Optimization of the Rate-Distortion Compromise for Stereoscopic Image Coding using Joint Entropy-Distortion Metric

Aysha-Khatoon Kadaikar

UNIVERSITÉ PARIS 13

Galilée Ecole doctorale

THESIS

in candidacy for the degree of

DOCTOR OF UNIVERSITY PARIS 13

Specialty **: Signals and Images**

defended by

Aysha KADAIKAR

on November 20, 2017

Optimization of the Rate-Distortion Compromise for Stereoscopic Image Coding using Joint Entropy-Distortion Metric

Director : Anissa MOKRAOUI
Co-Supervisor : Gabriel DAUPHIN

JURY

Marco CAGNAZZO	*Associate professor, TELECOM ParisTech*	**Reviewer**
François-Xavier COUDOUX	*Professor, Université de Valenciennes*	**Reviewer**
Maria TROCAN	*Professor, ISEP*	**Examinor**
Anissa MOKRAOUI	*Professor, Université Paris 13*	**Examinor**
Gabriel DAUPHIN	*Assistant professor, Université Paris 13*	**Examinor**

Abstract

Optimization of the Rate-Distortion Compromise for Stereoscopic Image Coding using Joint Entropy-Distortion Metric

During the last decades, a wide range of applications using stereoscopic technology has emerged still offering an increased immersion to the users such as video games with auto-stereoscopic displays, 3D-TV or stereo visio-conferencing. The raise of these applications requires fast processing and efficient compression techniques. In particular, stereoscopic images require twice the amount of information needed to transmit or store them in comparison with 2D images as they are composed of two views of the same scene. The contributions of our work are in the field of stereoscopic image compression and more precisely, we get interested in the improvement of the disparity map estimation.

Generally, disparities are selected by minimizing a distortion metric which is sometimes subjected to a smoothness constraint, assuming that a smooth disparity map needs a smaller bitrate to be encoded. But a smoother disparity map does not always reduce significantly the bitrate needed to encode it but can increase the distortion of the predicted view. Therefore, the first algorithm we have proposed minimizes a joint entropy-distortion metric to select the disparities. At each step of the algorithm, the bitrate of the final disparity map is estimated and included in the metric to minimize. Moreover, this algorithm relies on a tree where a fixed number of paths are extended at each depth of the tree, ensuring good rate-distortion performance. In the second part of the work, we have proposed a sub-optimal solution with a smaller computational complexity by considering an initial solution -the one minimizing the distortion of the predicted view- which is successively modified as long as an improvement is observed in terms of rate-distortion.

Then, we have studied how to take advantages of large search areas in which the disparities are selected as one can easily supposed that enlarging the search area will increase the distortion performance as there will be more choices of disparities. In the other hand, the larger is the range of the selected disparities, the higher is supposed to be the cost of the disparity map in terms of bitrate. We have proposed two approaches allowing to take advantage of a large search area by selecting only sets of disparities belonging to it enabling to achieve a given bitrate while minimizing the distortion of the predicted image.

The last part of the work concerns variable block sizes which undeniably allows to improve the bitrate-distortion performance as the block size suits to the image features. We have thus proposed a novel algorithm which jointly estimates and optimizes the disparity and the block length maps.

Keywords: optimization, stereoscopic images, compression, bitrate, distortion, entropy.

Résumé

Optimisation du compromis débit-distorsion pour le codage d'images stéréoscopiques utilisant une métrique conjointe entropie-distortion

Ces dernières décennies ont vu émerger de nombreuses applications utilisant la technologie 3D, telles que les écrans auto-stéréoscopiques, les écrans de télévisions 3D ou encore la visio-conférence stéréoscopique. Ces applications requièrent des techniques adaptées afin que leur flux de données soit compressé efficacement. En particulier, dans le cas des images stéréoscopiques, ces dernières étant composées de deux vues de la même scène, elles nécessitent à ce titre deux fois plus d'informations à transmettre ou à stocker que dans le cas des images 2D traditionnelles. Nos travaux se situent dans le cadre de la compression des images stéréoscopiques. Plus précisément, ils concernent l'amélioration de l'estimation de la carte de disparité dans le but d'obtenir un meilleur compromis entre le débit binaire nécessaire au codage de la carte de disparité et la qualité de l'image prédite.

Généralement, la carte de disparité est estimée en minimisant la distorsion de l'image prédite. Cette minimisation peut être sujette à une contrainte de lissage. L'idée étant qu'une carte de disparité plus lisse nécessitera un débit binaire moindre en supposant que les mêmes vecteurs de disparités seront sélectionnés plus souvent. Néanmoins cette contrainte de lissage ne permet pas toujours de diminuer le coût binaire de la carte. Le lissage peut entraîner par ailleurs une augmentation notable de la distorsion de l'image prédite. Dans le premier chapitre de la thèse, nous présentons un algorithme d'estimation de carte de disparité minimisant une métrique conjointe entropie-distorsion. Le coût binaire finale de la carte de disparité est estimée à chaque étape de l'algorithme et est intégré dans le calcul de la métrique. La distortion globale de la carte de disparité est aussi mise à jour au fur et à mesure du traitement de l'image. Par ailleurs, cette algorithme repose sur la construction séquentiel d'un arbre dont on ne garde qu'un nombre défini de branches à chaque profondeur de l'arbre. Ainsi, l'algorithme développé apporte une solution sous-optimale en minimisant le coût binaire de la carte de disparité tout en assurant une bonne qualité de l'image prédite. Le chapitre deux étend l'algorithme précédent au cas des images non rectifiées.
Dans le troisième chapitre, nous nous intéressons au fait de trouver une solution au problème d'optimisation du compromis débit-distorsion en réduisant la complexité numérique par rapport à l'algorithme précédent. De ce fait, nous avons développé le R-algorithme qui se base sur une solution initiale de Référence (celle minimisant la distorsion de l'image prédite) et la modifie successivement tant qu'une amélioration est constatée en termes de compromis débit-distorsion.
Le quatrième chapitre s'intéresse toujours au fait d'accroître les performances de l'algorithme développé tout en réduisant le coût en complexité numérique et donc en temps de traitement. Nous proposons deux approches afin de tirer profit d'un grand espace de recherche sans avoir à tester pour chaque bloc à apparier l'ensemble des disparités qui composent cet espace de

recherche. En effet, un espace de recherche plus grand permet plus de choix de disparités et donc potentiellement une meilleur reconstruction de l'image prédite. En contrepartie, il se peut que le coût binaire de la carte de disparité augmente si l'ensemble des disparités sélectionnées constituent un ensemble plus divers qu'auparavant. Les deux approches proposées permettent de restreindre l'espace de recherche à un ensemble composées de certaines disparités permettant de minimiser la distorsion de l'image prédite pour un débit donné.
Le dernier chapitre de la thèse s'intéresse à l'utilisation des blocs de taille variable pour la compression des images stéréoscopiques. Elle permet une meilleure description de l'image prédite. Cependant l'utilisation des blocs de tailles variables induit un nouveau coût : celui du signalement de la taille de chaque bloc. Nous proposons un nouvel algorithme d'estimation de carte de disparités optimisant conjointement les disparités choisies ainsi que la taille de chaque bloc.

Mots-clés : optimisation, images stéréoscopiques, compression, débit, distorsion, entropie.

Acknolegments

First of all, I would like to heartily thank my director Pr. Anissa Mokraoui and my co-supervisor Dr. Gabriel Dauphin for all they bring to me during this thesis personally and professionally. I feel that they are more than just my thesis supervisors but rather they are models to follow in all aspects of life. It's a great chance for me to have worked with them. Beyond all they did for me, I would like to thank both of them for their constant encouragement, their availability and their help every time I needed it.

I am also very grateful to my thesis committee members Pr. François-Xavier COUDOUX, Pr. Maria TROCAN and Dr. Marco Cagnazzo for their precious feedback on my work which helps improving it.

I would like to express all my gratefulness to my husband without who none of this would have been possible. Again thank you for always supporting me and helping me in all what I do.

A great thank you for my parents for their eternal support and encouragement and also to all my family for having looked at my children day after day as long as I needed to write this thesis.

My appreciation also goes to all the members of the Laboratoire de Traitement et Transport de l'Information (L2TI) with who I spent good time during work and breaks.

Finally, I address a special thank to my two loving babies Sakina and Ayan that came into my life during this PhD thesis. They have brought me so much joy and taught me to put all situations into perspective.

Contents

Notations **16**

1 Introduction **1**

1.1 Thesis context 1

1.2 Objectives of the work and contributions 2

1.3 Thesis outlines 4

1.4 Publications 5

2 State-of-the-art on stereo image coding **7**

2.1 Basic concepts of stereoscopic vision 7

2.1.1 Principles of stereo vision 7

2.1.2 Geometrical models for stereoscopic imaging 9

2.1.2.1 The pinhole camera model 9

2.1.2.2 Stereoscopic imaging systems representation 9

2.1.2.3 Epipolar geometry and rectification 10

2.1.3 Disparity estimation for 3D reconstruction 11

2.1.3.1 Local methods 12

2.1.3.2 Global methods 13

2.2 Overview on stereo image coding 13

2.2.1 Disparity compensated stereo image coding scheme 14

2.2.2 Standard for stereoscopic videos compression 15

2.2.3 Disparity estimation for stereoscopic image compression 16

2.3 Conclusion 18

3 Disparity map estimation algorithm using joint entropy-distortion metric: MMA **19**

3.1 Stereo-matching optimization problem 19

3.1.1 Assumptions and notations 20

3.1.2 Rate-distortion optimization problem formalization 21

3.1.3 Properties of the global cost function 22

3.2 Presentation of the modified M-algorithm (MMA) 23

3.2.1 Sequential MMA based on a tree structure 23

3.2.2 Entropy-distortion metric computation 24

3.2.3 Settings parameters in the MMA 25

3.3 Experimental results and discussions 28

3.3.1 Choice of the parameter β 28

3.3.2 Influence of the number of retained paths M on the performance 29

3.3.3 Comparison with the Block-Matching Algorithm 30

3.3.4 Comparison with the BMA including a smoothness constraint 33

3.3.5 Comparison of all algorithms using DPCM coding 36

3.3.6 Performance of the MMA estimating dense disparity map 38

3.3.7 MMA processing time and complexity 39

3.4 Conclusion 39

4 Extension of the MMA to non-rectified stereoscopic images: EMMA **51**

4.1 Notations and assumptions 51

4.2 Presentation of the extended MMA 52

4.3 Simulation results and discussions 53

4.3.1 Simulation results on non-rectified stereoscopic images 53

4.3.1.1 Impact of the window size choice 53

4.3.1.2 Performance evaluation of the EMMA 54

4.3.2 Simulation results on rectified stereoscopic images 55

4.3.2.1 Impact of the window size choice 56

4.3.2.2 Performance evaluation of the EMMA 56

4.4 Conclusion 57

5 Reference-based block matching algorithm: R-algorithm **59**

5.1 Rate-distortion optimization problem 59

5.1.1 Assumptions and notations 59

5.1.2 Formulation of the Rate-Distortion optimization problem 60

5.2 Description of the proposed R-algorithm 61

5.2.1 Raster scanning notations 61

5.2.2 Blockwise disparity map based on the joint entropy-distortion criterion . . 61

5.2.3 Entropy and distortion recursive equations 62

5.3 Performance evaluation 63

5.3.1 Rate-distortion performance of the R-algorithm 63

5.3.2 Processing time and complexity of the R-algorithm 67

5.4 Conclusion 68

6 Improving the BMA performance based on a selection of disparities sets: BMA_H and BMA_S **69**

6.1 Selection of disparities sets in a given search area 69

6.1.1 Notations and assumptions 70

6.1.2 Problem statement 70

6.2 Proposed sub-optimal BMA_H and BMA_S algorithms 71

6.2.1 Computational complexity of the optimal solution 72

6.2.2 Selecting the best set of disparities using BMA_H algorithm 72

6.2.3 Selecting the best set of disparities using BMA_S algorithm 73

6.3 Analysis and discussion of the simulation results 73

6.3.1 Comparing BMA_H, BMA_S and the optimal solution 75

6.3.2 Rate-distortion performance of sub-optimal algorithms 75

6.3.3 Comparing the predicted right view for a given bitrate 76

6.3.4 Performance evaluation using Bjøntegaard metric 76

6.3.5 Processing time and complexity of the proposed algorithms 79

6.4 Conclusion 80

7 Joint disparity and block-length maps optimization algorithm: JDBLMO 81

7.1 Rate-Distortion optimization problem 82

7.1.1 Notations and assumptions 82

7.1.2 Coding of the block layout 82

7.1.3 Formulation of the rate-distortion problem 84

7.2 Joint disparity and block-length maps optimization algorithm (JDBLMO) 86

7.2.1 Initialization process 86

7.2.2 Adapation of the R-Algorithm (ARA) 88

7.2.3 Block Division Algorithm (BDA) 89

7.3 Performance evaluation and discussions 90

7.3.1 JDBLMO performance evaluation compared to BMA and MIIBSC 91

7.3.2 JDBLMO average rate-distortion performance compared to BMA and MI-IBSC 92

7.3.3 JDBLMO rate-distortion performance compared to MMA and R-algorithm 94

7.3.4 Processing time and complexity of the JDBLMO 96

7.4 Conclusion 96

8 Conclusion and future research 97

Bibliography 101

List of Figures

2.1 Stereoscopic vision 8

2.2 Stereoscopic camera 8

2.3 Pinhole camera model. 9

2.4 Stereoscopic imaging system. 10

2.5 Epipolar geometry. 11

2.6 Stereoscopic image rectification. 11

2.7 Stereo-matching in area-based methods. 12

2.8 Open Loop structure. 14

2.9 Closed Loop structure. 15

3.1 MMA notations. 20

3.2 MMA: Entropy-distortion optimization algorithm 25

3.3 Stereoscopic matching optimization algorithm with $M = 2$ and $N_1 = 5$. 26

3.4 Selection of the block sizes in the MMA 27

3.5 Tree-structure to explore the block sizes. 27

3.6 MMA performance according to the parameter β. 29

3.7 Influence of the parameter M on the MMA performnce using "Barn2" with a block size 1×1; $M = 1$ and $M = 8$. 31

3.8 Influence of the parameter M on the MMA performance using "Wood2" with a block size 1×1; $M = 1$ and $M = 8$. 31

3.9 Comparison of the MMA and BMA performance on "Tsukuba" using block sizes from 1×1 to 3×3 for MMA and 1×1 to 5×5 for BMA; and bitrates belong to $[0, 5.8]$bpp. 40

3.10 Comparison of the MMA and BMA performance on "Tsukuba" using block sizes from 4×4 to 7×7 for MMA and 4×4 to 10×10 for BMA; and bitrates belong to $[0, 0.27]$bpp. 41

3.11 Comparison of the MMA and BMA performance on "Tsukuba" using block sizes from 8×8 to 11×11 for MMA and 8×8 to 17×17 for BMA; and bitrates belong to $[0, 0.06]$bpp. 41

