 for medium/large motion content and for small motion content, respectively.

The conjugate Directional Search (CDS) [Sri85] is advanced to offer less motion estimations errors since it considers multiple search directions. The search starts initially in the vertical and horizontal axis directions of the center of the search area. Each new minimum leads to a new search line joining the previous and the new minima. The average MSE for medium/high motion content and small motion content are of 190 and 16.7, respectively. The CDS algorithm complexity is of $2 \times (2p + 1)$.

Zeng and Liou [Li94] advanced the New Three Step Search (NTSS) algorithm (Figure 2.2.b) for fast block matching estimation and showed that it provides smaller motion compensation errors than the state of the art given by the TSS, the 4SS and the CDS. NTSS requires three search steps. In the first step, it considers two search patterns: nine-points at a grid of 1×1 and eight-points on a grid of 9×9 . The algorithm complexity is of $9 + 8 \times \log_2(p + 1)$ and the average search point number is 21 for a 16×16 pixel search area. The estimated MSE values for medium/large motion content and for small motion

content are of 180.92 and 16.58, respectively. The NTSS always provides small compensation errors for small motion content, but can result in a large average number of search points for large motion content.

Zhu and Kuang [Zhu00] designed a Diamond Search (DS) algorithm using a compact diamond search patterns. DS achieves performances close to the NTSS algorithm in terms of reconstructed image quality and alleviates the computational constraints by a factor of 1.5. Hence, the average search point number is 18.52, the complexity is $8 \times \log_2(p + 1)$ and the MSE values are of 188.87 and 16.5 for medium to large motion content and small motion content, respectively.

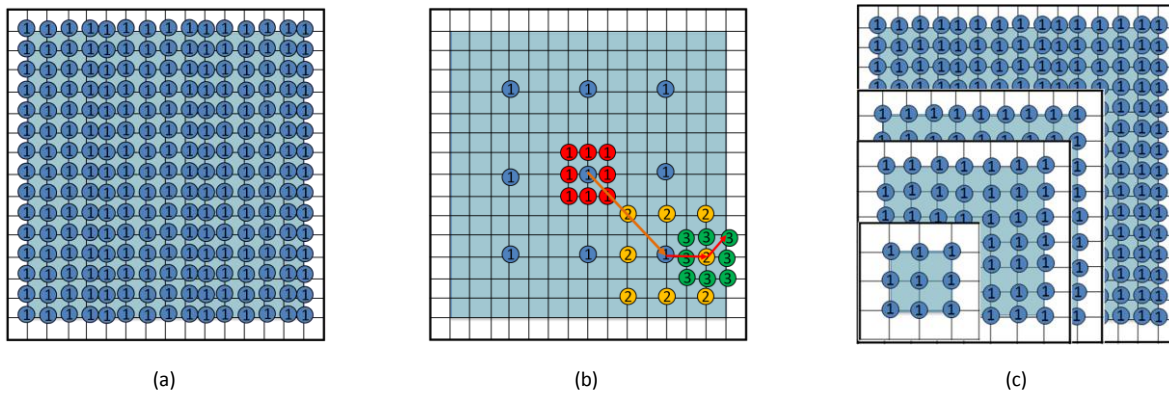


Figure 2.2 Block searching strategies: (a) ES, (b) NTSS, and (c) FS –MPEG.

These principles inherited from the 2D video are currently in use for stereoscopic video processing [Jia06] [Wan11] [Din06]. For instance, in [Wan11], Wang et al. uses a non-static binocular stereo vision system based on an improved diamond search algorithm helping a dynamic target tracking for a mobile robot. In [Din06], a new structure of prediction in stereo video coding system is proposed based on a modified full search block matching algorithm. The stereoscopic video compression was approached by the MPEG community via shape-adaptive block matching based on exhaustive search [Tan09], further referred to as FS-MPEG (full search MPEG). Hence, with respect to the basic ES algorithm, FS-MPEG is expected to increase the reconstructed image quality at the expense of the computational cost. The block matching starts from 9×9 pixels and is progressively reduced (e.g. 5×5 or 3×3 in [Sta08]) until the SAD reaches its local minimal value, Figure 2.2.c. During this matching procedure the vertical disparity component is assumed to be zero.

Table 2.1 and Figure 2.3 synoptically display the performances of the state-of-the-art studies. In Figure 2.3, the abscissa corresponds to the visual quality of the reconstructed image, expressed by the MSE values (the better the image quality, the lower the MSE value) while the ordinate stands for the computational complexity, expressed by the number of search points in a given 16×16 pixels block.

Table 2.1 Performance evaluation between the state of the art algorithms.

	Visual quality (MSE)		Average search points (16×16 pixels block)	Complexity (p pixel searching distance)	Advantage	Disadvantage
	Medium to high motion content	Small motion content				
ES Exhaustive search (a.k.a. FS – Full Search) [Kua99]	168.85	16.11	225	$(2p + 1)^2$	best picture quality for fixed window size and assuming no local minimum phenomenon	maximal computational cost
FS-MPEG with n iterations [Sta08]	Not available	Not available	$n \times 255$	$n \times (2p + 1)^2$	hierarchical computation increase picture quality	heavy computational cost
TSS Three Step Search (a.k.a. 3SS) [Kog81]	192.77	16.89	25	$1 + 8 \times \log_2(p + 1)$	less complex than ES, recommended in MPEG-2	cannot detect small motion
CDS (Conjugate Directional Search) [Sri85]	190	16.7	27	$2 \times (2p + 1)$	better quality than TSS	more complex than TSS
NTSS [Li 94]	180.92	16.58	21.33	$9 + 8 \times \log_2(p + 1)$	more efficient than TSS, CDS and 4SS in term of complexity and quality especially for small motion	more complex than TSS for large motion
4SS Four Step Search [Jai81]	190.27	16.68	20.1	$9 + 18 \times \log_2[(p + 1)/4]$	reduced initial step size, hence increased efficiency for small motion	more complex than TSS and NTSS
DS Diamond Search [Zhu00]	188.87	16.5	18.52	$8 \times \log_2(p + 1)$	works well in searching small motion	can be very complex for large motion

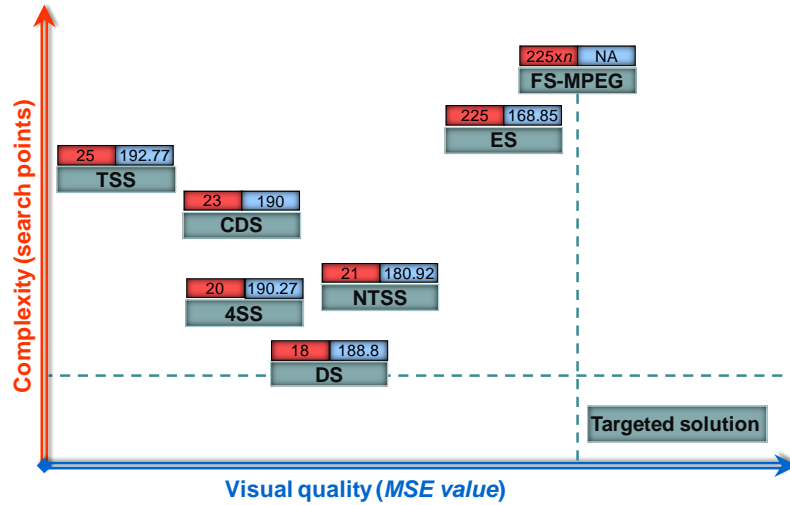


Figure 2.3. Synoptic representation of the state-of-the-art studies.

On the abscissa: the visual quality, expressed on MSE (the lower the MSE, the better the visual quality). On the ordinate: the computational complexity, expressed in number of search points per 16×16 block. The “best” solution should provide low complexity and high visual quality (see the “Target solution” block).

By inspecting the values in Tables 3.1 and 3.2, it can be noticed that the NTSS, 4SS and DS disparity map computation algorithms feature the best positioned solution in terms of complexity and reconstructed image visual quality.

However, as they were never benchmarked on the same data, we implemented them and assessed the quality of reconstructed image by five objective metrics: PSNR, IF, NCC, SC and SSIM (see Chapter 1.2.2.1). Each algorithm is individually run with the same matching similarity measures SAD. In this respect, we considered the two corpora (see Appendix A.1).

Table 2.2 presents the numerical results, corresponding to the 3DLive and MPEG corpora.

Table 2.2 Reconstructed image visual quality, expressed by PSNR, IF, NCC, SC and SSIM.
Each time (for each disparity map and each corpus), the similarity measures SAD has been considered for block matching.

		PSNR	IF	NCC	SC	SSIM
3DLive	NTSS	31.21	0.782	0.741	1	0.840
	DS	30.75	0.755	0.782	0.938	0.819
	4SS	27.36	0.671	0.698	0.812	0.805
MPEG	NTSS	32.56	0.684	0.889	1.025	0.917
	4SS	27.86	0.677	0.699	0.825	0.768
	DS	29.55	0.702	0.791	0.916	0.828

The quantitative results presented in Table 2.2 confirm the general view on the state-of-the-art presented in Figure 2.3: NTSS features the best trade-off between the reconstructed image quality and

the computational complexity among the fast algorithms and can be considered together with the FS-MPEG as comparative basis in our work.

2.1.3. Conclusion

The state-of-the-art approaches presented in the previous section can be suboptimal when computing the disparity map for live 3D HD television. First, the disparities between the left and the right views are predominant on the horizontal direction (where the very depth information is conveyed) while the vertical disparities, although non-zero, are mainly connected with the video shooting errors (the so-called rig errors). Secondly, basic metrics like MSE or SAD are unrelated to the human visual system and would *a priori* lack in achieving visually good block matching. Finally, note that none of the experiments reported in the above-mentioned studies considered HD stereoscopic image databases.

The present work takes the challenge of designing a new disparity map so as to reach the trade-off between visual quality and computational complexity when processing HD 3D TV (high definition stereoscopic television) content [3DI12], see “Targeted solution” in Figure 2.3.

2.2. Watermarking embedding techniques

The basic watermarking theoretical model (see Chapter 1.2.4) points at two main directions in watermarking: spread spectrum (SS) [Cox97] [Cox02] and side information (SI) [Cox02] [Cos83]. The former considers the original content as a random noise source and maximizes the quantity of inserted data under joint constraints of noise (original content and malicious attacks) and inserted signal power. The latter considers the original content as a noise source known at the embedder [Sha58] and maximizes the quantity of inserted data accordingly.

2.2.1. Principles

Spread spectrum

The SS methods have already been used in telecommunication applications (e.g. CDMA) and provided a good solution for very low power signal transmission over noisy channels [Cox97] [Cox02]. Their principle consists in spreading the signal over a very large band (e.g. 100 to 10000 times larger than the signal band), thus inserting a very low power signal in any frequency sub-band. Figure 2.4.a illustrates a simple spread spectrum watermarking system.

The spread spectrum communications have two major benefits. First, the signal energy inserted into any frequency sub-band is very small, thus reducing the signal-to-noise ratio and the risk of perceptible artifacts. Secondly, the redundant watermark spreading over such a large band of frequencies provides robustness to many common signal distortions such as band-pass filter or addition of band-limited noise [Cox97].

Side information

Side information – based watermarking techniques take advantage of the fact that the original content is known at the embedder side (but unknown at the detector). This knowledge can be exploited at two levels. First, the *informed coding* methods (see Figure 2.4.b) encode the information to be inserted into the original content by a codeword depending on that content, [Cox02] [Cos83]. Secondly, the *informed embedding* methods (see Figure 2.4.c) modulate the watermark according to the original content [Cox02] [Sha58] [Egg03]. In practice, informed coding and informed embedding can be individually or jointly deployed.

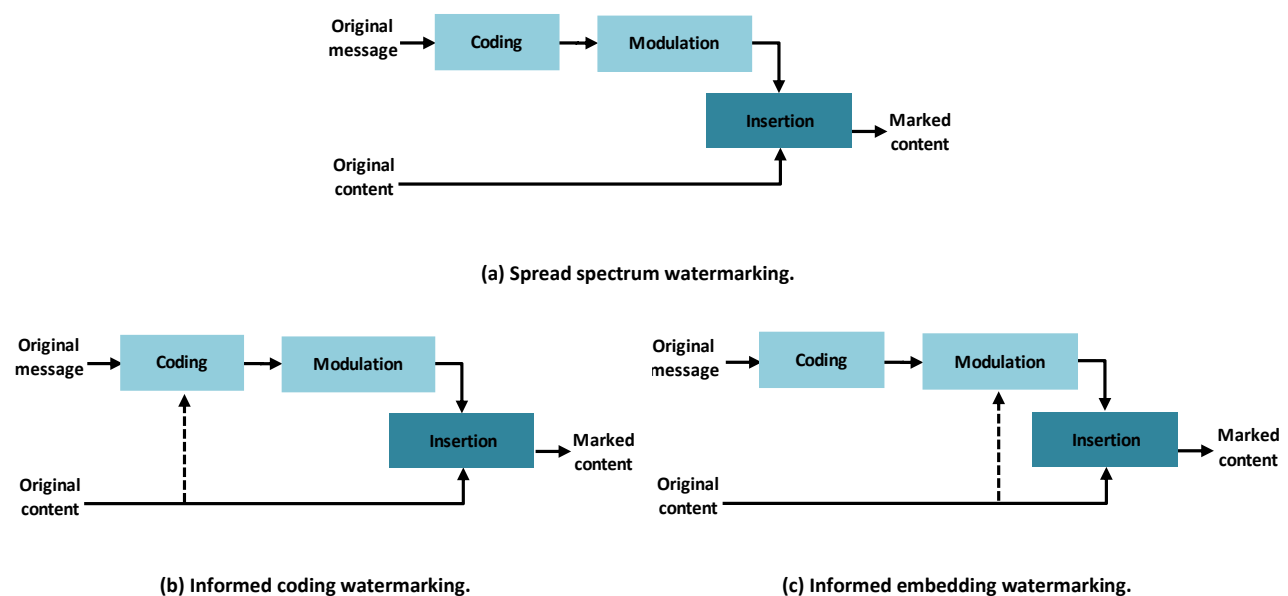


Figure 2.4 The watermarking embedder scheme.

(a) spread spectrum, (b) informed coding and (c) informed embedding watermarking techniques.

From the theoretical point of view, the side information watermarking is more sophisticated and should potentially outperform the blind coding and embedding methods. On the one hand, the informed coding would increase the transparency of the watermarking technique and, on the other hand the informed embedding would grant robustness against a large range of attacks. However, the studies reported in the literature show that with respect to the spread-spectrum method, the side information allows a significant increase of the data payload, generally at the expense of the robustness and/or transparency [Cox02].

2.2.2. State-of-the-art

While the insertion method itself is always directly inherited from still image/mono video (being either of spread spectrum or of informed embedding type), these studies can be structured into view-based and disparity-based methods, according to the domain in which the mark is actually inserted. Figure 2.5 gives an overview of these approaches.

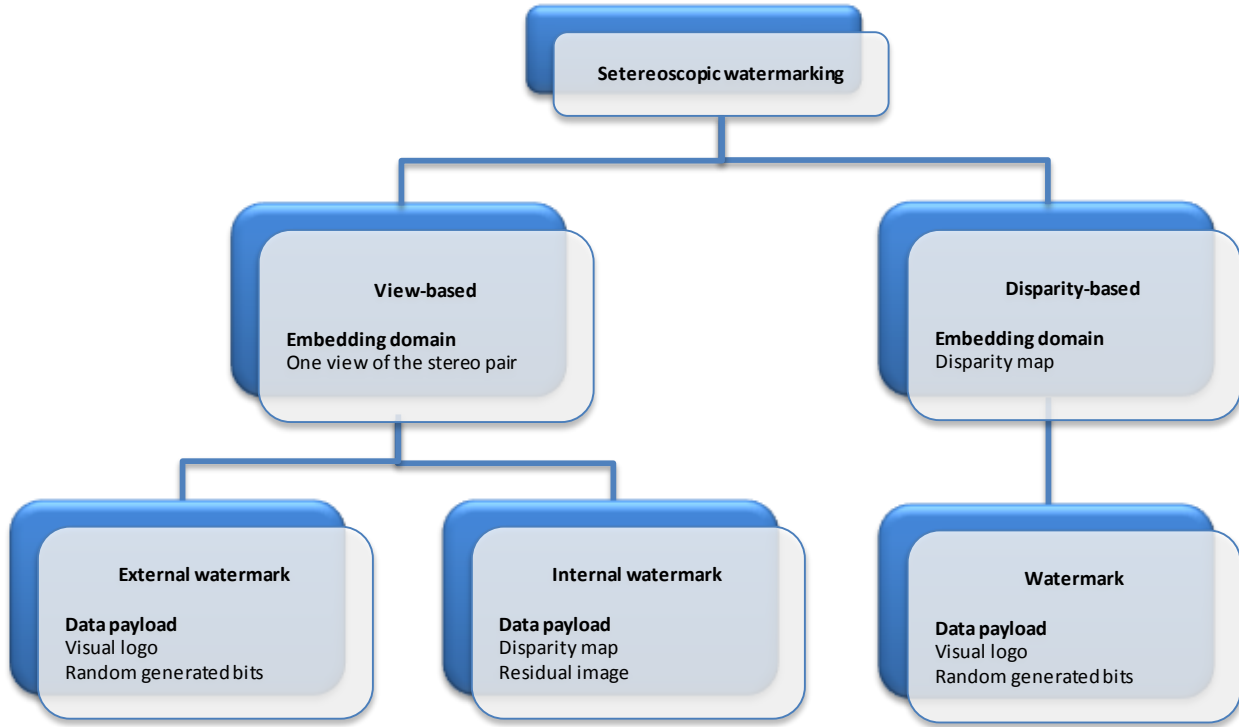


Figure 2.5 Synoptic overview of the state of the art of the stereoscopic watermarking approaches⁵.

2.2.2.1. View-based approaches

The predilection direction in the literature is represented by the view-based watermarking approaches, which are currently deployed for stereoscopic still images [Don03a-Cam08].

External watermark

In their incremental studies [Don03a-Don03b], Dong-Choon et al. address several issues connected to the spread spectrum stereo image watermarking. In [Don03a], a mark of 1024 bits (representing a visual logo) was embedded in the DCT domain of each of the right views of a stereo sequence of 25 image-pairs with a resolution of 256×256 pixels. Both transparency and robustness performances have been analyzed in terms of PSNR. Thus, the watermarked views feature an average PSNR of 34.89dB when compared to their corresponding original views. While no information is provided about the actual applied attacks, the robustness is evaluated with the PSNR measure computed between the recovered and the inserted logos which yielded an average value of 19.04dB.

In [Don03b], the experiment was resumed in the DWT domain. This time, a sequence of 3 image-pairs of 512×512 pixels was considered as the original content and a logo of 64×64 pixels as the watermark.

⁵ The DIBR is out of the scope of the present thesis.

While the average transparency was increased by 3.92 dB (reaching now the limit of 38.81 dB), the robustness was decreased by 2.36 dB (i.e. lowered at 16.68 dB).

Campisi [Cam08] advances a semi-fragile stereo watermarking approach based on the Quantization Index Modulation-QIM insertion method performed in the DWT domain. The embedding scheme relies on the use of the depth information, extracted from the stereo pair, to determine the region where the watermark will be embedded. The depth information is extracted from the low pass-subband (LL) of the selected DWT decomposition level. The watermark payload is 2000 bits per frame and is embedded in the high level subbands (HL, LH and HH) of the wavelet decomposition. Experiments show that the advanced method is robust to JPEG and JPEG2000 compression and fragile with respect to other signal manipulations (like Row/Column removal and small rotations). The fragility property was assessed by computing the values of BER (bit error rate); the following numerical values are obtained: 0.07 after Gaussian filtering, 0.11 after median filtering, 0.38 after row/column removal, 0.30 after a 0.25 degree rotation and 0.39 after 0.50 degree rotation. No information concerning the transparency property or the experimental database is provided.

The study advanced by Yu et al. [Yu11] embeds the watermark into both the left and the right views of the stereo pair, by taking into account some intra and inter-blocks statistical relationships established by combining the DWT and DCT. A parity quantization is also designed for handling the cases in which such a relationship does not hold. During the experiments, a binary logo of 64×64 pixels is embedded in a stereo frame of 640×480 pixels. The transparency is evaluated by the PSNR value between the original and the watermarked images; values of 52.14 dB and 51.99 dB are obtained for the left and the right views, respectively. The robustness was evaluated in terms of the Watermark Recovering Rates (WRR) (referred to as HC in the [Yu11]) given by: $WRR = 1 - \sum w \oplus w' / m \times m$, where \oplus is exclusive-OR, w denotes the original binary watermark of $m \times m$ size, and w' denotes the recovered watermark. Values of WRR of 0.94, 0.90, 0.94 and 0.81 are obtained after applying a JPEG compression, a salt and pepper noise, a median filtering, and a cropping, respectively. Note that $WRR = 1 - BER$.

In Ntalianis *et al.*'s study [Nta02], a grayscale image of 6×20 pixels is redundantly embedded in the coefficients of the highest energy of the video objects. Initially, the video objects are extracted using a segmentation process based on a depth segments map, and then are decomposed into three levels using the 2D-DWT. Later on, a Qualified Significant Wavelet Tree (QSWT) is constructed, including the highest energy coefficient given by the 2D-DWT. Once the QSWTs of the video sequence objects are created, the watermark of size $a \times b$ is repeatedly embedded in the top $a \times b$ of the QSWTs. The embedding process is spread spectrum based. The performances of the watermarking system are assessed in terms of imperceptibility and robustness. The obtained PSNR value of the watermarked video object is 44.3 dB. The robustness has been investigated under five different types of attacks: salt & pepper noise, Gaussian noise, blurring, sharpening and JPEG lossy compression. The obtained PSNR values of the recovered watermark are of 24.7, 26.43, 22.53, 30.26 and 32.06 dB, respectively.

Internal watermark

In this category of view-based stereo watermarking methods, the disparity map is used as a watermark and embedded in one view of the stereo pair. Two major contributions are discussed in the sequel.

Kumar et al [Kum09] also considered a spread spectrum based watermarking method. The disparity map is first estimated in the DWT domain and then embedded as a mark in the left view. The embedding procedure is optimized by a genetic algorithm and takes place in the singular value decomposition (SVD) domain of the left DWT image. Before the insertion process, the image undergoes an Arnold Transform. A genetic algorithm is used to estimate the optimal order of the Arnold Transform and the watermark's strength. The experiments have been performed on five gray level stereo images with a resolution of 512×512 pixels. The watermark image is one fourth of the host image. The obtained transparency was expressed by a PSNR of 42.88 dB. The robustness against average filtering, rotations, resizing and Gaussian noise addition was assessed by computing the normalized cross correlation (NCC) value between the original and the extracted watermarks, namely 0.91, 0.91, 0.90 and 0.89, respectively.

In Bhatnager et al.'s study [Bha09], the watermark is represented by the disparity map, computed this time directly in the pixel domain. This disparity map is further inserted by a SS technique applied in the SVD domain of the Discrete Fractional Fourier Transform (DFrFT) of the left image. Before the insertion procedure, the watermark's strength parameter α was defined according to an iterative experimental process. The selection is achieved between 25 different values going from 0.02 to 0.50 with a step size of 0.02. A Performance Index (PI) is calculated to define the appropriate α value. PI is based on the PSNR value computed between the original and the watermarked image and the NCC between the original and recovered watermark.

The experimental validation considered a corpus of three stereo images (two image pairs of 256×256 pixels and one image pair of 512×512 pixels). The transparency evaluation shows an average PSNR of 45.92 dB. This excellent value is obtained at the expense of the robustness which is now reduced at NCC (between the original and the extracted watermark) values of 0.64, 0.71, 0.69, 0.98, 0.57, 0.63 and 0.46 for the average filtering, median filtering, resizing, JPEG compression, cropping, rotation and additive noise attacks, respectively.

A particular case of the internal watermarking is represented by the reversible watermarking. Actually, reversible watermarking is rather a compression technique than a copyright protection tool. Its aim is to reduce the bandwidth consumption when transmitting/storing stereoscopic video content. In this respect, one of the two views is kept as reference and the difference with respect to the second view are inserted as a watermark. From the applicative point of view, such an approach strengthens the data payload requirements while completely relaxing the robustness constraints (compression is the only investigated attack).

In his study, Coltuc [Col07] embeds the compressed normalized residual in the left view. The watermarking scheme starts by partitioning the image into pairs of pixels, where the watermark will be inserted by a simple addition. The experimental test considered two images and measured the robustness against JPEG compression. The PSNR of the recovered images are of 31.56 dB and 38.64 dB, respectively. The bit rate needed to embed the lossy compressed residual image is 1.11 bpp (bit per pixel) for an image of 450×375 pixels. According to the obtained value of PSNR and the considered image content, it has been proven that, the less complex textured the image, the better the quality of the recovered image versus the compression ratio.

Ellinas 2009 [Ell09] embeds the normalized compressed residual frame and the disparity vector in the left view of the stereo pair. The considered embedding algorithm is Pixel Value Difference (PVD), which partitions the reference frame (the left view in this study) into blocks of two consecutive non-overlapped pixels according to a scan order. Then, it embeds the watermark in each pair of pixels. The experimental validation considered a corpus of two stereoscopic image sequences of 28 images each, with a resolution of 320×240 pixels. The algorithm performances are evaluated in terms of PSNR average values of the watermarked frames given by the PSNRw and the recovered right frames PSNRr. The obtained values are of 35 dB and 45 dB, respectively. The average embedding capacity is evaluated for the test sequences at 117 700 bits. However, the actual data payload accounts for 106 000 bits.

Chen *et al.* [Che12] embed the compressed residual error image and the compressed disparity vector field into the right frame. The stereo image is then coded as a watermarked 2D image and transmitted to the receiver. Two stereo images are used to evaluate the performance of the suggested scheme, with a size of 413×370 and 384×288 pixels. Their maximum embedding capacities are of 2.19 bpp and 1.36 bpp respectively. The obtained average PSNR for the watermarked right frame is 17.51 dB and for the reconstructed left frame is 45.84 dB.

2.2.2.2. Disparity-based approach

The *disparity-based* stereoscopic watermarking schemes can be represented by the study reported in [Zen07]. The insertion technique combines spread spectrum principles and Low Density Parity Check (LDPC) error correcting codes. The experiments are performed on four video sequences: three of them are composed of 22 stereo frames of 512×512 pixels while the forth one has 192 frames of 720×576 pixels. The transparency is assessed by computing the average image degradation index $\bar{\delta} = 0.06$. This index is defined by $\bar{\delta} = \bar{\alpha} - \bar{\beta}$, where $\bar{\alpha}$ and $\bar{\beta}$ denote the average PSNRs of the non-watermarked and watermarked video images respectively (no details are provided on how the non-watermarked sequence average PSNR is computed). The robustness is assessed in terms of the Watermark Detection Ratio (WDR), evaluated for each video and defined by: $WDR = e_0 e_1 / m_0 m_1$, where m_0 and m_1 are the numbers of "0" and "1" of the original watermark respectively, while e_0 and e_1 are the numbers of the extracted "0" and "1" from the recovered watermark, respectively. The reported WDR values are larger than 0.78 after recoding and 0.74 after both recoding and noise addition.

2.2.3. Discussion

Table 2.3 and Figure 2.6 present a general view on these various approaches. When considering them for real life applications (e.g. HD 3D TV real time content protection) the following main limitations can be identified:

- the structure and the size of the processed corpora are too small to ensure generality and statistical relevance for the results;
- the selection of the insertion domain was rather the experimenter's choice than the result of an objective study;

- the transparency evaluation was solely guided by the PSNR values; no other objective transparency method nor subjective evaluation are reported;
- the robustness against geometric attacks is not yet achieved;
- the computational cost was never investigated.

The present study addresses these above-mentioned issues:

- all the results consider two corpora, further referred to as 3DLive [3DI12] and MPEG [Hoi11];
- a new 3D video disparity map is considered and its watermarking usefulness is objectively investigated in terms of transparency, robustness and computational cost;
- the transparency evaluation is carried out on both subjective and objective basis;
- the robustness is objectively expressed by means of the watermarked detection BER against several classes of attacks, such as linear and non-linear filtering, compression and geometric transformations;
- the computational cost is estimated for each and every processing step involved in the watermarking chain.

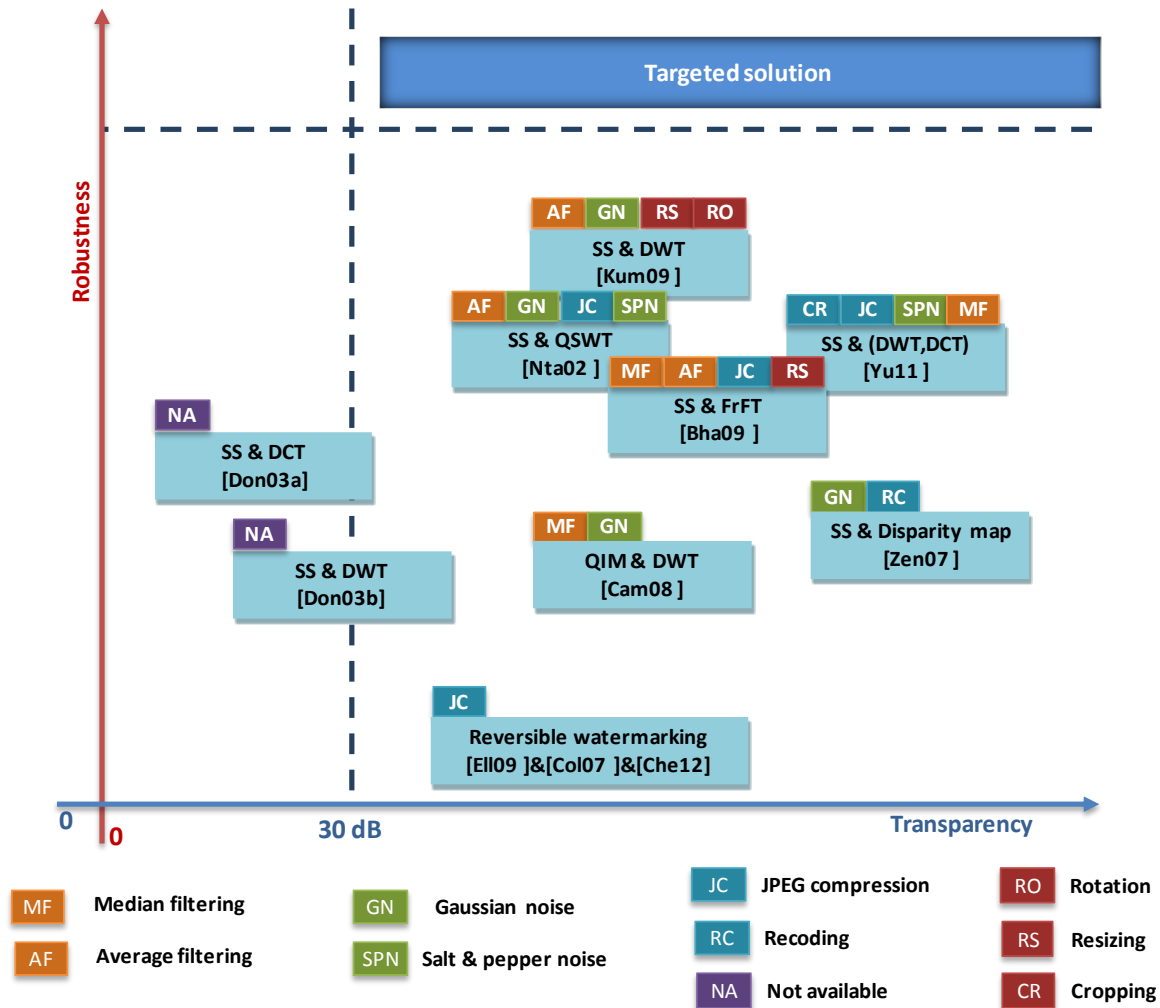


Figure 2.6 Synoptic representation of the state-of-the-art studies.

On the abscissa: the transparency of the watermarked, expressed by PSNR value. On the ordinate: the robustness against various types of attacks (the stronger the attack it withstands, the better the robustness). For the study in [Don03a] and [Don03b] no information is provided about the actual applied attacks (hence NA-Non Available label is assigned). The “best” solution should provide robustness against a widest class of attacks and transparency above a prescribed limit (see the “Targeted solution” area).