3.12 Comparison of the MMA and BMA performance on "Tsukuba" using blocksizes from 12×12 to 16×16 for MMA and 12×12 to 24×24 for BMA; and bitrates belong to $[0, 0.025]$bpp. 42

3.13 Original right image "Tsukuba". 42

3.14 Reconstruction using BMA (30.18dB; 0.022bpp). 43

3.15 Reconstruction using the proposed algorithm (30.81dB; 0.022bpp) with $\lambda = 9.9$. . 43

3.16 Reconstruction using BMA (34.12dB; 0.27bpp). 43

3.17 Reconstruction using MMA algorithm (33.89dB; 0.14bpp) with $\lambda = 1.8$. 44

3.18 Histogram of the disparity map based on the BMA (34.12dB; 0.27bpp). 44

3.19 Histogram of the disparity map based on the MMA algorithm (33.89dB; 0.14bpp) with $\lambda = 1.8$. 44

3.20 Disparity map using BMA (34.12dB; 0.27bpp). 45

3.21 Disparity map using the MMA algorithm (33.89dB; 0.14bpp) with $\lambda = 1.8$. . . . 45

3.22 Comparison of MMA and BMA performance using "Barn2" with block sizes from 1×1 to 3×3 for MMA and 1×1 to 5×5 for BMA. 46

3.23 Comparison of MMA and BMA-RA performance on "Wood2". 46

3.24 Comparison of the MMA-DPCM, BMA-DPCM and BMA-RA-DPCM performance on Wood2. 47

3.25 From left to right close-up views of the original right image "Barn2", reconstructed image using the BMA; and the reconstructed image using MMA. 48

3.26 Original right image. 49

3.27 Reconstructed image using dynamic programming (23.15dB; 3.05bpp). 49

3.28 Reconstructed image using the MMA (33.97dB; 2.88bpp). 49

3.29 Disparity map estimated using dynamic programming (23.15dB; 3.05bpp). 50

3.30 Disparity distribution using the MMA (23.81dB; 0.87bpp). 50

3.31 MMA processing time according to target bitrates. 50

4.1 EMMA notations. 52
4.2 BMA performance on "Sand" using different sliding matching window sizes. . . . 54
4.3 Rate-distortion optimization on the non-rectified stereo image "Sand". 55
4.4 BMA performance on "Tsukuba" using different sliding matching window sizes. . 56
4.5 Rate-distortion optimization on the rectified stereo image "Tsukuba". 57

5.1 Proposed optimization R-algorithm. 63
5.2 Rate-distortion performance on "Tsukuba". 64
5.3 BMA blockwise disparity map histogram. 66
5.4 R-algorithm blockwise disparity map histogram. 66
5.5 "Tsukuba" original right image. 67
5.6 Close-up views of the reconstructed image using the BMA (left figure) and the R-algorithm (right figure). 67
5.7 Rate-distortion performance on "Stereo_13". 68

6.1 BMA_H sub-optimal algorithm. 72
6.2 BMA_S sub-optimal algorithm. 74
6.3 RD performance on a small stereoscopic image extracted from "sand". 76
6.4 RD performance on the stereoscopic image "house2". 77
6.5 "house2" original left view (at left) and original right view (at right). 77
6.6 Close-up of the predicted right view using the BMA (left figure), the BMA_H (middle figure) and the BMA_S (right figure). 77
6.7 Best RD points under bitrate constraints. 78

7.1 Example of a block-length map and its corresponding block layout. 83
7.2 Tree structure associated with the block layout shown in Figure 7.1. 84
7.3 Description of the JDBLMO algorithm 87
7.4 Rate-distortion performance of the JDBLMO, MIIBSC and BMA performed on "rubik". 92
7.5 Above: original right view "rubik". Below, from left to right close-up views extracted from the three predicted views processed with the JDBLMO, MIIBSC and BMA. 93

7.6 Block-length map resulting from the JDBLMO 93

7.7 Block-length map resulting from the MIIBSC 94

7.8 Rate-distortion performance of the MMA, R-algorithm, JDBLMO and BMA on "stereo_13". 95

List of Tables

3.1 Comparison of rate-distortion performance of BMA, BMA-RA and MMA. 36

3.2 Comparison of rate-distortion performance of BMA, BMA-RA and MMA using DPCM. 38

5.1 Comparison of rate-distortion performance between the BMA and the R-algorithm. 65

6.1 Average PSNR gain of the BMA_S over the BMA and the BMA_H 79

7.1 Average PSNR gain of the JDBLMO over the BMA and the MIIBSC 95

Notations

I_L	Left view of the stereoscopic image
I_R	Right view of the stereoscopic image
$\widehat{I}_R$	Predicted right view
$I_{\tilde{L}}$	Left interpolated view
$K \times L$	Size of the stereoscopic image given in pixels
$X \times Y$	Size of the stereoscopic image given in blocks
T	Total number of blocks
$N_X \times N_Y$	Size of one block given in pixels
(i_P, j_P)	Position of the block in the pixel coordinates system
(i_B, j_B)	Position of the block in the block coordinates system
$d(i_B, j_B)$	Disparity of the block at position (i_B, j_B)
$\mathbf{d}$	Disparity map
W	Sliding matching window containing disparities
$W_X \times W_Y$	Size of the sliding matching window W
w	A disparity inside the matching window W
w_{min}	The minimum disparity inside the matching window W
w_{max}	The maximum disparity inside the matching window W
(w_x, w_y)	A two-dimensions disparity inside the matching window W
w_{x_max}	Maximum value of the vertical component w_x of the disparity (w_x, w_y)
w_{y_max}	Maximum value of the horizontal component w_y of the disparity (w_x, w_y)
α	Precision of the disparity
N_W	Number of disparities inside the matching window W
N_α	Number of disparities inside the matching window W depending on the disparities precision α, i.e. including non-integer disparities
$E_G(\mathbf{d})$	Global distortion of the predicted view associated to the map $\mathbf{d}$
E_B	Local distortion of a block
E_b	Local distortion of a branch in the MMA

E_t	Global distortion of a partial disparity map
$H(\mathbf{d})$	Entropy of the map $\mathbf{d}$
$J(\lambda, \mathbf{d})$	Cost function
λ	Lagrange multiplier of the cost function
λ_{max}	Maximum value of λ
M	Number of paths extended at each depth in the MMA
$\mathbf{d}_t^k$	Partial disparity map containing t disparities, corresponding to the k-th path in the MMA
d_t^k	Disparity of the k-path at the t-th depth (i.e. of the t-th block) in the MMA
k	Index of the path in the MMA
t	One dimension block-coordinate
p_M	Estimate of the disparity distribution probability given by a mixture distribution
p_U	Uniform distribution probability
p_E	Empirical disparity distribution probability
β	Parameter of the MMA
T_{bpp}	Target bitrate
s	Size of block
s^+	Size of block just superior to the size s
s^-	Size of block just inferior to the size s
Reg	Regularization constraint
$V(\mathbf{d}, w)$	Number of occurence of the disparity w in the disparity map $\mathbf{d}$
$\mathbf{d}^R$	Reference blockwise disparity map
d_t^R	Reference disparity of the t-th block
$\mathbf{d}^t(w)$	Disparity map where the reference disparity at the t-th block has been replaced by w
S_n	Set of n disparities selected inside W
$\mathrm{Co}(D)$	Set composed of all different disparities in the disparity map D
$D(S_n)$	Disparity map estimated by the BMA using the set S_n of disparities
$\mathcal{P}(W)$	All possible sets of disparities that can be made from W
W_0	Set of disparities composed of the disparities effectively setected by the BMA when it is applied using all the disparities in W
B	A block
$\mathbf{B}$	Map containing all the blocks of the image
$\mathbf{l}$	Block length map
l_B	Length of a square block
L_{max}	Maximum length of a block

L_{min}	Minimum length of a block
d_B	Disparity of the block B
B_f	Block at the front of the queue
$\mathbf{B}_t$	Tail of the queue

ARA	*Adaptation of the R-Algorihtm*
BDA	*Block Division Algorihtm*
BMA	*Block Matching Algorihtm*
$BMA-DPCM$	*Block Matching Algorihtm* using DPCM coding
BMA_H	*Block Matching Algorihtm based on Histograms*
$BMA-RA$	*Block Matching Algorihtm* with a regularization constraint
$BMA-RA-DPCM$	*Block Matching Algorihtm* with a regularization constraint and using DPCM coding
BMA_S	*Block Matching Algorihtm based on Sets of disparities*
CMU	Stereoscopic image dataset
$CMU-VASC$	Stereoscopic image dataset
DCT	*Discret Cosine Transform*
$DPCM$	*Differential Pulse-Code Modulation*
$EMMA$	*Extended Modified M-Algorihtm*
$JDBLMO$	*Joint Disparity and Block Length Maps Optimization*
MMA	*Modified M-Algorihtm*
$MMA-DPCM$	*Modified M-Algorihtm* using DPCM coding
$PSNR$	*Peak Signal to Noise Ratio*
RD	Rate-Distortion
SSD	*Sum of Squared Differences*

Chapter 1

Introduction

1.1 Thesis context

Three dimensions (3D) technology offers to the viewer a great immersive experience so they can have a more realistic representation of the scene that is viewed. From the past years, the number of applications using 3D-technology have been increasing in many fields. We can cite for exemple 3D visio-conferencing system [1]. In the field of medicine, the Da Vinci robot allows the surgeon to operate more precisely by providing him with a 3D high-definition vision of the part of the body to operate [2]. In the entertainment field also, new technologies have been developed such as 3DTV, 3D cinema, 3D video games using virtual reality [3]. Indeed, they provide the viewer a more pleasant experience.

A 3D/stereoscopic image is composed of two views acquired from two cameras capturing the same scene from two slightly different points of views. By making the left eye looking at the left view and the right eye looking at the right view simultaneously, the viewer can perceive depth in the scene. As these images are composed of two views, if no precaution is taken, they will require twice the amount of information needed to encode them for transmission or storage purpose compared to traditional two-dimensional images. The interest of the public for three-dimension contents and the rapid growth of applications using 3D images have made the issue of efficiently coding them taking importance.

As stereoscopic images are composed of two views of the same scene, rather than encoding them independently, in the literature, the proposed algorithms usually take advantage of the redundancies between the two views by estimating the spatial displacement between them which is called disparity. They generally proceed according to the following scheme: (a) first, one of the two views is taken as the reference one. Let's say the left one which is encoded independently; (b) then the second view (the right one) is predicted from the reference view by estimating a disparity map; (c) a residual image is computed between the original right view and its prediction. The disparity map is usually encoded using an entropy coder while the left view and the residual image

are encoded using transforms such as DCT (Discrete Cosine Transform) [4, 5, 6]. They are all sent to the decoder which reconstructs the right view using the left view and the disparity map, compensated with the residual image. This coding scheme is known as the disparity compensated scheme because of its ressemblance with the motion compensated scheme developed in video codec [7].

The disparity map estimation is therefore of primary importance in the coding process as a good disparity map in terms of distortion improves the quality of the predicted view but can be very expensive in terms of bitrate. It should be noted that a large number of research aims at estimating the true disparity map for three-dimension reconstruction purpose. In fact, knowing the true disparity, one can easily recover the depth in the scene using also the parameters of the camera used for the image acquisition. Our work is concerned with the issue of efficiently encoding stereoscopic images. So, we do not necessarily estimate true disparity values. For the considered compression purpose, a good disparity map is the one achieving the best reconstruction of the predicted view at a given bitrate.

1.2 Objectives of the work and contributions

The objective of this thesis is to develop efficient algorithms for compressing stereoscopic images. In particular, we are interested in the estimation of disparity maps improving the optimization of the trade-off between the bitrate needed to encode them and the distortion induced by these maps when predicting one of the two views. Another point to consider in this thesis is the computational complexity of the proposed solutions. In fact, let's consider for example a small image of size 64×64 and a searching window of size 3×61 (representing all possible disparities for each pixel of the image to be predicted). Using an exhaustive search, finding the optimal solution in terms of Rate-Distortion requires to compare about 10^{4100} different disparity maps.

This thesis aims at finding sub-optimal solutions to the Rate-Distortion optimization problem with a reduced complexity of the developed algorithms. Here is a list of our main contributions.

Usually the disparities of the map are estimated by minimizing a distortion metric. Sometimes this minimization is subjected to a smoothness constraint thanks to a coefficient used to give more or less weight to the constraint. A smoother disparity map is assumed to give a disparity map with a smaller bitrate as same or close values of disparities are more often chosen locally. It is generally done at the expense of an increase in the distortion of the predicted view. The reduction of the bitrate of the disparity map is not a proper criteria present in the metric to minimize. We propose a novel joint entropy-distortion metric to minimize taking directly into account the bitrate of the disparity map at the time of selecting the disparities. The metric is composed of a traditional term measuring the distortion of the predicted view and a weighted

term approximating the final bitrate of the disparity map by its entropy at each step of the map computation for each choices of disparities. Doing this allows to minimize not only the distortion of the predicted image but also the bitrate of the computed disparity map. Furthermore, the entropy of the disparity map is computed from the beginning of the map estimation without knowing the complete distribution of the disparities in the final map. We propose in one of the developed algorithm to approximate the entropy of the final disparity map thanks to a finite mixture probabilities. These probabilities rely on assumptions we made about the disparity map distribution which become more and more close to the true disparity distribution as the image is processed.

As mentioned earlier, finding the disparity map optimizing the bitrate-distortion compromise is a complex combinatorial problem. We propose to restrain the large amount of disparity maps to explore according to two approaches: in the first one, the disparity map is estimated sequentially block by block by constructing a tree where each path corresponds to a (partial) disparity map being explored. At each step of the tree, only the M-best paths are extended allowing to explore several paths and keep good rate-distortion performance. The second approach consists in considering an initial solution which is the one minimizing the distortion of the predicted image. Then this solution is successively modified as long as improvements are observed in terms of rate-distortion performance. This also allows to explore solutions close to the initial one minimizing the distortion but with a lower bitrate.

Another of our contributions concerns the exploitation of large search areas in which the disparities are selected. The use of larger search area usually increases the quality of the predicted view thanks to more adequate choices of disparities. But it can also induce an increase in the bitrate needed to encode the disparity map as the range of the selected disparities gets larger. Another point is that testing all the disparities inside a larger search area for all the blocks to match increases the processing time of the map estimation algorithms. We propose two methods to take advantage of large search areas without having to test all choices of disparities for each block to match. Two criteria are defined to select relevant sets of disparities inside the search area. They are then used in the disparity map estimation process. Each of the selected sets allows to minimize the distortion of the predicted image at a given bitrate.

Our last contribution concerns the use of variable block sizes for stereoscopic image compression. We have proposed a novel algorithm which jointly optimizes the choices of disparities as well as the division of the blocks. Using blocks of variable sizes allows to better suit the disparities to the image features. As for example, one disparity can be used to encode a large zone with roughly the same depth. But this is done at the expense of extra bitrate needed to encode the blocks sizes. The proposed algorithm estimates a disparity map such as to minimize the distortion of the predicted image but also the bitrate needed to encode the disparities and the block size map.

1.3 Thesis outlines

This thesis is organized as follows. Chapter 2 first presents some basic concepts about stereoscopic vision, then we give an overview of the methods used for disparity map estimation, first in the case of three-dimension reconstruction purpose and then techniques used for compression purpose are described with more details.

Chapter 3 describes the Rate-Distortion optimization problem. Then we present the developed M-algorithm bringing a sub-optimal solution to the optimization problem which rely on a joint entropy-distortion metric. As this method implies many setting parameters, we have included a discussion on the choices of the parameters values to obtain best performance with our algorithm. Finally experimental results are provided to compare the developed algorithm with conventional algorithms including or not a smoothing constraint. Chapter 4 extends the method developed in the previous chapter to non rectified stereoscopic images. We also discuss the impact of the search area size on the distortion performance using several images. These works have been presented in [8, 9, 10, 11].

In Chapter 5, the purpose still consists in finding the best disparity map optimizing the Rate-Distortion trade-off. We explore different disparity maps with a new approach. It consists in considering an initial solution (which is the one minimizing the distortion of the predicted view) and to modify it as long as improvements are observed. A detailed presentation of the method is given completed with simulation results to evaluate the method. This algorithm shows to perform better than the block-matching algorithm and also than the previous developed M-algorithm. These results have been presented in [12, 13].

Because of the complexity of the problem of finding the optimal solution to the Rate-Distortion optimization problem, in Chapter 6, we propose to consider rather the problem of selecting inside a search area the best sets of disparities that improves the Rate-Distortion trade-off. Doing this allows to take advantage of large search areas without having to test for each block to match all possible disparities in the search area. Two methods allowing to select relevant sets of disparities belonging to the search area are presented. Each of these sets minimizes the distortion of the predicted view at a given bitrate. The two mentioned methods have been presented in [14].

To still ameliorate the performance of the developed algorithms, in Chapter 7 we consider the use of variable block sizes as they can reduce the bitrate needed to encode the disparity map. But this is done at the expense of encoding the size of each block. We describe the proposed JDBLMO algorithm which jointly optimize the disparity as well as the block length maps while minimizing the distortion of the predicted view to select the disparities. This algorithm has been depicted in [15].

Finally, Chapter 8 concludes our work and possible futur investigations are addressed.

Note that the Rate-Distortion optimization problem is reminded at the beginning of each chapter to allow a better understanding of the specific issues raised in each of the chapters.

1.4 Publications

International journal papers

1. A. Kadaikar, G. Dauphin, A. Mokraoui, Sequential Block-Based Disparity Map Estimation Algorithm for Stereoscopic Image Coding. Elsevier Signal Processing : image communication, September 2015, DOI 10.1016/j.image.2015.09.007.

2. A. Kadaikar, G. Dauphin, A. Mokraoui, Joint Disparity and Variable Size-Block Optimization Algorithm for Stereoscopic Image Compression. Elsevier Signal Processing : image communication, 2017 (under revision).

International conference papers

3. A. Kadaikar, A. Mokraoui , G. Dauphin, Entropy-constrained Dense Disparity Map Estimation Algorithm for Stereoscopic Images, European Signal Processing Conference, EUSIPCO, 1-5 September 2014, Lisbon Portugal.

4. A. Kadaikar, G. Dauphin and A. Mokraoui, Extended Disparity Map Estimation Algorithm Using Joint Entropy-Distortion Metric for Non-Rectified Stereoscopic Images, Symposium on Signal Processing and Information Technology, ISSPIT, December 7-10, 2015, Abu Dhabi, UAE.

5. A. Kadaikar, G. Dauphin and A. Mokraoui, Modified Block Matching Algorithm Improving Rate-Distortion Performance for Stereoscopic Image Coding, Symposium on Signal Processing and Information Technology, ISSPIT, December 7-10, 2015, Abu Dhabi, UAE.

6. A. Kadaikar, G. Dauphin and A. Mokraoui, Improving block-matching algorithm by selecting disparity sets minimizing distortion for stereoscopic image coding, European Workshop on Visual Information Processing, EUVIP, October 25-27, 2016, Marseille, France.

National conference papers

7. A. Kadaikar, G. Dauphin, A. Mokraoui, Estimation de cartes de disparité basée sur l'algorithme revisité d'appariement de blocs pour la compression d'images stéréoscopiques, Proceedings of the GRETSI Conferences, 8-11 septembre 2015, Lyon, France.

8. A. Kadaikar, G. Dauphin, A. Mokraoui, Méthode d'optimisation pour l'appariement de pixels d'images stéréoscopiques basée sur une métrique conjointe entropie-distorsion, Compression et Représentation des Signaux Audiovisuels, CORESA, 26-27 Novembre 2014, Reims.

Award

Second price for the best paper at CORESA 2014.

Chapter 2

State-of-the-art on stereo image coding

Due to the growing number of applications using stereoscopic images, it has become important to find an efficient way of encoding them. They are composed of two images corresponding to a left and a right view of the same scene, and thus requires twice the amount of information needed to be encoded in comparison to traditional two-dimensional images. In this chapter, we recall some basic concepts that are used in our work concerning stereoscopic images, and then we give an overview of main techniques that have been deployed to encode them.

2.1 Basic concepts of stereoscopic vision

2.1.1 Principles of stereo vision

The term stereopsis is commonly used to refer to the three-dimensional perception resulting from the human visual system. Our left and right eyes each capture a different view of the same scene. These two views are sent to the brain for processing which merged them into one single image where we can better perceive the depth. This stereoscopic vision is represented in Fig. 2.1.

Stereoscopic systems have tried to reproduce this depth perception: basically, they are composed of two cameras placed close to each other in the scene. They capture the same scene but from two points of view slightly different as represented in Fig. 2.2. When the left image is viewed by the left eye and the right image is viewed by the right eye simultaneously, the viewer can perceive the depth in the scene and see the objects in 3D.

Some additional cues are used by the human visual system to analyze the depth in the scene. Amongst them, we can cite:

- the linear perspective: parallel lines converge with the distance;
- the texture gradient: the closer is an object from the viewer, the more its details can be seen;

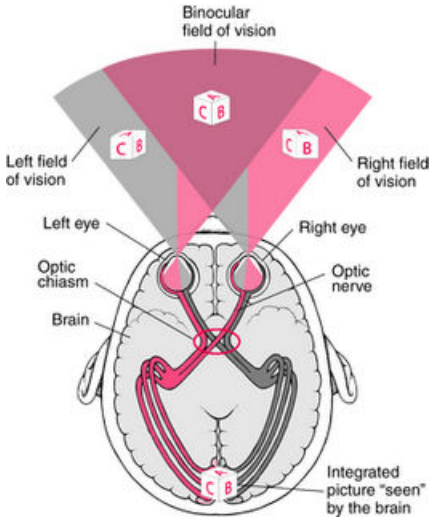

Figure 2.1: Stereoscopic vision

Figure 2.2: Stereoscopic camera

- the motion parallax: moving objects that are closer to the viewer will seem to move faster than objects that are far;

- the relative size: having an idea on the size of two objects or two people in an image, one can guess how long is the distance between them;

- the superposition: when one object partially hides the view of another object, we can guess that the first one is in front of the second one;

- the shadow: the shadow of an object can give a cue for the brain on the position and the depth of the object in the scene.

2.1.2 Geometrical models for stereoscopic imaging

Before discussing about stereoscopic image coding, we first introduce some geometrical models to better understand some notions of the image coding.

2.1.2.1 The pinhole camera model

The acquisition of an image using the pinhole camera model is shown in Fig. 2.3. It is considered as the simplest camera model but which can still accurately approximate the geometry and optics of modern cameras [16]. This model is composed of a retinal plane, a focal plane and an image plane. It represents the relation between a 3D object in the scene and its 2D projection into the image plane: this mapping from 3D to 2D is called the perspective projection. Considering the

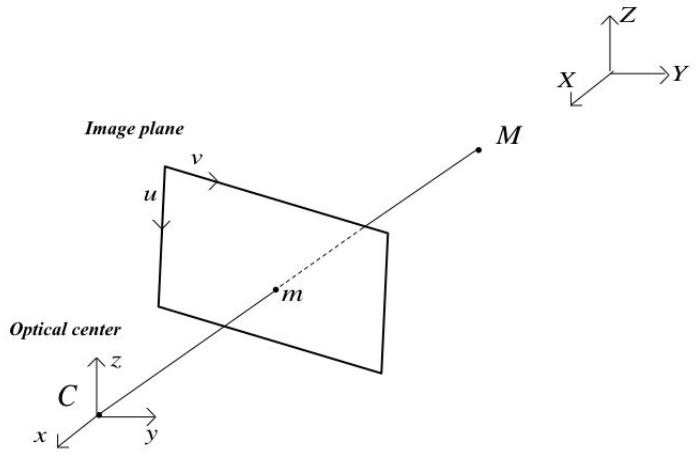

Figure 2.3: Pinhole camera model.

3D point M in the scene, a relationship can be established with this point and its projection m in the image plane given by :

$$m = P \times M. \tag{2.1}$$

where P is the perspective projection matrix of the camera.

2.1.2.2 Stereoscopic imaging systems representation

Using the pinhole camera model, a stereoscopic imaging system can be represented as it is composed of two cameras capturing the same scene from two points of view slightly different as shown in Fig. 2.4. The left and right cameras are represented by their optical centers $\mathbf{C}_l$ and $\mathbf{C}_r$ and their corresponding perspective projection matrices $\mathbf{P}_l$ and $\mathbf{P}_r$ respectively. Given a 3D point in the scene, the relation of this point with its two projections m_l and m_r on the image

planes of the two cameras can be expressed as:

$$\begin{aligned} m_l &= P_l \times M \\ m_r &= P_r \times M. \end{aligned} \tag{2.2}$$

These two projections of the same 3D point on the image planes are called homologous points.

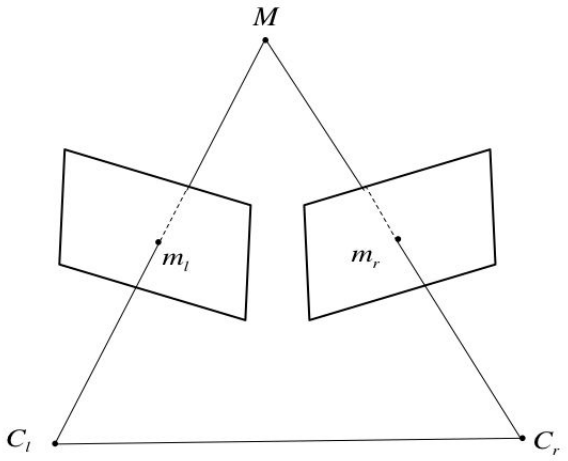

Figure 2.4: Stereoscopic imaging system.

2.1.2.3 Epipolar geometry and rectification

The relation between the two images can be represented by the so-called epipolar geometry as shown in Fig. 2.5. The plane which passes through the object point M and the optical centers $\mathbf{C}_l$ and $\mathbf{C}_r$ of the cameras is called the epipolar plane. The two points which are the intersections of the line joining the optical centers with the two image planes are denoted $\mathbf{e}_l$ and $\mathbf{e}_r$ respectively and called the epipoles. Finally, the lines joining the epipole $\mathbf{e}_l$ (resp. $\mathbf{e}_r$) with the projection $\mathbf{m}_l$ (resp. $\mathbf{m}_r$) in the image plane is called the left (resp. right) epipolar line. This geometry implies that knowing the projection point m_l on the left epipolar line, its corresponding point m_r must lye on the conjugate right epipolar line. This constraint can be expressed as:

$${m_r}^T F m_l = 0. \tag{2.3}$$

where F is the fundamental matrix depending on the parameters of the cameras that can be found as in [17, 18].

Basically, two configurations of cameras are used for stereoscopic image acquisition. In the first one, as shown previously in Fig. 2.5, the two cameras are placed side by side and rotated towards each other. In this case, the epipolar lines are inclined. In the second configuration shown in Fig. 2.6, the two cameras are still side by side but with parallel optical axes. In this case,

the epipolar lines are parallel also. As they coincide with the images scan lines, the homologous points can be found on the two images at the same line. If we are in the first configuration, the epipolar lines can be made parralel by a process called rectification. This is made by finding a transformation which projects the two images onto a common plane. Different methods for rectifications can be found in [19, 20, 21].

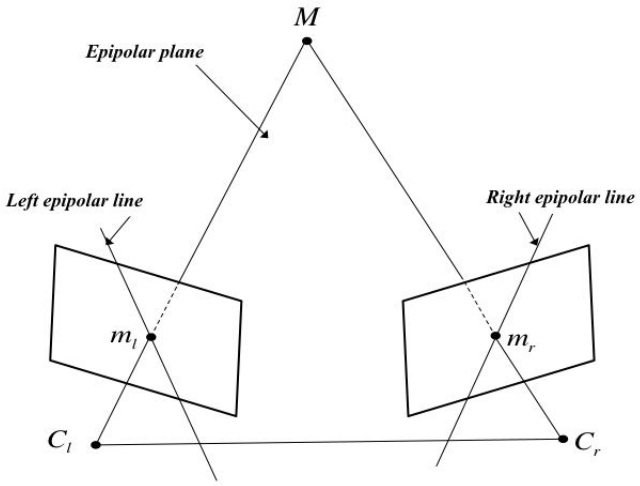

Figure 2.5: Epipolar geometry.

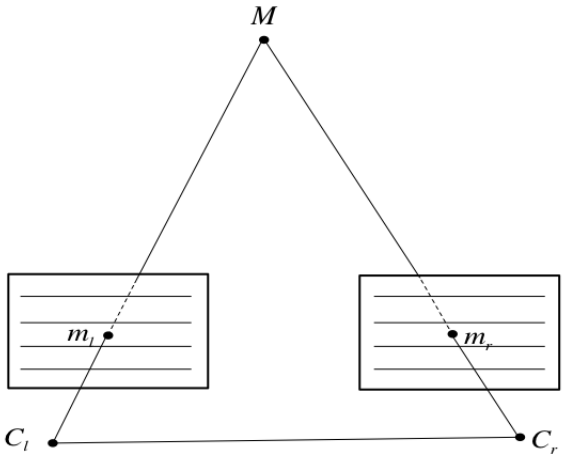

Figure 2.6: Stereoscopic image rectification.

2.1.3 Disparity estimation for 3D reconstruction

Many works have been proposed in the field of stereo-matching with the particular goal of reconstructing the scene in three dimensions. In fact, matching two homologous pixels results

in a disparity vector corresponding to the relative spatial displacement between the pixels in the two views. Performing this stereo-matching for every pixel leads to a dense disparity map. Then, the depth in the scene can be deduced from the disparity as they are proportional. For this purpose, several methods have been proposed that can mainly be put into two categories described below: the local and the global methods.

2.1.3.1 Local methods

Local methods compare the similarity between two sets of pixels. Amongst them, we can distinguish two main types of methods: the feature-based methods and the area-based methods.