Table 2.3 A comparative view on the state of the art of stereoscopic image/video watermarking methods.

		Embedding technique	Corpus	Transparency marked vs. original views	Robustness recovered vs. original watermark	Data payload (per frame)	Processing time
View-based	External watermarking	SS in the DCT domain [Don03a]	25 images of 256x256 pixels	Average PSNR	Average PSNR	Visual logo	NA
				34.89dB	19.04 dB	32x32 pixels binary image	
		SS in the DWT domain [Don03b]	3 images of 512x512 pixels	Average PSNR	Average PSNR	Visual logo	NA
				38.81dB	16.68dB	64x64 pixels binary image	
		QIM in the DWT domain [Cam08]	NA	NA	BER (bit error rate)	Random generated bits	NA
					Gaussian filtering : 0.07 Median: 0.11 Row/Column removal: 0.38 Rotation 0.25°: 0.30 Rotation 0.5°: 0.39	2000 bits	
		SS in the combined DWT-DCT domain [Yu11]	1 image of 640x480 pixels	Average PSNR	WRR (watermark recovering rate)	Visual logo	NA
				52.07dB	JPEG compression: 0.94 Salt & pepper noise: 0.90 Median filtering: 0.94 Cropping: 0.81	64x64 pixels binary image	
		SS in the QSWt [Nta02]	Eye2eye stereo video sequence (88 frames)	Average PSNR	Average PSNR	Visual logo	NA
				44.3 dB	Gaussian noise : 26.43 dB JPEG compression: 32.06 dB Salt & pepper noise: 24.7 dB Gaussian filtering: 22.53 dB Sharpening : 30.26 dB	6x20 grayscale image	

Table 2.3 (continued) A comparative view on the state of the art of stereoscopic image/video watermarking methods.

		Embedding technique	Corpus	Transparency marked vs. original views	Robustness recovered vs. original watermark	Data payload (per frame)	Processing time
View-based	Internal watermarking	SS in the SVD-DWT domain [Kum09]	5 images of 512x512 pixels	Average PSNR 44.05dB	Average NCC (normalized cross correlation) Average filtering: 0.91 Rotation: 0.91 Resizing: 0.90 Gaussian noise: 0.89	Disparity map 128x128 pixels image	Advantage Disparity map computed in the DWT domain Drawback Genetic algorithm
		SS in the FrFT domain [Bha09]	2 images of 256x256 pixels 1 image of 512x512 pixels	Average PSNR 48.5dB	Average NCC JPEG compression: 0.98 Gaussian noise: 0.46 Median filtering: 0.71 Average filtering: 0.64 Rotation: 0.63 Resizing: 0.69 Cropping: 0.57	Disparity map 64x64 pixels image 128x128 pixels image	NA
		SS in the DWT domain [Fra11]	1 video sequence of 1024x768 pixels (number of frames not available)	NA	Correlation score Low pass filtering (3x3): 0.81 Gaussian noise (AWGN): 0.72 MPEG4 compression at high (3 Mbps): 0.62 MPEG4 compression at low bit rates (500 kbps): 0.29	NA	NA
		Linear addition PVD [Col07]	2 stereo images	Average PSNR 35 dB	JPEG compression	Compressed normalized residual image 1.11 bpp 0.89 bpp (bits per pixel)	NA
		Linear addition PVD algorithm [Eli09]	2 video sequences with images of 320x240 pixels (number of frames not available)	Average PSNR 35 dB	NA	The compressed residual image and the disparity vector 106 000bits	NA

Table 2.3 (continued) A comparative view on the state of the art of stereoscopic image/video watermarking methods.

		Embedding technique	Corpus	Transparency marked vs. original views	Robustness recovered vs. original watermark	Data payload (per frame)	Processing time
View-based	Internal watermarking	Linear addition [Che12]	1 image of 413×370 pixels	Average PSNR	JPEG compression	Compressed residual image & compressed disparity vector	NA
			1 image of 384×288 pixels	17.5 dB		2.19bpp for 413×370 pixels image 1.36bpp for 384×288 pixels.	
Disparity-based		SS in the disparity map [Zen07]	3 video sequences of 22 frames of 512×512 pixels	Average degradation index	WRR	NA	NA
			1 video sequence of 192 frames of 720×576 pixels	0.06dB	Recoding: 0.86 Recoding and noise addition: 0.81		

2.3. Conclusion and thesis objectives

This Chapter presents the main trends of disparity map computation and stereo watermarking techniques given by the literature.

Concerning the disparity map, the challenge is to advance a method devoted to watermarking purposes. Seven algorithms have been investigated and it has been noticed that the disparity maps are directly inherited from 2D video and simply neglect the stereoscopic video peculiarities. For such content, the disparities between the left and the right views are predominant on the horizontal direction (where the very depth information is conveyed) while the vertical disparities, although non-zero, are mainly connected to the video shooting errors. Moreover, in the block matching stage, basic visual quality metrics like mean squared error – MSE or sum of absolute differences - SAD are considered. As such metrics are unrelated with the human visual system, the obtained results are sub-optimal from a quality of experience perspective, see Table 2.3. Consequently, specifying a disparity map, jointly exploiting the horizontal/vertical peculiarities of the stereoscopic content and a visual quality metric related to the HVS remains a challenging research topic and will be addressed in Chapter 3.

Concerning the stereoscopic watermarking techniques, several classes of insertion techniques, that already proved their efficiency for 2D video watermarking applications, have been extended for stereoscopic watermarking system. However, they still lack in achieving good transparency, they deal

only with a restricted class of attacks and fail in embedding the required amount of data payload. Consequently, specifying a watermarking method able to reach this three-folded trade-off is still an open issue, which will be considered in Chapters 4 and 5. Such a study should also be accompanied by an evaluation of the computational cost.

The thesis main challenges and their underlying current limitations are presented in Table 2.3.

Table 2.4 The thesis main challenges, the underlying current limitations and the thesis objectives.

Constraints	Challenges	Current limitations	Thesis objectives
Embedding domain	Disparity map for watermarking application	2D video inherited disparity map <ul style="list-style-type: none"> ignoring the vertical disparities block matching metrics unrelated to the human visual system prohibitive computational cost for HD 3D content 	New disparity map <ul style="list-style-type: none"> content adaptive (discriminating weights between the horizontal and vertical disparities) human visual system metrics
Embedding technique	Transparency/robustness/data payload trade-off Low computational cost	Transparency <ul style="list-style-type: none"> no subjective evaluation a single objective metric (PSNR) Robustness <ul style="list-style-type: none"> fragility to geometric random bending transformations Computational cost <ul style="list-style-type: none"> Never investigated 	Optimal watermarking technique <ul style="list-style-type: none"> Transparency Robustness Computational cost
Performance evaluations	Benchmarking <ul style="list-style-type: none"> statistical relevance of the results standard recommendations 	No statistical background <ul style="list-style-type: none"> limited structure and small size in processed corpus no statistical relevance for the results Application-driven benchmarking <ul style="list-style-type: none"> no referenced standards 	Corpora design Statistical relevance for results Standards compliance

Chapter 3 A new insertion domain: 3D Video-New Three Step Search

Abstract

This Chapter presents the newly designed disparity map (the 3D video-New Three Step Search - 3DV-NTSS). 3DV-NTSS reconsiders some 2D video inherited approaches and adapts them to the stereoscopic video content and to the human visual system peculiarities. The inner relation between the left and the right views is modeled by some weights discriminating between the horizontal and vertical disparities. First, the Chapter introduces the advanced disparity map. Second, it describes the experimental validations and demonstrate the 3DV-NTSS effectiveness in both reconstructed image quality (average gains between 3% and 7% in both PSNR and SSIM) and computational cost (search operation number reduced by average factors between 1.3 and 13).

3.1. Method presentation

3.1.1. Principles

The general idea of the 3DV-NTSS algorithm, Figure 3.1.b, is based on the NTSS procedure (see Figure 3.1.a) while taking into account the spatial constraints brought by the stereoscopic context and the human visual system peculiarities. As the right and left cameras are located on the same horizontal plane (given by the rig), the horizontal disparities are to be preponderantly considered in depth computation, while vertical disparities are mainly linked to the rig alignments errors. Hence, the 3DV-NTSS algorithm assigns discriminative weights for the horizontal and vertical disparities which are subsequently used to adapt the vertical and horizontal sizes of the search area according to the content on which the disparity is estimated (see Figure 3.1.b). The block matching is achieved by the NCC – Normalized Cross Correlation similarity measure.

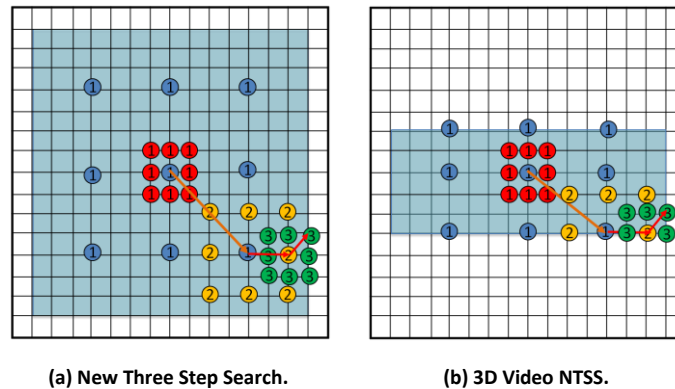


Figure 3.1 Block searching strategies: (a) NTSS, and (b) 3DV-NTSS.

3.1.2. Pseudo-code and flowchart

The 3DV-NTSS algorithm, illustrated in the flowchart in Figure 3.2, is further presented for groups of T adjacent pairs of frames, each of which is divided in N blocks of 16×16 pixels. Details related to each step are illustrated in figure 3.3.

Step1 Parameter initialization

Step1.1 Global parameters

Search area size $H \times V = 15 \times 15$ pixels, $frame_index = 1$, $T = 750$,
 $similarity_measure = NCC$,
 $reference_frame = \text{right frame}$.

Step1.2 Block index initialization

$block_index = 1$.

Step1.3 Block level parameters

Search area center $c(0,0)$, search distance $s = 4$ pixels, $disparity_h = 0$, $disparity_v = 0$.

Step2 Disparity computation

WHILE $block_index \leq N$ DO:

Step2.1 Similarity computation

Compute the $similarity_measure$ between the block located at c in the target image and its homologous in the reference image, then between it and its 8 neighbors located at distance 1 on the vertical/horizontal axis, and finally between it and its 8 neighbors located at distance s on the vertical/horizontal axis.

Step2.2 Disparity updating

Step2.2.1 Search area center similarity

IF the largest $similarity_measure$ value is provided by the search area center,
 THEN set $block_index = block_index + 1$ and GOTO Step1.3.

Step2.2.2 Search step iteration

IF the largest $similarity_measure$ value is provided by a block located at (x, y) where x or $y \in \{-s, -1, 0, 1, s\}$
 THEN set: $c(x, y)$, $s = s/2$, $disparity_h = disparity_h + x$,
 $disparity_v = disparity_v + y$ and GOTO Step2.1.

Step3 Searching area adaptation

IF $frame_index = 1$,

THEN

Step3.1 Weights estimation

$disparity_average_h$ = the average of horizontal disparities obtained for each block

$disparity_average_v$ = the average of vertical disparities obtained for each block

$$\alpha = 2 \times disparity_average_h / H$$

$$\beta = 2 \times disparity_average_v / V$$

Step3.2 Searching area resizing

set: $H = \alpha H, V = \beta V$

$frame_index = frame_index + 1$

GOTO Step1.2.

Step4 Exit condition

$frame_index = frame_index + 1$

IF $frame_index \leq T$ THEN GOTO Step1.2

ELSE STOP.

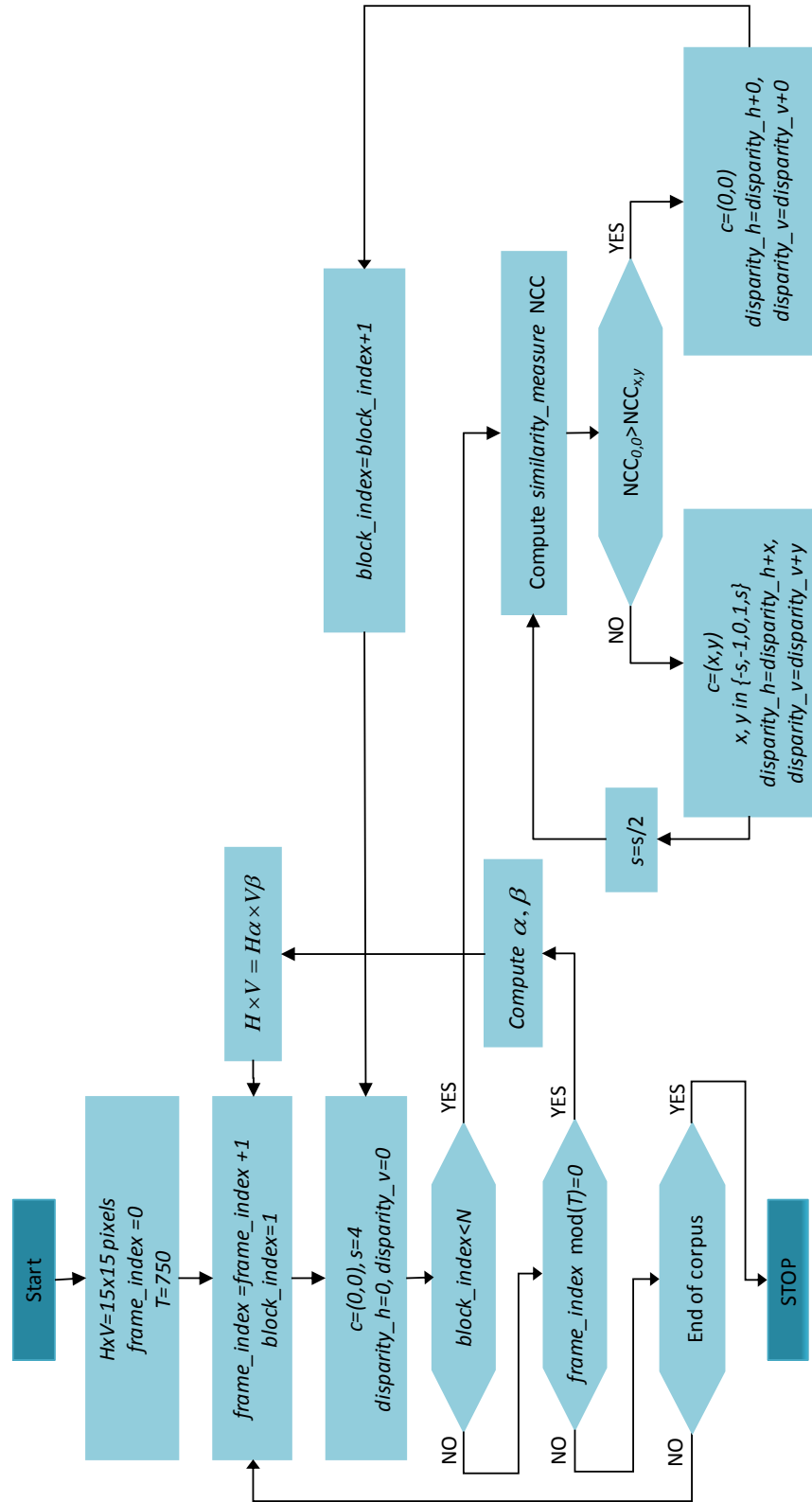


Figure 3.2 The 3DV-NTSS algorithm flowchart.

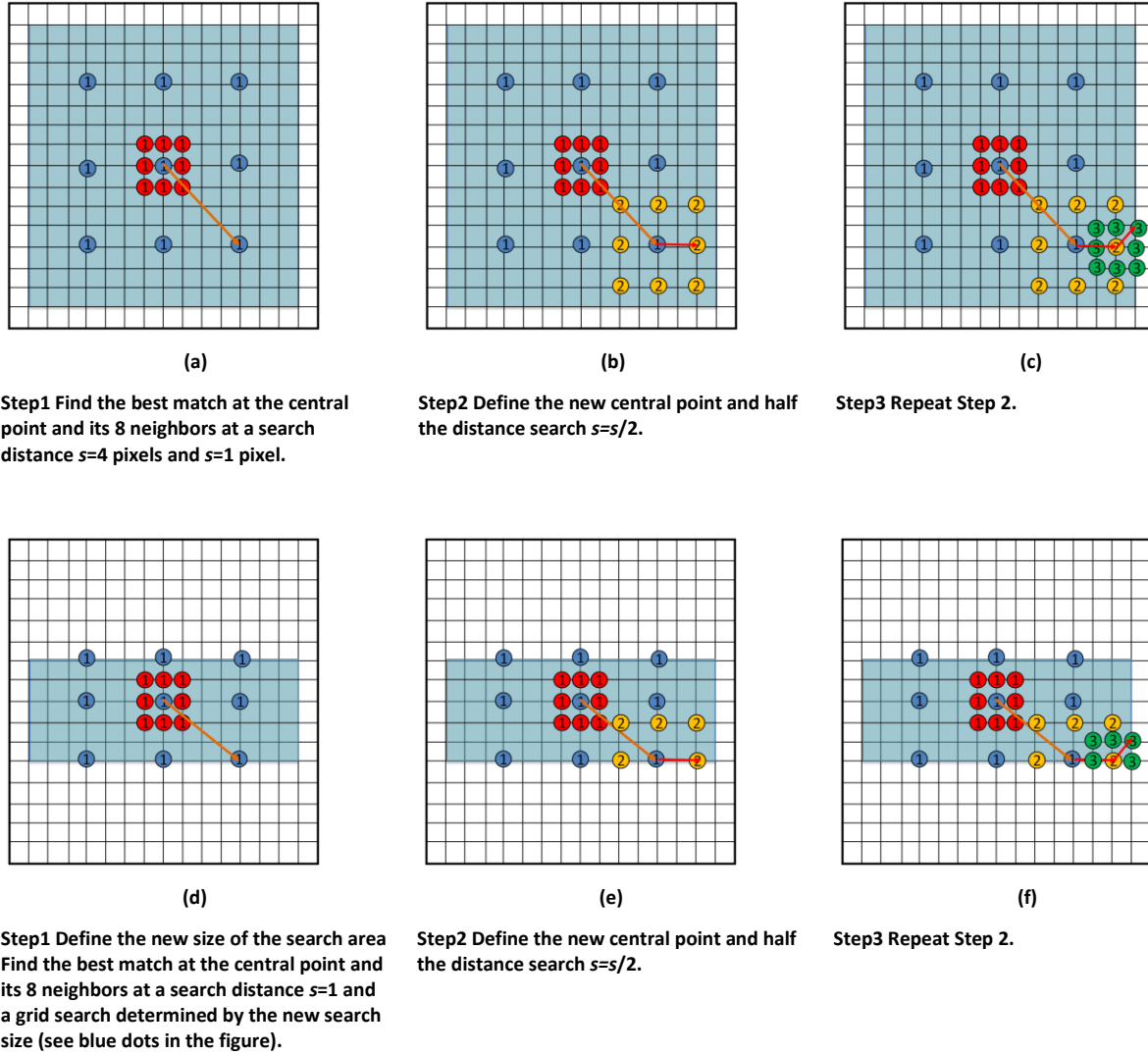


Figure 3.3 The main three steps of the 3DV-NTSS algorithm.
 Illustrations, for a block of 16×16 pixels size, before and after the search area adaptation.

Figure 3.3 is an illustration of the main steps of the 3DV-NTSS algorithm. For the first frame in a sequence of T frames, the search area size $H \times V = 15 \times 15$ is initially defined (step a). Then, the best match is searched for at the central point of the search area and in a grid search of 9×9 and 1×1 (step b). Once the best match is found, a new central point is defined and the search distance is halved. The third step is identical with the second step, but for a search distance $s = 1$ pixels. These three steps are repeated for all the blocks of the 1^{st} frame. The resulting average values of the vertical and horizontal disparities define the discriminative weights for the horizontal and vertical directions which are subsequently considered to adapt the search area sizes. Steps (d), (e) and (f) are then processed for the next $T - 1$ frames.

3.2. Experimental validation

3.2.1. Test bed

All the experiments reported in the present Chapter are carried out on two corpora, further referred to as 3DLive [3DI12] and MPEG [Hoi11], respectively. Each of these two corpora combines indoor/outdoor, unstable and arbitrary lighting conditions, still and high motion scenes, as illustrated in Figures 3.4 and 3.5 and as detailed in Appendix A.1.



Figure 3.4 Left and right views sampled from the 3DLive corpus.
The Rugby and Dancing sequences.

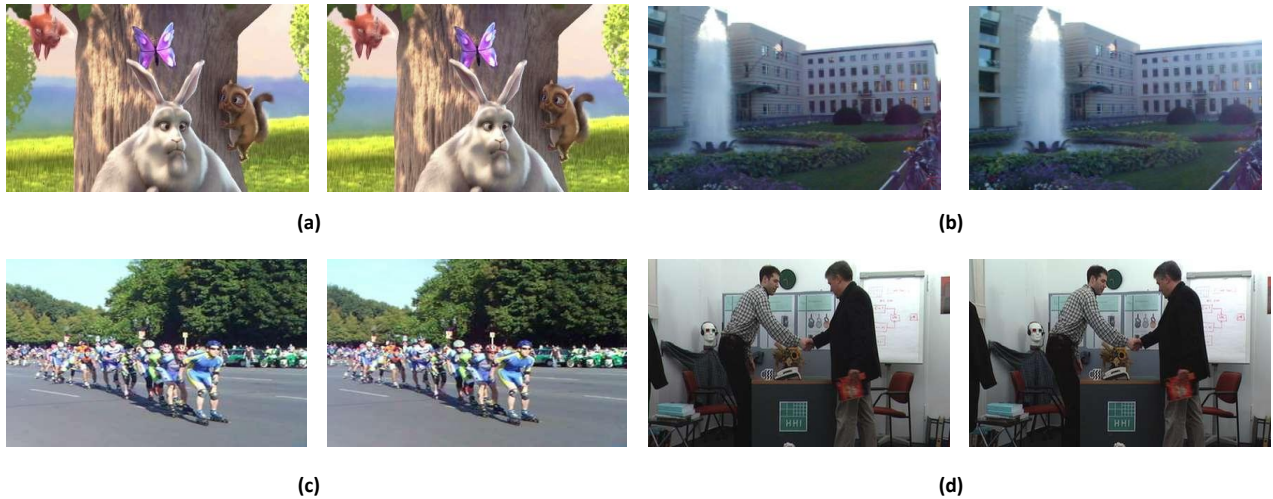


Figure 3.5 Left and right views sampled from the MPEG corpus.
Cartoons, city tours, rollerblade, and indoor content.

Note that for the uniformity purposes, the frames in Figures 3.4 and 3.5 are represented at the same size, although their actual sizes are very different, as explained in Appendix A.1.

In the sequel, the new 3DV-NTSS method will be benchmarked against two state-of-the-art algorithms, namely the NTSS and FS-MPEG. These two algorithms were implemented according to their references in [Li94] and [Sta08], respectively. Such an approach was compulsory, as no reference software was available on the MPEG software repository and as it gives us the possibility to customise these algorithms for an accurate benchmarking (e.g. changing the similarity measures).

3.2.2. Horizontal and vertical displacements tendency

The first experiment investigates the *a priori* potential of the adaptive search window size, see Step3.1 in the 3DV-NTSS algorithm. In this respect, Figure 3.6 illustrates the α and β adaptation weight variation as a function of time for the 3DLive and MPEG corpora, respectively. It can be seen that α and β continuously vary and that the variations are even more important on the MPEG corpus. This preliminary result shows that the adaptation mechanism included in the 3DV-NTSS is justified for all the stereoscopic pairs included in our corpora. Note that in the disparity map computation, the α and β weights are involved indirectly, by means of H and V searching area sizes, respectively. Their variations illustrate the intimate relationship between the horizontal and vertical disparities in the considered stereoscopic video corpora, given not only by the camera set-up but also by the natural content itself. However, when considering different types of content (e.g. computer augmented medical data), such a behavior is expected to change.

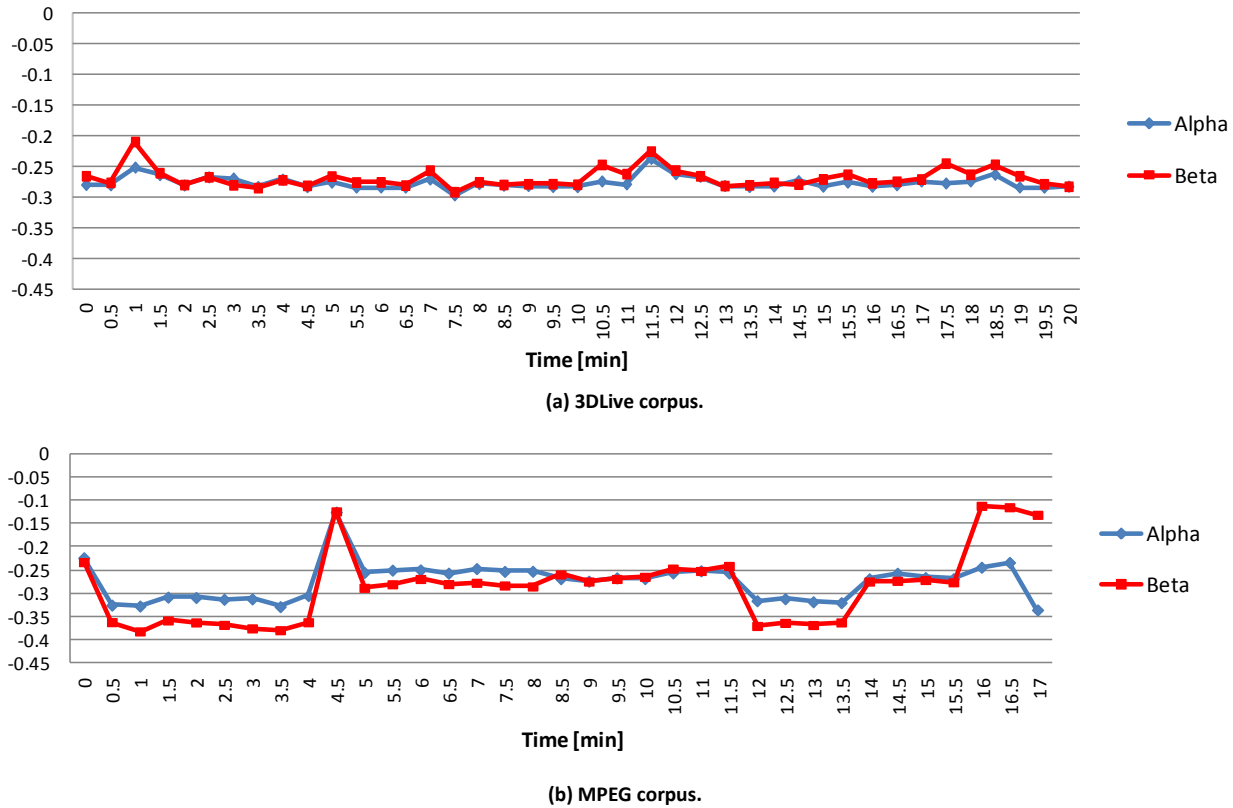


Figure 3.6 Search area weights variation (α and β) for 3DLive and MPEG corpora.

The α and β variations are plotted as functions of time (expressed in minutes). The time axis was sampled with a 0.5 minute step.

3.2.3. Visual investigation of the disparity maps

The disparity map computation for the four stereo pairs in Figure 3.4.a, 3.4.b and Figure 3.5.a and 3.5.b are presented in Figures 3.7, 3.8, 3.9 and 3.10 respectively. Three disparity map algorithms have been considered: the NTSS (see Figures 3.7.a, 3.8.a, 3.9.a and 3.10.a), the FS-MPEG (see Figures 3.7.b, 3.8.b, 3.9.b and 3.10.b) and the 3DV-NTSS (see Figures 3.7.c, 3.8.c, 3.9.c and 3.10.c). For each of these three algorithms, three similarity measures were alternatively employed, namely the MSE, the SAD and the NCC.

While there is neither a standardized procedure nor an objective basis for visually assessing the quality of the disparity maps, a panel composed of 5 image processing experts agreed that: (1) for a given disparity map computation algorithm, the use of the NCC as a similarity metric tends to give results closer to the intuitive ones; (2) assuming NCC as similarity measure, the use of 3DV-NTSS provides a finer granularity than the NTSS and FS-MPEG.

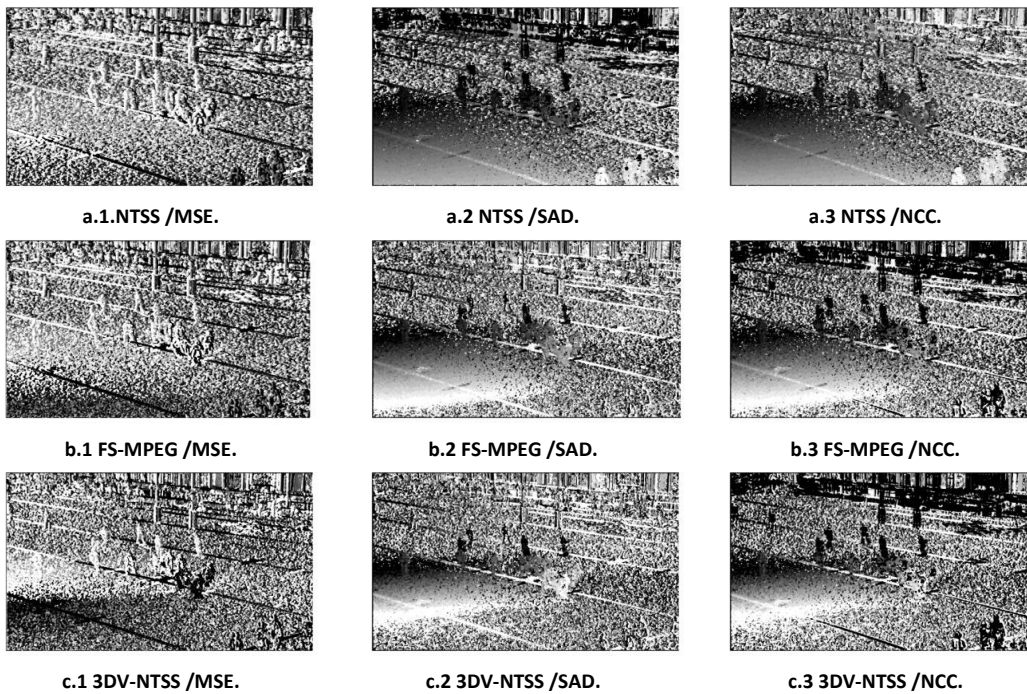


Figure 3.7 NTSS (a), FS-MPEG (b) and 3DV-NTSS (c) disparity maps.
Disparity maps samples corresponding to the stereo pair in Figure 3.4.a.

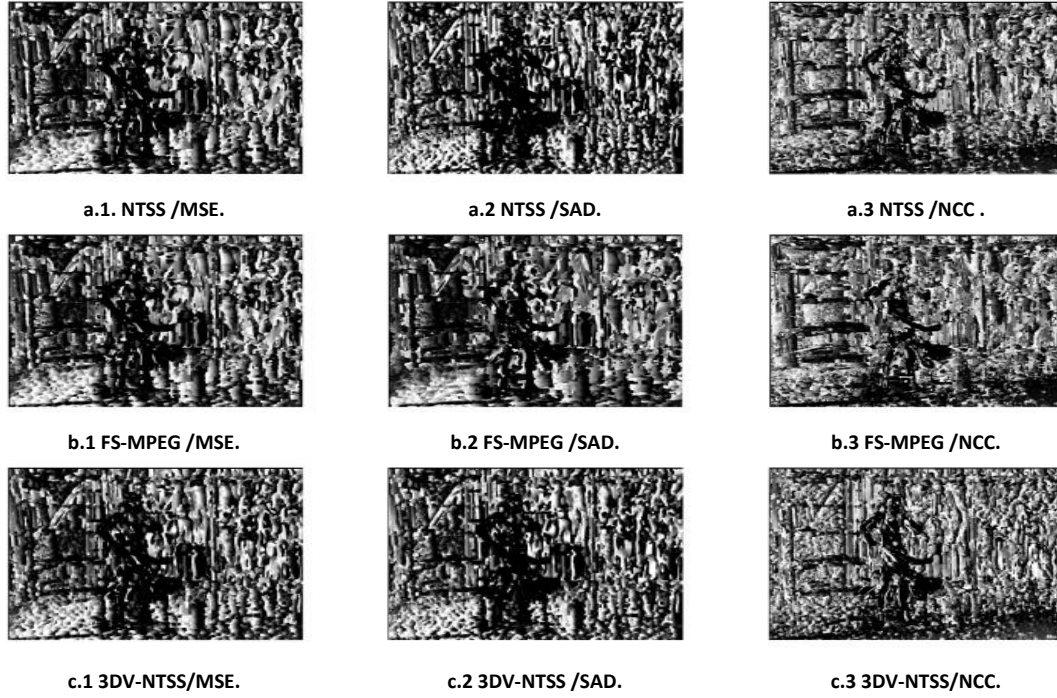


Figure 3.8 NTSS (a), FS-MPEG (b), and 3DV-NTSS (c) disparity maps.
Samples corresponding to stereo pair in Figure 3.4.b.