In the first category of methods, the sets of pixels are compared using different features of the images like edges [22], segments [23] or curves [24]. These methods have the advantages that features are less sensitive to the noise than pixels. What's more, they are fewer in numbers and so more easily distinguishable than pixels. But their main drawbacks is that these methods leads to sparse disparity maps and thus need a backward step of interpolation to complete the disparity map. Furthermore, they also need an additional step before the stereo-matching to compute the features of the images which increase the complexity of the proposed methods.

The second category corresponds to the area-based methods. The idea consists in defining a window around the pixel to be matched and comparing it to the same window surrounding each possible candidate pixel inside a defined searched area in the other view. This is represented in Fig 2.7. These kinds of methods allow to compute directly a dense disparity

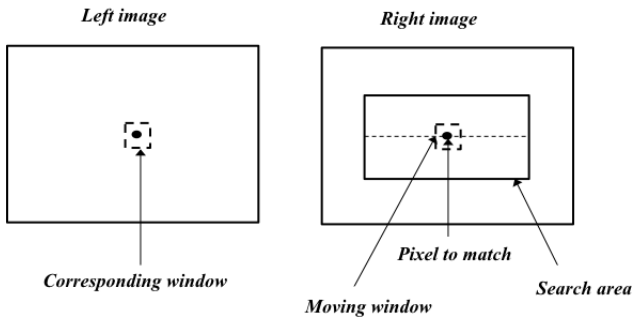

Figure 2.7: Stereo-matching in area-based methods.

map but they are sensitive to locally ambiguous regions as occlusions or untextured regions. Another difficulty is the choice of the window size and shape. Many works have been proposed which do not consider a constant window but rather, the size and the shape of the window is adaptively modified according to the intensity variations in the image [25, 26]. Another work proposes to try a set of windows and to select the one which maximizes the correlation measure

[27].

2.1.3.2 Global methods

The second main category of methods aims at minimizing one global cost function over the entire image contrarily to local methods. Generally the cost function is expressed as the sum of a term $E_d(\mathbf{d})$ measuring the dissimilarity between the corresponding pixels and a term $E_S(\mathbf{d})$, called smoothing prior, that enforces the smoothness in the estimated disparity map:

$$E(\mathbf{d}) = E_d(\mathbf{d}) + \alpha E_S(\mathbf{d}). \tag{2.4}$$

$E_d(\mathbf{d})$ and $E_S(\mathbf{d})$ both depend on the estimated disparity map $\mathbf{d}$. α is a positive constant which controls the smoothness of the estimated disparity map.

Amongst the large numbers of global methods that were proposed, we can cite the dynamic programming: first use in [22] for stereo-matching, this technique is composed of two steps. Considering two corresponding scan lines of size N, the first step consists of computing a path cost matrix of size $N \times N$ by associating a cost to each possible pairing of pixels between the two scan lines. Then, using the ordering and the smoothness constraint, the path composed of disparities minimizing all the path costs is selected. Processing like this, the smoothness of the disparities are enforced along the same scan line. Other works have tried to enforce the smoothness both in horizontal and vertical directions [28, 29]. Amongst the global methods we can also cite the ones based on graph cuts [30, 31]: the stereo matching problem is seen as a labeling problem. Two sets are defined: a set of sites and a set of labels. The problem consists in associating one unique label to each site of the first set. In our case, the sites correspond to the pixels to be matched and the set of labels corresponds to a finite set of disparities. The mapping problem consists in finding the minimal cut through a constructed graph.

Other global methods have been proposed such as belief propagation [32, 33] or variational methods [34, 35].

It is important to note that all previously mentioned methods (local and global) aim at estimating the true disparity map in order to reconstruct the view in three dimensions. The goal is different from the one in this thesis which is the stereoscopic image compression. Values of disparities achieving good compression performance are not necessarily close to the true disparities.

2.2 Overview on stereo image coding

Stereoscopic images are composed of two views generated from two cameras capturing the same scene. Thus, the amount of information to store them or to transmit them is twice compared to the case of traditional two-dimensional images. Efficient coding techniques taking advantages

of the redundancies between the two views have been developed. The next section presents the principles of these coding schemes. Then an overview of the developed methods to improve stereoscopic image coding is given.

2.2.1 Disparity compensated stereo image coding scheme

A simple way of coding stereoscopic images would be to encode them separately using traditional coders like JPEG or JPEG2000. However, this technique known as independent coding does not take advantages of the redundancies that exist between the two views and thus is not so efficient in comparison of more recent methods using the strong correlations between the two views. The basic scheme on which they rely is called the disparity compensated coding scheme because of its similarity with the motion compensated coding scheme used in video coding from which it is inspired [36]. This scheme is composed of the following steps:

- at first, one of the two views is independently encoded, for example, the left one. This view is called the reference image;
- a disparity map is estimated to predict the right view which is the target image;
- a residual image is computed between the original right image and its prediction.

The residual image and the left image are generally coded after a transformation while the disparity map is generally encoded using an entropy coder.

Two schemes can be found in the literature: the schemes based on a open loop structure and the ones based on a closed loop structure. Both structures are represented respectively in Fig 2.8 and Fig 2.9. The difference between the two structures is in the estimation of the

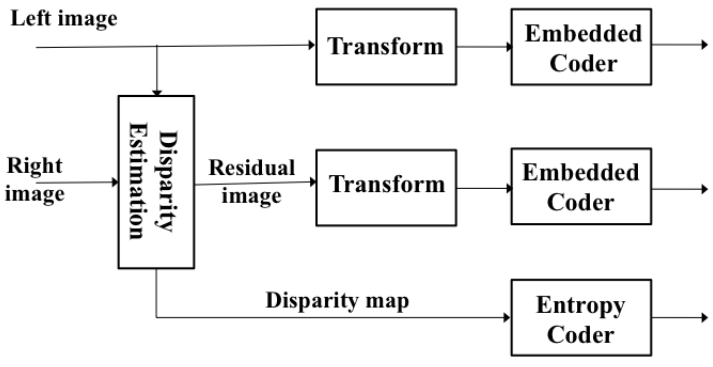

Figure 2.8: Open Loop structure.

disparity map: in the open loop structure, the disparity map is estimated with the original left

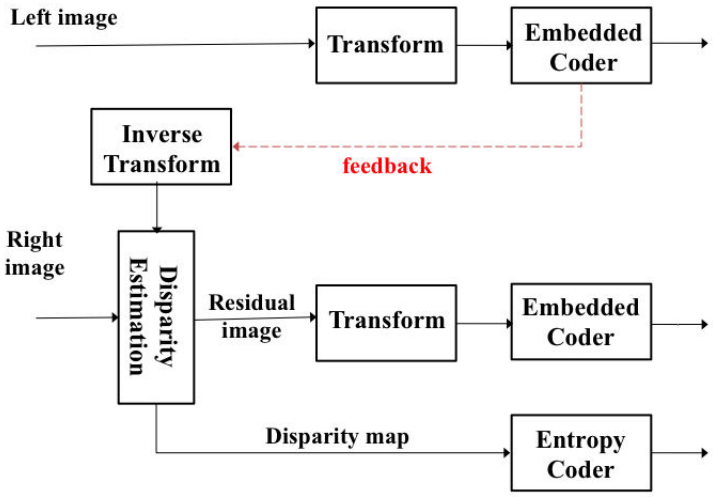

Figure 2.9: Closed Loop structure.

image whereas in the closed loop structure, the disparity map is estimated using the decoded left image. At the decoder side, no matter the structure that is used, the left image is decoded first. The right image is predicted from the left decoded image and the disparity map. Then, the residual image is added to the predicted image to obtain the final reconstructed right image. One can clearly see that when using the open loop structure, the computation of the residual image (using the decoded left image) is inconsistent with the disparity compensation at the decoder side (as the right image is predicted using the original left image). This made this structure sub-optimal in comparison to the second one. The second closed loop structure does not have this drawback as the encoder side and the decoder side both use the decoded left image. However, this structure is computationally more expensive compared to the first one: considering different target bitrates and different configurations of the left image coding, one disparity map should be estimated for each of these decoded left images. In the open loop structure, the disparity map is estimated only once using the original left image and is often more preferred for real time applications.

2.2.2 Standard for stereoscopic videos compression

Recently a standard has emerged to encode multiview and more specifically stereoscopic videos. Both of them can be used to encode stereoscopic image also. The Video Coding Expert Group (VCEG) and the Moving Pictures Experts Group (MPEG) has jointly developed two extensions of the video coding standard H.265 also called HEVC (High Efficiency Video Coding) adapted for the coding of 3D videos [37]. The first one MV-HEVC has been developed for multiview videos and the second one 3D-HEVC is more specific to stereoscopic videos.

In the two extensions, the base view is independently encoded using an HEVC encoder. Additional coding tools have been added to the extensions to encode the dependent view(s).

The major coding tool that have been added in the case of stereoscopic video coding is the disparity-compensated prediction (DCP): DCP is used as an alternative to the motion-compensated prediction (MCP). A block in a frame of the dependent view can either be encoded using blocks of the same view at other instant (motion compensation) or using blocks of the frame in the same access unit (i.e corresponding to the same instant) in the base view (disparity compensation). Amongst the additional tool is the derivation of the disparity vector of a block considering the disparities of its spatially and temporally neighboring blocks. Moreover, the motion vectors of the dependent view can be deduced from the motion vectors of the base view.

2.2.3 Disparity estimation for stereoscopic image compression

Concerning the specific case of stereoscopic images, several approaches have been developed from which we present the most important. One of the major axes of research in the field of stereoscopic image coding is the estimation of the disparity map. It is usually done using blocks instead of pixels as one disparity is send for a whole block. This step plays an important role in the whole coding scheme as a good disparity map will lead to a better predicted image but will be more expensive in terms of bitrates. In a same way, a disparity map with a very small amount of information will cost less in terms of bitrate but can lead to a very poor quality of the predicted image. The estimation of the disparity map can thus be seen as a trade-off between the quality of the predicted image and the bitrate needed to encode it.

Some of the works estimate the disparity map in the spatial domain like Frajka et al. [38] who distinguish the occluded blocks from the non-occluded ones. If the distortion of a block is higher than a threshold, it is considered as occluded, so the block is left intact without being matched. A mixed residual image results from this process with residual of blocks that have been matched and original blocks for the occluded parts. The authors propose to encode these two kinds of blocks differently (using a Haar transform for the occluded blocks and the DCT for the non-occluded).

In [39], Aydinoglu and al. proposes to perform the disparity estimation by distinguishing the occluded regions, the edges and the smooth regions. The occluded regions are independently coded while the non-occluded regions are segmented into edges and smooth regions first. These regions are divided into blocks of fixed sizes and then matched according to their type. The disparity is then quantized using a uniform scalar quantization before being coded with an adaptive arithmetic coder based on the segmentation information. Another area-based method is proposed in [40]. Jiang and al. estimate also a blockwise disparity map by first making an edge detection in the image to be matched. Only the disparities of the blocks corresponding to an edge are sent to the decoder. No disparities are sent for the blocks of the internal area of an object. The drawback of these methods is that they request to send side information about the segmentation which impact the performance at low bitrates. In [41], Palfner and al. use a full search block-matching algorithms to match the blocks of the predicted views. He then applies a

wavelet transformation to the disparity map and the residual image, and encodes the resulting coefficients with the SPIHT coder (Set Partitioning In Hierarchical Trees)[42].

Other works propose to improve the disparity estimation by varying the size and shape of the matching window. For example, Woo and al. [26] estimate the disparity vectors for each block of the image using an adaptive window whose shape depends on the position of the block inside a search area. A measure of confidence is attributed for each computed disparity vector. For the blocks with a measure of confidence two low, the disparity vectors is re-affined by using the Overlapped Block Disparity Compensation technique including a smoothness constraint: the disparity of the block is re-evaluated considering the disparities of neighboring blocks. After this, a decision is taken by analyzing the corresponding blocks in the residual image: if the energy level of a block is superior to a threshold, no disparity will be sent for this block and it will be finally coded in Intra mode

Some works estimate the disparity map in the transform domain, mainly after having applied the DCT and the DWT to the images. Pagliari and al. [43] propose to first apply the DCT to the two views. Then the disparity map is estimated using a genetic algorithm. A pseudo-random disparity map is taken as an initial solution to the matching problem. Thereafter, the genetic algorithm is used to match the blocks and improves the initial solution. The process is repeated iteratively until no improvements are observed. More recently, Maalouf and al. [44] have first computed the bandelet transform of the left and the right images. Each image is thus segmented into dyadic squares where the pixels in each square share the same geometrical properties. Similar blocks are matched by minimizing a metric taking into account the size of the two blocks to match and their optimal geometrical direction. At the end of the process, the blocks for which no corresponding blocks were found are encoded directly in the residual image which is composed of only such blocks. A similar method have been proposed in [45] where the authors first apply a wavelet transform to both images. The matching step is performed again with the dyadic squares after a segmentation of the subbands using a quadtree. Two dyadic squares are matched if they minimize an energy function taking into account the dimensions of the square and its energy.

We have seen that many techniques have been deployed to improve the stereoscopic image coding. Many of them proposes novel way of achieving the disparity map estimation step to better exploit the cross-correlations between the two views, by proposing new metrics to minimizing during the stereo-matching process, using variable block sizes of variable matching window size and/or shape. Some works also get interested in the improvement of the entropy coding of the transformed coefficients or the choice of the transform that should be applied to the images.

Our contributions mainly focus on the improvement of the disparity map estimation.

2.3 Conclusion

In this chapter, we have recalled some basic notions about the stereoscopic images before presenting the most known methods that have been developed to encode them. In the next chapters, we will present our contributions to improve the disparity map estimation process.

Chapter 3

Disparity map estimation algorithm using joint entropy-distortion metric: MMA

This chapter presents the proposed disparity map estimation algorithm using a joint entropy-distortion metric to select the disparities. This metric models the trade-off between the minimization of the predicted image distortion and the minimization of the bitrate needed to encode the disparity map assuming that it is coded using and entropy coder. Furthermore, the algorithm builds a tree where each path corresponds to a disparity map. At each depth of the tree, only the M best paths are retained and extended with all possible disparities. This algorithms shows to perform better in terms of rate-distortion than the Block-Matching Algorithm (BMA), even when using a regularization constraint.

This chapter is organized as follows: the next section presents the rate-distortion optimization problem of estimating the disparity map. Then, we present principles of the developed algorithm which is a sub-optimal solution of the problem. The last section provides simulation results and discuss the performance of the proposed algorithm. The contributions of this chapter have been presented in [8, 9, 11].

3.1 Stereo-matching optimization problem

The optimization problem of the stereo-matching can be seen as the problem of minimizing a global rate-distortion function over the entire image. Finding the optimal solution minimizing the global cost is a complex combinatorial problem. Before presenting this rate-distortion optimization and its complexity, let's first introduce some basic assumptions and notations.