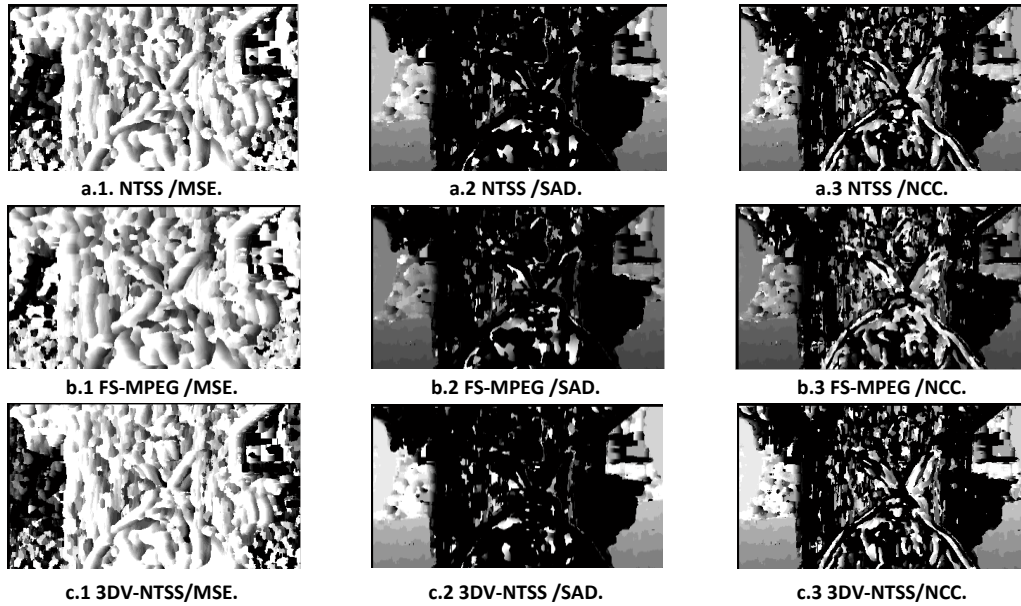


Figure 3.9 NTSS (a), FS-MPEG (b), and 3DV-NTSS (c) disparity maps.
Samples corresponding to Figure 3.5.a.

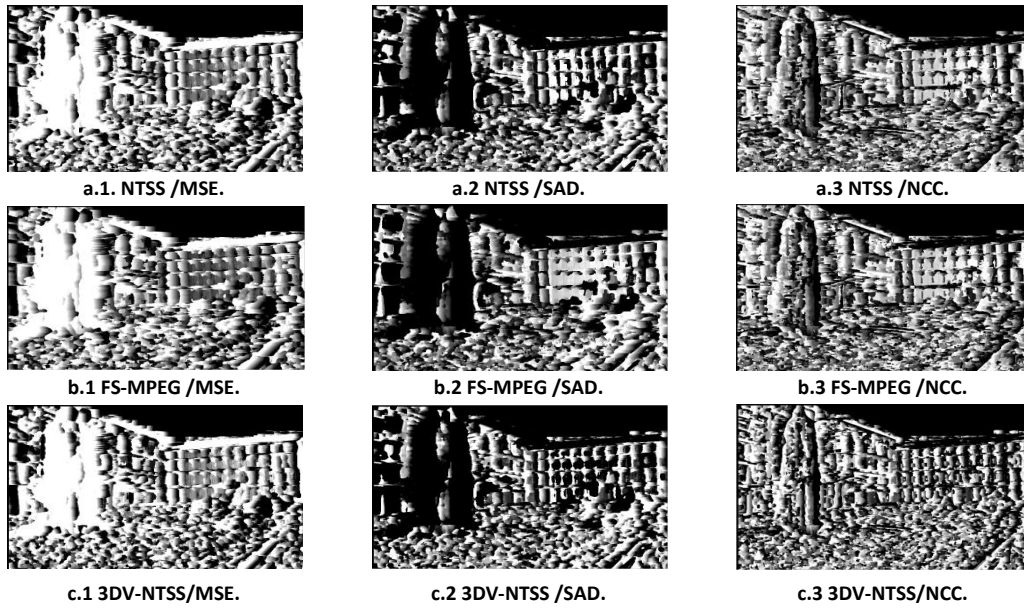


Figure 3.10 NTSS (a), FS-MPEG (b), and 3DV-NTSS (c) disparity maps.
Samples corresponding to Figure 3.5.b.

3.2.4. Reconstructed image quality assessment

The quality of the reconstructed images was assessed for the same above-mentioned three algorithms (NTSS, FS-MPEG and 3DV-NTSS), each of which being individually run with the same three different matching similarity measures (MSE, SAD and NCC).

The quality of the reconstructed images was evaluated by five objective metrics: PSNR, IF, NCC, SC and SSIM (see Chapter 1.2.2.1).

These measures are individually computed on each and every frame in the reconstructed video sequence and subsequently averaged over all the frames in the sequence.

Table 3.1 presents the numerical results, corresponding to the Rugby and Dancing sequences, as well as to the 3DLive and MPEG corpora. The following conclusions can be drawn:

3DV-NTSS vs. NTSS

(1) The search area adaptivity involved in the 3DV-NTSS leads on the 3DLive corpus to average relative gains of: 4% in PSNR, 4% in IF, 18% in NCC, 1% in SC and 4% in SSIM. However, when considering the MPEG corpus, NTSS outperforms 3DV-NTSS with an average relative gain of 2% in PSNR. When considering IF, NCC, SC and SSIM, the two methods provide quite similar results (average relative gains lower than 1%). Note: for a given quality measure (PSNR, IF, NCC, SC or SSIM), these average relative gains are computed by averaging the three relative gains corresponding to the three types of the block-matching measures (MSE, SAD and NCC). As for example, in the 3DLive and PSNR case, the average gain of 4% was computed as:

$$\frac{1}{3} \left(\frac{32.35 - 31.21}{31.21} + \frac{32.29 - 31.21}{31.21} + \frac{32.98 - 31.42}{31.42} \right) = 0.0402 \cong 0.04$$

(2) The joint use of search area adaptivity and visual quality based similarity metric in 3DV-NTSS leads on the 3DLive corpus to relative gains of: 6% in PSNR, 5% in IF, 14% in NCC, 0% in SC and 5% in SSIM. When considering the MPEG corpus, 3DV-NTSS and NTSS provide quite similar results in terms of PSNR and SC (average relative gains lower than 1%), while 3DV-NTSS outperforms NTSS with gains of 4% in IF, 2% in NCC, and 2% in SSIM. Note: these gains are computed as relative gains between the advanced 3DV-NTSS method based on the NCC block-matching criterion and the state-of-the-art reference given by the NTSS method based on the SAD block-matching criterion. As for example, in the 3DLive and PSNR case, the relative gain 6% was computed as:

$$\left(\frac{32.98 - 31.21}{31.21} \right) = 0.0567 \cong 6\%$$

3DV-NTSS vs. FS-MPEG

(1) The search area adaptivity involved in the 3DV-NTSS leads on the 3DLive corpus to average relative gains of: 4% in PSNR, 7% in IF, 19% in NCC, 1% in SC and 7% in SSIM. When considering the MPEG corpus, the values of gains are lower than in the previous case: 2% in PSNR, 5% in NCC, and 3% in SSIM; when considering IF and SC, the average relative gains between 3DV-NTSS and NTSS are lower than 1%. Note: these gains are computed as explained above.

(2) When considering the 3DLive corpus, the joint use of search area adaptivity and visual quality based similarity metric in 3DV-NTSS leads to relative gains of: 6% in PSNR, 8% in IF, 18% in NCC, 0% in SC and 8% in SSIM. On the MPEG corpus, these relative gains become: 4% in PSNR, 9% in IF, 8% in NCC, 2% in SC and 6% in SSIM. Note: these gains are computed as explained above.

The statistical relevance of the numerical results reported in Table 3.1 was investigated by computing the 95% error limits for each and every quality metric and for each and every investigated case (video sequence, disparity map computation algorithm and similarity metric). These error limits are presented in Table 3.2, which keeps the same structure as Table 3.1. When considering the NTSS algorithm applied on the *Rugby* sequence with an SAD block matching metric, an average PSNR value of 29.59 dB was obtained (see Table 3.1, first row and second column). This average value is the center of the 95% confidence interval (29.59 – 0.0115; 29.59 + 0.0115), cf. Table 3.2 (first row and second column).

By inspecting the values in Tables 3.1 and 3.2, it can be noticed that:

1) the sizes of the investigated video sequences and corpora were large enough so as to ensure the statistical relevance of the results, with a singular exception: the PSNR values computed on the MPEG corpus for the NTSS/SAD and 3DV-NTSS/NCC algorithms. In this case, the two corresponding 95% confidence intervals overlap: (32.56 – 0.0517; 32.56 + 0.0517) and (30.50 – 0.0323; 30.50 + 0.0323).

2) the 95% errors corresponding to the MPEG corpus are larger than the ones corresponding to the 3DLive corpus. As the sizes of the two corpora are quite equal, this difference brings to light a larger variation in the results corresponding to the MPEG corpus. The heterogeneity in both MPEG content and its representation (frame sizes, ...) may explain this situation.

Table 3.1 Reconstructed image visual quality, expressed by PSNR, IF, NCC, SC and SSIM. Each time (for each disparity map and each video sequences), three similarity measures have been considered for block matching: MSE, SAD and NCC.

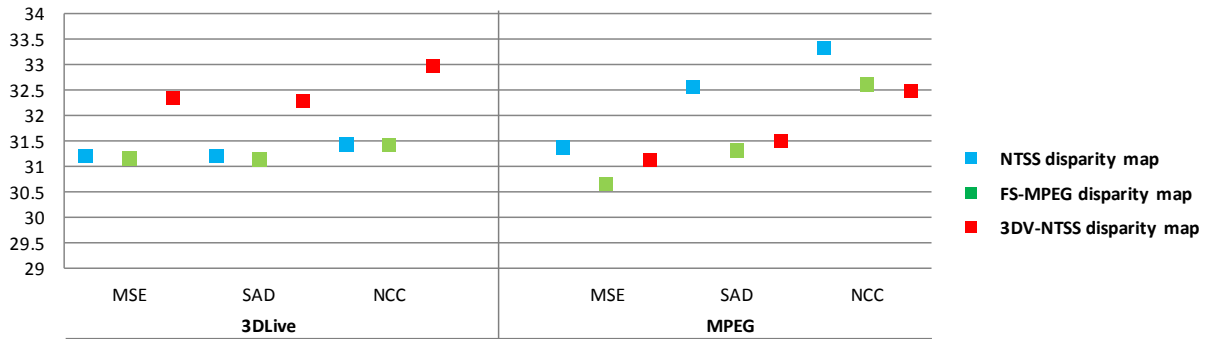
Algorithm	PSNR (dB)			IF			NCC			SC			SSIM		
	MSE	SAD	NCC	MSE	SAD	NCC	MSE	SAD	NCC	MSE	SAD	NCC	MSE	SAD	NCC
Rugby	NTSS			29.59	29.59	29.60	0.711	0.711	0.711	0.686	0.686	0.686	0.800	0.800	0.800
	FS-MPEG			29.58	29.58	29.59	0.710	0.711	0.711	0.682	0.683	0.683	0.799	0.798	0.798
	3DV-NTSS			30.71	30.61	31.20	0.738	0.732	0.752	0.789	0.789	0.828	0.815	0.822	0.828
Dancing	NTSS			32.83	32.83	33.25	0.853	0.853	0.857	0.796	0.797	0.797	0.879	0.880	0.879
	FS-MPEG			32.74	32.69	33.22	0.811	0.809	0.823	0.712	0.709	0.728	0.829	0.827	0.838
	3DV-NTSS			33.99	33.97	34.77	0.887	0.887	0.897	0.868	0.868	0.868	0.921	0.899	0.935
3DLive	NTSS			31.21	31.21	31.42	0.782	0.782	0.784	0.741	0.741	0.741	0.839	0.840	0.839
	FS-MPEG			31.16	31.13	31.40	0.760	0.760	0.767	0.697	0.696	0.705	0.814	0.812	0.818
	3DV-NTSS			32.35	32.29	32.98	0.812	0.809	0.824	0.828	0.828	0.848	0.868	0.860	0.881
MPEG	NTSS			31.36	32.56	33.32	0.719	0.684	0.743	0.872	0.889	0.887	0.902	0.917	0.938
	FS-MPEG			30.65	31.30	32.60	0.736	0.654	0.743	0.838	0.844	0.867	0.874	0.880	0.902
	3DV-NTSS			31.13	31.51	32.50	0.711	0.702	0.711	0.881	0.876	0.909	0.906	0.907	0.931

Table 3.2 95% error limits in reconstructed image quality assessment.
Each time (for each disparity map and each video sequences), three similarity measures have been considered for block matching: MSE, SAD and NCC. (In this table, a 0 error limit stands for values lower than E-05).

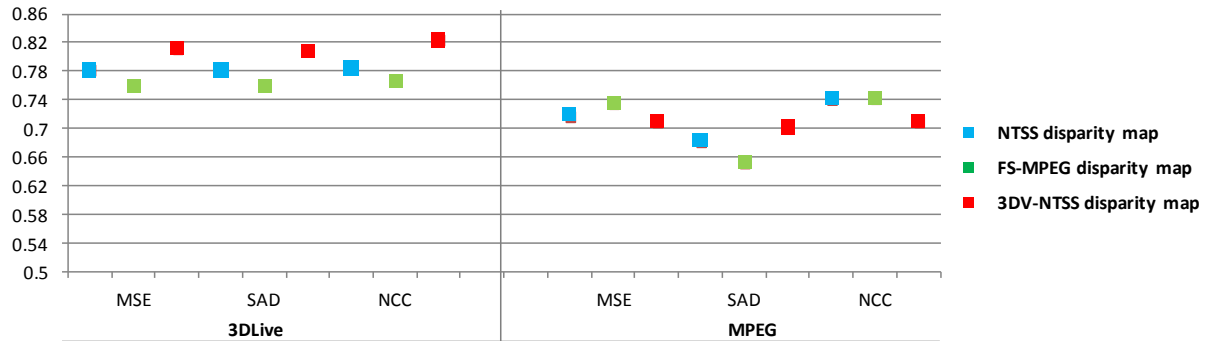
Algorithm	PSNR (dB)			IF			NCC			SC			SSIM		
	MSE	SAD	NCC	MSE	SAD	NCC	MSE	SAD	NCC	MSE	SAD	NCC	MSE	SAD	NCC
Rugby	NTSS	0.0108	0.0115	0.0140	0.0005	0.0008	0.0007	0.0008	0.0008	0.0002	0.0004	0.0004	0.0004	0.0005	0.0004
	FS-MPEG	0.0151	0.0163	0.0149	0.0007	0.0007	0.0007	0.0001	0.0008	0.0008	0.0003	0.0004	0.0005	0.0004	0.0005
	3DV-NTSS	0.0149	0.0155	0.0149	0.0007	0.0007	0.0007	0.0008	0.0008	0.0004	0.0004	0.0004	0.0005	0.0004	0.0005
Dancing	NTSS	0.0133	0.0129	0.0100	0.0003	0.0005	0.0005	0.0007	0.0007	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005
	FS-MPEG	0.0101	0.0132	0.0128	0.0007	0.0007	0.0007	0.0005	0.0004	0.0008	0.0004	0.0004	0.0005	0.0005	0.0005
	3DV-NTSS	0.0011	0.0139	0.0157	0.0005	0.0005	0.0005	0.0007	0.0007	0.0005	0.0003	0.0005	0.0005	0.0005	0.0005
3D Live	NTSS	0.0335	0.0329	0.0323	0	0	0	0.0011	0.0011	0.0009	0	0.0001	0.0001	0.0008	0.0012
	FS-MPEG	0.0324	0.0325	0.0149	0	0	0	0.0002	0.0008	0.0008	0	0.0001	0.0001	0.0009	0.0008
	3DV-NTSS	0.0109	0.0149	0.0127	0	0	0	0.0008	0.0008	0.0021	0	0	0.0008	0.0008	0.0011
MPEG	NTSS	0.0584	0.0517	0.0651	0.0019	0.0018	0.0019	0.0020	0.0018	0.0021	0	0	0.0015	0.0013	0.0016
	FS-MPEG	0.0355	0.0323	0.0323	0.0006	0.0014	0.0006	0.0020	0.0033	0.0012	0	0	0.0016	0.0024	0.0009
	3DV-NTSS	0.0319	0.0323	0.0323	0.0014	0.0014	0.0014	0.0031	0.0033	0.0033	0	0	0.0024	0.0024	0.0022

A synoptic view on the detailed information filled in Table 3.1 and 3.2 is provided by Figure 3.11 where the abscissa is decrementally divided at two levels in order to represent all the investigated cases. First, the two corpora (3DLive and MPEG) are figured out on the left and right sides, respectively. Secondly, for each corpus, the three similarity metrics (MSE, SAD and NCC) are presented from left to right. The ordinate gives the obtained average values of the five considered metrics (PSNR, IF, NCC, SC and SSIM) represented in squares and the 95% confidence limits in their estimation (represented in vertical lines centered on the related values).

As the corresponding error is always lower than 0.06, the related confidence limits are lower than the printing resolution for the average values and cannot be represented in Figure 3.11.

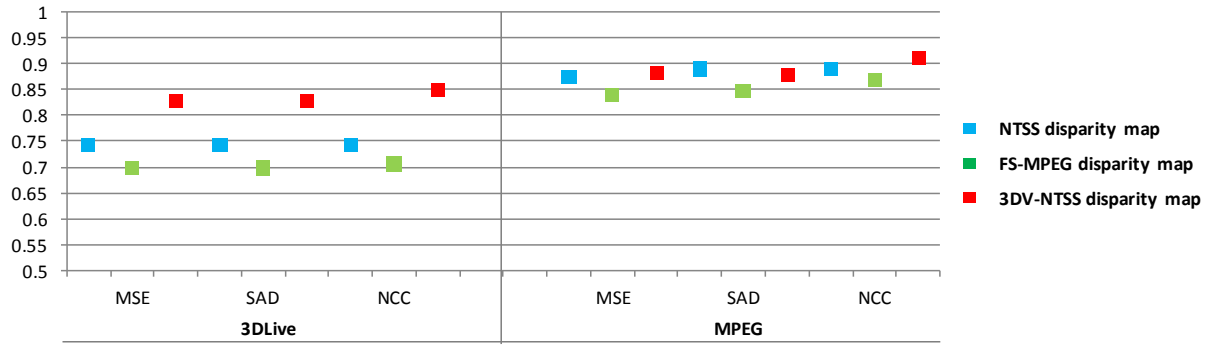


(a) PSNR average values.

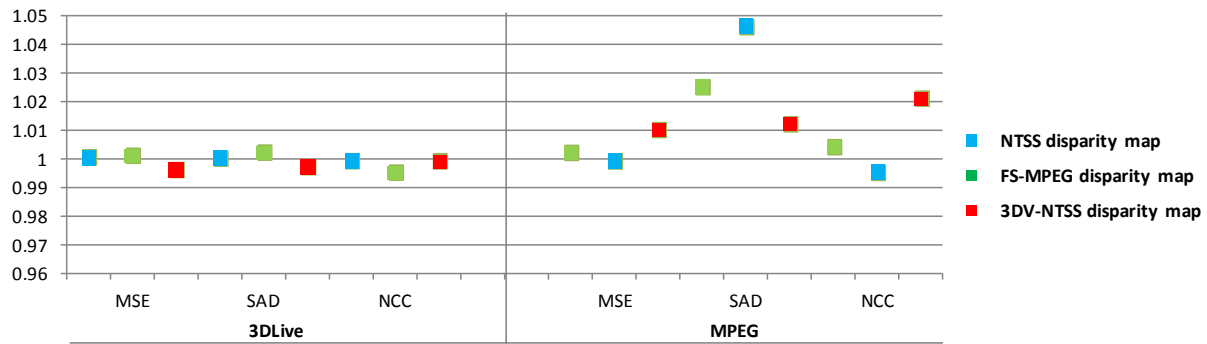


(b) IF average values.

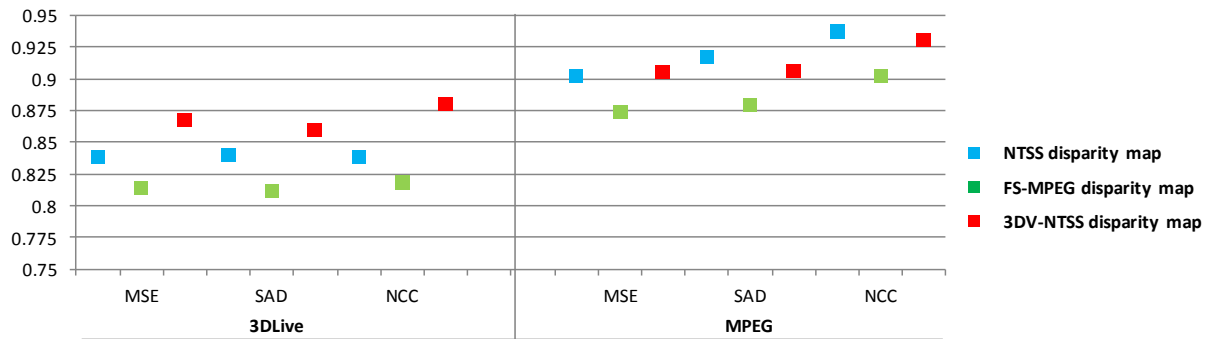
Figure 3.11 Objective evaluation of the reconstructed image quality.



(c) NCC average values.



(d) SC average values.



(e) SSIM average values.

Figure 3.11 (continued) Objective evaluation of the reconstructed image quality.

3.2.5. Computational cost

The computational cost of the considered algorithms was expressed in number of search points per 16×16 block and in the related gain factor, see Tables 3.3 and 3.4, respectively. The same three algorithms (NTSS, FS-MPEG and 3DV-NTSS) and the same three matching similarity measures (MSE, SAD and NCC) have been considered.

The values in Table 3.3 and 3.4 show that, when compared to the NTSS and FS-MPEG algorithm, the 3DV-NTSS features a computational cost decreased by a factor between 1.3 and 13, on both the 3DLive and MPEG corpora.

Table 3.3 Computational cost expressed in number of search points per 16x16 block.
Each time (for each disparity map and each video sequences), three similarity measures have been considered for block matching: MSE, SAD and NCC.

	Rugby			Dancing			3DLive			MPEG		
	NTSS	FS-MPEG	3DV-NTSS	NTSS	FS-MPEG	3DV-NTSS	NTSS	FS-MPEG	3DV-NTSS	NTSS	FS-MPEG	3DV-NTSS
MSE	22.08	222.48	16.81	24.96	220.38	16.82	23.52	221.43	16.81	23.16	216.09	16.80
SAD	23.16	222.48	16.81	25.84	220.38	16.82	24.5	221.43	16.81	22.48	216.09	16.80
NCC	21.82	222.48	16.81	21.66	220.38	16.82	21.74	221.43	16.81	21.60	216.09	16.80

Table 3.4 Gain factor on computational cost 3DV-NTSS vs. NTSS and 3DV-NTSS vs. FS-MPEG.

Each time (for each disparity map and each video sequences). Three similarity measures have been considered for block matching: MSE, SAD and NCC.

	Rugby		Dancing		3DLive		MPEG	
	3DV-NTSS vs. NTSS	3DV-NTSS vs. FS-MPEG	3DV-NTSS vs. NTSS	3DV-NTSS vs. FS-MPEG	3DV-NTSS vs. NTSS	3DV-NTSS vs. FS-MPEG	3DV-NTSS vs. NTSS	3DV-NTSS vs. FS-MPEG
MSE	1.31	13.23	1.37	13.10	1.34	13.10	1.48	12.86
SAD	1.37	13.23	1.30	13.10	1.33	13.10	1.53	12.86
NCC	1.29	13.23	1.29	13.10	1.29	13.10	1.28	12.86

3.3. Discussion

The present study advances 3DV-NTSS, a new method for disparity map computation for stereoscopic video. When applied on HD 3D TV content, it outperforms state-of-the-art algorithms (like NTSS or FS-MPEG) in terms of reconstructed image quality, computational cost and watermarking effectiveness. When considering stereoscopic video content encoded at lower quality (e.g. the MPEG stereoscopic video dataset), 3DV-NTSS features significant gains in computational cost and watermarking performances at the expense of a slight reduction in reconstructed image quality.

In order to obtain these results, two mechanisms were considered in the 3DV-NTSS design: the adaptation of the searching area according to the stereoscopic content itself and the use of a human visual system based metric in the matching decision. The former alleviates the problem of local minima reported in [Zhu00] where it is stated that: “Since the error surface is usually not monotonic, multiple local minimum points generally exist in the search window especially for those image sequences with large motion content. Therefore, searching with small search pattern, such as the one used in DS with size 3×3 , is quite likely to be trapped into local minimum for those video sequences with large motion content. On the other hand, a large search pattern with size 8×8 and sparse checking points as exploited in the first step of TSS is most likely to mislead the search path to wrong direction and hence misses the optimum point.” Our experiments confirmed this statement: from the visual quality point of view, 3DV-NTSS clearly outperforms NTSS when considering HD content, with sharp details and high motion content. The latter mechanism is devoted to a better identification of the blocks to be matched, in the sense of the human visual system.

The quantitative results presented in sections 3.2.3 and 3.3.4 also confirm the general view on the state-of-the-art presented in Figure 2.3 (Chapter 2), where NTSS and FS-MPEG were reported to feature an image quality expressed by MSE larger than 168.85, a computational cost of 21 and 255 search points per 16×16 block, respectively. Our experiments show that NTSS and FS-MPEG feature MSE values of 45.19 and 47.43 and computational cost of 23 and 221, respectively. Finally, note that at least for the HD 3D TV, 3DV-NTSS can be figured in Figure 2.3 at the same position as the “Targeted solution”, featuring an MSE of 32.73 and a computational cost of 16 search points.

3.4. Conclusion

With this Chapter, a new disparity map for HD 3D TV is advanced. By jointly exploiting the horizontal/vertical peculiarities of this kind of content and a visual quality metric in the block matching procedure, gains in both the reconstructed image quality and the computational cost are obtained with respect to the state-of-the-art algorithms like NTSS and FS-MPEG. This disparity map was validated for watermarking purposes under the framework of the 3DLive project. Automatically adapting the 3DV-NTSS parameters (T , block size ...) and cross-checking its effectiveness on test (synthetic) content are the main directions of our future work.

Chapter 4 A reliable watermarking technique: Fast-IProtect

Abstract

The watermarking state of the art exhibits the hybrid methods combining spread spectrum and side information principles as the most efficient approaches. The present study is focussed on speeding up such an algorithm (jointly patented by SFR – Vodafone Group and Institut Telecom), by deploying Monte Carlo generators accurately representing the watermarking attacks. A gains factor of 80 in computational speed is thus obtained. In this respect, two difficulties should be overcome. First, accurate statistical models for the watermarking attacks should be obtained. Secondly, efficient Monte Carlo simulators should be deployed for these models. The last part of the study is devoted to the experimental validations.

4.1. IProtect: hybrid embedding technique

Common state of the art watermarking methods are based either on spread spectrum or on side information principles, see Chapter 2.2.1 and Appendix A.2. The study in [Mit07a] advanced IProtect, to the best of our knowledge the first hybrid embedding techniques establishing synergies between SS and SI techniques.

This way, the three folded trade-off between the transparency, the robustness and the data payload constraints is reached for the 2D video content [Mit07a] [Mit06]. However, the main drawback of this technique is its heavy computational cost.

The present Chapter reconsiders IProtect, with a view to increasing its speed [Cha10].

4.1.1. IProtect embedding scheme

IProtect is a hybrid watermarking method [Mit05-07], combining the spread spectrum and side information concepts. It considers the watermarking procedure as an optimization problem, where the robustness is maximized under transparency and data payload constraints, see Figure 4.1.

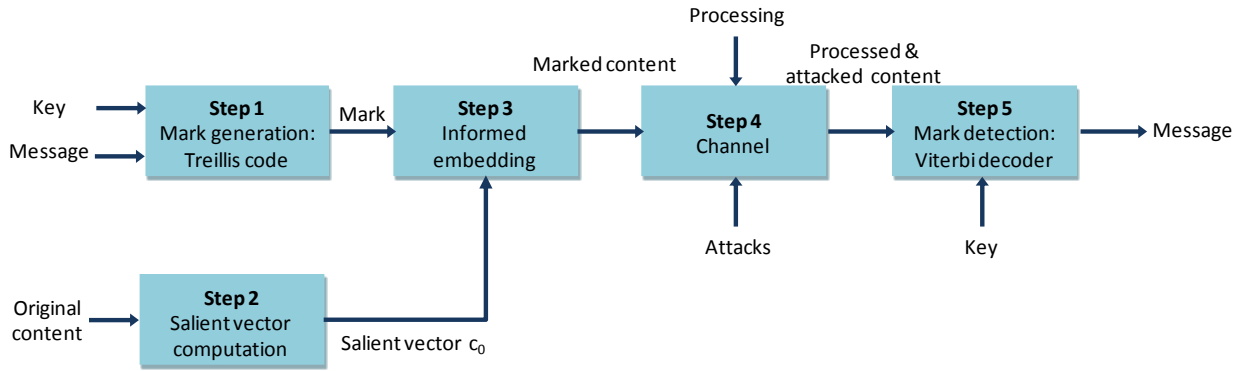


Figure 4.1 The watermarking method synopsis.

Step 1: Mark generation

Be there a message of M bits (the copyright information) and be there the secret encoding key. Starting from the message and the secret key, this step computes the watermark \tilde{m} to be embedded.

In order to make the mark \tilde{m} fit the original content during the embedding process, the M bits are encoded by means of a modified *treillis* code [Cox02] [Lin83]. The *treillis* codes are ones of the sophisticated error correcting codes, known to have good performance and serve as a foundation for some interesting research in watermarking [Cox02]. In the following we describe the encoding process for our example code.

The *treillis* has K states and 2 arcs exiting each state (each transition codes only one bit). Each arc is labelled with an N length vector whose components are real numbers unlike the basic *treillis* encoders where the label's components are bits. These labels are computed starting from the secret key, which means that only the true owner knows them.

Note: The output of a *treillis* encoder depends on the input bit and on the previous $\log_2 K$ ones. Each combination of $\log_2 K + 1$ adjacent bits from the message to be embedded is replaced by an N length label. Consequently, the mark is a vector denoted by g , with real components, having an $M \times N$ length.

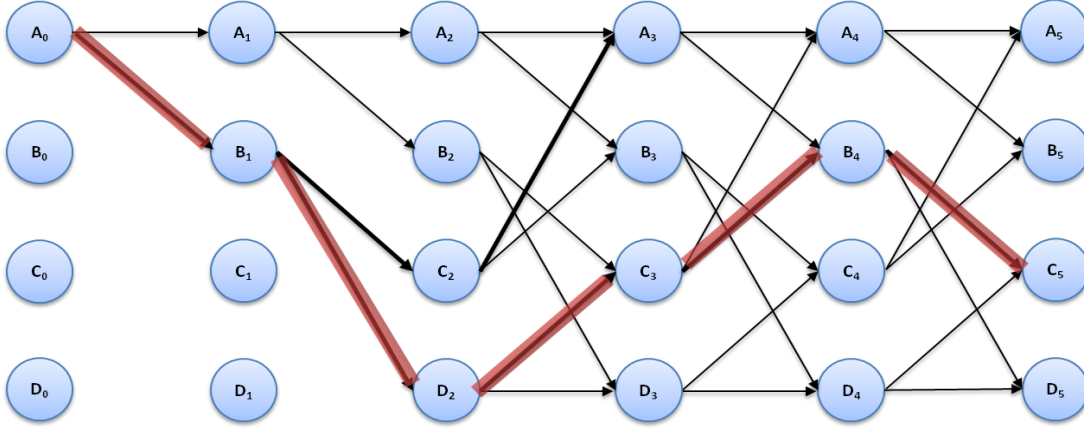


Figure 4.2 *Treillis* representation. Each row corresponds to one of the K states.
Each column corresponds to an iteration of the encoding.