3.1.1 Assumptions and notations

The left and the right images are supposed to be rectified thanks to a special care during the acquisition time or by having applied a rectification algorithm. The left image is taken as the reference and the right view as the one to predict. Each view is divided into non-overlapping blocks of equal size and a unique disparity is estimated for each block. That is to say all the pixels inside a block have the same disparity.

As the two images are considered rectified, the blocks will be matched between the same scan lines, so that the disparity is a one-dimensional horizontal vector and does not have a vertical component. For better accuracy, non-integer disparities are considered: values of sub-pixels are obtained by a weighed combination of the neighboring pixels intensities.

We introduce some notations before presenting the stereo-matching optimization problem. The left and the right image (respectively I_L and I_R) are of size $K \times L$ pixels. They are divided into $X \times Y$ non-overlapping blocks, each of size $N_X \times N_Y$ pixels. A block position can be defined by its block-coordinates (i_B, j_B) or by its pixel coordinates (i_P, j_P). These coordinates are linked by the following equations: $i_P = i_B \times N_X$ and $j_P = j_B \times N_Y$. These notations are illustrated in Figure 3.1.

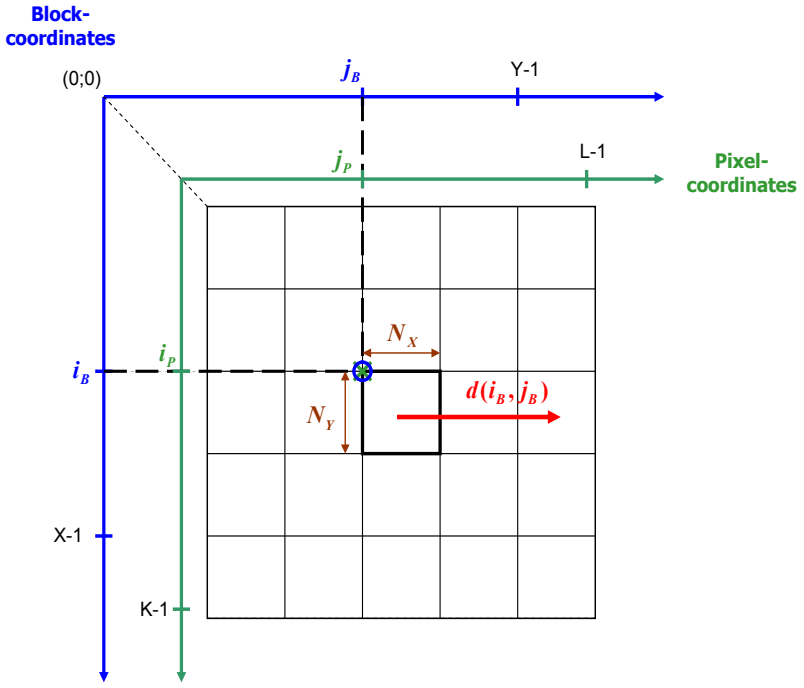

Figure 3.1: MMA notations.

Inside a block located at position (i_P, j_P), other pixels coordinates are defined as $(i_P + u, j_P + v)$ where $u = 0, \ldots, N_X - 1$ and $v = 0, \ldots, N_Y - 1$.

Considering these notations, in the two images, pixels intensities can be expressed as $I_L(i_P + u, j_P + v)$ and $I_R(i_P + u, j_P + v)$. The disparity of the block at position (i_B, j_B) is denoted by $d(i_B, j_B)$. Thus, we can represent the whole blockwise disparity map by $\mathbf{d}$ as:

$$\mathbf{d} = \{d(i_B, j_B) \text{ with } i_B = 0, ..., X-1; j_B = 0, ..., Y-1\}. \tag{3.1}$$

The disparities are not necessarily integer values and are defined with a precision of $\frac{1}{\alpha}$ where $\alpha \in \{1, 2, 4, 8\}$. They belong to a set W considered as the search area given by:

$$W = \{w_{\min}, w_{\min} + 1/\alpha, ..., w_{\max} - 1/\alpha, w_{\max}\}, \tag{3.2}$$

where $w_{\min}$ and $w_{\max}$ are the searching region bounds. It should be noted here that the size of the search area $N = w_{\max} - w_{\min} + 1$ is different from the number of possible disparities than can be selected in this area which is equal to $N_\alpha = (N-1) \times \alpha + 1$. The pixels intensities of the right image are predicted from the pixels intensities in the left interpolated image and their associated disparities as:

$$\begin{aligned} \widehat{I}_R(i_P + u, j_P + v) &= I_{\tilde{L}}(i_P + u, \alpha(j_P + v + d(i_B, j_B))) \\ &\text{with } u = 0, ..., N_X - 1 \text{ and } v = 0, ..., N_Y - 1. \end{aligned} \tag{3.3}$$

where $I_{\tilde{L}}$ is the left interpolated image of size $K \times \alpha L$ pixels.

From this, the global distortion of the predicted image is expressed as:

$$E_G(\mathbf{d}) = \sum_{i_B=0}^{X-1} \sum_{j_B=0}^{Y-1} E_B(d(i_B, j_B)) \tag{3.4}$$

$$\text{with } E_B(d(i_B, j_B)) = \sum_{u=0}^{N_X-1} \sum_{v=0}^{N_Y-1} \left(\widehat{I}_R(i_P + u, j_P + v) - I_R(i_P + u, j_P + v) \right)^2$$

$$\text{and } \widehat{I}_R(i_P + u, j_P + v) = I_{\tilde{L}}(i_P + u, \alpha(j_P + v + d(i_B, j_B))).$$

An estimate of the bitrate needed to encode the disparity map is given by:

$$H(\mathbf{d}) = -\sum_w p_E(d = w|\mathbf{d}) \log_2 \left(p_E(d = w|\mathbf{d}) \right). \tag{3.5}$$

where $p_E(d = w|\mathbf{d})$ is the empirical distribution probability (knowing the disparity map $\mathbf{d}$) computed as the ratio between the number of blocks having the disparity w to the total number of blocks $X \times Y$.

3.1.2 Rate-distortion optimization problem formalization

The stereo-matching optimization considered in this thesis consists in finding the best compromise between the quality of the predicted view and the bitrate needed to encode the disparity

map. This problem can be reformulated as finding the best disparity map minimizing the distortion of the predicted view subjected to a given bitrate. Using the Lagrangian formulation, a global rate-distortion cost function is defined as:

$$J(\lambda, \mathbf{d}) = E_G(\mathbf{d}) + \lambda H(\mathbf{d}), \tag{3.6}$$

where λ is the Lagrange multiplier, $E_G(\mathbf{d})$ is the global distortion of the predicted view defined in Equation 3.1.1 and $H(\mathbf{d})$ is the entropy of the disparity map defined in Equation 3.5.

Minimizing this cost function for every values of λ leads to the optimal curve formed by all the Rate-Distortion (RD) points $(H(\widehat{\mathbf{d}}_\lambda), E_G(\widehat{\mathbf{d}}_\lambda))$. Each of these points is obtained using the optimal disparity map $\widehat{\mathbf{d}}_\lambda$ achieving the best reconstruction of the predicted view at a given bitrate. The global cost function appears clearly to be a joint entropy-distortion metric to minimize.

3.1.3 Properties of the global cost function

This global cost function presents some features that need to be discussed.

Indeed, for a given value of λ, finding the optimal disparity map minimizing the global cost $J(\lambda, \mathbf{d})$ is a complex combinatorial problem. For stereoscopic images composed of $X \times Y$ non-overlapping blocks, there is approximately $(N_\alpha)^{(XY)}$ disparity maps that can be computed among which is the optimal one. As an example to illustrate this gigantic search space, let's consider an image of size 100×100 pixels. Assume that the disparities are selected in a set of size 60 with a quater-pel accuracy (i.e. $\alpha = 4$). The optimal disparity is to be found amongst 10^{23802} possible disparity maps.

Another aspect of this global cost concerns the relationship between the selected disparities. For $\lambda = 0$, minimizing the global cost is equivalent to minimize the global distortion of the predicted view, just as the Block-Matching Algorithm (BMA). Minimizing the global distortion $E_G(\mathbf{d})$ over the whole image can in fact be done by minimizing independently all the local distortions in the image as $E_G(\mathbf{d})$ is nothing else than a sum of the independent local distortions $E_B(d(i_B, j_B))$. So for $\lambda = 0$, the optimal disparity map can simply be computed by selecting for each block, the disparity minimizing its local distortion. But for $\lambda \neq 0$, i.e. when taking into account the bitrate of the disparity map, the estimation of the optimal disparity map is more complex, as the choices of disparities for each block are all interrelated. Indeed, the choice of a disparity at one block tends to favor the choice of the same disparity for the other blocks. This is explained by the fact that the entropy which approximates the bitrate needed to encode the disparity map get reduced as much as the disparity map is composed of similar disparities.

3.2 Presentation of the modified M-algorithm (MMA)

This section describes the proposed solution to the rate-distortion optimization problem. The developed algorithm has been called the Modified M-Algorithm (MMA) in reference to the original M-Algorithm from which it is inspired. Originally, the M-Algorithm has been developed in [46] to estimate the transmitted data stream through a noisy channel using the maximum likelihood criterion. The algorithm sequentially builds a tree, and at each depth of the tree, only the M-best paths according to the maximum likelihood metric are being extended. This allows to reduce the computational complexity while ensuring good performance.

In order to adapt it to our stereo-matching optimization problem, we have brought several modifications to this algorithm which are described below.

3.2.1 Sequential MMA based on a tree structure

The developed algorithm estimates the disparity map sequentially by building a tree where each path correspond to a different disparity map being explored. The blocks of the image to match are processed in a raster scanning order. For each block, a disparity is selected by taking into accounts the already chosen disparities for previous blocks but also by making an assumption on the distribution probability on the disparities to be chosen for the remaining blocks. The selection process is detailed in the next subsection.

As the MMA processes the blocks in a raster scanning order, we have simplified some notations for sake of simplicity. The block coordinates (i_B, j_B) are now replaced by a one-dimension coordinate t defined as:

$$t = i_B \times Y + j_B + 1 \text{ with } i_B = 0, ..., X-1 \text{ and } j_B = 0, ..., Y-1. \tag{3.7}$$

t ranges from 1 to T where T is the total number of blocks in the image (i.e. $X \times Y$). In the same way, the disparity $d(i_B, j_B)$ is now denoted d_t and corresponds to the disparity of the t-th block in the considered raster scanning order.

Assume that the MMA has already processed the $t-1$ first blocks of the image. The M-best paths that have been retained by the MMA are denoted $\mathbf{d}_{t-1}^k$, where k corresponds to the index of the path. They correspond each to a partial disparity map defined as:

$$\mathbf{d}_{t-1}^k = \left[d_1^k, d_2^k, ..., d_{t-1}^k\right] \quad \text{with } k = 1, \ldots, M. \tag{3.8}$$

where d_{t-1}^k is the disparity of the $(t-1)$-th block in the k-th path. When processing the t-th block, the MMA extends each of the M previous best paths by N_α new branches corresponding to the N_α choices of possible disparities $w \in W$ for the current block. Each of these extended new paths is denoted $(\mathbf{d}_{t-1}^k, w)$. At this step of the algorithm, amongst the MN_α extended paths, the MMA selects again the M best paths minimizing the global cost function. Note that after the process of the t-th block, the k-th retained path may not necessarily stem from the k-th path retained for the previous block.

3.2.2 Entropy-distortion metric computation

The global cost function defined in Equation 7.6 is composed of the global distortion of the predicted view $E_G(\mathbf{d})$ and the entropy of the disparity map $H(\mathbf{d})$. Thus, its computation requires the knowledge of the disparity distribution probability over the entire map. As the MMA estimates the disparity map sequentially, when processing the t-th block, this distribution probability is unknown at this step of the algorithm. So we propose some modifications to the global cost function to adapt it to the sequential processing of the MMA. Concerning the computation of the global distortion $E_G(\mathbf{d})$, it is a sum of disconnected local distortions. So when processing the t-th block, it can easily be replaced by $E_t(\mathbf{d}_{t-1}^k, w)$ which is the global distortion of the partial disparity map $(\mathbf{d}_{t-1}^k, w)$. It is defined as:

$$E_t(\mathbf{d}_{t-1}^k, w) = \sum_{\tau=1}^{t-1} E_B(d_\tau^k) + E_B(w) \text{ with } k = 1, ..., M. \tag{3.9}$$

This replacement does not impact the choice of the M best paths at the t-th block. Things are different for the computation of the disparity map entropy in the global cost function. As said previously, it requires the knowledge of the whole disparity map which is unknown when processing an intermediate block of the image. So we propose to estimate this distribution probability at each step of the algorithm by making an adaptive assumption over it. To allow a sequential disparity map estimation by the MMA, the entropy term $H(\mathbf{d})$ is replaced by an adapted entropy $H_{\beta,t}(\mathbf{d}_{t-1}^k, w)$ when processing the t-th block. It is define as follows:

$$H_{\beta,t}(\mathbf{d}_{t-1}^k, w) = -\sum_{d \in W} p_M(d_t = d|\mathbf{d}_{t-1}^k, w) \log_2 \Big(p_M(d_t = d|\mathbf{d}_{t-1}^k, w) \Big) \text{ with } k = 1, ..., M. \tag{3.10}$$

where $p_M(d_t = d|\mathbf{d}_{t-1}^k, w)$ is the disparity distribution probability of the whole disparity map which is estimated thanks to a finite mixture distribution given by:

$$p_M(d_t = d|\mathbf{d}_{t-1}^k, w) = a(\beta, t) p_U(d_t = d) + (1 - a(\beta, t)) p_E(d_t = d|\mathbf{d}_{t-1}^k, w), \tag{3.11}$$

where the parameter $a(\beta, t)$ is set as:

$$a(\beta, t) = \frac{\beta(T - t)}{\beta(T - t) + t}. \tag{3.12}$$

$p_U(d_t = d)$ is a uniform distribution defined as:

$$p_U(d_t = d) = \frac{1}{N_\alpha}. \tag{3.13}$$

because the set W of all possible disparities contains N_α disparities.

$p_E(d_t = d|\mathbf{d}_{t-1}^k, w)$ is the empirical distribution of the partial disparity map $(\mathbf{d}_{t-1}^k, w)$ which is computed as the ratio between the number of blocks having the disparity d to the total number of processed blocks t. It is expressed as:

$$p_E(d_t = d|\mathbf{d}_{t-1}^k, w) = \frac{V(\mathbf{d}_{t-1}^k, d)}{t}. \tag{3.14}$$

The parameter β allows to choose to what extent the estimated disparity distribution should be close to a uniform distribution (obtained when $\beta = 1$) or to the empirical distribution computed with the partial disparity map (obtained when $\beta = 0$).

It should be noted that the function $a(\beta, t)$ decreases linearly with the number of blocks already processed, such as it is equal to zero when processing the last block of the image. In this way, as the image is being processed, the uniform probability has less and less importance to the detriment of the empirical probability which becomes closer to the true disparity distribution. At the end of the image process, the estimated distribution probability is exactly equal to the empirical probability.