Step 2: Salient characteristic vector representing the document

The aim of this block is to extract a vector denoted by c_0 which has the same $M \times N$ length as the mark \tilde{m} and which contains salient information representing the original content.

The watermark is inserted into the DWT – Discrete Wavelet Transform coefficients of the document. The wavelet decomposition proves its efficiency when protecting video content [Xia97] [Pot05] [Bar01b]. In practice, it is appropriate to identify the suitable coefficient for watermark insertion. For instance, the high frequencies should be avoided, since they are damaged in most forms of image processing and result in poor robustness. The lowest frequencies are very sensitive to modifications, thus resulting in poor transparency. Consequently, the mark is to be embedded in coefficients that have medium level of perceptual significance and reliability, like the *HL* and the *LH* sub-bands [Cox02].

The particular way in which this transform is applied and the salient coefficients are selected is described in as follow:

In order to obtain the c_0 salient vector the following steps should be performed:

1. The $(9, 7)$ $2D - DWT$ is individually applied to each frame in the video sequence, at an N_r resolution level. Figure 4.3.a gives an example of $2D - DWT$, third decomposition level $N_r = 3$.
2. The coefficients belonging to the HL_{N_r} and LH_{N_r} low frequency sub-bands (blue-shaded in Figure.4.3.b) are sorted in a decreasing order of their values. The largest D coefficients in each frame are (randomly) shuffled and then recorded into the c_0 vector.
3. The original locations of the c_0 vector components are stored into a v vector.

Let now the numerical values involved in the video watermarking be precised.

Note that, the $2D - DWT$ is applied at an $N_r = 6$ resolution level for 1920×1080 pixels images and $N_r = 4$ for 640×480 pixels image and $N_r = 3$ for images smaller than 320×192 pixels.

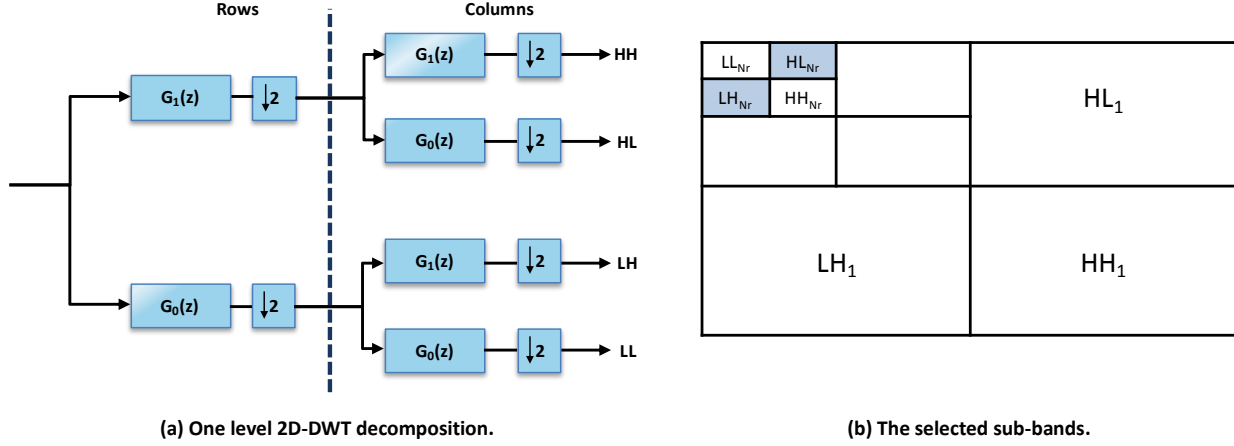


Figure 4.3 The selected sub-bands.

The original message to be inserted corresponds to the binary ARTEMIS logo, see Figure 4.4. Each bit from this message is *treillis* encoded by $N = 360$ real number labels. These numbers are extracted from a random generator obeying a Gaussian distribution of $\mu = 0$ mean and $\sigma = 0.005$ standard deviation.



Figure 4.4 The embedded ARTEMIS logo.

The D number of DWT coefficients selected from each frame is $D = M \times \frac{N}{L} = 360$.

The R_t parameter involved in the embedding scheme was set to $R_t = 2$.

The noise generator considers an n_g Gaussian noise of $\mu = 0$ mean, $\sigma = 0.2$ standard deviation and the result of a geometric random bending attack n_a .

Step 3: Informed embedding

This scheme is designed by adapting the principles in [Mil04] [Lin83]. Its aim is to embed the mark (the g vector) into the original content (represented by the c_0 vector). Under the informed watermarking framework, the crucial issue is to find a c_w vector which is as close as possible to the c_0 vector and for which the *Viterbi* decoder produces the same output as for the g vector.

The c_w vector is computed by an iterative algorithm, see Figure 4.5. In the first iteration, c_w is initialized with c_0 . Further on, a vector denoted by b is computed applying the *Viterbi* decoder to $c_w + n$, and by

treillis encoding the resulting bits. n is an $M \times N$ vector, whose components are sampled from a noise source modelling the channel perturbations. This noise is computed as a sum of a Gaussian noise considered in most watermarking application as a universal model for the channel noise and a noise that corresponds to the non-Gaussian effects [Mit06] [Mit07b] of some transformations or attacks (e.g. the geometric StirMark bending attack).

The c_w vector is then modified according to:

$$c_w = c_w + \alpha \cdot (g - b)/|g - b|$$

The scalar value α is computed as follow:

$$\alpha = R_t - R(g, b, c_w)$$

where $R(g, b, c_w) = c_w \cdot (g - b)/|g - b|$ and R_t is a scalar parameter. The dot product between the c_w and the $(g - b)$ vectors is the un-normalized correlation coefficient.

The loop of b computation and c_w modification is repeated until the condition $R(g, b, c_w) \geq R_t$ is reached several times successively (e.g. 200 times- $N_j = 200$). If the equality between the g and the b vectors is reached before the $R(g, b, c_w) \geq R_t$ condition is verified, then the b vector is computed without modifying c_w . If such a situation is encountered many times successively (e.g. 200 times- $N_j = 200$), then we consider that the g mark was successfully embedded into the c_w vector: regardless of the added noise, the decoder is able to recover the embedded message.

The computed c_w vector replaces the c_0 salient vector and the marked document is obtained by performing the inverses of the operations in the Step 2.

The modification of c_w can be seen as an attempt to remove the “bad” components of the host vector, those that would yield a different message, replacing them with “good” components, leading to the correct decoding. The operation is targeting both the original host components and the noise components. Each iteration leads to a relationship $c_w \cdot g = (c_w \cdot b) + R_t$, with R_t being seen as a “safety distance” taking into account untested noise configurations.

In the final step, an independent attack is performed on the watermarked vector. If the error rate after this attack is higher than an acceptable threshold (e.g. 20%) the watermarking procedure is restarted with a new $R_t = R_t + R_\Delta$. It should be noted that in practice this re-watermarking is rarely necessary.

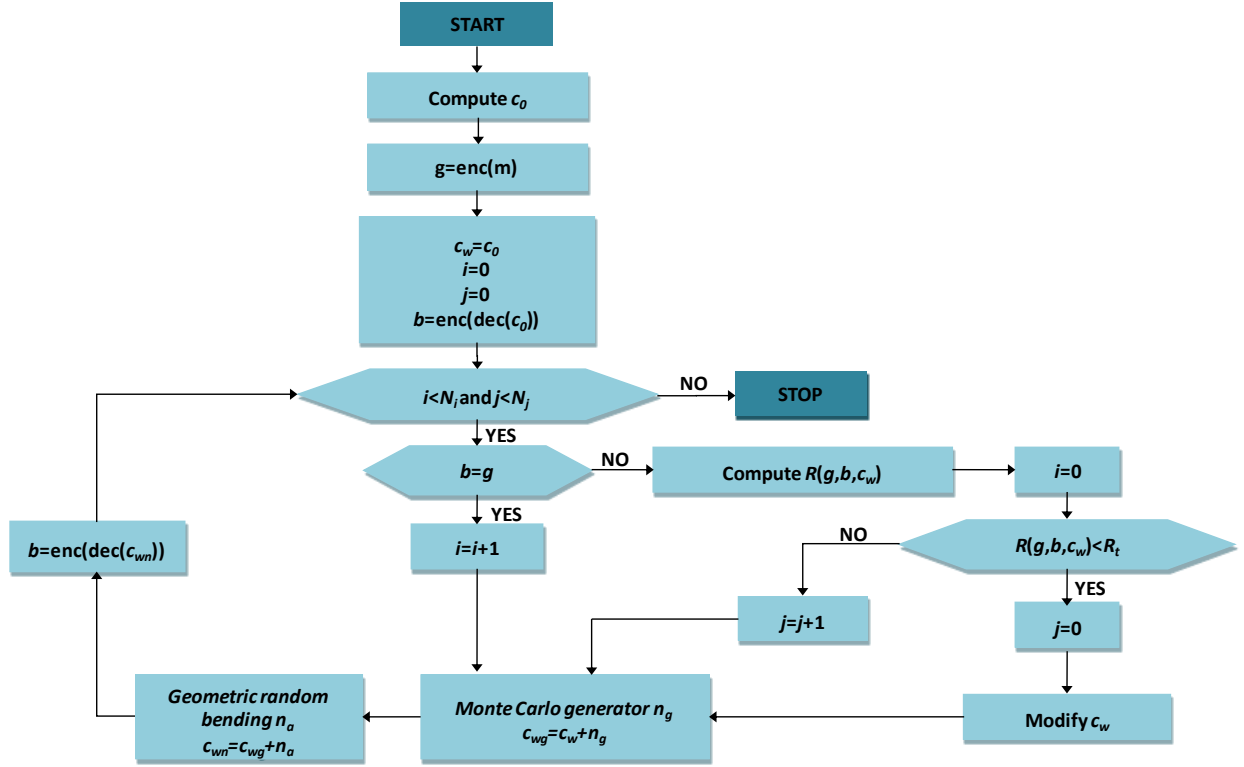


Figure 4.5 The IProtect embedding algorithm synopsis.

The enc and dec functions denote the *treillis* encoder and the *Viterbi* decoder, respectively. The n_g and n_a terms represent the Gaussian and non Gaussian noise components, respectively, while m denotes the inserted (public) message.

Step 4: The channel

The marked document withstands a large variety of transformations divided in two categories: generic and malicious.

Generic transformations include format or representation changes, compression, document editing and changing. The malicious attacks are dependent on the media type on which they are performed. Generally, in watermarking, all these transformations are implicitly assumed to be Gaussian distributed. However, recent studies on multimedia data statistical behaviour brought into evidence that this Gaussian assumption does not hold for challenging attacks, like the video geometric random bending attack, for instance. Consequently, in our watermarking scheme we consider two types of perturbations: (1) Gaussian noise (denoted by n_g in Figure 4.5), which can model the generic transformations, and (2) non Gaussian noise (denoted by n_a in Figure 4.5), which represents the malicious transformations. In the practical implementation, the former is sampled with random number generator software while the latter is the effect of directly applying a geometric random bending attack [Pet98] [Pet00].

4.1.2. IProtect detection

Be there a video sequence that is supposed to be marked. The aim of this step is to establish whether the M bits message has been embedded into the considered video sequence or not.

The first task is to extract from that sequence the salient vector susceptible to convey the mark, see Step 2 above. Then the coefficients corresponding to the locations where the mark might have been inserted are recorded, thus obtaining a c_w vector with $M \times N$ real components.

This vector is the input of a *Viterbi* decoder [Lin83]. The decoder is pair designed with the *treillis* encoder and gives the most likely path through the *treillis* that leads to the highest correlation between the recovered message vector and the initially encoded message vector g . The algorithm relies on the fact that the most likely path through each node in the *treillis* always contains the most likely path up to that node. Thus, once the most likely path from A_0 to some further node is found, we can forget about the other possible paths A_i . In each iteration, the paths and its cost are updated. The cost involved in the *Viterbi* algorithm is the (un-normalised) correlation coefficient between the input sequence and the transition labels. This cost is to be compared to a detection threshold τ to decide whether the watermark is present.

4.2. Towards Fast-IProtect

When evaluating the IProtect performances, the experimental results showed that the method features good transparency with a PSNR > 35 dB and a resistance against a large range of attacks, such as linear and non-linear filtering, noise addition, small rotations and especially the StirMark random bending attack [Mit05] [Mit06]. The study in [Cha10] also investigated the speed of the embedding method. The analysis of the average processing time required by the different parts of the watermarking chain (pre-processing, embedding, post-processing, detection), shows that the insertion step accounts for 90% (Figure 4.6), from the total. When investigating the insertion, it can be noticed that more than 99% of time is spent on the non-Gaussian attacks.

This is a consequence of the fact that applying the real attacks on the c_w vector obtained in an iteration i of the Step 3, requires several time-consuming operations:

- apply the 2D – DWT inverse at an N_r resolution level at the c_w vector and store the intermediate sequence S ;
- apply the attack to the sequence S thus obtaining the attacked sequence S_a ;
- apply the 2D – DWT to the luminance component of S_a at the same initially considered N_r resolution level;
- record in a vector the 2D – DWT coefficients corresponding to the HL_{N_r} and LH_{N_r} lowest frequency sub-band;
- build up the coefficient hierarchy by sorting in a decreasing order the vector obtained in the previous step, and record the largest D ranks in a vector denoted c_{w+a} .

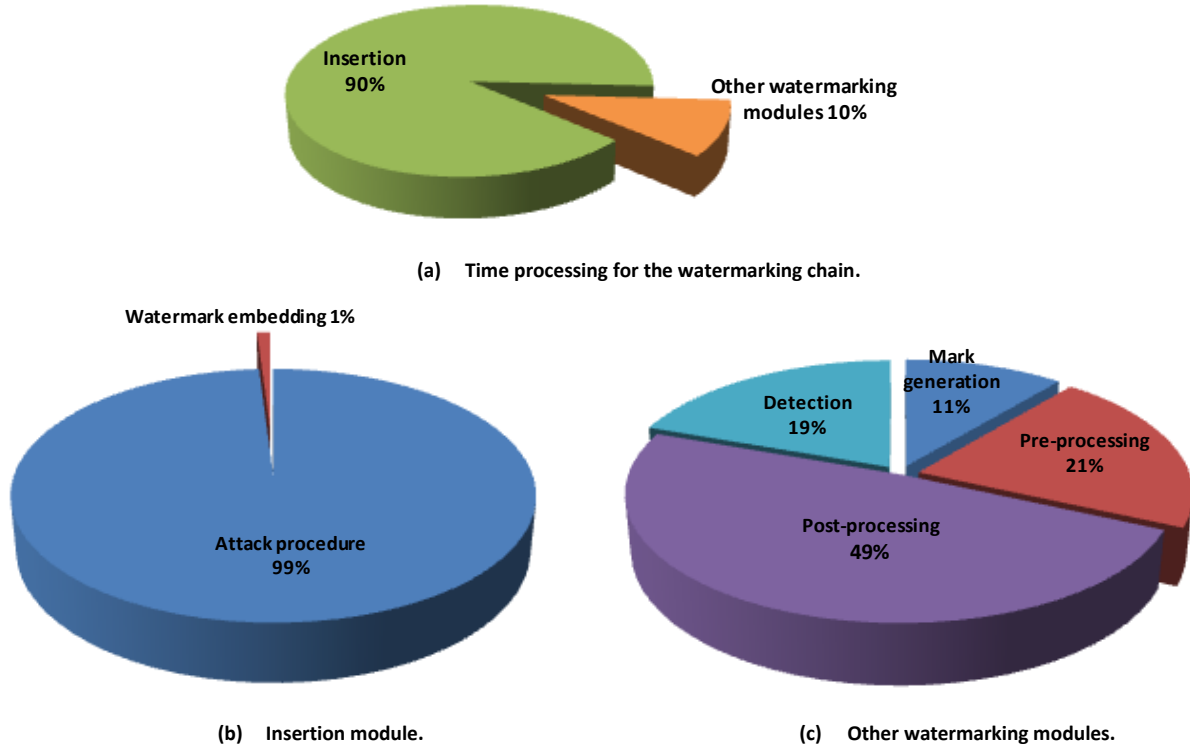


Figure 4.6 IPProtect time processing.

Time consumption rates for different watermarking operations in the original hybrid watermarking method: IPProtect.

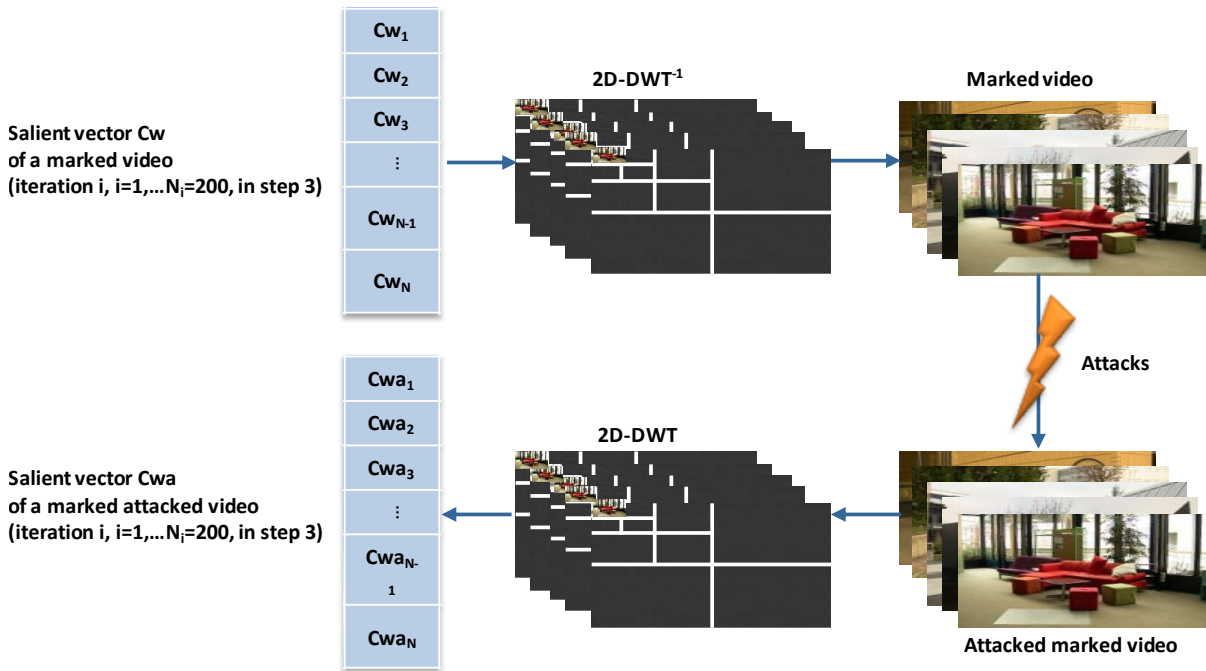


Figure 4.7 The five steps undergone when applying an attack procedure.

These five tasks are intrinsically avoided, assuming the effects of the attacks can be accurately modelled by some Monte Carlo generators.

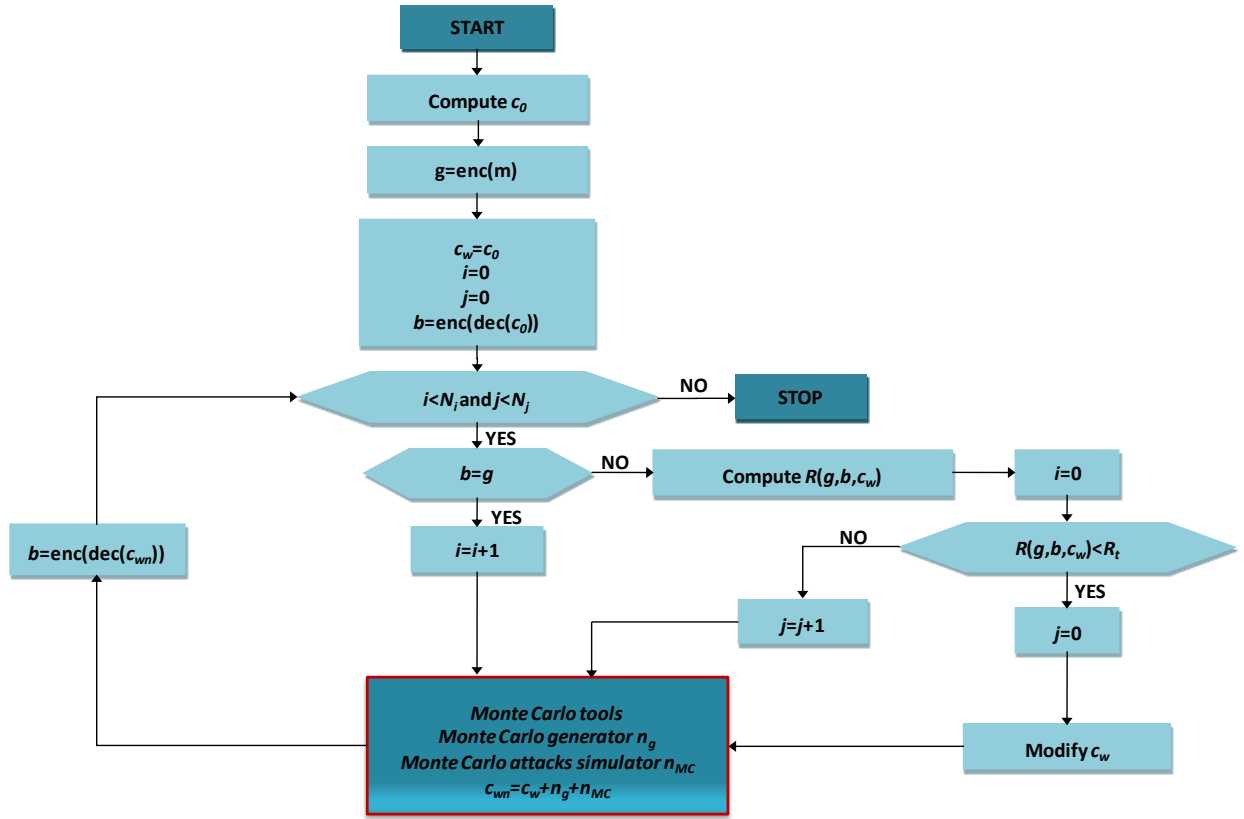


Figure 4.8 Fast-IPProtect embedding algorithm synopsis.

The n_g and n_{MC} terms represent the Gaussian noise components generated by the Monte Carlo generator and non Gaussian noise resulting from the Monte Carlo attack simulation, respectively.

4.2.1. Statistical models for watermarking attacks

Generally, in the literature, the video watermarking attacks are by default assumed to be Gaussian distributed, although no firm support is available. A study [Dum08] carried out in the ARTEMIS department on this hypothesis has brought into evidence that for most of the attacks the Gaussian behaviour has been refuted. However there are some cases when it can be accepted but just as a limit approximation and not as a fine model. This in-depth study has been made for eight types of attacks: filtering (median, Gaussian, Frequency Model Laplacian Removal (FMLR) [Pea98], sharpening), rotations, JPEG compression, row & column removal, and StirMark geometric random bending attack.

The study also estimates the probability density function (*pdf*) for attacks. In order to obtain these models, the original and generic statistical approach for *pdf* estimation based on Gaussian mixtures has been considered.

Finite Gaussian mixture can approximate any continuous *pdf*, provided the model has a sufficient number of components and provided the parameters of the model are chosen correctly [Bis95]. The true *pdf* is approximated by a linear combination of K component densities:

$$p(x) = \sum_{k=1}^K P(k) \frac{1}{\sqrt{2\pi}\sigma(k)} \exp\left(-\frac{(x - \mu(k))^2}{2\sigma(k)^2}\right),$$

In this equation $p(x|k)$ is the probability of x given the component distribution k and $P(k)$ are the mixture proportions or weights. The weights are non-negative and sum up to one. A popular technique for approximating iteratively the maximum likelihood estimates of the model parameters $P(k)$, $\mu(k)$ and $\sigma(k)$ is the *Expectation-Maximization* (EM) algorithm [Dem77]. The likelihood function is given by:

$$L = \prod_{n=1}^N p(x_n),$$

Maximizing the likelihood function is equivalent to finding the most probable *pdf* estimate provided the data set. The *EM* operates in two stages. First, in the *E – step*, the expected value of some unobserved data are the data labels of the samples. They correspond to the identification number of the different mixture components and specify which one generated each datum. Subsequently, during the *M – step*, they are used to update the model parameters accordingly. Each iteration step can be summarized as follows [Pee00]:

- a) E-step, $p^{(i)}(k/x_n) = \frac{p^{(i)}(x_n/k)P^{(i)}(k)}{p^{(i)}(x_n)}$
- b) M-step, $\mu(k)^{(i+1)} = \frac{\sum_{n=1}^N p^{(i)}(k/x_n)x_n}{\sum_{n=1}^N p^{(i)}(k/x_n)}$
- c) $(\sigma(k)^2_k)^{(i+1)} = \frac{\sum_{n=1}^N p^{(i)}(k/x_n)(x_n - \mu(k)^{(i+1)})^2}{\sum_{n=1}^N p^{(i)}(k/x_n)}$
- d) $P^{(i+1)}(k) = \frac{1}{N} \sum_{n=1}^N p^{(i)}(k/x_n)$

Table 4.1 gives an exemple of statistical models for some watermarking attacks (i.e. Gaussian filtering, sharpening and Stirmark random bending) in the (9,7) DWT hierarchy.

Table 4.1 Attacks statistical models in the (9,7) DWT hierarchy.

The models are given for the Gaussian filtering, sharpening and StirMark random bending watermarking attacks.

The model parameters $P(k)$, $\mu(k)$ and $\sigma(k)$ are the *pdf* weights, the mean and the standard deviation, respectively.

Attacks	Rank	Model parameters											Error
Gaussian filtering	1	$P(k)$	0.015	0.199	0.199	0.076	0.027	0.046	0.078	0.338	0.060	0.042	0.036
		$\mu(k)$	-0.021	-0.061	-0.067	-0.082	-0.153	-0.082	-0.078	-0.046	-0.115	-0.082	
		$\sigma(k)$	0.001	0.015	0.015	0.020	0.023	0.019	0.019	0.012	0.003	0.020	
	150	$P(k)$	0.016	0.199	0.119	0.075	0.027	0.046	0.078	0.338	0.060	0.042	0.017
		$\mu(k)$	-0.021	-0.061	-0.067	-0.082	-0.153	-0.081	-0.078	-0.046	-0.115	-0.082	
		$\sigma(k)$	0.001	0.015	0.015	0.020	0.023	0.020	0.019	0.012	0.003	0.020	
	300	$P(k)$	-0.023	-0.058	-0.153	-0.044	-0.385	-0.117	-0.119	0.078	0.057	0.071	0.009
		$\mu(k)$	0.029	0.002	-0.002	-0.001	-0.001	-0.003	-0.059	-0.002	-0.001	-0.002	
		$\sigma(k)$	0.015	0.009	0.003	0.009	0.002	0.008	0.011	0.003	0.009	0.008	
Sharpening	1	$P(k)$	0.022	0.024	0.107	0.076	0.194	0.014	0.110	0.087	0.100	0.274	0.036
		$\mu(k)$	0.101	-0.029	0.078	0.071	0.040	0.189	0.148	0.158	0.142	0.059	
		$\sigma(k)$	0.001	0.001	0.104	0.104	0.056	0.002	0.092	0.090	0.015	0.052	
	150	$P(k)$	0.047	0.133	0.117	0.041	0.262	0.061	0.062	0.101	0.143	0.015	0.012
		$\mu(k)$	0.044	0.031	0.032	0.045	0.004	0.034	0.042	0.009	-0.034	0.045	
		$\sigma(k)$	0.051	0.006	0.025	0.051	0.008	0.050	0.051	0.039	0.034	0.051	
	300	$P(k)$	0.415	0.006	0.151	0.114	0.019	0.053	0.014	0.073	0.016	0.136	0.010
		$\mu(k)$	0.001	0.035	-0.016	0.029	-0.041	-0.011	0.035	-0.079	0.044	0.021	
		$\sigma(k)$	0.004	0.072	0.010	0.021	0.044	0.001	0.071	0.041	0.070	0.009	
StirMark random bending	1	$P(k)$	0.045	0.097	0.120	0.016	0.097	0.102	0.092	0.122	0.185	0.124	0.015
		$\mu(k)$	-0.039	-0.277	-0.155	-0.586	-0.012	-0.101	-0.019	-0.047	0.005	-0.065	
		$\sigma(k)$	0.122	0.146	0.084	0.253	0.013	0.140	0.123	0.062	0.047	0.060	
	150	$P(k)$	0.215	0.081	0.022	0.043	0.287	0.046	0.102	0.138	0.038	0.028	0.008
		$\mu(k)$	0.001	0.003	-0.092	-0.032	-0.017	-0.002	-0.003	0.003	-0.020	-0.055	
		$\sigma(k)$	0.024	0.068	0.092	0.053	0.020	0.067	0.067	0.002	0.060	0.090	
	300	$P(k)$	0.014	0.372	0.053	0.023	0.031	0.007	0.290	0.181	0.001	0.027	0.008
		$\mu(k)$	0.200	0.006	0.056	-0.060	-0.040	-0.021	-0.002	-0.011	-0.044	-0.052	
		$\sigma(k)$	0.026	0.017	0.017	0.029	0.036	0.001	0.001	0.014	0.035	0.032	

4.2.2. Monte Carlo simulation for watermarking attacks

The simulation of a random variable with known probability law is generally achieved by inverting the corresponding cumulative distribution function (*cdf*). This inversion method is based upon the following theorem:

Be there $f: R \rightarrow R$ a one-to-one mapping and be X a random variable defined on a given probability field; then, the $Y = f(X)$ random variable will be defined on the same probability field and will be characterised by the following *pdf*:

$$p_Y(y) = \frac{p_X(x)}{|f'(x)|} \Big|_{x=f^{-1}(y)}$$

When considering the particular case of $f(x) = F_X(x)$, *i.e.* when the transform function is the very *cdf* of X , Y becomes uniform distributed.

From the practical point of view, this means that arbitrary random variables of $p_X(x)$ *pdf* can be simulated by applying to a uniform random variable a transform described by $F_X^{-1}(x)$:

$$X = F_X^{-1}(Y).$$

Although very simple from the conceptual point of view, this method cannot be directly deployed in the watermarking case, where the attack statistical models are available as finite Gaussian mixtures. On the one hand, in such a case, the *cdf* cannot be analytically computed. On the other hand, the *cdf* numerical computation should be properly handled, as it requires integral evaluation and function inversions. Consequently, the Monte Carlo generator developed in our study follows a different approach: It is based on the very principle of the finite Gaussian mixture: which combines two random phenomena: the choice of a Gaussian law among the K composing the mixture then the selection of one value for that law.

From the applicative point of view, to apply the attacks effects on the c_w vector obtained in the i^{th} iteration of step 3 of Chapter 4.1.1, the following steps are to be performed:

Step1 Parameters initialization

Step1.1 Global parameters

The attacked and watermarked c_{wn} vector size $N = 360$, a new vector n_{MC} representing the attack component is created.

Step1.2 Component index initialization

Component index $C_index = 1$

Gaussian pdf index $k=0$

Step2 Watermarked and attacked vector c_{wn} computation

WHILE $C_index \leq N$ DO:

Step2.1 Attack components computation

Step2.1.1 pdf selection

Generate a uniform random number denoted by α , where $\alpha \in [0,1]$

IF $\alpha \leq P(1)$ THEN $k = 1$

ELSE

FOR ($i = 2, i \leq K, i++$)

IF $\sum_{j=1}^{i-1} P(j) < \alpha \leq \sum_{j=1}^i P(j)$ THEN $k = i$.

Step2.1.2 Gaussian Mixture sample estimation

Generate two uniform random numbers x_1 and x_2 , where x_1 and $x_2 \in [0,1]$

Be $\mu(k)$ and $\sigma(k)$ the selected pdf's parameters. The C_index^{th} component $n_{MC}(C_index)$ is:

$$p(C_index) = \sigma(k)\sqrt{-2\ln x_1} \cos(2\pi x_2) + \mu(k)$$

$$n_{MC}(C_index) = p(C_index)$$

Step2.2 c_{wn} computation

$$c_{wn}(C_index) = c_w(C_index) + n_{MC}(C_index)$$

$$C_index = C_index + 1$$

Step3 Exit condition

IF $C_index = N + 1$ THEN STOP

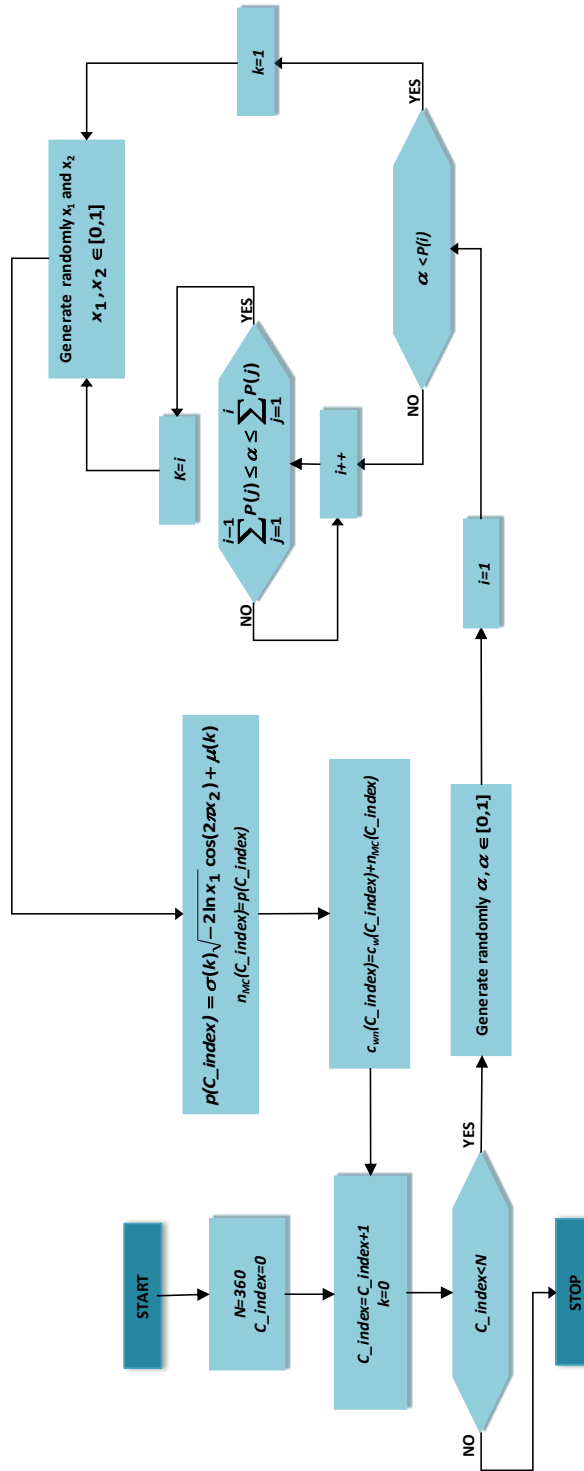


Figure 4.9 The Monte Carlo attack simulation algorithm flowchart.

4.2.3. Experimental validation

The first experiment considers the quality of the Monte Carlo generators implemented according to the methodology in Chapter 4.1.5. Figure 4.10 brings into light a good visual concordance between the Gaussian mixture estimations corresponding to the StirMark random bending attack (in solid line) and the histograms computed on 2000 data extracted from the corresponding Monte Carlo generators. Three coefficient hierarchy ranks that have been investigated are $r = 1$, $r = 150$, and $r = 300$. The upper row corresponds to the effects of the attacks applied on high quality video watermarking, while the lower row corresponds to attacks applied on low quality video watermarking.

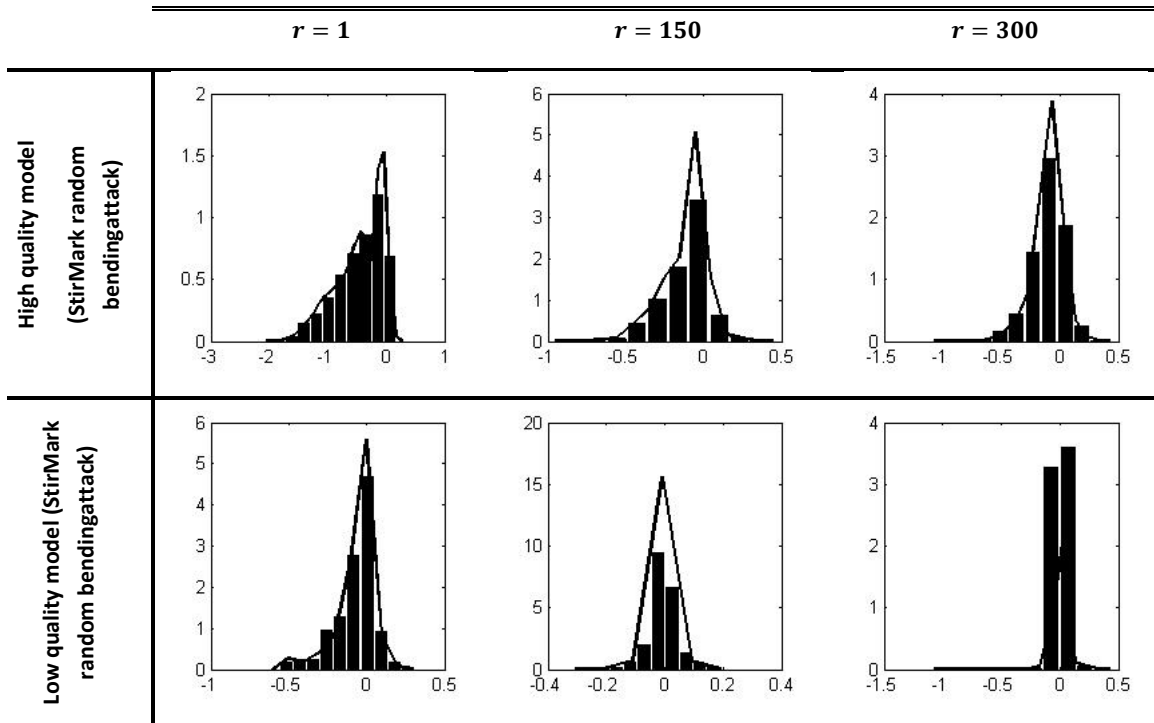


Figure 4.10 The quality of the Monte Carlo generators for the StirMark attack.

The theoretical models (continuous line) vs. normalised histograms computed on the data sampled from the generators. Three DWT coefficient hierarchy ranks have been considered $r = 1$, $r = 150$, and $r = 300$.

The statistical quality of the generator has also been checked by applying the Chi-square goodness of fit tests. For each rank $r = 1 \dots 360$, such tests are run at an $\alpha = 0.05$ significance level and consider data set of $N = 750$ values.

The last evaluation is devoted to the processing time. The results presented in Figures 4.6 and 4.12 are obtained on a PC with a Centrino processor and 1 GB RAM. Speed performance evaluations show first a gain in processing time by a factor of 80 compared to the basic hybrid method. Indeed, after introducing the Monte Carlo attack simulation module (Figure 4.11 right) the time devoted for embedding module represents only 13% of the total time required by the chain of watermarking, which is very close to the detection module (11%) and seven times less than the pre- and post-processing modules (accounting for 75%).

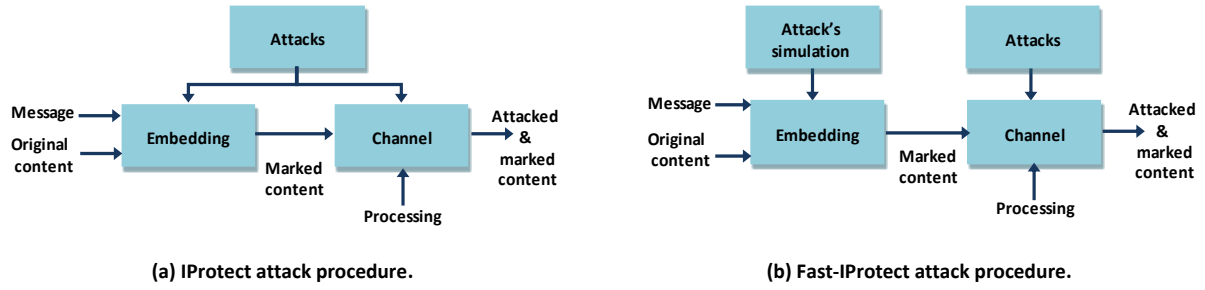


Figure 4.11 The attack procedure during the embedding process.
(Before and after the Monte Carlo generator).

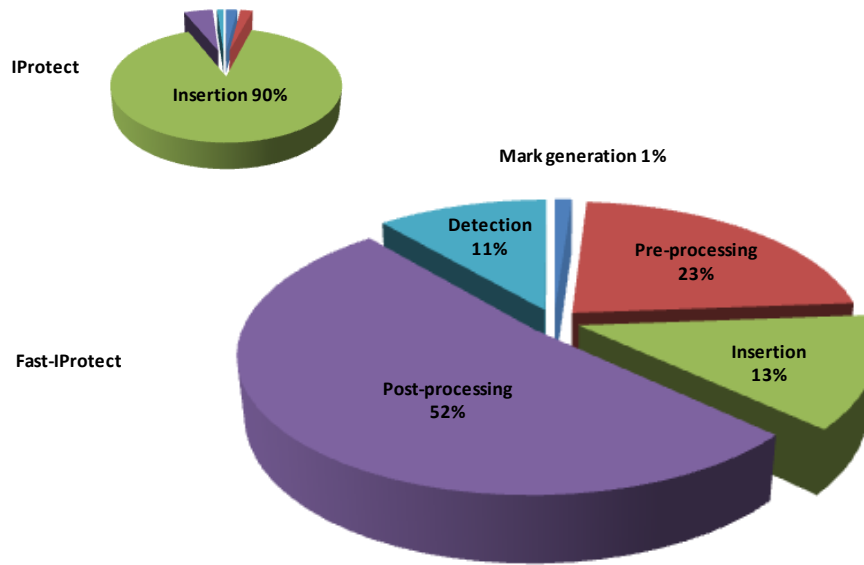


Figure 4.12 Time processing for different watermarking operations.
The new hybrid watermarking method: Fast-IProtect vs. IProtect.

4.3. Conclusion

This Chapter reconsiders IProtect, a hybrid SS-SI watermarking methods, with the aim of accelerating its insertion step. In this respect an algorithm for simulating finite Gaussian mixture is advanced and integrated in the Fast-IProtect algorithm. Experiments carried out on 2h29min of video content show a gain factor of 80 in the algorithm speed.

Chapter 5 Stereoscopic watermarking benchmark

Abstract

In this Chapter, a comparative study on the main classes of 2D inherited watermarking methods and on their related optimal stereoscopic insertion domains is carried out. Four insertion methods are considered; they belong to the SS, SI (binary QIM and 5-symbols QIM) and hybrid (Fast-IProtect) families. First, this Chapter describes the experimental protocol. Second, it presents the obtained results. Thirdly, it discusses the benchmarked methods performances. It was thus demonstrated that Fast-IProtect applied on the 3DV-NTSS is the best solution because it is the only insertion method ensuring: (1) the imperceptibility of the watermark according to subjective tests and objective metrics with limits $PSNR > 35dB$ and IF, NCC, SC and SSIM larger than 0.95; (2) robustness expressed by BER lower than 0.05 after filtering and JPEG compression and lower than 0.1 after the geometric random bending attacks. Finally, Fast-IProtect features a non-prohibitive computational complexity, compatible to real time applications.

A general view on the watermarking assessment procedure carried out in our study is presented in Figure 5.1.

Four insertion methods are considered; they belong to the SS, SI (binary QIM and 5-symbols QIM) and hybrid (Fast-IProtect) families, see Chapter 5.1.

Each of these four methods is successively applied on the left view of the video sequences as well as on three disparity maps, computed according to the NTSS, FS-MPEG and 3DV-NTSS algorithms (all of them considering the NCC as block matching criterion). In our study, the mark insertion actually takes place in the DWT representation of each of these four insertion domains; however, for simplicity, these domains will be further referred to as left view, NTSS, FS-MPEG and 3DV-NTSS, respectively, see Chapter 5.2.

The watermarking properties are evaluated in terms of transparency (both subjective and objective procedures), robustness and computational cost. The transparency is assessed by both subjective and objective procedures (see Chapter 5.3). The former relies on the ITU-R BT 710 – 4 [BT98], 500 – 12 [BT02] and BT 1438 [BT00] recommendations and concerns the image quality, the depth perception and the visual comfort. The latter is performed based on five objective image quality metrics, namely PSNR, IF, NCC, SC and SSIM (see Chapter 5.4). The robustness is assessed by computing the BER in the watermark detection after five types of attacks, namely the Gaussian filtering, sharpening, JPEG compression, and geometric (small rotations and StirMark random bending). These attacks are selected so as to represent the main classes of attacks mentioned by the DCI standards. The computational cost is not only expressed by the computational time needed to insert the mark but also by an analysis of the computation complexity (see Chapter 5.5). The quantity of inserted information is kept unchanged in all the cases, namely 1 bit per frame (i.e. 25 bits per second); note that this value is 200 times larger than the lower limit imposed by the DCI standards (35 bits per 5 min of video).

All the experiments considered both the 3DLive and the MPEG stereoscopic video corpora (cf. Appendix A.1).

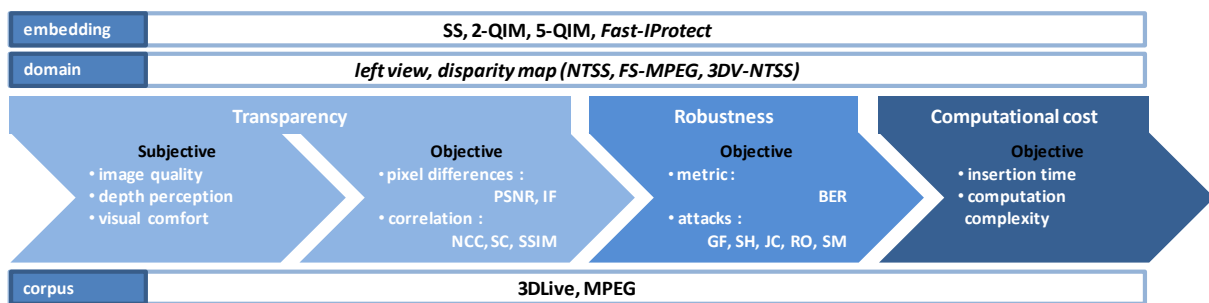


Figure 5.1 The watermarking assessment procedure.

5.1. Embedding procedure

The embedding procedure in our study inserts the watermark in the 2D-DWT coefficients of the stereoscopic content. The insertion is performed according to the four pre-mentioned insertion methods: SS, binary QIM (referred to 2-QIM), 5-symbols QIM (referred to 5-QIM) and Fast-IProtect, see Appendix A.2 and Chapter 4. All these insertion methods have the same input, the c_0 salient vector extracted from the original video sequence to be watermarked.

The c_0 salient vector computation algorithm is given by:

Step 1 Parameter initialization

The DWT resolution level N_r , Number of frames T , the size of the salient vector vf extracted from each frame $Size_vf = 360$, $frame_index = 1$.

Step 2 Video salient vector computation

WHILE $frame_index < T$ Do

Step 2.1 Coefficients extraction

Apply the (9, 7) 2D – DWT at the N_r resolution level.

Extract the coefficients belonging to the HL_{N_r} and LH_{N_r} low frequency sub-bands and store them in a vector vf_0

Sort in a decreasing order the vf_0 components.

Step 2.2 Coefficients storage

The largest $Size_vf$ coefficients from vf_0 are (randomly) shuffled and then recorded into the vf vector.

The original locations of the c_0 vector components are stored into a v vector.

The considered insertion techniques have different parameters setting, whose values are synoptically displayed in Table 5.1.

Table 5.1 The considered insertion methods and their main parameters.

Insertion technique	Main Parameters		Fixed values
SS	α_w	Watermark power	$\alpha_w = 0.5$
2-QIM	m	The alphabet size	$m = 2$
	D	The alphabet	$D = \{0, 1\}$
	Δ	A fixed quantization step size (depends on the coefficients range)	$\Delta = 50$
	α	A fixed parameter where $0 < \alpha \leq 1$, $\alpha \geq \frac{m-1}{m}$	$\alpha = 0.6$
	k	A random key	k is randomly selected where $0 < k \leq 1$
5-QIM	m	The alphabet size	$m = 5$
	D	The alphabet $D = \{-(m-1)/2, -(m-2)/2, \dots, 0, \dots, (m-2)/2, (m-1)/2\}$	$D = \{-2, -1, 0, 1, 2\}$
	Δ	A fixed quantization step size (depends on the coefficients range)	$\Delta = 50$
	α	A fixed parameter where $0 < \alpha \leq 1$, $\alpha \geq \frac{m-1}{m}$	$\alpha = 0.83$
	k	A random key	k is randomly selected where $0 < k \leq 1$
Fast-IProtect	σ_g	Gaussian noise standard deviation	$\sigma_g = 0.2$
	Param[0]	Viterbi's parameters: the size of input character in bits	Param[0] = 1
	Param[1]	Viterbi's parameters: the size of register (number of bits in each shift register)	Param[1] = 3
	Lcode	Code word length	Lcode = 360

Table 5.1 (continued) The considered insertion methods and their main parameters.

Insertion technique	Main Parameters		Fixed values
Fast-IProtect	α_w	Watermark power	$\alpha_w = 0.5$
	Ncod_mesaj	Number of bits per frame	Ncod_mesaj=1
	R_t	Robust_target	$R_t = 2$
	R_Δ	Delta_robust	$R_\Delta = 0.2$
	τ	Detection threshold ($\frac{1}{\sqrt{M}}$ where M is the embedded message size)	$\tau = 0.527$
	N_j	Counter for searching through reference marks	$N_j = 200$

5.2. Embedding domains

Each of these four methods is successively applied on the left view of the video sequences as well as on three disparity maps, computed according to the NTSS, FS-MPEG and 3DV-NTSS algorithms (all of them considering the NCC as block matching criterion). The watermarking schemes for each type of embedding domain (i.e. view-based and disparity based) are described in Figure 5.2 and 5.3.

At the detection side for the view-based insertion scheme, the watermark is directly recovered from the host view. For the disparity-based scheme, the detection is performed at the reconstructed right view level.

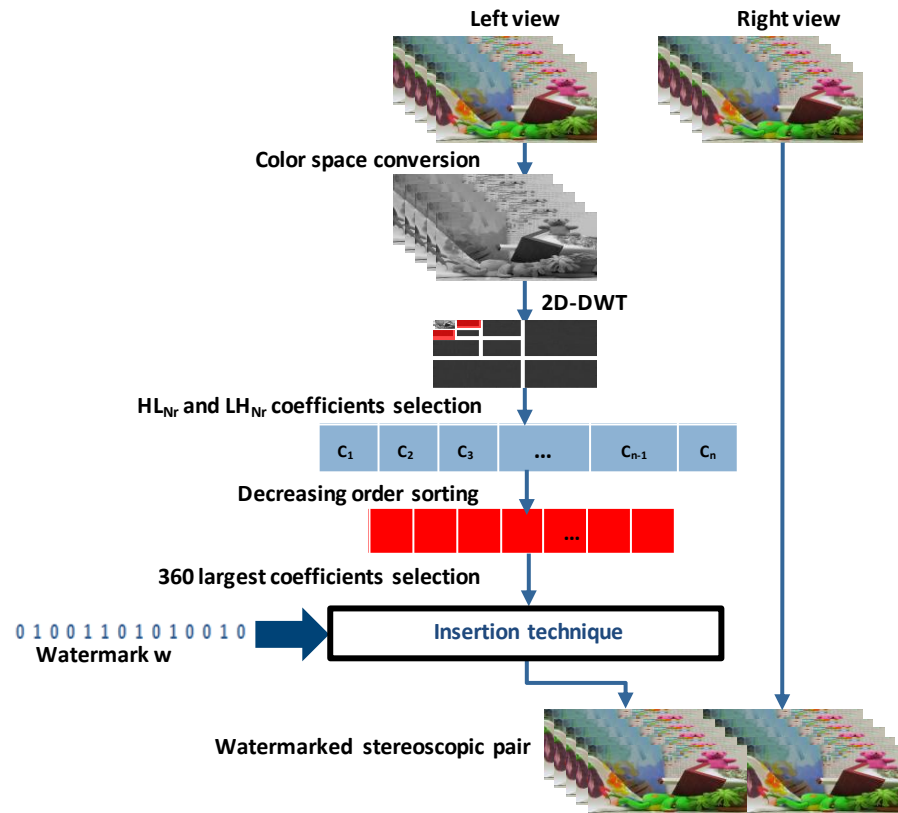


Figure 5.2 View-based watermarking scheme.

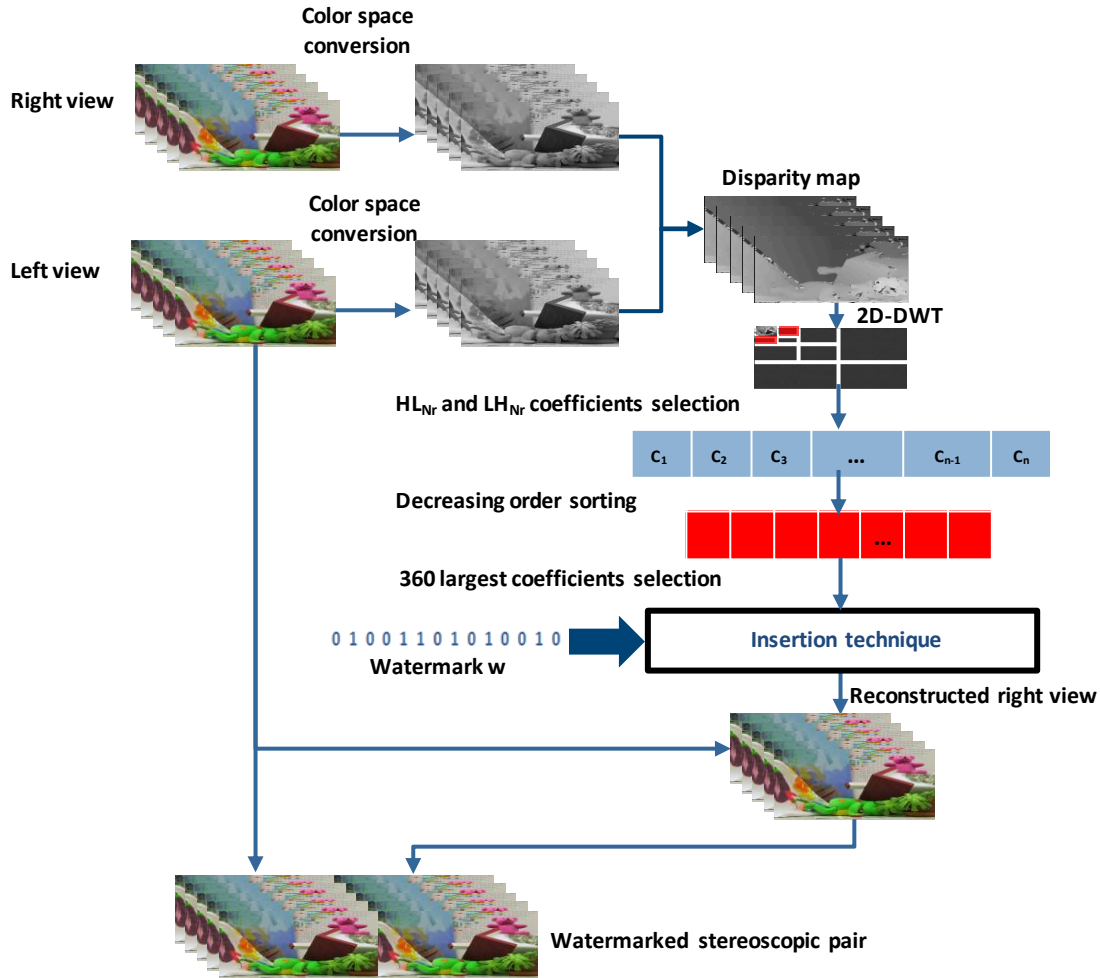


Figure 5.3 Disparity-based watermarking scheme.

5.3. Transparency evaluation

5.3.1. Subjective protocol

During the subjective testing procedure, the ITU-R BT 710 – 4 [BT98], 500 – 12 [BT02] and ITU-R BT 1438 [BT00] recommendations were followed. The testing conditions are described below.

Subject

At least 15 observers should be used. They should be non-expert, in the sense that they are not directly concerned with television picture quality as part of their normal work, and are not experienced assessors. Prior to a session, the observers should be screened for (corrected-to) normal visual acuity on the Snellen or Landolt charts, and for normal colour vision using specially selected charts (Ishihara, for instance). Observers must have also normal and dynamic stereopsis. Eight main vision tests are recommended for this goal and are described in Appendix A.3. Tests VT-04 and VT-07 are compulsory while the remaining six tests are for more detailed characterization.

In our study, the test was conducted on a total number of 25 non-expert viewers (hence, larger than the ITU-R lower limit set at 15), with marginal knowledge on the image quality. The age distribution ranged from 20 to 45 with an average of 28. All the subjects are screened for fine and dynamic stereopsis, visual acuity using Snellen chart and color vision using the Ishihara test [BT02].

Assessors have been carefully introduced to the method of assessment, the quality factors, the grading scale, the sequence and timing. Four training sequences demonstrating the range and the type of the impairments to be assessed are used in the test.

Laboratory environment: viewing conditions

The evaluation has been conducted at two locations: in professional testing conditions, at Cesson Sévigné 3D theater and in laboratory conditions, at the Advanced Research & TEchniques for Multimedia imaging Systems (ARTEMIS) Department. In the latter case, a 47" LG LCD, full HD 3D monitor (1920 × 1080 pixels) and a 400cd/m² maximum brightness is used in the experiments. Table 5.2 gives more details about the laboratory test conditions. The experiments involved 2 subjects per session. The subjects were seated in line with the center of the monitor, at a distance D equal to the height of the screen multiplied by factor 3 and defined as the Preferred Viewing Distance PVD, see Table 5.3.

Table 5.2 General viewing conditions for subjective assessments at the ARTEMIS laboratory environment.









Rec. ITU-R BT.500-12	ARTEMIS
Ratio of luminance of inactive screen to peak luminance ≤ 0.02	
Ratio of the luminance of the screen, when displaying only black level in a completely dark room, to that corresponding to peak white: ≈ 0.01	
Display brightness and contrast: set up via PLUGE software	
The viewing distance and the screen sizes are to be selected in order to satisfy the Preferred Viewing Distance PVD, see Table 5.3.	
Maximum observation angle relative to the normal (this number applies to CRT displays, whereas the appropriate numbers for other displays are under study): 30	
Ratio of luminance of background behind picture monitor to peak luminance of picture: ≈ 0.15	
Chromaticity of background: D65	
Other room illumination: low	

Table 5.3 Preferred Viewing Distance-PVD for moving image.

Screen diagonal (in)		Screen height (H)	PVD
4/3 ratio	16/9 ratio	(m)	(H)
12	15	0.18	9
15	8	0.23	8
20	24	0.30	7
29	36	0.45	6
60	73	0.91	5
> 100	> 120	> 1.53	3-4

Test method and assessment session

Test method

A DSCQS (double stimulus continuous quality scale) method has been adopted. The image quality, depth perception and visual comfort are scored on a quality scale with 5 levels going from 1 to 5 (bad, poor, fair, good, and excellent), see Table 5.4. For the result analysis, the MOS (mean opinion score) is computed for each test condition as the average of the individual score.

Assesement

The three main characteristics of stereoscopic video content which are assessed (image quality, depth perception and visual comfort) are scored according to the sheet presented in Table 5.4.

Table 5.4 A sample of a transparency: subjective evaluation sheet.

	Image quality	Depth perception	Visual comfort
Excellent
Good
Fair
Poor
Bad

Assessment session

At the beginning of the first session, from 2 to 5 training presentations are introduced to stabilize the observers' opinion, see Figure 5.2. The data issued from these presentations are not taken into account in the results of the test. If several sessions are required, only two training presentations are done at the beginning of the next session.

Each observer evaluates 34 randomly chosen video excerpts of 40 seconds each. These excerpts represent the two corpora and all the possibilities investigated in the experiments: original/watermarked video content obtained through any of the four methods applied on any of the four insertion domains.

During the testing session the subjects evaluates the suggested sequences in a random order. Each sequence is shown once or twice and a break between the presentations is necessary to give the scores.

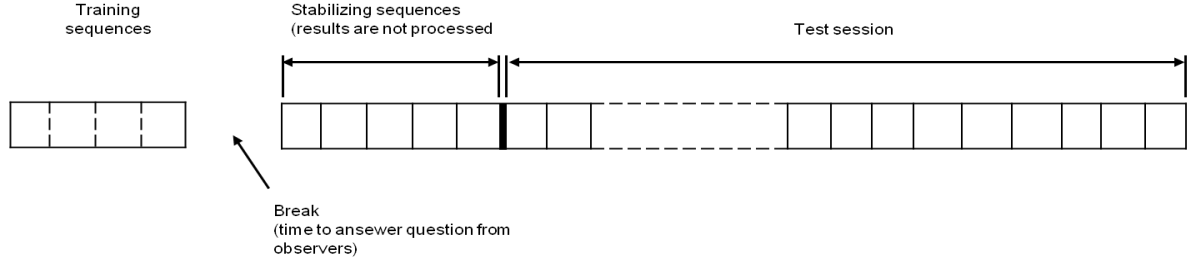


Figure 5.4 Presentation structure of the assessment session.

Results analysis

For the results analysis, the mean opinion score is computed for each test condition j as: $MOS_j = \frac{1}{N_j} \sum_{i=1}^{N_i} S_{ij}$, where N is the number of valid subjects and S_{ij} is the score by subject i for the condition j .

Watermarked samples of stereo pairs are represented in Figures 5.5 to 5.8.



Figure 5.5 Side-by-side spread spectrum watermarked samples.

Stereoscopic images from the MPEG corpus (Cartoon) and the 3DLive corpus (Rock band).



Figure 5.5 (continued) Side-by-side spread spectrum watermarked samples.
Stereoscopic images from the MPEG corpus (Cartoon) and the 3DLive corpus (Rock band).