The new joint entropy-distortion metric is finally defined as:

$$J\left(\lambda, \beta, \mathbf{d}_{t-1}^{k}, w\right) = E_t\left(\mathbf{d}_{t-1}^{k}, w\right) + \lambda H_{\beta,t}\left(\mathbf{d}_{t-1}^{k}, w\right) \text{ with } k = 1, ..., M. \tag{3.15}$$

The steps of the MMA are summarized in Figure 3.2.

An illustration of the algorithm is also given in Figure 3.3 as an example taking $M = 2$ and $N_1 = 5$.

Input: Left image I_L and right image I_R of size $X \times Y$
Output: Estimated disparity map associated with I_R
1. Set initial values: λ; M; $w_{\min}$; $w_{\max}$; α; β; $i_B = -1$; $j_B = -1$;
2. Set all paths to empty sets.
3. Increment by 1 the row index i_B;
4. Increment by 1 the column index j_B;
5. Define $M \times N_\alpha$ branches by extending each path with a disparity;
6. Estimate the disparity probabilities of all extended paths;
7. Deduce the disparity entropy of all paths;
8. Compute the modified cost functions $J(\lambda, \beta, \mathbf{d}_{t-1}^{k}, w)$ of all paths;
9. Sort the paths in an increasing order of the modified cost functions;
10. Select among the $M \times N_\alpha$ paths, the first M paths;
11. Update the M disparity maps;
12. Start again from step 3 if $j_B < Y$ otherwise continue;
13. Start again from step 2 if $i_B < X$ otherwise continue;
14. Select the best dense disparity map associated with I_R.

Figure 3.2: MMA: Entropy-distortion optimization algorithm

3.2.3 Settings parameters in the MMA

The MMA relies on many parameters: the number of paths M that are extended at each depth, the disparity range W, the size of the block to match $N_X \times N_Y$, the precision of the disparities α, the parameters λ and β. To reduce the computational load, we chose to set the following

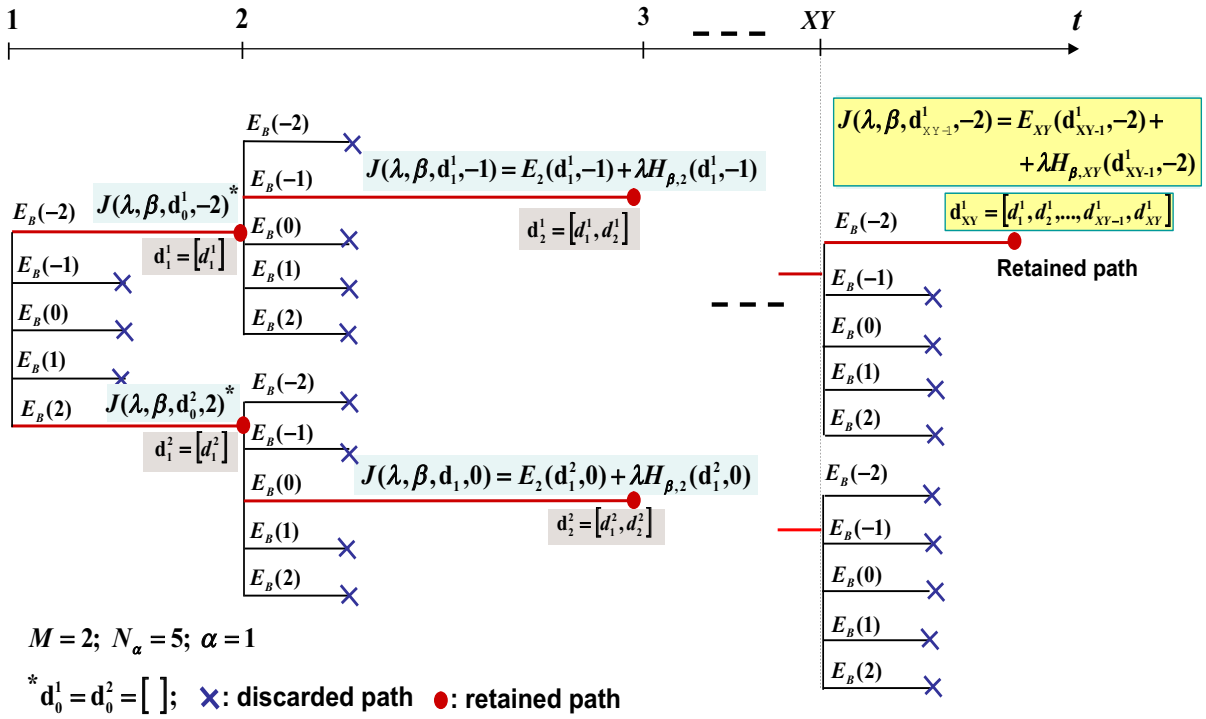

Figure 3.3: Stereoscopic matching optimization algorithm with $M = 2$ and $N_1 = 5$.

parameters: the number of retained paths M, the disparity range W, α and β (the choice of β is discussed in section 3.3.1).

Indeed, at a given bitrate, we restrict the parameter optimization to the choice of the size of blocks $N_X \times N_Y$ and the selection of the parameter λ. For this, we consider only square blocks of size ranging from 1×1 to 16×16. The value of λ is also chosen amongst a finite set of values ranging from 0 to λ_{max}.

The optimized values of $N_X \times N_Y$ and λ are computed by an iterative algorithm able to determine amongst the predefined sets, values of parameters minimizing the global distortion at a given bitrate.

The parameter optimization algorithm relies on two tree structure: the first one allows to explore the sizes of blocks (see Figure 3.5) and the second one, similar to the first one, allows to explore the values of λ. Both tree structure share the same feature: going upwards in the tree (i.e selecting a smaller size of block or a smaller value of λ) tends to increase the bitrate of the disparity map and reduce the distortion of the predicted view. The developed algorithm uses the knowledge of this feature to reject some combinations of parameters without trying them for

a given bitrate, allowing to reduce the computational load of the MMA.

We introduce some notations to present the optimization parameter of this algorithm. T_{bpp} is the target bitrate, $bpp(\lambda = \lambda_i, s)$ is the bitrate of the disparity map estimated by the BMA using a size of block equal to s and $\lambda = \lambda_i$. Furthermore, s^+ represents the size of the block just superior to s in the tree structure and s^- represents the size of the block just inferior to s (see Figure 3.5). For example, if $s = 8 \times 8$, then $s^+ = 12 \times 12$ and $s^- = 4 \times 4$. The algorithm first determines the block sizes that can satisfy the target bitrate by testing the two extreme values of λ (i.e. $\lambda = 0$ and $\lambda = \lambda_{max}$) as in Figure 3.4:

Input: Left image I_L and right image I_R of size $X \times Y$
Output: Sizes of blocks that can satisfy the target bitrate
1. Initialization: set $s = 8 \times 8$ and $\lambda = 0$;
2. Compute $bpp(\lambda = 0, s)$. If $bpp(\lambda = 0, s) < T_{bpp}$, go to step 3. Otherwise go to step 4.
3 Prune all branches related to a block size s_i such as $s_i \geq s$, set $s = s^-$ and return to step 2.
4 Compute $bpp(\lambda = \lambda_{max}, s)$. If $bpp(\lambda = \lambda_{max}, s) > T_{bpp}$, go to step 5, otherwise go to step 6.
5 Prune all branches related to a block size s_i such as $s_i \leq s$, set $s = s^+$ and return to step 2;
6 Retain the block size s.
7 Start again from step 2 by setting $s = s^-$ and $s = s^+$ as long as this is possible.

Figure 3.4: Selection of the block sizes in the MMA

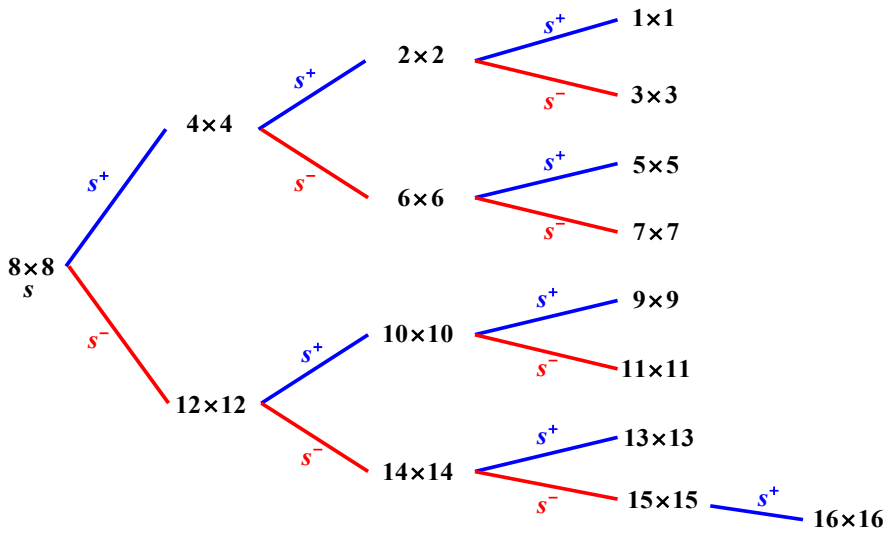

Figure 3.5: Tree-structure to explore the block sizes.

A dichotomy algorithm is then applied for each retained size of blocks to determine amongst a predefined set, the value of λ leading to the best reconstruction under the bitrate constraint. Finally, amongst all the retained couples of parameters (block size and λ value), the best one in terms of distortion is selected. Note that with limited modifications to this algorithm, a different

goal could have been set: finding the optimal set of parameters minimizing the bitrate for a given quality of the predicted view.

3.3 Experimental results and discussions

This section presents the simulations results that have been conducted on stereoscopic images taken from the "Middlebury" dataset [47] ("Tsukuba", "Barn2", "Baby1", "Wood2", "Bull", "Sawtooth" and "Venus") and the "Deimos" dataset [48] ("Stereo_2", "Stereo_13" and "Stereo_67").

For all the simulations the left view is taken as the reference one and the right view as the one to be predicted. The bitrate needed to encode the disparity map is approximated by the entropy computed using the empirical disparity distribution of the disparity map. It is expressed in bits per pixel (bpp). The distortion of the predicted view is computed using the Peak Signal-to-Noise Ratio (PSNR) between the original right view and its prediction. It is expressed in decibel (dB).

The performance of the MMA are compared to the BMA with and without using a regularization constraint to estimate the disparity maps. We also analyze the performance of all algorithms using a DPCM (Differential Pulse-Code Modulation) encoder.

The rest of this section is organized as follows. We first present how the choice of the parameter β is performed. Then we analyze the influence of the number of the retained paths M on the MMA rate-distortion performance. We present some detailed simulation results comparing the MMA with the BMA. Then we discuss the performance of the MMA compared to the BMA using a regularization constraint, which is denoted BMA-RA (BMA with Regularization Algorithm). The impact of using a DPCM encoder on all algorithms is also studied. The particular case of dense disparity map estimation is also considered and finally we end this section with a discussion on the MMA processing time.

3.3.1 Choice of the parameter β

To discuss the choice of the parameter β, we present an experiment that have been conducted on the four stereoscopic images "Tsukuba", "Stereo_13", "Wood2" and "Baby1". The disparity map is estimated by the MMA using values of β ranging from 0 to 1 with a step equal to 0.1. Disparities are selected amongst the set $[-15, \ldots, 14]$. Figure 3.6 presents the simulations results as the average PSNR computed from the reconstruction of the four images versus the value of β taken for the simulation. This simulation has been carried out for six different target bitrates: 0.01; 0.03; 0.06; 0.2; 0.5; 1.5 bpp. For each target bitrate, the performance are represented using a different color. We can note that for all the curves, the best performance in terms of distortion are achieved taking $\beta \geq 0.1$. $\beta = 0$ leads to the worse performance in all cases. This corresponds

to the case where the mixture probability does not consider the *a priori* uniform distribution of the disparities in Equation (3.11), and is only composed of the empirical disparity distribution. Furthermore, we can note that β greater than 0.2, the distortion performance remains fairly similar even if a little decrease can be noticed as the value of β increases until 1. Considering all these observations, we have set the parameter β equal to 0.2 for the rest of the simulations presented in this chapter.

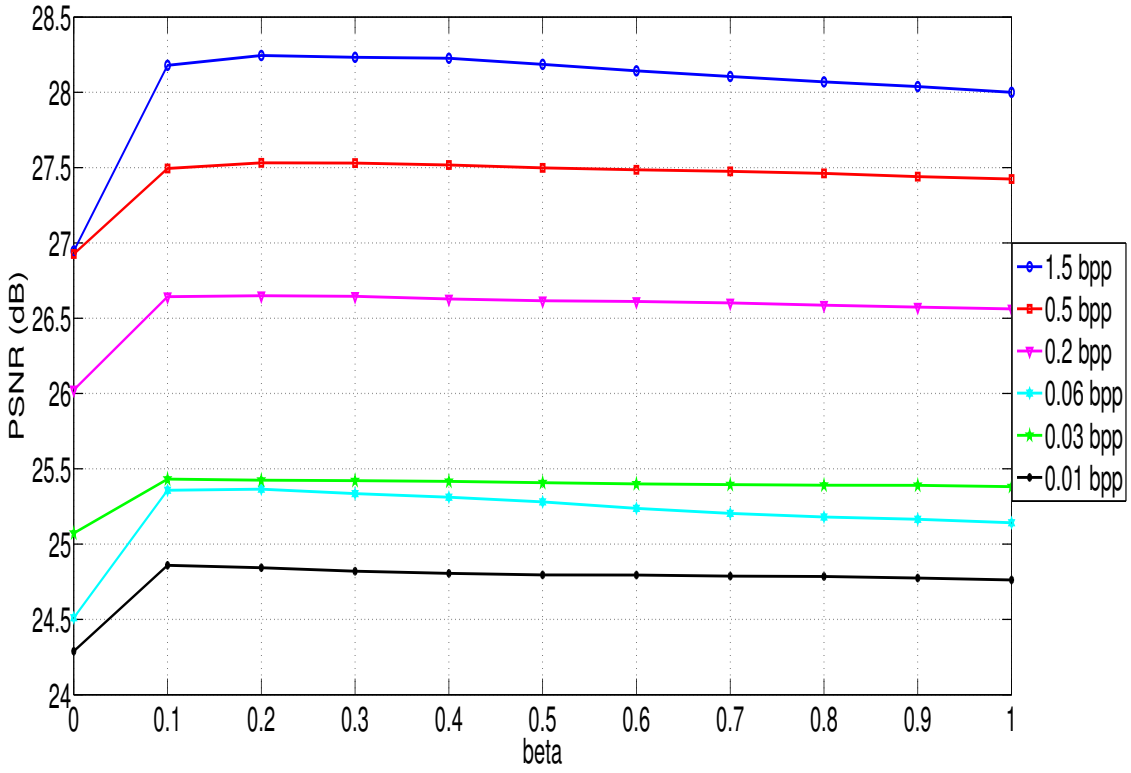

Figure 3.6: MMA performance according to the parameter β.