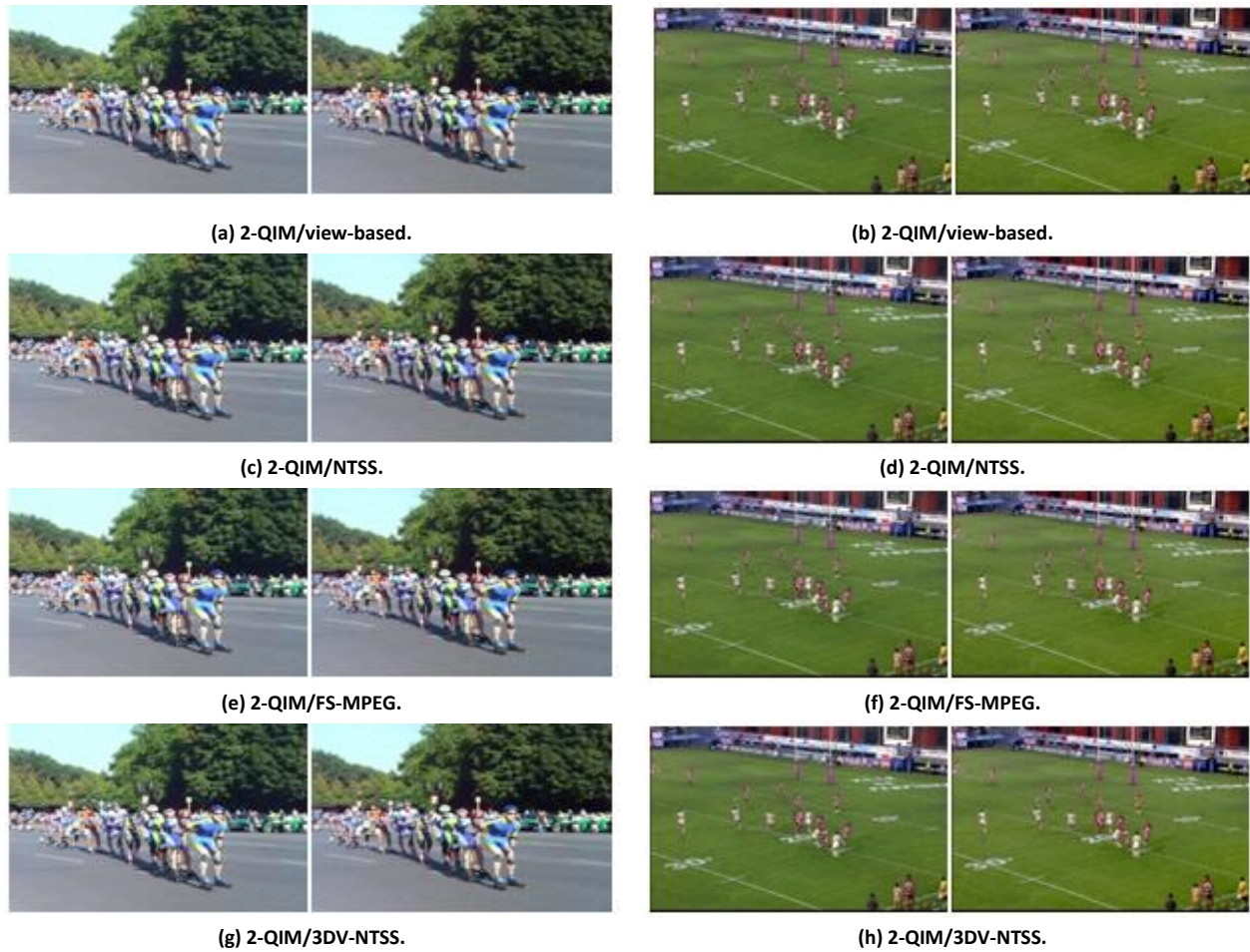


Figure 5.6 Side-by-side 2-QIM watermarked samples.
Stereo images from the MPEG corpus (Roller) and the 3DLive corpus (Rugby).



(a) 5-QIM/view-based.



(b) 5-QIM/view-based.



(c) 5-QIM/NTSS.



(d) 5-QIM/NTSS.



(e) 5-QIM/FS-MPEG.



(f) 5-QIM/FS-MPEG.



(g) 5-QIM/3DV-NTSS.



(h) 5-QIM/3DV-NTSS.

Figure 5.7 Side-by-side 5-QIM watermarked samples.
Stereo images from the MPEG corpus (City tour) and the 3DLive corpus (Scapin).



(a) Fast-IPProtect/view-based.



(b) Fast-IPProtect/view-based.



(c) Fast-IPProtect/NTSS.



(d) Fast-IPProtect/NTSS.

Figure 5.8 Side-by-side Fast-IPProtect watermarked samples.
Stereo images from the MPEG corpus (Office) and the 3DLive corpus (Volley).

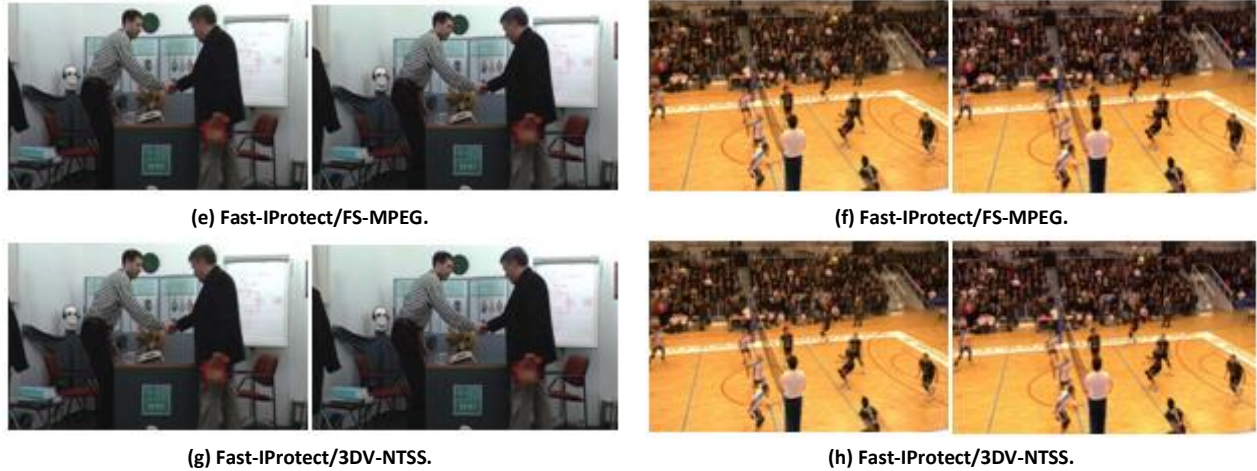


Figure 5.8 (continued) Side-by-side Fast-IProtect watermarked samples.
Stereo images from the MPEG corpus (Office) and the 3DLive corpus (Volley).

The experimental results concerning the image quality, the depth perception and the visual comfort are synoptically presented in Figure 5.9, where the MOS values are displayed alongside their 95% confidence intervals [Wal02].

Figure 5.9 is organized as follows. The abscissa is decrementally divided at three levels in order to represent all the investigated cases. First, the two corpora (3DLive and MPEG) are figured out on the left and right sides, respectively. Secondly, for each corpus, the four watermarking methods (SS, 2-QIM, 5-QIM and Fast-IProtect) are presented from left to right. Finally, for each method, each of the four insertion domains (left view, NTSS, FS-MPEG and 3DV-NTSS) are depicted from left to right. The ordinate gives the MOS values (represented in squares) and the 95% confidence limits in its estimation (represented in vertical lines centered on the related MOS values). The 95% confidence limits obtained when evaluating the original content are also presented (in horizontal, continuous red lines).

The values reported in Figure 5.9 allow us to formulate the following general conclusions:

- The 3DLive watermarked content results in better visual experience than the MPEG watermarked content, with an average MOS difference of 0.16 (this average value is computed for a given corpus, over all the insertion methods and all the insertion domains and for the image quality, depth perception and visual comfort scores). Such a behavior can be explained by the difference in the quality of the original content which may influence the viewers in assessing the quality of the watermarked content;
- When comparing among them the four classes of watermarking methods, it can be noticed that Fast-IProtect offers the best visual quality, with an average MOS larger by 0.18 (in the 3DLive case) and by 0.11 (in the MPEG case) with respect to the SS which is the second best method; (this average value is computed for a given corpus and the corresponding insertion method, over all the insertion domains and for the image quality, depth perception and visual comfort scores).

This result is unexpected, as for the 2D video content, the SS methods were reported to have the best visual quality;

- When comparing among them the four insertion domains, it can be noticed that 3DV-NTSS offers the best visual quality, with an average MOS larger by 0.06 (in the 3DLive case) and by 0.12 (in the MPEG case) with respect to the NTSS which is the second best domain (this average value is computed for a given corpus and the corresponding insertion domain, over all the insertion methods and for the image quality, depth perception and visual comfort scores). This result enforces the usefulness of the 3DV-NTSS disparity map for watermarking applications;
- The Fast-IProtect method applied in the 3DV-NTSS domain is the only solution for achieving visually imperceptibly watermarking insertion. Actually, the Fast-IProtect/3DV-NTSS is the only combination ensuring for both corpora (3DLive and MPEG) and for the three evaluation criteria (image quality, depth perception and visual comfort) confidence limits inside the confidence limits corresponding to the original content. There is only one exception (the 3DLive corpus and the visual comfort) for which the lower limit of the Fast-IProtect/3DV-NTSS confidence interval is outside the confidence limits of the original content; even in this case, the Fast-IProtect/3DV-NTSS provides the best results.

Note that the conclusions above are meant to be general. However, several types of methods/insertion domains may be alternatively considered in order to solve particular applications defined by a particular type of content/targeted quality criterion. For instance, the protection of some low quality MPEG content can be achieved under the depth perception constraints by three types of solutions: SS/NTSS, Fast-IProtect/FS-MPEG and Fast-IProtect/3DV-NTSS (cf. Figure 5.9.b).

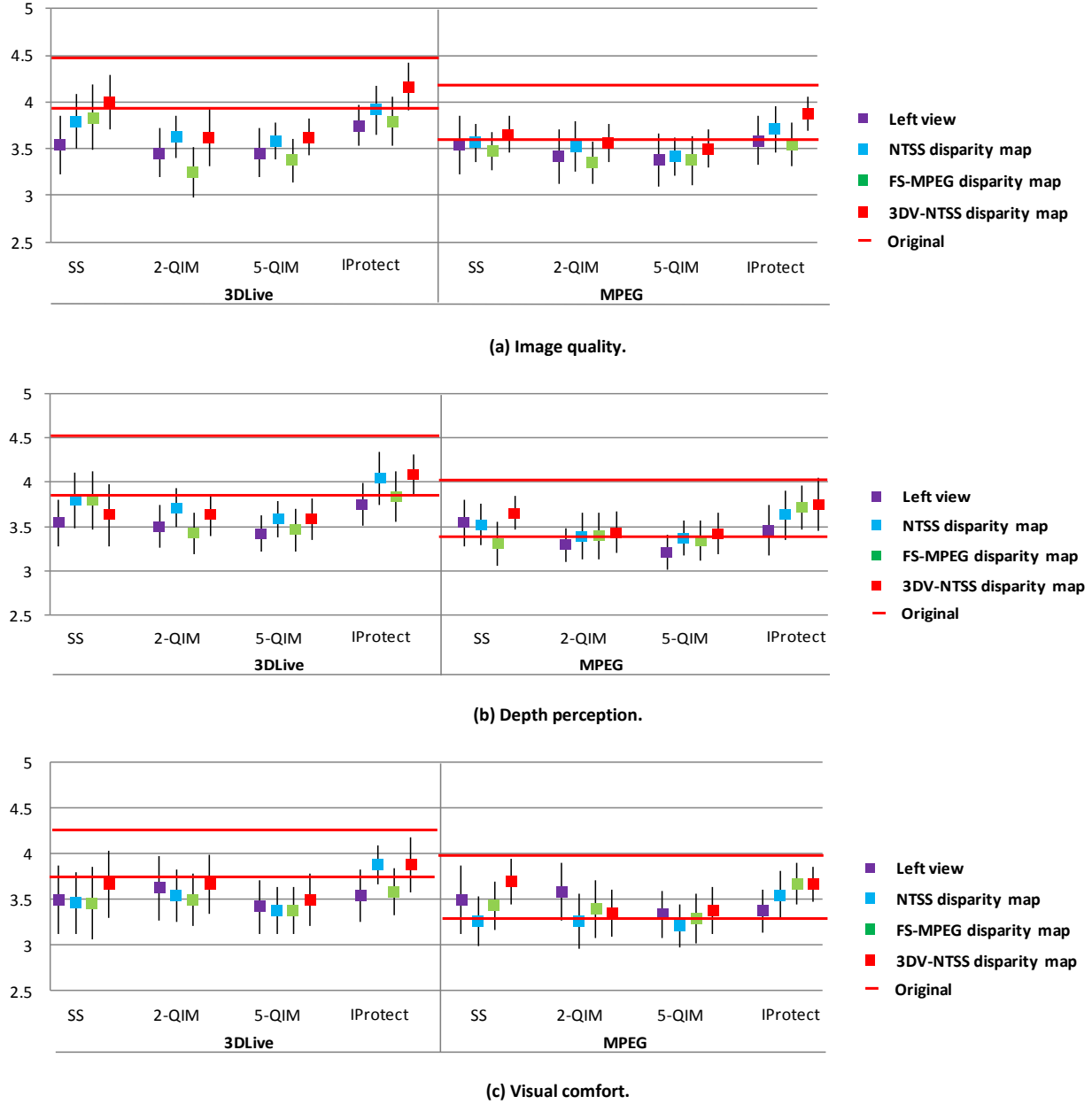


Figure 5.9 Subjective evaluations for image quality, depth perception and visual comfort.

MOS values (in squares) and the related 95% confidence limits (in vertical lines centered on the MOS) for watermarked content. The original content subjective evaluation is represented by its 95% confidence limits (in horizontal red lines).

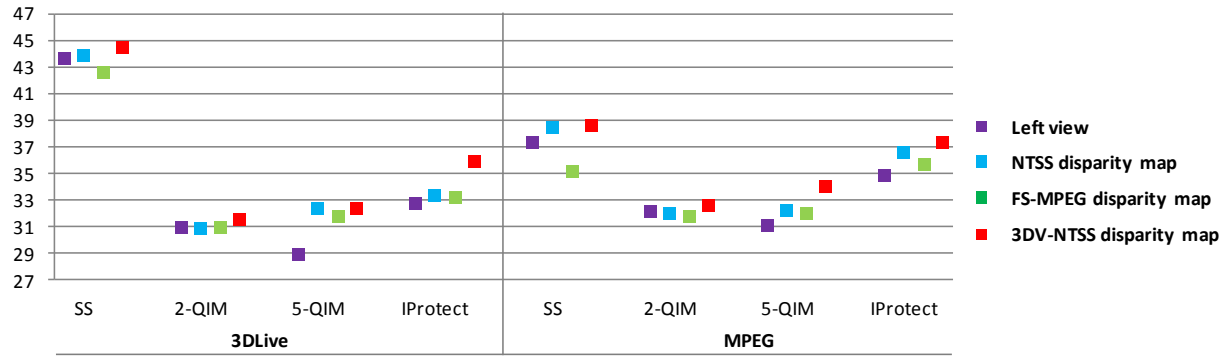
5.3.2. Objective assessment

The visual quality of the watermarked content is objectively evaluated by the five measures described in Chapter 1.1.2.1, namely the PSNR, IF, NCC, SC and SSIM. For each watermarking method and insertion domain, each of these five measures is first computed at the view level, then averaged at the corpus level. Figure 5.10 represents the corresponding average values. This figure is organized in the same way as Figure 5.9.

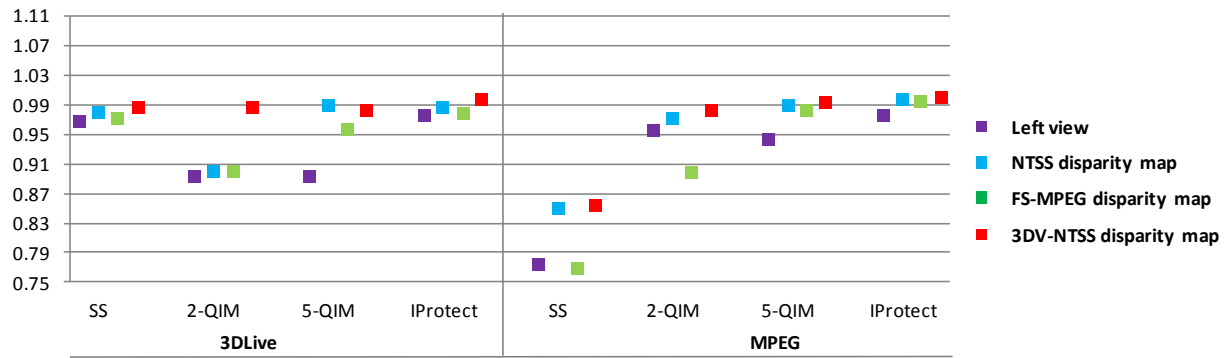
The 95% confidence limits are also computed for each experiment (each corpus, watermarking method and insertion domain). As the corresponding error is always lower than 0.45dB in the PSNR case and 0.001 in the IF, NCC, SC and SSIM cases, the related confidence limits cannot be represented in Figure 5.10, being lower than the printing resolution for the average values. There is only one exception: for the 3DLive corpus, the Fast-IProtect method and the view-based domain, the 95% error in SC estimation is 0.015 and is represented in Figure 5.10.d.

The values reported in Figure 5.10 allow us to formulate the following general conclusions:

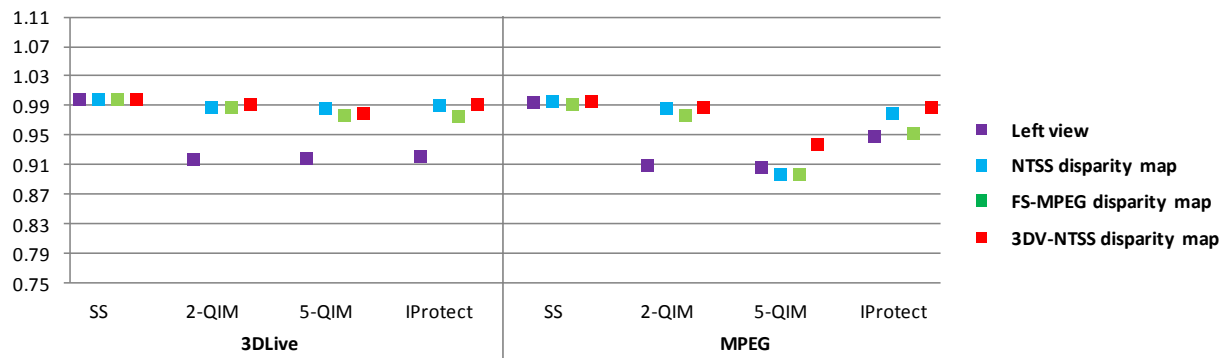
- For a given measure and insertion domain, very few differences can be noticed between the 3DLive and MPEG corpora. This result validates the fairness of the benchmarking conditions (i.e. the parameters of the investigated watermarking methods were set so as to ensure a transparency independent with respect to the original data and dependent only on the method/insertion domain);
- The PSNR average values are always larger than 30dB (with a singular exception, namely the 5-QIM method applied to the left views of the 3DLive corpus). Consequently, all the considered watermarking methods/insertion domains can ensure basic transparency properties. However, very good transparency (larger than 35dB) can be achieved only by the SS (for all insertion domains and for the two corpora) and by the Fast-IProtect (3DV-NTSS in the 3DLive case and all the four insertion domains in the MPEG case). According to the PSNR values, SS would be the best watermarking method, followed by Fast-IProtect;
- The IF, NCC, SC and SSIM values also support the idea that basic transparency (i.e. values between 0.95 and 1.05) can be virtually ensured by all the considered watermarking methods (with some constraints in the choice of the insertion domain). There is one exception, represented by the SS method which is refuted by the IF measures estimated on the MPEG corpus. Here again, the SS and Fast-IProtect identified themselves as the best solutions;
- For each watermarking method and for each corpus, all the five objective quality metrics select the 3DV-NTSS disparity map as the optimal insertion domain, with a singular exception (the NCC values computed for the 5-QIM insertion method applied to the 3DLive corpus).



(a) Average values: PSNR.

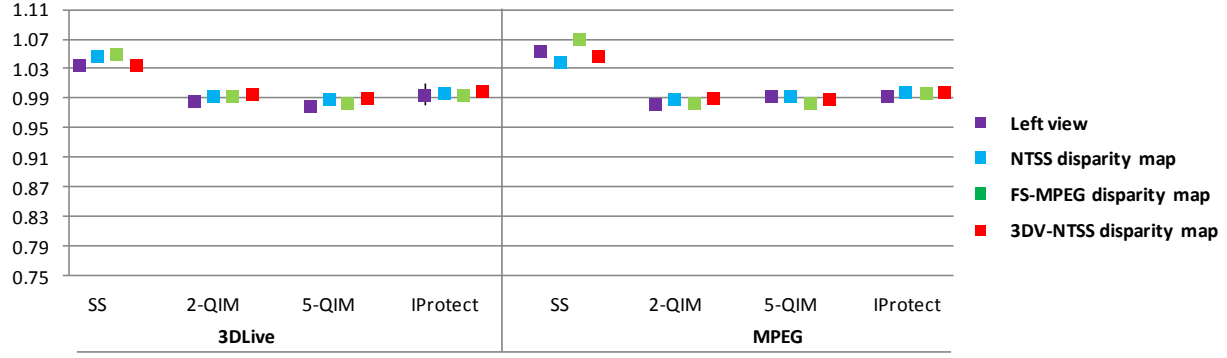


(b) Average values: IF.

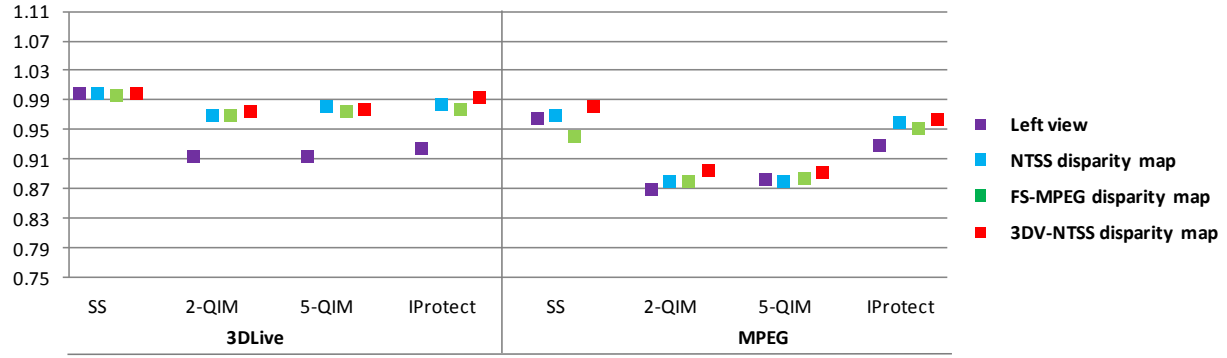


(c) Average values: NCC.

Figure 5.10 Objective evaluation of the watermarked content visual quality.



(d) Average values: SC.



(e) Average values: SSIM.

Figure 5.10 (continued) Objective evaluation of the watermarked content visual quality.

5.4. Robustness evaluation

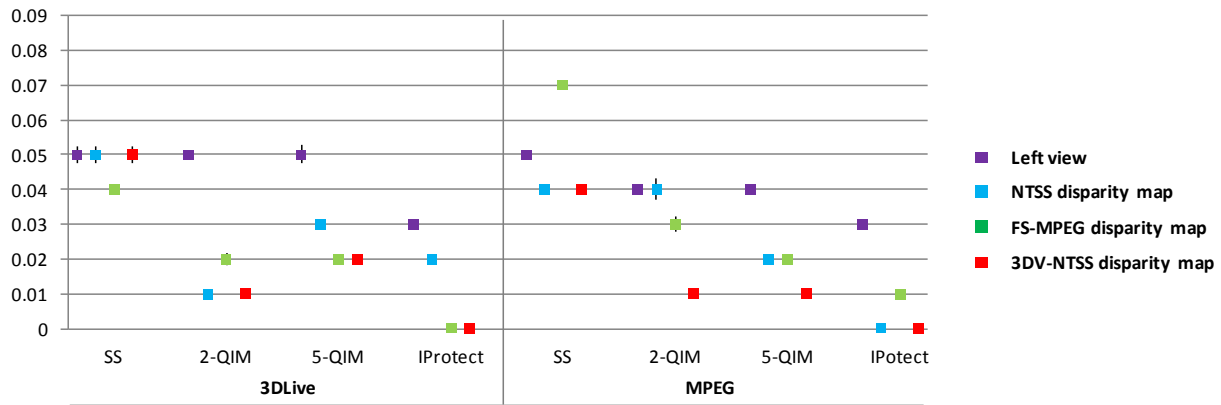
The robustness is evaluated by the BER in the watermark detection after the attacks. Three types of attacks are considered: filtering (Gaussian and sharpening), JPEG compression and geometric (both small rotations and StirMark random bending). Each attack is individually applied at the frame level, then the corresponding BER are averaged at the corpus level. The current section presents a synthesis of the results, obtained for Gaussian and sharpening attacks applied with a 3×3 convolution kernel, JPEG compression at a quality factor $Q_f = 60$, a rotation with an angle of ± 0.5 and the StirMark random bending attack applied at its default parameters [Pet98] [Pet00].

The BER average values and their related 95% confidence limits are reported in Figure 5.11, which is organized in the same way as Figure 5.10.

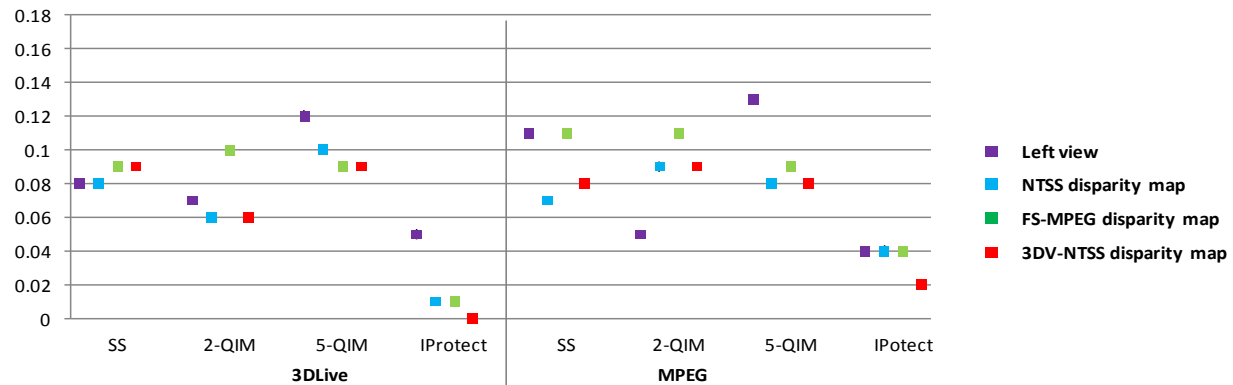
When inspecting the results reported in Figure 5.11, it can be noticed that Fast-IProtect/3DV-NTSS is the only combination ensuring a robustness expressed by BER lower than 0.05 after filtering and JPEG compression and lower than 0.1 after the geometric attacks, irrespective to the corpus.

The 2-QIM method applied on the 3DV-NTSS domain features the same good values of the BER against Gaussian filtering, compression and rotation attacks but fails in meeting the robustness requirements against sharpening (on both 3DLive and MPEG corpora) and against StirMark random bending (only in the case of the MPEG corpus).

Also note that the SS method does not succeed in meeting the robustness requirements, irrespective of the insertion domain and/or processed corpus.

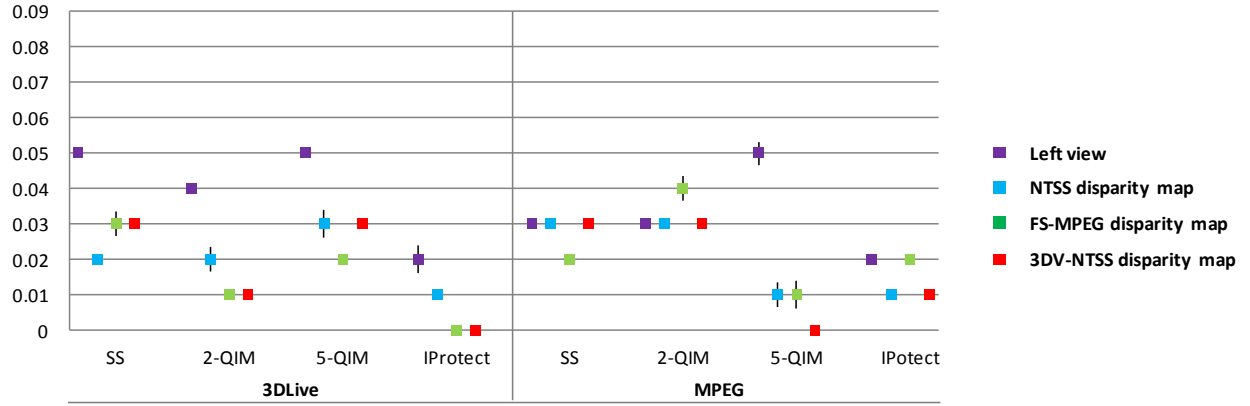
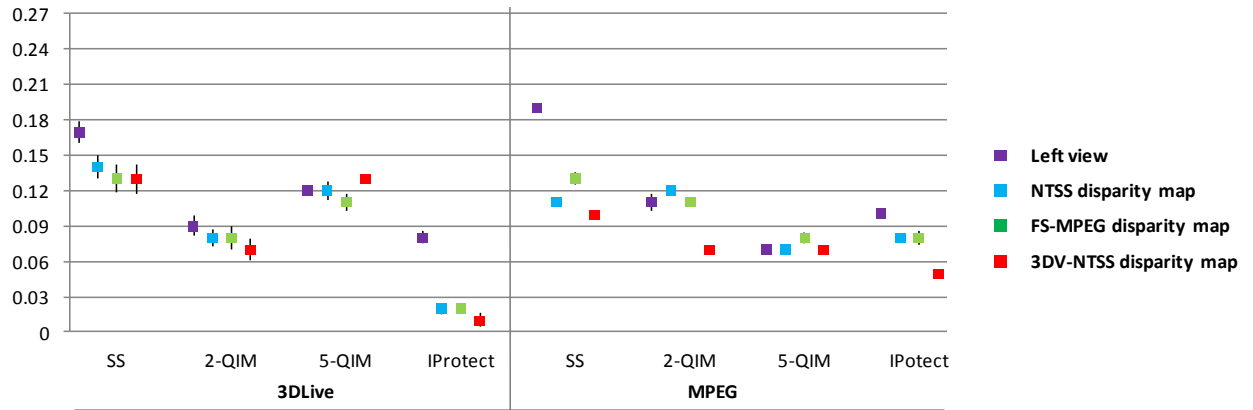
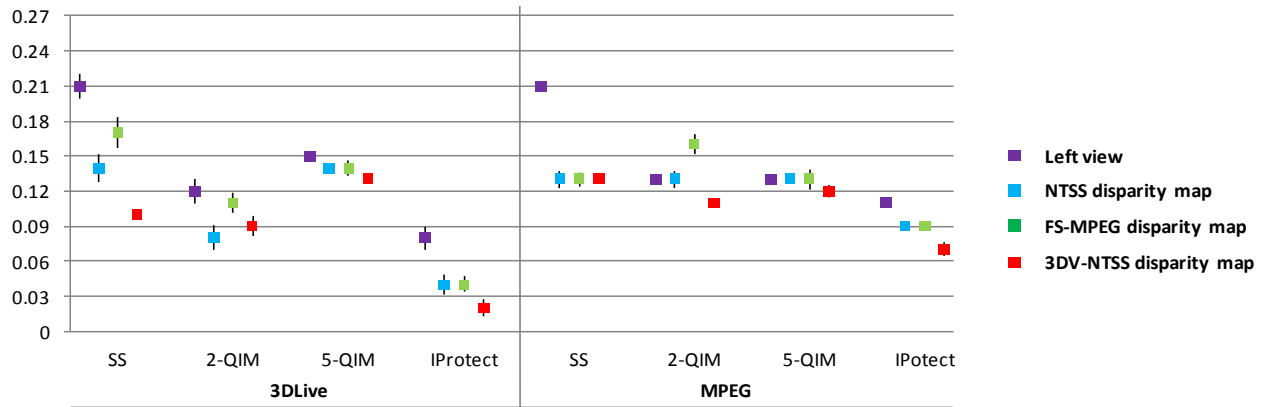


(a) Gaussian filtering (3×3 convolution kernel).



(b) Sharpening (3×3 convolution kernel).

Figure 5.11 Watermark robustness against five different attacks.
BER average value and the related 95% confidence limits.

(c) JPEG compression (Quality factor $Q_f = 60$).(d) Small rotations $\pm 0.5^\circ$.

(e) Geometric attacks (StirMark random bending).

Figure 5.11 (continued) Watermark robustness against five different attacks.
BER average value and the related 95% confidence limits.

5.5. Computational cost

5.5.1. Processing time

The results already presented hint to the Fast-IProtect watermarking method performed in the 3DV-NTSS disparity map as the most effective solution for stereoscopic video protection, when considering transparency and robustness constraints, for a fixed data payload. However, for several real life applications (e.g. live HD 3D TV content protection), the computational cost should also be investigated. In this respect, we evaluate the processing time required by the three main steps in the watermarking chain: DWT/disparity computation, mark insertion and inverse DWT/image reconstruction. Such values are evaluated at view level, than averaged at the corpus level. The values illustrated in Figure 5.12 are expressed in milliseconds (ms); they are obtained on a PC Core2 CPU @ 2.13GHz, 2GB de RAM.

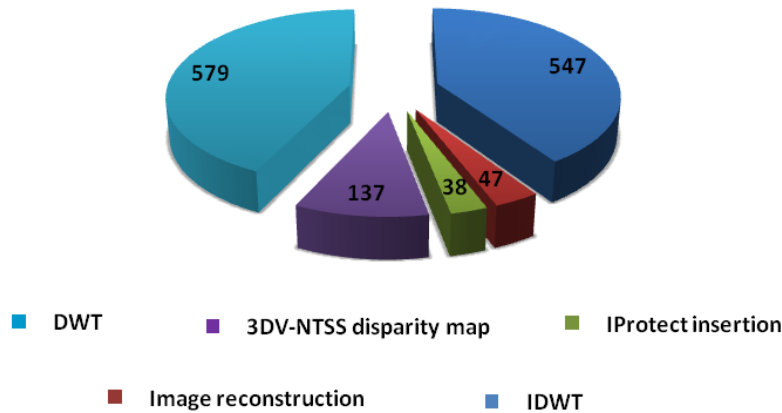


Figure 5.12. Watermarking chain computational cost.

It can be noticed that for the 3DV-NTSS disparity map computation, the watermark insertion and the image reconstruction are 4, 15 and 12 times faster than a DWT computation, respectively. Consequently, for real life solutions implementing the DWT/IDWT in real time [Gal11] the Fast-IProtect/3DV-NTSS watermarking can be also carried out in real time. Of course, this is not case of our software implementations, which should be considered in the present study only as an investigation tool.

5.5.2. Complexity

In order to obtain an *a priori* estimation of the computational cost, independent of the software peculiarities, the complexity of the underlying algorithm should be investigated.

Table 5.5 Computational complexity of the watermarking modules involved in Fast-IProtect/3DV-NTSS.

Watermarking module	Data size	Complexity
2D-DWT	$h \times w$	$O(h \times w)$
3DV-NTSS disparity map computation	$K = h \times w / (Bs \times Bs)$	$K \times O(\log_2(p + 1))$
Fast-IProtect insertion	$N \times m$	$N \times O(m)$
Image reconstruction	$h \times w$	$O(h \times w)$
2D-IDWT computation	$h \times w$	$O(h \times w)$

In Table 5.5 the following notations are made:

h and w are the height and width of the frame for which the wavelet is computed. m is the watermark size and N is the iteration number. p is the pixel searching distance in block of size $Bs \times Bs = 16 \times 16$ pixels.

5.6. Discussion

This Chapter reports on a comparative study the possibility of using 2D inherited watermarking methods for stereoscopic video protection. The comparison is carried out on the watermarking method (belonging to the SS, IE and hybrid SS-IE classes) and on the underlying insertion domain (left view, NTSS, FS-MPEG and 3DV-NTSS).

The experimental results brought to light that the Fast-IProtect (a hybrid SS-IE method) performed in a new disparity map domain (3DV-NTSS) would be generic enough to as to serve a large variety of applications when:

- it is the only insertion method ensuring the imperceptibility of the watermark according to subjective tests (TU-R BT 710 – 4, 500 – 12 and BT 1438 recommendations) and three criteria (image quality, depth perception and visual comfort);
- this subjective transparency evaluation is reinforced by offering limits PSNR > 35 dB and limit IF, NCC, SC and SSIM larger than 0.95;
- it is the only investigated method ensuring robustness expressed by BER lower than 0.05 after filtering and JPEG compression and lower than 0.1 after the geometric attacks;
- a non-prohibitive computational complexity compatible to the real time application (in the sense discussed in Chapter 5.5.1).

The generality of the results is ensured by the size and composition of the two corpora (a total of 2 hours, 11 minutes and 24 seconds of heterogeneous content) and by the statistical error control (95% confidence limits) for all the results.

Chapter 6 Conclusion and future work

6.1. Conclusion

Nowadays, cinemas and TVs production companies are competing to release high quality 3D movies and make customers enjoy a new multimedia experience. 3D devices ownership is consequently increasing rapidly all over the world. The boom of stereoscopic video applications and the proliferation of powerful duplication/manipulation tools, have raised concerns about content tracking and copyright protection and created an urgent need to protect ownership and to prevent the content from tampering. Digital watermarking has been proposed to address these needs since it potentially supports all the requirements set by real life applications without imposing any constraints for a legitimate user.

The watermarking applicative issue is to reach the trade-off between the properties of transparency, robustness and data payload. Selecting the optimal insertion domain is an additional challenge imposed by the stereoscopic contents peculiarities. Establishing an innovative and efficient stereoscopic video watermarking a specific benchmarking reinforced by statistical relevance and fostered by standard recommendations.

The present thesis tackles these three challenges (as synoptically presented in the table included in the Abstract page 4).

First, by reconsidering some 2D video inherited approaches and by adapting them to the stereoscopic video content and to the human visual system peculiarities, a new disparity map (3DV-NTSS) is designed. The performances of the 3DV-NTSS are evaluated in terms of visual quality of the reconstructed image and computational cost. When compared with state of the art methods (New three step search NTSS and FS-MPEG) average gains of 2 dB in PSNR and 0.10 in SSIM are obtained. The computational cost is reduced by average factors between 1.3 and 13.

Second, a comparative study on the main classes of 2D inherited watermarking methods and on their related optimal stereoscopic insertion domains is carried out. Four insertion methods are considered; they belong to the SS, SI (binary QIM and 5-symbols QIM) and hybrid (Fast-IProtect) families. The Fast-IProtect establishes synergies between SS and SI in order to achieve the transparency/robustness/data payload trade-off and relays on Monte Carlo generators (following the attack theoretical models) in order to meet the time constraints. Each of these four methods is successively applied on the left view of the video sequences as well as on three disparity maps (computed according to the NTSS, FS-MPEG and 3DV-NTSS algorithms). The experiments brought to light that the Fast-IProtect performed in the new disparity map domain (3DV-NTSS) would be generic enough so as to serve a large variety of applications:

- it ensures the imperceptibility according to subjective tests preformed according to three different criteria: image quality, depth perception and visual comfort;
- it offers PSNR > 35 dB and IF, NCC, SC and SSIM values larger than 0.95;
- it features robustness expressed by a BER lower than 0.05 after filtering and JPEG compression and lower than 0.1 after the geometric random bending attacks;
- it is compatible with real time applications (e.g. insertion time of $T_{insertion} = 38$ ms, lower than the frame rate in video, results obtained on a Core2 PC, CPU@2.13GHz, 2 Go de RAM).

Finally, concerning the performance evaluation, all the quantitative results are obtained out of processing two corpora (3DLive and MPEG) of stereoscopic visual content, organized according to three general criteria: significance, acceptability and exploitability. Each of these two corpora combines indoor/outdoor, unstable and arbitrary lighting, still and high motion scenes. The 3DLive corpus sums up about 2 hours of HD 3D TV content captured by French professionals. The MPEG 3D video reference corpus is composed of 17 minutes of video provided by both academic/industry and encoded at different resolutions (from 320×192 to 640×480 pixels).

The statistical relevance of the results is given by the 95% confidence limits computed for all the values reported in our study (for both transparency and robustness), and by their underlying the relative errors which are lower than $\varepsilon_r < 0.1$.

Two standards have been considered in our study. The transparency of the watermarked content is subjectively assessed according to the ITU-R BT 710 – 4, 500 – 12 and BT 1438 recommendations. The robustness and data payload are considered so as to comply with the Digital Cinema Initiatives (DCI) prescriptions.

6.2. Future work

This thesis represents the proof of concepts for stereoscopic video robust watermarking: by developing a new disparity map and by combining spread spectrum with side information principles, the transparency/robustness/data payload balance can be reached in real time.

The perspectives open towards the specification of the theoretical model for stereoscopic watermarking. The 2D-video inherited model is defined by a noisy channel where the attacks act as a general noise source while the original content stands for a side information (a noise source known at the embedder). The issue is to maximize the data payload on such a channel, under an additional power constraint set by the human visual system.

When extending this model to the stereoscopic video case, three issues should be dealt with. First, the theoretical probability density functions modeling the various transforms the stereoscopic video suffers during its distribution are not yet investigated. Secondly, the two stereo views represent correlated side information noise sources; no theoretical result is nowadays available for handling such a situation. Finally, the human stereoscopic vision is not yet modeled with precision, at least not so as to be directly integrated as a power limit constraint for the inserted mark.

Appendix

A.1 Processed corpora

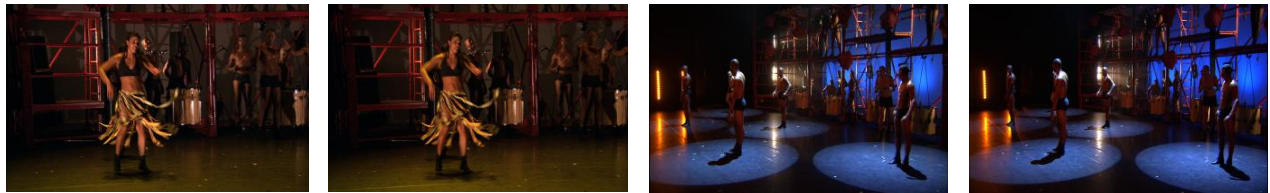
All the experiments reported in the present paper are carried out on two corpora further referred to as 3DLive and MPEG, respectively. Each of these corpora combines indoor/outdoor, unstable and arbitrary lighting, still and high motion scenes, as illustrated in Figures A.1.1 to A.1.4.

A.1.1 3DLive corpus

The 3DLive French national project was meant to create expertise in France for shooting and live TV transmission of 3D stereo contents. The 3DLive corpus sums-up to 2 hours, 11 minutes and 24 seconds of stereoscopic video sequences (197000 stereoscopic pairs encoded at 25 frames per second), representing 10 minutes of a rugby match (further referred to as the Rugby sequence), 10 minutes of a dancing performance (Dancing), 1 minute of a private gig of rock band “Skip the Use” (Skip the Use), one hour and 45 minutes and 24 seconds of a volley-ball match (Volley) and 5 minutes of a theater play “les Fourberies de Scapin” (Scapin). These sequences are full HD encoded (1920×1080 pixels).



(a) Rugby match sequence (Rugby).



(b) Dancing experience sequence (Dancing).



(c) Volley match sequence.

Figure A.1.1 Left and right views sampled from the 3DLive corpus.



(d) A theater play "les Fourberies de Scapin" sequence.



(e) Rock band concert sequence.

Figure A.1.1 (continued) Left and right views sampled from the 3DLive corpus.

A.1.2 MPEG corpus

The MPEG corpus [Hoi11], is composed of 41 sequences and sums-up to 17 minutes and 29 seconds (29908 stereoscopic pairs, as several frame rates have been considered). Various resolutions are represented; they range from 320×192 to 640×480 pixels. These videos were provided by the Heinrich-Hertz-Institute (HHI), KUK Film produktion and the Technical University of Berlin - Communication Systems Group. The content corresponds to street events, like roller and biking races (10 minutes and 6 seconds, 640×480 pixels), indoor (office) scenes (1 minutes and 22 seconds, 320×192 pixels), city tours (2 minutes and 13 seconds, 360×288 pixels), cartoons (22 seconds, 432×240 pixels), etc.

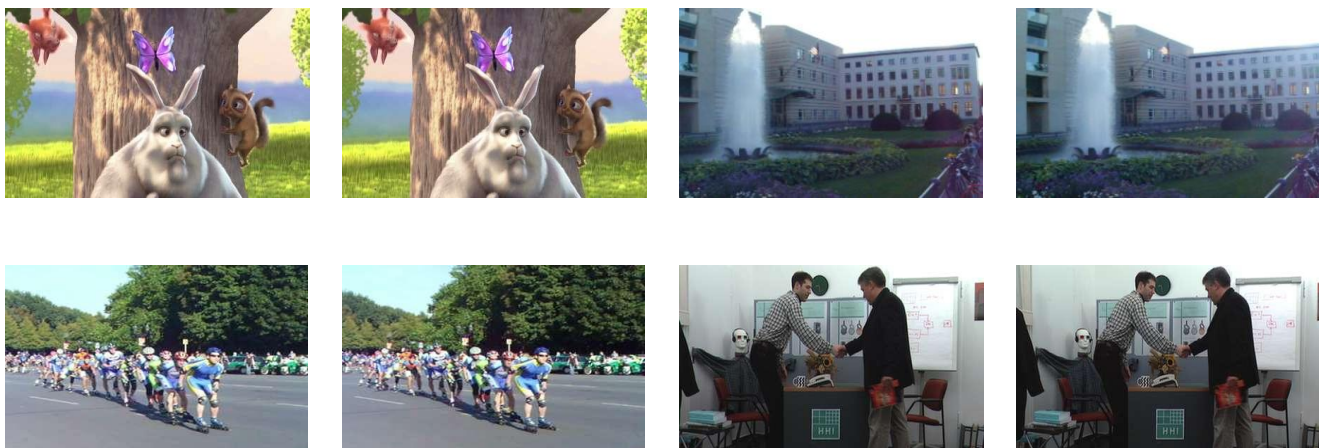


Figure A.1.2 Left and right views sampled from the MPEG corpus.
Cartoons, city tours, rollerblade, and indoor content.

A.1.3 Cross-checking content

The generality of the result has been cross-checked on content of completely different types: special effects, still images and computer generated stereoscopic medical images.

HD3D² corpus

This corpus is composed by an anaglyph video sequence provided by Mikros Image (<http://www.mikrosimage.eu/>) in the context of HD3D² project. This corpus contains 5650 anaglyph images (720×576 pixels) of a promotional clip, see Figure A.1.3.

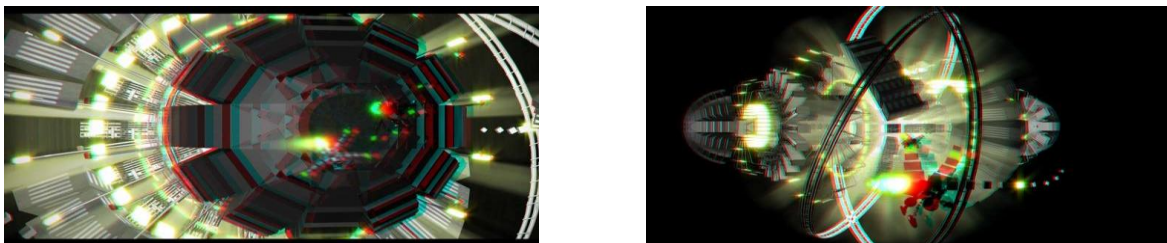


Figure A.1.3 Experimental HD3D² database.

EPFL corpus

The stereoscopic image database compiled at Ecole Polytechnique Fédérale de Lausanne by Prof. T. Ebrahimi and Dr. L. Goldman (<http://mmspl.epfl.ch/page38841.html>) [Gol10] contains 100 stereoscopic images with a resolution of 1920×1080 pixels. Various indoor and outdoor scenes with a large variety of colors, textures, and depth structures are included, see Figure A.1.4.



Figure A.1.4 Experimental EPFL database.

ARTEMIS corpus

This corpus contains a medical image database jointly provided by Prof. P. Grenier from Pitié-Salpêtrière Hospital and Dr. Catalin Fetita from the ARTEMIS Department (www.it-sudparis.eu/artemis) [Gre04] [Fet09]. It consists of 360 images of 500×500 pixels corresponding to bronchial tree reconstruction from CT images.



Figure A.1.5 Experimental ARTEMIS database.

Middlebury corpus

The Middlebury corpus [Sch02] is composed of 58 stereoscopic pairs. Various resolutions are represented; they range from 450×375 to 1800×1500 pixels. These frames were created by Anna Blasiak, Jeff Wehrwein, Brad Hiebert-Treuer, Sarri Al Nashashibi, and Daniel Scharstein at Middlebury College (<http://vision.middlebury.edu>).



Figure A.1.6 Experimental Middlebury database.

The experiments detailed in Chapter 5 are resumed in the sequel for these three corpora.

Table A.1.1 gives the PSNR, IF, NCC, SC and SSIM average values obtained when considering the Fast-IProtect/3DV-NTSS for watermarking.

Table A.1.2 gives the average BER values obtained after applying five different attacks when considering the Fast-IProtect/3DV-NTSS for watermarking.

The values reported in Tables A.1.1 and A.1.2 confirm that the Fast-IProtect performed in the new disparity map domain (3DV-NTSS) features good transparency while ensuring robustness expressed by BER lower than 0.05. While the size of these three corpora does not allow an accurate statistical investigation, the overall average values point to even better watermarking results then obtained on the 3DLive and MPEG corpora.

Table A.1.1 Watermarking transparency.

Corpus	PSNR	IF	NCC	SC	SSIM
HD3D ²	43.62	0.999	0.995	0.998	0.998
EPFL	45.62	0.999	0.997	0.997	0.998
Artemis	37.28	0.989	0.991	0.998	0.997
Middlebury	45.06	0.997	0.993	0.990	0.999

Table A.1.2 BER (Bit Error Rate) in mark detection after several attacks.

Corpus	Gaussian filtering	Sharpening	JPEG compression Q=60	Rotation +0.5°	Rotation -0.5°	StirMark random bending
HD3D ²	0	0	0	0.027	0.032	0.045
EPFL	0	0	0	0.020	0.010	0.020
Artemis	0	0	0	0.021	0.019	0.013
Middlebury	0	0	0	0	0	0.047

A.2 Considered embedding techniques

This Appendix presents details related to the SS and QIM insertion techniques.

A.2.1 Spread spectrum

Amongst the large family of watermarking techniques, our study considers the method reported in [Cox97], because of its well recognized transparency and robustness properties. Let $M = \{m_i/m_i \in \{1,0\}\}$ be the additional information to be inserted in the original frame. Each bit m_i is spread by a large factor cr (the so-called chip-rate), to obtain the spread sequence: $= \{b_i\}$. The watermark w_i is added to image $C = \{c_i\}$ yielding a watermarked image $\hat{c}_i = c_i(1 + \alpha \cdot b_i)$, where α is the watermark strength.

A.2.2 Quantization index modulation

The Quantization Index Modulation (QIM) is an informed embedding watermarking method which proved its effectiveness for mono video watermarking techniques, by achieving good robustness while keeping the practical limits of the perceptual distortions [Has11] [Bel10]. The method starts by modulating an index or a sequence of indexes with the message m to be embedded and then quantizes the host data c_0 by using the associated quantizer or sequence of quantizers.

Initially designed under the binary framework, the QIM methods were generalized to *multisymbol QIM* in [Has11] Be there a binary message to be inserted; instead of directly inserting it, a message m encoded into an s -ary alphabet $D = \{-(s-1)/2, -(s-2)/2, \dots, 0, \dots, (s-2)/2, (s-1)/2\}$ is considered so as to increase the data payload by a factor of $\log_2(s)$.

For a host signal c_0 and a message m , the watermarked signal sample y is computed by:

$$y = c_0 + \alpha \cdot q$$

$$w = \alpha \cdot q \text{ and } q = Q_\Delta \left(c_0 - \Delta \left(\frac{m}{s} + k \right) \right) - \left(c_0 - \Delta \left(\frac{m}{s} + k \right) \right)$$

where Δ is a fixed quantization step size, k a random key and α a fixed parameter, $0 < \alpha \leq 1$. The standard quantization operation is defined as:

$$Q_\Delta(c_0, \Delta) = \Delta \text{Round}(c_0/\Delta)$$

At the decoder, the embedded message bit is recovered by a scalar quantization of the received signal sample, r (a corrupted version of y).

The $Y(m)$ detection variable is computed as follows:

$$Y(m) = Q_\Delta(r - k\Delta) - r + k\Delta$$

The decision rule is given by:

$$\begin{cases} -I_{sup}(m) \leq Y(m) \leq -I_{inf}(m) \\ \alpha > \frac{s-1}{s} \end{cases}$$

where $I_{sup}(m) = \frac{\Delta((1-\alpha)s+2m)}{2s}$ and $I_{inf}(m) = \frac{\Delta((\alpha-1)s+2m)}{2s}$.

Example:

For $m = 5$ we have $\alpha^* = 0.8$. The decision regions are obtained as illustrated in Figure A.2.1.

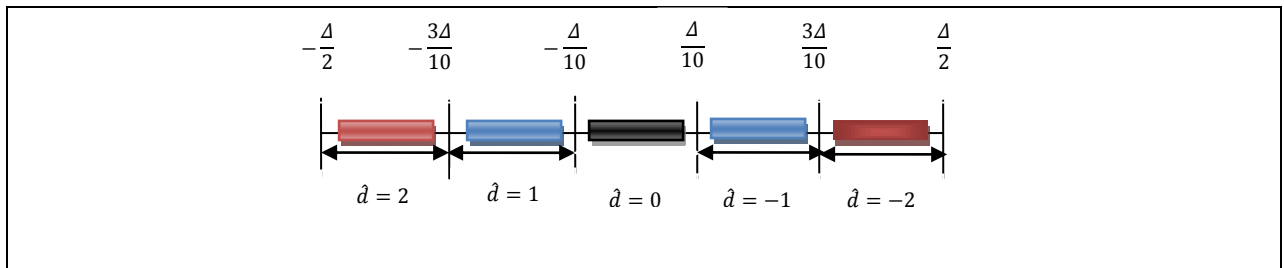


Figure A.2.1 Decision regions for $m=5$ and $\alpha \geq \alpha^*$.

A.3 Transparency evaluation

Subjective evaluation

Table A.3.1 describes the main 8 vision tests (VTs) for good stereopsis as recommended by ITUR-1438. Observers must have normal stereopsis, meaning that they must pass test VT-04 for fine stereopsis and test VT-07 for dynamic stereopsis. The remaining six tests are for more detailed characterization. The test charts should be viewed from three times the height of the display screen. Below, right and left thumbnail images are put side by side for crossed free fusion for explanatory purposes.

Tables A.3.2 and A.3.3 give the Mean Opinion Score values for a valid number of 25 subjects and their 95% error limits, respectively. These values are given for image quality, depth perception and visual comfort criteria.

Table A.3.1 The considered attacks and their main parameters.




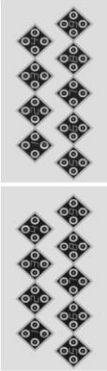


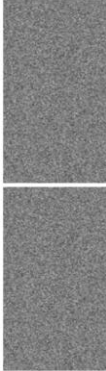

No	Item	Test		Illustration
1	Simultaneous perception	Test for	The ability to perceive dichoptically presented images simultaneously and in the correct position	
		Content	A cage image is presented to one eye and a lion image to the other eye	
2	Binocular fusion	Test for	The ability to perceive two dichoptic images in left and right eyes as one image	
		Content	The image for one eye has two dots, and the image for the other eye has three dots, with one dot in common	
3	Coarse stereopsis	Test for	The ability to perceive dichoptically presented images with a parallax as one image with a coarse depth	
		Content	The image for two eyes are a stereopair of images of a dragonfly with its wings spreading	
4	Fine stereopsis	Test for	The ability to perceive dichoptically presented images with a parallax as one image with a fine depth	
		Content	Nine test lozenge patches are provided and each of them has four circles in which one circle has a small parallax	
5	Crossed fusion limit	Test for	The ability to perceive dichoptically presented images with crossed disparities as one image	
		Content	A stereopair of bars is presented with its crossed parallax changing by 10'/s	
6	Uncrosses fusional limit	Test for	The ability to perceive dichoptically presented images with uncrossed disparities as one image	
		Content	A stereopair of bars is presented with its uncrossed parallax changing by 11'/s	
7	Dynamic stereopsis	Test for	The ability to perceive depth in moving random dot stereogram images	
		Content	Dynamic random dot stereogram	
8	Binocular acuity	Test for	The binocular acuity, including any imbalance of monocular acuity which might prevent good stereopsis	
		Content	E characters with a variety of orientation and size	

Table A.3.2 The Mean Opinion Score values for a valid number of 25 subjects.
MOS values given for image quality, depth perception and visual comfort, see Figure 5.9.

				Image quality	Depth perception	Visual comfort
3DLive	Original			4.20	4.16	4.00
	SS	View-based		3.54	3.54	3.50
		Disparity-based	NTSS	3.79	3.79	3.45
			FS-MPEG	3.83	3.79	3.45
			3DV-NTSS	4.00	3.65	3.66
	2-QIM	View-based		3.45	3.5	3.65
		Disparity-based	NTSS	3.65	3.70	3.54
			FS-MPEG	3.25	3.41	3.50
			3DV-NTSS	3.65	3.65	3.66
	5-QIM	View-based		3.45	3.41	3.41
		Disparity-based	NTSS	3.58	3.58	3.35
			FS-MPEG	3.37	3.45	3.35
			3DV-NTSS	3.65	3.58	3.5
	Fast-IProtect	View-based		3.75	3.75	3.54
Disparity-based		NTSS	3.97	4.04	3.87	
		FS-MPEG	3.79	3.83	3.58	
		3DV-NTSS	4.16	4.08	3.85	
MPEG	Original			3.91	3.70	3.62
	SS	View-based		3.54	3.54	3.5
		Disparity-based	NTSS	3.56	3.52	3.26
			FS-MPEG	3.47	3.30	3.43
			3DV-NTSS	3.65	3.65	3.69
	2-QIM	View-based		3.41	3.29	3.583
		Disparity-based	NTSS	3.52	3.39	3.26
			FS-MPEG	3.34	3.39	3.39
			3DV-NTSS	3.56	3.43	3.34
	5-QIM	View-based		3.375	3.20	3.33
		Disparity-based	NTSS	3.41	3.3	3.20
			FS-MPEG	3.37	3.33	3.29
			3DV-NTSS	3.5	3.41	3.37
	Fast-IProtect	View-based		3.53	3.45	3.37
		Disparity-based	NTSS	3.78	3.65	3.51
			FS-MPEG	3.51	3.78	3.66
			3DV-NTSS	3.85	3.75	3.66

Table A.3.3 95% error of the MOS for a valid number of 25 subjects.
MOS values given for image quality, depth perception and visual comfort, see Figure 5.9.

				Image quality	Depth perception	Visual comfort
3D Live	Original			0.288478	0.347311	0.263802
	SS	View-based		0.311670	0.263252	0.373072
		Disparity-based	NTSS	0.288478	0.311670	0.333251
			FS-MPEG	0.347311	0.333251	0.390911
			3DV-NTSS	0.288980	0.350221	0.366802
	2-QIM	View-based		0.263252	0.235951	0.350221
		Disparity-based	NTSS	0.230355	0.220054	0.288478
			FS-MPEG	0.270310	0.233481	0.288980
			3DV-NTSS	0.307926	0.230355	0.326660
	5-QIM	View-based		0.263252	0.201482	0.286967
		Disparity-based	NTSS	0.201482	0.201482	0.258808
			FS-MPEG	0.230355	0.235336	0.258808
			3DV-NTSS	0.197851	0.233481	0.288980
	Fast-IProtect	View-based		0.212684	0.243213	0.288478
Disparity-based		NTSS	0.261594	0.300298	0.214719	
		FS-MPEG	0.263252	0.280838	0.261594	
		3DV-NTSS	0.254857	0.233481	0.296410	
MPEG	Original			0.311670	0.307926	0.333251
	SS	View-based		0.311670	0.263252	0.373072
		Disparity-based	NTSS	0.202786	0.237288	0.275532
			FS-MPEG	0.204340	0.254028	0.264998
			3DV-NTSS	0.194830	0.194830	0.254028
	2-QIM	View-based		0.286967	0.185757	0.310271
		Disparity-based	NTSS	0.266189	0.262600	0.300780
			FS-MPEG	0.229150	0.262600	0.313146
			3DV-NTSS	0.202786	0.235951	0.258961
	5-QIM	View-based		0.284429	0.203629	0.254857
		Disparity-based	NTSS	0.201482	0.197851	0.235336
			FS-MPEG	0.258808	0.225906	0.276153
			3DV-NTSS	0.204340	0.233481	0.258808
	Fast-IProtect	View-based		0.261590	0.288478	0.230355
Disparity-based		NTSS	0.249684	0.284429	0.263252	
		FS-MPEG	0.235336	0.249684	0.225906	
		3DV-NTSS	0.179404	0.294939	0.192653	

Objective evaluation

Tables A.3.4 and A.3.5 give the average values of PSNR, IF, NCC, SC and SSIM and their 95% error limits, respectively.

Table A.3.4 Watermarking transparency: 3DV-NTSS vs. NTSS vs. FS-MPEG, see Figure 5.10.

				PSNR	IF	NCC	SC	SSIM
3DLive	SS	View-based		43.62	0.967	0.999	1.034	0.998
		Disparity-based	NTSS	43.82	0.979	0.999	1.046	0.999
			FS-MPEG	42.51	0.971	0.999	1.049	0.997
			3DV-NTSS	44.42	0.985	0.999	1.035	0.999
	2-QIM	View-based		30.96	0.893	0.917	0.986	0.915
		Disparity-based	NTSS	30.89	0.899	0.987	0.991	0.968
			FS-MPEG	30.99	0.899	0.987	0.991	0.969
			3DV-NTSS	31.57	0.985	0.991	0.995	0.975
	5-QIM	View-based		28.90	0.893	0.919	0.980	0.915
		Disparity-based	NTSS	32.36	0.989	0.985	0.987	0.981
			FS-MPEG	31.78	0.957	0.977	0.983	0.975
			3DV-NTSS	32.41	0.981	0.980	0.990	0.978
	Fast-IProtect	View-based		32.73	0.975	0.921	0.995	0.925
		Disparity-based	NTSS	33.30	0.985	0.989	0.996	0.984
			FS-MPEG	33.21	0.978	0.975	0.994	0.978
			3DV-NTSS	35.89	0.997	0.991	0.999	0.993
MPEG	SS	View-based		37.33	0.774	0.994	1.054	0.965
		Disparity-based	NTSS	38.47	0.849	0.997	1.038	0.968
			FS-MPEG	35.12	0.768	0.993	1.069	0.941
			3DV-NTSS	38.52	0.854	0.996	1.046	0.981
	2-QIM	View-based		32.15	0.955	0.910	0.981	0.869
		Disparity-based	NTSS	32.01	0.971	0.985	0.987	0.881
			FS-MPEG	31.78	0.897	0.977	0.983	0.880
			3DV-NTSS	32.56	0.981	0.987	0.990	0.895
	5-QIM	View-based		31.08	0.943	0.907	0.991	0.882
		Disparity-based	NTSS	32.21	0.989	0.897	0.991	0.881
			FS-MPEG	31.98	0.981	0.897	0.983	0.885
			3DV-NTSS	33.98	0.992	0.937	0.989	0.893
	Fast-IProtect	View-based		34.87	0.975	0.948	0.992	0.928
		Disparity-based	NTSS	36.54	0.997	0.98	0.998	0.961
			FS-MPEG	35.62	0.995	0.953	0.996	0.951
			3DV-NTSS	37.24	0.999	0.988	0.998	0.964

Table A.3.5 95% error in watermarked transparency: 3DV-NTSS vs. NTSS vs. FS-MPEG, see Figure 5.10.