3.3.2 Influence of the number of retained paths M on the performance

The parameter M defines the number of paths that should be retained and then extended at each depth of the tree.

We present simulations performed on two stereo images ("Barn2" and "Wood2") to discuss the influence of the variation of this parameter. In the simulations, β is set to 0.2, the disparities are integer values ($\alpha = 1$) and are selected amongst the set $[-15, \ldots, 14]$.

The MMA estimates a dense disparity map (the size of blocks used is equal to 1×1). We compare the rate-distortion performance achieved by the MMA using $M = 1$ and $M = 8$. Figure 3.7 presents the results on the stereoscopic image "Barn2" as the PSNR of the predicted view versus the bitrate of the disparity map. The curve joining cross symbols (in red color)

corresponds to the performance of the MMA using $M = 1$ and the curve joining circles (in blue color) corresponds to $M = 8$.

For each curve, the RD points are obtained by varying the values of λ from 0 to 8000. Both curves start from the same point (obtained using $\lambda = 0$) and join again on the bottom left (highest values of λ). The first point actually corresponds to the performance that would have achieved the BMA, as for $\lambda = 0$, the entropy term plays no role in the cost function. The two curves remains below that point. This was expected because increasing the value of λ can only lead in a reduced performance in terms of distortion as we add a constraint on the optimization problem. From this experiment, we can note that the curve associated to $M = 8$ remains above the curve associated to $M = 1$. That is to say, taking $M = 8$ allows to achieve better rate-distortion performance in this example for bitrates, especially between 0.6 bpp and 2 bpp. Taking $M = 8$ allows to achieve a gain of up to 3.2 dB as compared to $M = 1$ at the bitrate of 0.62 bpp.

Our understanding is that the solution found by the MMA becomes closer to the optimal solution of the stereo-matching optimization problem when λ is relatively small and when M is increased. Indeed when λ is equal to 0, the optimal solution is the one found by the BMA, and the complexity of finding the optimal disparity map increases with the value of λ. When M is increased, a higher number of disparity maps are explored. So combining a small value of λ and a higher value of M leads to a disparity map closer to the optimal solution.

However, in most of the experiments that we have conducted, no significative gains have been observed by varying reasonably the value of M. Figure 3.8 shows the rate-distortion performance of the MMA conducted on the stereoscopic image "Wood2" in the same conditions as for the previous experiment. No significant difference can be seen between the performance obtained using $M = 8$ and $M = 1$. Our understanding is that the stereo-matching optimization problem is so complex and the search space so gigantic that increasing M with a reasonable values does not allow to see a gain in the performance of the MMA. For this reason, the parameter M is set to 1 in most of the following simulations.

3.3.3 Comparison with the Block-Matching Algorithm

This section discusses the rate-distortion performance of the MMA in comparison to the BMA. Before giving an overview of the simulations results, we first present an analysis of the obtained results on two stereoscopic images.

The first experiment have been conducted on the image "Tsukuba" using blocks of size ranging from 1×1 up to 16×16 for the MMA and up to 24×24 for the BMA. The left image has been interpolated according to [49] to allow half disparities ($\alpha = 2$). The latter is chosen amongst the set $[-15, -14.5, \ldots, 14, 14.5]$ and the parameter β is set to 0.2 while M is set to 1.

Rate-Distortion performance of both algorithms are given from Figure 3.9 to Figure 3.12.

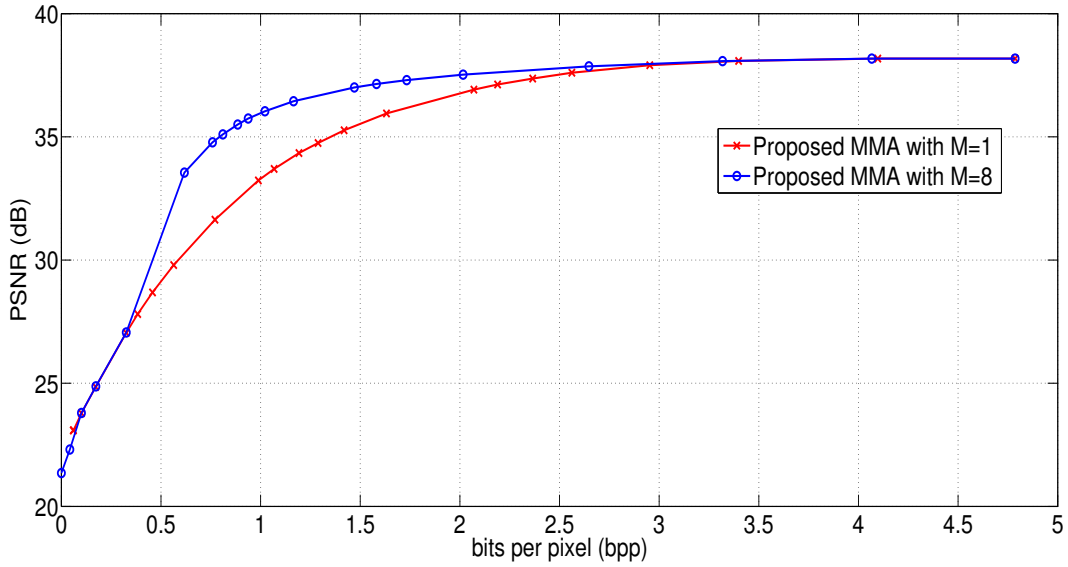

Figure 3.7: Influence of the parameter M on the MMA performnce using "Barn2" with a block size 1×1; $M = 1$ and $M = 8$.

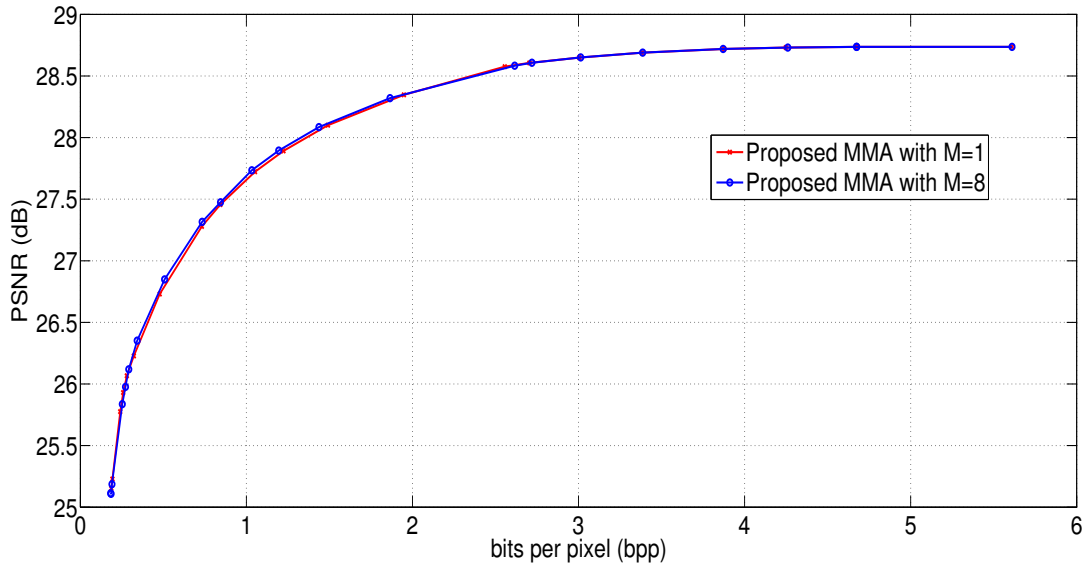

Figure 3.8: Influence of the parameter M on the MMA performance using "Wood2" with a block size 1×1; $M = 1$ and $M = 8$.

Each figure corresponds to a specific bitrate range and has been obtained using different block sizes. Figure 3.9 (respectively Figure 3.10, Figure 3.11 and Figure 3.12) represents the performance of the MMA obtained using blocks of size ranging from 1×1 to 3×3 (respectively 4×4 to 7×7, 8×8 to 11×11 and 12×12 to 16×16). Simulation results have been conducted using a specific block size for different values of λ ranging from 0 to 8000 and are represented by a curve of specific color (one color for each block size). All the curves start on the top right at a black

circle obtained with $\lambda = 0$. The performance of the BMA are given by the dashed line joining circles (in black color) where each circle is obtained using a different block size.

These figures clearly show that the MMA performs better than the BMA in terms of rate-distortion. For example, using blocks of size 4×4, the BMA leads to a disparity map with a bitrate of 0.27 bpp for a predicted image quality of 34.12 dB. Using the same block size, the MMA yields to a predicted image with a slightly lower quality of 33.89 dB but the estimated disparity has a lower bitrate of 0.14 bpp (for $\lambda = 1.8$). Although the bitrate needed to encode the disparity map is much lower with the MMA than with the BMA, no noticeable differences can be seen between the reconstructions of the predicted image obtained with the BMA and the MMA (respectively given in Figure 3.16 and Figure 3.17). For comparison, the original right view is presented in Figure 5.5. This can be explained as the MMA tends to favor the use of identical disparities to reduce the entropy of the disparity map while ensuring good performance thanks to the joint entropy-distortion metric. The observation of the histograms of the disparity maps estimated by the BMA and the MMA (respectively given in Figure 3.18 and Figure 3.19) is in agreement with this understanding. Indeed the large amount of small bars in the BMA-histogram is replaced by fewer bars of bigger size in the histogram of the MMA. This transformation reduces significantly the entropy of the disparity map as they are less different values of disparities, but the selected ones appear more often in the disparity map.

Figure 3.20 and Figure 3.21 show the disparity maps estimated by the BMA and the MMA respectively. One can clearly see that the one computed by the MMA is smoother than the map computed by the BMA. This is coherent with the analysis of the histograms.

At the same bitrate, the MMA achieves also better performance than the BMA. For example, we can see in Figure 3.11 that using blocks of size 13×13, the BMA yields the RD point (30.18dB, 0.022bpp) while the MMA yields to the point (30.81dB, 0.022bpp) using smaller blocks of size 9×9 with $\lambda = 9.9$. We can observe this resulting gain of 0.63dB in the reconstruction of the BMA (given in Figure 3.14) and the one of the MMA (given in Figure 3.15). On can see that the edges of the lamp are better reconstructed with the MMA thanks to the use of smaller blocks.

In what follows we discuss the performance of the two algorithms on the stereoscopic image "Barn2". Experiments have been conducted using blocks of size ranging from 1×1 to 3×3 for the MMA and up to 5×5 for the BMA. The disparities are selected in the set $[-15, -14, \ldots, 13, 14]$, the parameter β is set to 0.2, values of λ are taken from 0 to 8000. The MMA has been tested for $M = 1$ and $M = 8$.

The performance of both algorithms are shown in Figure 3.22. The results obtained by the MMA using the same size of blocks are represented using the same color, while the performance of the BMA are given by the dashed black line joining circles. Each circles has been obtained using a different block size. Two curves representing the performance of the MMA starts from each circle: the one joining cross symbols "x" is obtained taking $M = 1$ and the one joining

plus symbols "+" is obtained taking $M = 8$. The presented results shows that allowing a little decrease in the distortion of the predicted image allows to achieve a significant gain in terms of bitrate. For example, using blocks of size 2×2, the BMA yields to the RD point (34.81dB, 0.98bpp). By allowing a little decrease in the distortion performance, the MMA can reduce the bitrate and leads to the RD point (34.59dB, 0.43bpp) by taking $\lambda = 1.2$. Improvements can also been observed in terms of distortion at the same bitrate. The MMA yields to a reconstructed image of quality 35.91dB at the bitrate of 0.98bpp using blocks of size 1×1 with $M = 8$ and $\lambda = 11.9$. This results in a gain of 1.1dB in comparison to the BMA. The reconstructions of the right view by the two algorithms confirm this gain. Figure 3.25 shows 3 close-up views of the area outlined by the white box on the original image represented on the top of the figure. From left to right, the close-up views correspond to the original right image, the reconstruction of the BMA and the reconstruction of the MMA. The face at the bottom of the triangle is better predicted by the MMA, as for the dotted line at the center of the triangle.

Table 5.1 provides additional simulation results performed on different stereoscopic images ("Tsukuba", "Barn2", "Baby1", "Wood2", "Bull", "Sawtooth" , "Venus", "Stereo_2", "Stereo_13" and "Stereo_67") using different size of blocks (4×4, 6×6 and 8×8) and taking $M = 1$. The disparities are selected amongst the set $[-15, -14.5, \ldots, 14, 14.5]$. Concerning the MMA, λ values ranges from 0 to 8000 and β is set at 0.2. The performance of the BMA and the MMA are given in Table 5.1 respectively by the column 1 and 3. Results are given as the PSNR of the reconstructed view in dB with the bitrate needed to encode the associated disparity map in bpp. For the MMA, the results provided are given for fairly small values of λ, allowing to reduce the bitrate as compared to the BMA to the detriment of a small decrease in the distortion performance. Results shows that for all images and considered block sizes, the MMA allows to achieve a significant gain in the bitrate needed to encode the disparity map, at the expense of a little decrease in the distortion performance as expected. It should be noted that this small decrease in the distortion performance does not induce noticeable modifications visually on the predicted view.

3.3.4 Comparison with the BMA including a smoothness constraint

In this section, we compare the performance of the MMA to the Block-Matching Algorithm in which a regularization constraint has been added (called BMA-RA). The BMA-RA processes the blocks in raster scanning order and the disparity for each block is selected by minimizing the following local cost function:

$$J = E_B(d) + \lambda Reg(d, p). \tag{3.16}$$

where λ is the Lagrange multiplier. In this equation, $E_B(d)$ represents the local distortion induced by the disparity d. $Reg(d, p)$ is a regularization constraint allowing to smooth the disparity map and consequently reduce its entropy. This term takes into account the disparities

of the three nearest neighboring blocks that have already been processed around the current block (i.e. the left, upper and left upper corner blocks). This regularization term is expressed as: $Reg(d,p) = \sqrt{(d^2 - p^2)}$ where p is the median value of the disparities computed over the three neighboring blocks previously cited.

In the provided simulation results, the disparities are selected in the set $[-15, -14.5, \ldots, 13.5, 14]$ for the two algorithms. Values of λ ranges from 0 to 8000 for the MMA and 0 to 500 for the BMA-RA. Figure 3.23 shows the performance of both algorithm on the stereoscopic image "Wood2" using blocks of size equal to 4×4, 6×6 and 8×8. Results obtained using the same size of block are represented with the same color, with "+" symbol for the MMA and "*" symbol for the BMA-RA. For each size of blocks, the MMA and the BMA-RA starts from the same point at the upper right of the curves. This point has been obtained for the two algorithms by taking $\lambda = 0$ which corresponds to the minimization of the predicted view distortion only. Considering the size of block 4×4, both algorithm leads to the RD point (27.13 dB, 0.34 bpp) for $\lambda = 0$.When allowing a little decrease in the predicted image quality, the MMA leads to a reconstructed image of quality 26.84 dB for a bitrate of 0.08 bpp using $\lambda = 6.8$, while the BMA-RA leads to a reconstructed image of the same quality 26.83 dB for a higher bitrate of 0.25 bpp using $\lambda = 3$. This analysis can be extended to the results obtained with the other block sizes, showing the advantage of the MMA on the BMA-RA.