				PSNR	IF	NCC	SC	SSIM
3DLive	SS	View-based		0.111519	0.000023	0.000020	0.000197	0.000214
		Disparity-based	NTSS	0.097677	0.000049	0.000020	0.000021	0.000010
			FS-MPEG	0.082060	0.000103	0.000010	0.000082	0.000010
			3DV-NTSS	0.068595	0.000175	0.000010	0.000055	0.000040
	2-QIM	View-based		0.070989	0.000600	0.000010	0.000185	0.000110
		Disparity-based	NTSS	0.089869	0.000076	0.000010	0.000052	0.000010
			FS-MPEG	0.074140	0.000268	0.000009	0.000916	0.000020
			3DV-NTSS	0.105730	0.000146	0.000010	0.000633	0.000020
	5-QIM	View-based		0.107414	0.000008	0.000108	0.000001	0.000352
		Disparity-based	NTSS	0.103359	0.000227	0.000009	0.000564	0.000020
			FS-MPEG	0.114411	0.000240	0.000009	0.000740	0.000020
			3DV-NTSS	0.098495	0.000720	0.000011	0.000290	0.000020
	Fast-IProtect	View-based		0.135111	0.000013	0.000066	0.000010	0.000125
		Disparity-based	NTSS	0.102422	0.000301	0.000010	0.000208	0.000060
			FS-MPEG	0.119970	0.000271	0.000010	0.000509	0.000020
			3DV-NTSS	0.082060	0.000103	0.000010	0.000080	0.000010
MPEG	SS	View-based		0.123043	0.000009	0.000006	0	0.000006
		Disparity-based	NTSS	0.085950	0.000233	0.000010	0.000889	0.000026
			FS-MPEG	0.063188	0.000093	0.000011	0.000780	0.000023
			3DV-NTSS	0.180423	0.000007	0.000007	0.000009	0.000029
	2-QIM	View-based		0.094204	0.000170	0.000006	0.000020	0.000007
		Disparity-based	NTSS	0.130989	0.000600	0.000017	0.000985	0.000001
			FS-MPEG	0.100156	0.000290	0.000007	0.000051	0.000007
			3DV-NTSS	0.120051	0.000316	0	0.000019	0.000041
	5-QIM	View-based		0.077566	0.000278	0.000011	0.000021	0.000034
		Disparity-based	NTSS	0.110104	0.000303	0.000003	0.000821	0.000024
			FS-MPEG	0.111227	0.000336	0.000011	0.000870	0.000014
			3DV-NTSS	0.127478	0.000296	0.000004	0.000798	0.000013
	Fast-IProtect	View-based		0.116524	0.000009	0.000065	0	0.000005
		Disparity-based	NTSS	0.157098	0.000286	0.000013	0.000217	0.000046
			FS-MPEG	0.098159	0.000330	0.000065	0.000050	0.000008
			3DV-NTSS	0.137452	0.000286	0.000013	0.0000320	0.000004

A.4 Robustness evaluation

Tables A.4.1 and A.4.2 give the average BER values obtained after applying five different attacks and their 95% error limits, respectively.

Table A.4.1 BER (Bit Error Ratio) in mark detection after several attacks.
Gaussian filtering (3×3 convolution kernel), Sharpening, (3×3 convolution kernel), JPEG compression (Quality factor $Q_i=60$), rotations (0.5) and geometric (StirMark random bending), see Figure 5.11.

			Gaussian filtering	Sharpening	JPEG compression Q=60	Rotation +0.5	Rotation -0.5	StirMark random bending
3DLive	SS	View-based	0.05	0.08	0.05	0.17	0.18	0.21
		Disparity-based	NTSS	0.05	0.08	0.02	0.14	0.11
			FS-MPEG	0.04	0.09	0.03	0.13	0.10
			3DV-NTSS	0.05	0.09	0.03	0.13	0.09
	2-QIM	View-based	0.05	0.07	0.04	0.09	0.09	0.12
		Disparity-based	NTSS	0.01	0.06	0.02	0.08	0.07
			FS-MPEG	0.02	0.10	0.01	0.08	0.09
			3DV-NTSS	0.01	0.06	0.01	0.07	0.07
	5-QIM	View-based	0.05	0.12	0.05	0.12	0.11	0.15
		Disparity-based	NTSS	0.03	0.10	0.03	0.12	0.13
			FS-MPEG	0.02	0.09	0.02	0.11	0.17
			3DV-NTSS	0.02	0.09	0.03	0.13	0.11
	Fast-IProtect	View-based	0.03	0.05	0.02	0.08	0.08	0.08
		Disparity-based	NTSS	0.02	0.01	0.02	0.01	0.04
			FS-MPEG	0.00	0.01	0.02	0.02	0.04
			3DV-NTSS	0.00	0.00	0.01	0.01	0.02
MPEG	SS	View-based	0.05	0.11	0.03	0.19	0.19	0.21
		Disparity-based	NTSS	0.04	0.07	0.03	0.11	0.12
			FS-MPEG	0.07	0.11	0.02	0.13	0.14
			3DV-NTSS	0.04	0.08	0.03	0.10	0.10
	2-QIM	View-based	0.04	0.05	0.03	0.11	0.11	0.13
		Disparity-based	NTSS	0.04	0.09	0.03	0.12	0.11
			FS-MPEG	0.03	0.11	0.04	0.11	0.09
			3DV-NTSS	0.01	0.09	0.03	0.07	0.08
	5-QIM	View-based	0.04	0.13	0.05	0.07	0.13	0.13
		Disparity-based	NTSS	0.02	0.08	0.01	0.07	0.10
			FS-MPEG	0.02	0.09	0.01	0.08	0.11
			3DV-NTSS	0.01	0.08	0.00	0.07	0.09
	Fast-IProtect	View-based	0.03	0.04	0.02	0.10	0.10	0.11
		Disparity-based	NTSS	0.00	0.04	0.01	0.08	0.07
			FS-MPEG	0.01	0.04	0.02	0.08	0.09
			3DV-NTSS	0.00	0.02	0.01	0.05	0.05

Table A.4.2 95% error of BER (Bit Error Ratio) in mark detection after several attacks.
Gaussian filtering and sharpening (3×3 convolution kernel), JPEG compression (Quality factor $Q_f=60$), rotations (0.5) and (StirMark random bending), see Figure 5.11.

				Gaussian filtering	Sharpening	JPEG compression Q=60	Rotation +0.5	Rotation -0.5	StirMark random bending
3DLive	SS	View-based		0.002265	0.000130	0.000076	0.009144	0.011731	0.010214
		Disparity-based	NTSS	0.002536	0.000180	0.000100	0.010100	0.000534	0.011588
			FS-MPEG	0.000233	0.002919	0.003356	0.011697	0.000116	0.012898
			3DV-NTSS	0.002362	0.000231	0.000135	0.012292	0.000692	0
	2-QIM	View-based		0.000283	0.000257	0.000160	0.008259	0	0.010252
		Disparity-based	NTSS	0	0.002916	0.003525	0.007674	0	0.010607
			FS-MPEG	0.001904	0	0.000102	0.010314	0	0.008185
			3DV-NTSS	0	0.000068	0.000035	0.009170	0	0.008367
	5-QIM	View-based		0.002630	0.003387	0.000234	0.001668	0.000918	0
		Disparity-based	NTSS	0.000231	0.000144	0.003717	0.008285	0.000135	0.000545
			FS-MPEG	0.000137	0.000059	0.000350	0.006655	0.000176	0.006319
			3DV-NTSS	0.000036	0.000054	0.000080	0.000135	0.000110	0.000075
	Fast-IProtect	View-based		0	0.003236	0.003723	0.005175	0.000093	0.009796
		Disparity-based	NTSS	0.000038	0.000101	0.000100	0.004382	0.000239	0.008645
			FS-MPEG	0	0	0	0	0.000265	0.006598
			3DV-NTSS	0	0	0	0.005990	0.000036	0.006986
MPEG	SS	View-based		0.000105	0.002723	0.000364	0.000080	0.001001	0.000266
		Disparity-based	NTSS	0.000108	0.000119	0.000343	0.000294	0.000175	0.007008
			FS-MPEG	0.001276	0.000049	0.000143	0.005064	0	0.006238
			3DV-NTSS	0.000107	0.000103	0.000235	0.001002	0.001004	0.004059
	2-QIM	View-based		0.000119	0.000232	0.000158	0.007315	0.000896	0.000199
		Disparity-based	NTSS	0.003126	0.003336	0.000564	0.001382	0	0.007018
			FS-MPEG	0.002111	0	0.003517	0.001636	0	0.008146
			3DV-NTSS	0	0.000361	0.000228	0.000232	0.000035	0.000176
	5-QIM	View-based		0.000020	0.000363	0.003241	0.000063	0.000363	0.000279
		Disparity-based	NTSS	0.000151	0.000365	0.003578	0.000831	0.000924	0.000232
			FS-MPEG	0.000013	0.000052	0.003743	0.005076	0.000639	0.008543
			3DV-NTSS	0	0.000131	0	0.000066	0.000113	0.005010
	Fast-IProtect	View-based		0.000293	0.000387	0.000725	0.000135	0.000100	0.000085
		Disparity-based	NTSS	0.001225	0.003323	0.000035	0.000056	0.000103	0.000106
			FS-MPEG	0	0.000167	0.000030	0.006093	0.000462	0.000150
			3DV-NTSS	0	0.000100	0	0	0.000427	0.006253

References

- [3DF13] 3D Film Factory, "Cameras Rigs, 3D Production, 3D Training", (2013).
<http://www.3dfilmfactory.com-3D>
- [3DI12] <http://3dlive-project.com>.
- [Ala03] Alattar, A. M., Lin, E. T., Celik, M. U., "Watermarking low bit-rate advanced simple profile MPEG-4 bitstreams", IEEE International Conference on Acoustics, Speech, and Signal Processing, Proc.(ICASSP'03), Vol. 3, pp. III-513, (2003).
- [Bar00] M. Barni, F. Bartolini, N. Checcacci, V. Cappellini, "Object watermarking for MPEG4 video streams copyright protection in Security and Watermarking of Multimedia Contents", Proc. SPIE Vol. 3971, (2000).
- [Bar01a] Bartolini F., Tefas A., Barni M., Pitas I., "Image authentication techniques for surveillance applications", IEEE Proc., Vol. 89 No. 10, pp. 1403-1418, (2001).
- [Bar01b] Barni M., Bartolini F., Piva A. "Improved wavelet-based watermarking through pixel-wise masking," IEEE Trans on Image Processing, Vol.10 No.5, pp. 783-791, (2001).
- [Bel10] Belhaj M., Mitrea M., Duta S., Prêteux F., "MPEG-4 AVC robust video watermarking based on QIM and perceptual masking", Int. Conf. on Communication, Bucharest, (2010).
- [Bha09] Bhatnager G., Kumar S., Raman B., Sukavanam N., "Stereo coding via digital watermarking", Journal of Electronic Imaging, Vol. 18 No. 3, (2009).
- [Bis95] Bishop C. M., Neural Networks for Pattern Recognition, Oxford University Press, Oxford, (1995).
- [Ble08] Blender Foundation / www.bigbuckbunny.org, copyright 2008, (2008).
- [Blo99] Bloom J., Cox I., Kalker T., Linnartz J., Miller M., Traw C., "Copy protection for dvd video", IEEE Proc., Vol. 87 No. 7, pp. 1267-1276, (1999).
- [BT00] ITU-R, "Subjective assessment of stereoscopic television pictures", Tech. Rep. BT.1438, (2000).
- [BT02] ITU-R, "Methodology for the subjective assessment of the quality of television pictures", Tech.Rep. BT.500-11, ITU-R, (2002).
- [BT98] ITU-R, "Subjective assessment methods for image quality in high-definition television", Tech. Rep. BT.710-4, ITU-R, (1998).
- [Cha10] Chammem A., Mitrea M., Prêteux F., "Speeding-up the hybrid video watermarking techniques in the DWT domain", Proc. SPIE Electronic Imaging (2010), Vol. 6845, (2010).

- [Cha11] Chammem A., Mitrea M., Prêteux F., "DWT based stereoscopic image watermarking", Proc. SPIE, Vol. 7863 No. 26, (2011).
- [Che12] Chen H., Zhu Y., "An Effective Stereo Image Coding Method with Reversible Watermarking", Journal of Computational Information Systems, Vol. 8 No.7, pp. 2761-2768, (2012).
- [Chi11] Chikkerur S., Sundaram V., Reisslein M., Karam L. J., "Objective video quality assessment method: A Classification, Review, and Performance Comparison", School of Electrical, Computer & Energy Engineering, Journal & Magazines, Vol. 57 No. 2, pp. 165-182, (2011).
- [Col07] Coltuc D., "On stereo embedding by reversible watermarking", IEEE Int. Symposium on Signals, Circuits and Systems, Vol. 2, (2007).
- [Cos83] Costa M. H. M., "Writing on dirty paper", IEEE Trans. on Information Theory, Vol. 29 No. 3, pp. 439-441, (1983).
- [Cox02] Cox I., Miller M., Bloom J., Digital watermarking, Morgan Kaufmann, (2002).
- [Cox07] Cox, I., Miller, M., Bloom, J., Fridrich, J., & Kalker, T., Digital watermarking and steganography. Morgan Kaufmann, (2007).
- [Cox97] Cox I., Klian .J., Leighton F.T., Shamoon T., "Secure Spread Spectrum watermarking for multimedia", IEEE Trans. on Image Processing, Vol. 6 No. 12, (1997).
- [Dem77] Dempster A. P., Laird N. M., Rubin D.B., "Maximum Likelihood from Incomplete Data via the EM algorithm", J. Roy. Stat. Soc. (B), pp. 38-39, (1977).
- [Dig03] Digimarc Corporation, "Digimarc Comments to USPTO Regarding Technological Protection Systemd for Digitized Copyrighted Works", (2003).
- [Din06] Ding L., Chien S., Chen L., "Joint prediction algorithm and architecture for stereo video hybrid coding systems", IEEE Trans. on Circuits Syst. Video Technol., Vol.16 No. 11, pp. 1324-1337, (2006).
- [Dod05] Dodgson N. A., "Autostereoscopic 3D displays", Computer, Vol. 38 No. 8, pp. 31-36, (2005).
- [Don03a] Dong-Choon H., Kyung-Hoon B., Eun-Soo K., "Stereo image watermarking scheme based-on discrete wavelet transform and adaptive disparity estimation", Mathematics of data/image coding, compression, and encryption, with applications. Conf. No. 6, pp. 196-205, (2003).
- [Don03b] Dong-Choon H., Kyung-Hoon B., Eun-Soo K. , "Real-time stereo image watermarking using discrete cosine transform and adaptive disparity maps", Proc. SPIE, Multimedia Systems and Applications VI, Vol. 5241 No. 233, (2003).
- [Dum08] Dumitru O., Mitrea M., Prêteux F., "Video Modeling in the DWT Domain", Proc. SPIE, Vol. 7000, (2008).

- [Egg03] Eggers J. J., Bäuml R., Tzschoppe R., Girod B., "Scalar Costa scheme for information embedding", IEEE Trans. on Signal Processing, Vol. 51 No. 4, (2003).
- [Ell09] Ellinas J. N., "Reversible Watermarking on Stereo Image Sequences", Int. journal of Signal processing, (2009).
- [Esk95] Eskicioglu M. A., Fisher P. S., "Image Quality Measures and their performance", IEEE Trans. on Communications, Vol. 43 No. 12, (1995).
- [Feh03] Fehn C., "A 3D-TV approach using depth-image-based rendering (DIBR)", Proc. of IASTED Conf. on Visualization, Imaging and Image Processing, pp. 482-487, (2003).
- [Fet09] Fetita, C., Ortner, M., Brillet, P.-Y., Preteux, F. and Grenier, P., "A morphological-aggregative approach for 3D segmentation of pulmonary airways from generic MSCT acquisitions", Second International Workshop on Pulmonary Image Analysis, MICCAI 09, pp.215-226 (2009).
- [Fra02] François, A. R., Medioni, G. G., & Waupotitsch, R., "Reconstructing mirror symmetric scenes from a single view using 2-view stereo geometry". IEEE Proc. of the 16th Int. Conf. In Pattern Recognition, Vol. 4, pp. 12-16, (2002).
- [Fra11] Franco-Contreras J., Baudry S., Doërr G., "Virtual view invariant domain for 3D video blind watermarking", IEEE Int. Conf. on Image Processing, (2011).
- [Gal11] Galiano V., Lopez O., Malumbres M. P., Migallon H., "Improving the discrete wavelet transform computation from multicore to GPU-based algorithms", Proc. of the 11th Int. Conf. on Computational and Mathematical Methods in Science and Engineering, (2011).
- [Gol10] Goldmann L., De Simone F., Ebrahimi T., "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video", Proc. SPIE, Vol. 7526, (2010).
- [Gre04] Grenier P., Beigelman-Aubry, C., Fetita, C., Martin-Bouyer, Y., "Multi-Detector-row CT of the airways in U.J. Schoepf (Ed.)", Multi-detector-Row CT of the Thorax, Springer-Verlag, pp.63-80, (2004).
- [Gro01] Grosbois R., Ebrahimi T., "Watermarking in the JPEG 2000 domain". IEEE 4th Workshop In Multimedia Signal Processing, pp. 339-344, (2001).
- [Hal05] Halle M., "Auto-stereoscopic displays and computer graphics", ACM SIGGRAPH 2005 Courses. ACM, (2005).
- [Har00] Hartung F., Ramme F., "Digital rights management and watermarking of multimedia content for m-commerce applications", IEEE Communications Magazine, Vol. 38 No. 11, pp. 78-84, (2000).
- [Has11] Hasnaoui M., Belhaj M., Mitrea M., Prêteux F., "mQIM principles for MPEG-4 AVC watermarking", SPIE Photonics West, (2011).

- [Hoi11] Hoikkala S., "Mobile3DTVcontent delivery optimization over DVB-H system", (2011). <http://sp.cs.tut.fi/mobile3dtv/stereo-video/>.
- [Hol00] M. Holliman, W. Macy, and M. Y. Yeung, "Robust frame dependent video watermarking," in Security and Watermarking of Multimedia Contents, Proc. SPIE Vol. 3971, (2000).
- [Jai81] Jain J. R., "Displacement measurement and its application in inter-frame image coding", IEEE Trans. Communication, Vol. 29 No. 12, pp. 1799-1808, (1981).
- [Jia06] Jiang G, Yu M., Shao F., Yang Y., Dong H., "Fast multi-view disparity estimation for multi-view video systems", in 8th Int. Conf. Advanced Concepts for Intelligent Vision Systems, ACIVS, Vol. 4179, pp. 493-500, Springer, (2006).
- [Kal99] Kalker T., Depovere G., Haitsma J., Maes M., "A video watermarking system for broadcast monitoring", Proc. SPIE in Security and Watermarking of Multimedia Contents, Vol. 3657, pp. 103-112, (1999).
- [Kaw02] Kawai T., "3D displays and applications", Displays, Vol. 23 No. 1, pp. 49-56, (2002).
- [Kin11] King P., "Global 3D Enabled Device Sales Forecast", Strategy Analytics, (2011). <http://www.strategyanalytics.com>
- [Koc10] Kockoth B., "White paper for digital Stereo System', media systems, (2010).
- [Kog81] Koga T., "Motion compensated inter-frame coding for video conferencing", Proc. in National Telecommunications Conf., NTC '81, pp.G5.3.1–G5.3.5, (1981).
- [Kua99] Kuang K., Hosur P. I., "Core experimental results of fast motion estimation based on new test conditions (Q4a)", ISO/IEC JTC1/SC29/WG11/M4934, (1999).
- [Kum09] Kumar S., Raman B., Thakur M., "Real coded genetic algorithm based stereo image watermarking", IJSDA Int. Journal of Secure Digital Information Age, Vol. 1 No. 1, (2009).
- [Kun11] Kunic S., Sego Z., "3D television", Proc. IEEE ELMAR, (2011).
- [Lat38] Lateltin H. "Stereoscope", U.S. Patent No. 2, pp. 131-444, (27 Sep. 1938).
- [Lee02] Lee J.J., Cho N.I., Kim J.W., "Watermarking for 3D NURBS graphic data". In Proc of the IEEE Workshop on Multimedia Signal Processing, pp. 304-307, (2002).
- [Lee12] Lee J. W., Kim H. D., Choi H. Y., Choi S. H., Lee H. K., "Stereoscopic watermarking by horizontal noise mean shifting.", SPIE Society of Photo-Optical Instrumentation Engineers Conf. Series. Vol. 8303, (2012).
- [Lev01] Levy K. L., Hiatt R. S., Rhoads G. B. (2001). U.S. Patent Application 10/017,679.

- [Li94] Li R., Zeng B., Liou M. L., "A new three-step search algorithm for block motion estimation", IEEE Trans. on Circuits Syst. Video Technol., Vol. 4 No. 4, pp. 438-442, (1994).
- [Lin00] Lin E. T., Podilchuk C. I., Delp E. J., "Detection of image alterations using semi-fragile watermarks", Int. Society for Optics and Photonics In Electronic Imaging, pp. 152-163, (2000).
- [Lin83] Lin S., Costello D. J. Jr., Error Control Coding: Fundamentals and Applications, Prentice-Hall, (1983).
- [May81] Mayhew J. E., Frisby J. P., "Psychophysical and computational studies towards a theory of human stereopsis". Artificial Intelligence, Vol. 17 No. 1, pp. 349-385, (1981).
- [Mil04] Miller L., Doërr G., Cox I., "Applying Informed Coding and Embedding to Design a Robust High-Capacity Watermark", IEEE Trans. on Image Processing, Vol. 13 No. 6, (2004).
- [Mit05] Mitrea M., Prêteux F., Petrescu M., Vlad A., "The StirMark Watermarking Attack in the DWT Domain", Proc. of the 12th Int. Workshop on Systems, Signals and Image Processing, Vol. 2, (2005).
- [Mit05-07] Mitrea M., Prêteux F., Nunez J., (for SFR and GET), "Procédé de tatouage d'une séquence video", French patent no. 05 54132 (29/12/2005), and EU patent no. 1804213 (04/07/2007).
- [Mit06] Mitrea M., Duta S., Prêteux F., "A unified approach to multimedia content watermarking", Proc. of 3rd Taiwanese-French Conf. on Information Technology (TFIT'2006), pp. 275-289, (2006).
- [Mit07a] Mitrea M., Prêteux F., "Tatouage robuste des contenus multimédias", in H. Chaouchi, M. Laurent-Maknavicius, La sécurité dans les réseaux sans fil et mobiles, Lavoisier, Paris (2007).
- [Mit07b] Mitrea M., Dumitru O., Prêteux F., Vlad A., "Zero Memory Information Sources Approximating to Video Watermarking Attacks", Lecture Notes in Computer Science, Vol. 4705, pp. 409-423, (2007).
- [Nik98]. Nikolaidis N., Pitas I., "Robust image watermarking in the spatial domain". Signal Processing, Vol. 66 No. 3, pp. 385-403, (1998).
- [Nta02] Ntalianis K. S., Doulamis N. D., Doulamis A. D., Kollias S. D., "Automatic stereoscopic video object-based watermarking using qualified significant wavelet trees", Int. Conf. Of Consumer Electronics, IEEE Int. Conf. on Digest of Technical Papers, pp. 188-189, (2002).
- [Ped13] Peddie J., "Stereo 3D - content market value 2010-2014", Statista, (2013).
<http://www.statista.com>
- [Pet00] Petitcolas F., "Watermarking schemes evaluation", IEEE Signal Processing, Vol. 17 No. 5, pp. 58-64, (2000).
- [Pet98] Petitcolas F., Ross J. A., Markus G. K., "Attacks on copyright marking systems", Proc. on Information Hiding, Second Int. Workshop, IH'98, , LNCS 1525, Springer-Verlag, ISBN 3-540-65386-4, pp. 219-239, (1998)

- [Pog84] Poggio, G. F., & Poggio, T., "The analysis of stereopsis". Annual review of neuro-science, Vol. 7 No. 1, pp. 379-412, (1984).
- [Pot05] Potdar V. M., Han S., Chang E., "A survey of digital image watermarking techniques", 3rd IEEE International Conference on Industrial Informatics, INDIN'05, pp. 709-716, (2005)
- [Rog82] Rogers B., Graham M., "Similarities between motion parallax and stereopsis in human depth perception". Vision research, Vol. 22 No. 2, pp. 261-270, (1982).
- [Saf11] Safonova A., "Cost of filming 3D Movie vs. 2D movie", False Greek white Paper, (2011).
<http://falsecreekproductions.com>
- [Sam09] Samtani, R., "Ongoing innovation in digital watermarking." Computer 42, No. 3, pp. 92-94, (2009).
- [Sch02] Scharstein D., Szeliski R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", Int. journal of Computer Vision, Vol. 47 No. 1-3, pp. 7-42, (2002).
<http://vision.middlebury.edu>
- [Sha02] Sharma R. K., Decker S., "Practical challenges for digital watermarking applications", EURASIP Journal on Applied Signal Processing, Vol.1, pp. 133-139, (2002).
- [Sha58] Shannon C.E., "Channel with Side Information at the Transmitter", IBM Journal, pp. 289-293, (1958).
- [Sri85] Srinivasan R., Rao K. R., "Predictive coding based on efficient motion estimation", IEEE Trans. Commun. Vol. 33 No. 8, pp. 888-896, (1985).
- [Sta08] Stankiewicz O., Wegner K., "Depth map estimation software", ISO/IEC JTC1/SC29/WG11 MPEG/M15175, Antalya, (2008).
- [Sub06] Subramanya S. R., Byung K. Yi, "Digital rights management", IEEE Potentials, Vol. 25 No. 2, pp. 31-34, (2006).
- [Tan09] Tanimoto M., Fujii T., Suzuki K., "Depth Estimation Reference Software (DERS) with Image Segmentation and Block Matching", ISO/IEC JTC1/SC29/WG11 M16092, Lausanne, (2009).
- [Teu10] Teulade V., Bothun D. K., "3D Here and Now... a goose that lays a golden egg?", PricewaterhouseCoopers Communications, Entertainment & Media and Technology Practice Leaders.
http://www.pwc.de/de_DE/de/technologie-medien-und-telekommunikation/assets/3d-studie-2010.pdf
- [Wad87] Wade N.J. "On the late invention of the stereoscope", Perception, Vol.16 No. 6, pp. 785-818, (1987).
- [Wal02] Walpole R. E., Myers R. H., Myers S. L., Ye K., "Probability & Statistics for Engineers and Scientists", Pearson Educational Int., (2002).

- [Wan04] Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P., "Image quality assessment: from error visibility to structural similarity", *IEEE Trans. on Image Processing*, Vol. 13, No. 4, (2004).
- [Wan11] Wang H. Y., Chung H.Y., "Design and implementation of dynamic target tracking with stereo vision system", in *Proc. of the Int. MultiConf. of Engineers and Computer Scientist, IMECS*, (2011). Int. Association of Engineers, (2011).
- [Whe12] Wheatley M., "3D TV Predicted To Spur 3-D Revenue Growth", (2012).
<http://www.hdtvtest.co.uk/news/3d-tv-revenue-growth-201210012242.htm>
- [Wol99] Wolfgang R. B., Delp E.J., "Fragile watermarking using the VW2D watermark", *Electronic Imaging'99*, International Society for Optics and Photonics, (1999).
- [Xia97] Xia X. G., Boncelet C. G., Arce G. R., A, "multi-resolution watermark for digital images", *Proc. IEEE. International Conference on In Image Processing*, Vol. 1, pp. 548-551, (1997).
- [Xu05] Xu X., Tomlinson M., Ambroze M., Ahmed M., "Techniques to provide robust and high capacity data hiding of id badges for increased security requirement", *3rd International Conference: Science of Electronics Technologies of Information and Telecommunication*, (2005).
- [Yan06] Yang Y. L., Cai N., Ni G. Q., "Digital image Scambling Technology Based on the Symmetry of Arnold transform", *Journal of Beijing Institute of Technology*, Vol. 15 No. 2, pp. 216-219, (2006).
- [Yeo99] B-L Yeo, and M. Yeung, "Watermarking 3D object for verification", in *IEEE Computer Graphics and Applications*, Vol. 19, No. 1, pp. 36-45, (1999).
- [You03] Kim Y. W., Moon K. A., Oh I. S., "A Text Watermarking Algorithm based on Word Classification and Inter-word Space Statistics", *IEEE Proc. of the 7th Int. Conf. on Document Analysis and Recognition*, (2003).
- [Yu11] Yu M., Wang A., Luo T., Jiang G., Li F., Fu S. "New Block-Relationships Based Stereo Image Watermarking Algorithm", in *ICSNC 2011, the 6th Int. Conf. on Systems and Networks Communications*, pp. 171-174, (2011).
- [Zen07] Zhang Z., Zhu Z., Xi L., "Novel Scheme for Watermarking Stereo Video", *Int. Journal of Nonlinear Science*, Vol. 3 No. 1, pp.74-80, (2007).
- [Zhu00] Zhu S., Kuang K., "A new diamond search algorithm for fast block matching motion estimation", *IEEE Trans. on Image Process.*, Vol. 9 No. 2, pp. 287-290, (2000).

List of acronyms

3DV-NTSS	3 Dimension Video New Three Step Search
4SS	Four Step Search
bpp	bit per pixel
cdf	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CDS	Conjugate Directional Search
DAM	Digital Asset Management
DCT	Discrete Cosine Transform
DfFrFT	Discrete Fractional Fourier Transform
DIBR	Depth Image Based Rendering
DRM	Digital Rights Management
DSCQS	Double Stimulus Continuous Quality Scale
DSIS	Double Stimulus Impairment Scale
DWT	Discrete Wavelet Transform
EM	Expectation-Maximization
ES	Exhaustive Search
FMLR	Frequency Model Laplacian Removal
FS-MPEG	Full Search MPEG
HD 3D TV	High Definition 3 Dimension Television
HH	High High
HL	High Low
HSV	Human Visual System
IF	Image Fidelity
IMAX	Image Maximum

ISBN	International Standard Book Number
ITU	International Telecommunication Union
JPEG	Joint Photographic Experts Group
LCD	Liquid crystal Display
LDPC	Low Density Parity Check
LH	Low High
LL	Low Low
MC	Monte Carlo
MOS	Mean Opinion Scores-
MSE	Mean Squared Error
NCC	Normalized Cross Correlation
NTSS	New Three Step Search
NURBS	Non-uniform rational basis spline
pdf	Probability Density Function
PI	Performance Index
PSNR	Signal to Noise Ratio
PSNRr	Signal to Noise Ratio recovered
PSNRw	Signal to Noise Ratio watermarked
PVD	Pixel Value Difference
QIM	Quantization Index Modulation-
QSWT	Qualified Significant Wavelet Tree
RAM	Random Access Memory
SAD	Sum of Absolute Differences
SC	Structural Content
SI	Side Information

SS	Spread Spectrum
SSCQE	Single Stimulus for Continuous Quality Evaluation
SSIM	Structural SIMilarity
SVD	singular value decomposition
TSS	Three Step Search
USD	United States dollar
USSR	Union of Soviet Socialist Republics
WDR	Watermark Detection Ratio
WRR	Watermark Recovering Rates

List of publications

Journal

- ▶ A. Chammem, M. Mitrea, F. Prêteux, "High-definition three-dimensional television disparity map computation", Journal of Electronic Imaging, vol 21, no 4, pp 043024-043024, (December 2012).
- ▶ A. Chammem, M. Mitrea, F. Prêteux, "Stereoscopic video watermarking: a comparative study", Annals of telecommunications, special issue on 3D video technologies and Services, guest editors: Béatrice Pesquet-Popescu, Touradj Ebrahimi, Shipeng Li, Frédéric Dufaux, DOI: 10.1007/s12243-013-0384-5, (2013).

Book chapter

- ▶ M. Mitrea, A. Chammem, F. Prêteux, "Tatouage stéréoscopique", Chapter 13 in "Vidéo et TVHD 3D ... de la capture à la diffusion Principe, tendances et perspectives", Laurent LUCAS, Céline LOSCOS et Yannick REMION Editors, Lavoisier, pp 251-271, Paris 2013.

International conferences

- ▶ A. Chammem, M. Mitrea, F. Prêteux, "Speeding-up the hybrid video watermarking techniques in the DWT domain", Proc. SPIE, vol. 6845, pp. 75350E-75358E, January 2010.
- ▶ A. Chammem, M. Mitrea, F. Prêteux, "DWT-based stereoscopic image watermarking, " Proc. SPIE, vol. 7863, pp. 786326-786334, January 2011.
- ▶ A. Chammem, M. Mitrea, F. Prêteux, "Adaptive disparity map computation for stereoscopic video watermarking", IEEE International Conference on Electronics, Circuits, and Systems (ICECS), Seville, Spain, December 2012.

Other contribution

- ▶ A. Chammem, M. Mitrea, F. Prêteux, "Tatouage stéréoscopique : du développement méthodologique au protocole d'évaluation", Journée d'étude Vidéo 3D SI2D (Image/Transmission/Compression), January 2012.

R&D projects deliverables (contribution under the supervision of M. Mitrea)

- ▶ HD3D2 project (April 2010 – October 2011)

D6.1.2 Sélection des méthodes de tatouage relief en conformité avec le CDC (October 2010)

D6.1.3 Briques logicielles pour le tatouage relief (May 2011)

D6.1.4 Méthode originale de tatouage relief (October 2011)

► **3DLive project (January 2010 – March 2012)**

D4.4.2 Premier logiciel de tatouage stéréo non temps-réel (December 2010)

D4.4.3 Logiciel hybride fingerprint/tatouage temps réel et spécifications d'une approche tatouage hybride et contextuelle (December 2011)