The table 5.1 is extended with the performance of the BMA-RA on different stereoscopic images (see column 2). λ values tested range from 0 to 8000 for the MMA and from 0 to 500 for the BMA-RA. The performance of the BMA-RA are also provided for fairly small values of λ when it allows a reduction of the bitrate needed to encode the disparity map at the expense of a little increase in the predicted image distortion. The comparison of the rate-distortion performance of the two algorithms shows the advantage of the MMA over the BMA-RA as it allows a higher decrease in the bitrate needed to encode the disparity map for an equivalent or higher quality of the predicted view.

Setting parameters	BMA: $d \in [-15, -14.5, \ldots, 13.5, 14]$ BMA-RA: $d \in [-15, -14.5, \ldots, 13.5, 14]$ MMA: $d \in [-15, -14.5, \ldots, 13.5, 14], \beta = 0.2, M = 1$

Stereoscopic		**BMA**		**BMA-RA**			**MMA**		
Images	**Block Size**	**PSNR**	**Bitrate**	**PSNR**	**Bitrate**	λ	**PSNR**	**Bitrate**	λ
Wood2	4x4	27.13	0.34	26.83	0.25	3	26.84	0.08	6.8
	6x6	26.7	0.143	26.41	0.109	3	26.39	0.027	10.6
	8x8	26.54	0.077	26.25	0.06	3	26.25	0.013	14.3
Stereo_2	4x4	32.7	0.313	32.24	0.258	1	32.4	0.219	4.6
	6x6	31.86	0.135	31.3	0.108	1	31.48	0.087	6.2
	8x8	31.59	0.074	31.25	0.061	1	31.25	0.049	6.2
Stereo_67	4x4	41.95	0.347	41.95	0.347	0	41.6	0.169	0.1
	6x6	40.68	0.1507	40.68	0.1507	0	40.4	0.056	0.1
	8x8	40.73	0.083	40.73	0.083	0	40.48	0.028	0.1
Tsukuba	4x4	34.12	0.269	34.12	0.269	0	33.82	0.133	2.7
	6x6	32.21	0.11	32.21	0.11	0	32.02	0.059	3.6
	8x8	31.48	0.059	31.48	0.059	0	31.29	0.032	4.5
Stereo_13	4x4	29.99	0.331	29.69	0.247	2	29.74	0.216	3.1
	6x6	28.96	0.139	28.68	0.103	3	28.68	0.08	4.6
	8x8	28.54	0.077	28.35	0.056	3	28.4	0.051	3.1
Baby1	4x4	26.98	0.347	26.98	0.347	0	26.73	0.21	7.2
	6x6	26.02	0.151	26.02	0.151	0	25.8	0.084	7.2
	8x8	25.5	0.082	25.5	0.082	0	25.3	0.043	7.2
Barn2	4x4	34.51	0.247	34.51	0.247	0	34.3	0.16	1.2
	6x6	34.03	0.101	34.03	0.101	0	33.77	0.067	1.8
	8x8	33.08	0.055	33.08	0.055	0	32.88	0.039	1.8
Venus	4x4	31.6	0.307	31.6	0.307	0	31.41	0.194	3
	6x6	30.48	0.131	30.48	0.131	0	30.18	0.083	6.6
	8x8	29.75	0.071	29.75	0.071	0	29.51	0.049	6.6
Sawtooth	4x4	31.7	0.273	31.7	0.273	0	31.46	0.215	3
	6x6	30.82	0.117	30.82	0.117	0	30.63	0.098	3
	8x8	30.21	0.063	30.21	0.063	0	30.21	0.054	3
Bull	4x4	35.35	0.284	35.35	0.284	0	35.13	0.186	1.2

Stereoscopic		BMA		BMA-RA			MMA		
Images	Block Size	PSNR	Bitrate	PSNR	Bitrate	λ	PSNR	Bitrate	λ
	6x6	34.79	0.118	34.79	0.118	0	34.52	0.079	1.8
	8x8	34.18	0.064	34.18	0.064	0	33.93	0.044	1.8

Table 3.1: Comparison of rate-distortion performance of BMA, BMA-RA and MMA.

3.3.5 Comparison of all algorithms using DPCM coding

This section analyzes the impact of the DPCM coding on the MMA, BMA and BMA-RA performance. The DPCM coder is lossless and is used here to encode the disparity maps estimated by all the algorithms. The entropy is then computed on all the coded disparity maps and the approximated bitrates are given in bits per pixel (bpp). Figure 3.24 compares the performance of MMA-DPCM, BMA-DPCM and BMA-RA-DPCM algorithms. Tests have been conducted using blocks of size equal to 4×4, 6×6 and 8×8.

Results obtained with a same size of blocks have been represented with the same color. The values of λ tested ranges from 0 to 8000 for MMA-DPCM and 0 to 500 for BMA-RA-DPCM. For each block size, note that the curves corresponding to the MMA-DPCM and the BMA-RA-DPCM start from the same point obtained with $\lambda = 0$. This point is also the one achieved by the BMA-DPCM as for $\lambda = 0$, the selection of the disparities takes only into account the minimization of the predicted image distortion. It can be noted that in the presented figure, the prediction quality (in dB) achieved by the MMA-DPCM and the BMA-RA-DPCM are the same as the one obtained by the MMA and the BMA-RA. Only the bitrate of the disparity map has been changed due to the DPCM encoder. An analysis of these curves shows that the MMA-DPCM still achieves better rate-distortion performance than the BMA-RA-DPCM. For example, using blocks of size 4×4 and $\lambda = 0$, both algorithms gives the R-D point (26.84 dB; 0.4 bpp). Allowing a little increase in the predicted image distortion, the MMA-DPCM yields to the point (26.84 dB; 0.109 bpp) for $\lambda = 6.8$ while the BMA-RA-DPCM provides the point (26.83 dB; 0.257 bpp) with $\lambda = 3$. For equivalent prediction quality, the MMA-DPCM still permits a higher reduction of the bitrate of the disparity map as compared to the BMA-RA-DPCM.

Additional simulation results comparing the MMA-DPCM, the BMA-RA-DPCM and the BMA-DPCM performance are given in Table 3.2. The parameters considered for all the algorithms are provided at the top of the table.

Setting parameters	BMA-DPCM: $d \in [-15, -14.5, \ldots, 13.5, 14]$ BMA-RA-DPCM: $d \in [-15, -14.5, \ldots, 13.5, 14]$ MMA-DPCM: $d \in [-15, -14.5, \ldots, 13.5, 14], \beta = 0.2, M = 1$

Stereoscopic		**BMA-DPCM**		**BMA-RA-DPCM**			**MMA-DPCM**		
Images	**Block Size**	**PSNR**	**Bitrate**	**PSNR**	**Bitrate**	λ	**PSNR**	**Bitrate**	λ
Wood2	4x4	27.13	0.4	26.83	0.257	3	26.84	0.109	6.8
	6x6	26.7	0.175	26.41	0.112	3	26.39	0.034	10.6
	8x8	26.54	0.095	26.25	0.06	3	26.25	0.016	14.3
Stereo_2	4x4	32.7	0.273	32.7	0.273	0	32.4	0.18	4.6
	6x6	31.86	0.102	31.86	0.102	0	31.75	0.066	3.1
	8x8	31.59	0.057	31.59	0.057	0	31.43	0.037	4.6
Stereo_67	4x4	41.95	0.363	41.95	0.363	0	41.6	0.223	0.1
	6x6	40.68	0.157	40.68	0.157	0	40.4	0.065	0.1
	8x8	40.73	0.088	40.73	0.088	0	40.48	0.035	0.1
Tsukuba	4x4	34.12	0.23	34.12	0.23	0	33.82	0.095	2.7
	6x6	32.21	0.089	32.21	0.089	0	31.97	0.044	4.5
	8x8	31.48	0.047	31.48	0.047	0	31.29	0.025	4.5
Stereo_13	4x4	29.99	0.328	29.63	0.215	3	29.58	0.212	4.6
	6x6	28.96	0.135	28.71	0.09	2	28.68	0.091	4.6
	8x8	28.54	0.075	28.38	0.049	2	28.31	0.05	4.6
Baby1	4x4	26.98	0.384	26.98	0.384	0	26.73	0.285	7.2
	6x6	26.02	0.17	26.02	0.17	0	25.8	0.12	7.2
	8x8	25.5	0.095	25.5	0.095	0	25.3	0.063	7.2
Barn2	4x4	34.51	0.20	34.51	0.20	0	34.3	0.126	1.2
	6x6	34.03	0.069	34.03	0.069	0	33.77	0.05	1.8
	8x8	33.08	0.034	33.08	0.034	0	32.88	0.03	1.8
Venus	4x4	31.6	0.233	31.6	0.233	0	31.41	0.153	3
	6x6	30.48	0.085	30.48	0.085	0	30.18	0.068	6.6
	8x8	29.75	0.041	29.75	0.041	0	29.74	0.039	0.3
Sawtooth	4x4	31.7	0.183	31.7	0.183	0	31.69	0.156	0.3
	6x6	30.82	0.069	30.82	0.069	0	30.81	0.065	0.3
	8x8	30.21	0.036	30.21	0.036	0	30.21	0.034	3
Bull	4x4	35.35	0.21	35.35	0.21	0	35.13	0.134	1.2

Stereoscopic		BMA-DPCM		BMA-RA-DPCM			MMA-DPCM		
Images	Block Size	PSNR	Bitrate	PSNR	Bitrate	λ	PSNR	Bitrate	λ
	6x6	34.79	0.073	34.79	0.073	0	34.76	0.055	0.3
	8x8	34.18	0.036	34.18	0.036	0	34.17	0.029	0.3

Table 3.2: Comparison of rate-distortion performance of BMA, BMA-RA and MMA using DPCM.

3.3.6 Performance of the MMA estimating dense disparity map

This section focuses on the performance of the MMA when estimating dense disparity map (using blocks of size 1×1). Results are compared with those obtained with the Dynamic Programing Algorithm (DPA). We used the one provided in the computer vision system toolbox of Matlab [50] which estimates blockwise disparity map.

For each block, this algorithm estimates a disparity by minimizing the Sum of Squared Differences (SSD) between the pixels of the blocks to be matched. Furthermore, a smoothness constraint is added to the cost function so that the selected disparity tends to be similar to the disparity of the neighboring already processed blocks.

Simulations results presented in this section have been performed on the stereoscopic image "Barn2". For both algorithms, disparities are selected amongst the set $[-15, -14, \ldots, 13, 14]$. The parameter of the MMA are set to $M = 1$ and $\beta = 0.02$. The performance of the MMA are compared to those of the DPA using blocks of size 3×3 providing the best quality of the predicted view for this algorithm. Using the DPA, the evaluated PSNR is equal to 23.15 dB for a bitrate of the disparity map equal to 3.05 bpp. The MMA can produce a predicted image with a much higher quality of 33.97 dB at a bitrate inferior to the previous one (2.88 bpp) taking $\lambda = 6.6$.

Figure 3.26 shows the original right view of the image "Barn2". The reconstruction obtained using the DPA and the MMA are given respectively in Figure 3.28 and Figure 3.27. The gain in the prediction quality can be observed here as the newspaper is better reconstructed in the case of the MMA.

For an equivalent quality of the predicted view (23.15 dB for the DPA and 23.81 dB for the MMA), the MMA achieves a significant gain in the reduction of the bitrate needed to encode the disparity map. While the DPA needs 3.05 bpp, the MMA needs only 0.83 bpp to achieve an equivalent quality. This is explained once again by the fact that the MMA tends to favor the use of identical disparities over the whole map, not only locally, and still ensures good rate-distortion performance by minimizing the joint entropy-distortion metric. This analysis can be confirmed by having a look at the disparity maps computed by the DPA and the MMA which are given respectively in Figure 3.29 and Figure 3.30.

3.3.7 MMA processing time and complexity

This section discusses the processing time of the MMA. For this, simulations have been performed under Matlab environment (version R2012a) on a Personal Computer having the following characteristics: Intel Core i5 processor (3.20 GHz) and a RAM of 4 GB. We present the results obtained using the stereoscopic image "Sawtooth" of size 380×434 pixels from the "Middlebury" dataset.

For a given target bitrate, the MMA has to find adequate values of λ and of block size to have the best prediction quality. Figure 3.31 shows the MMA processing time in function of given target bitrates from 0.006 to 1.5 bpp. In this simulation, the MMA selects the disparities in the set $[-15, -14, \ldots, 14]$. The block size are chosen amongst sizes from 1×1 to 16×16 and the adequate value of λ is chosen amongst a defined set of values ranging from 0 to 8000. It can be observed that the MMA processing time globally decreases with the target bitrate, starting from 1677 seconds for a bitrate equal 1.5 bpp up to 85 seconds for a low bitrate of 0.006 bpp. It can be explained by the fact that for higher bitrates, the MMA tests smaller sizes of blocks, resulting in an increase in the computation time. The processing time of the MMA is relatively high compared to that of the BMA. In this thesis, we have focused on finding efficient disparity map estimation algorithms in terms of Rate-Distortion performance. No special care have been taken to optimize or reduce the processing time. We are optimist that investigations dedicated to the reduction of the processing time of the MMA could improve its rapidity by for example executing tasks of the MMA in parallel.

The complexity of the MMA corresponding to the number of elementary operations needed to compute one Rate-Distortion point (i.e. for a specific value of λ) is given by:

$$N_w \left[(XY)(3N_xN_y - 1) + M(XY - 1)\right] + M(XY)(14N_w - 1) + 2MXYN_w \tag{3.17}$$

This complexity increases with the number of disparities in the search area, the size of the blocks, the size of the image and the parameter M.

3.4 Conclusion

In this chapter, we have presented a new stereoscopic matching optimization algorithm called MMA. The MMA sequentially estimates a disparity map reducing the distortion of the predicted view as well as the bitrate needed to encode the disparity map. This is achieved by the minimization of a joint entropy-distortion metric.

The MMA assumes that the disparities of the whole map follow a specific distribution to compute this metric. Furthermore, the MMA sequentially builds a tree where each path corresponds to a partial disparity map. The MMA also extends only the M best paths in terms of rate-distortion performance at each depth of the tree, thus reducing the computational load of

the algorithm. The analysis of the simulation results supports the idea that the MMA reduces the bitrate of the disparity map by favoring similar disparities.

Simulation results have also shown the advantage of the MMA over the BMA using or not a smoothness constraint. All the images used for these simulations were assumed to be rectified. The next chapter presents an extension of the MMA that have been adapted to the case of non-rectified stereoscopic images but still works with rectified images. The extended MMA shows to improve the performance of the MMA on both types of images.

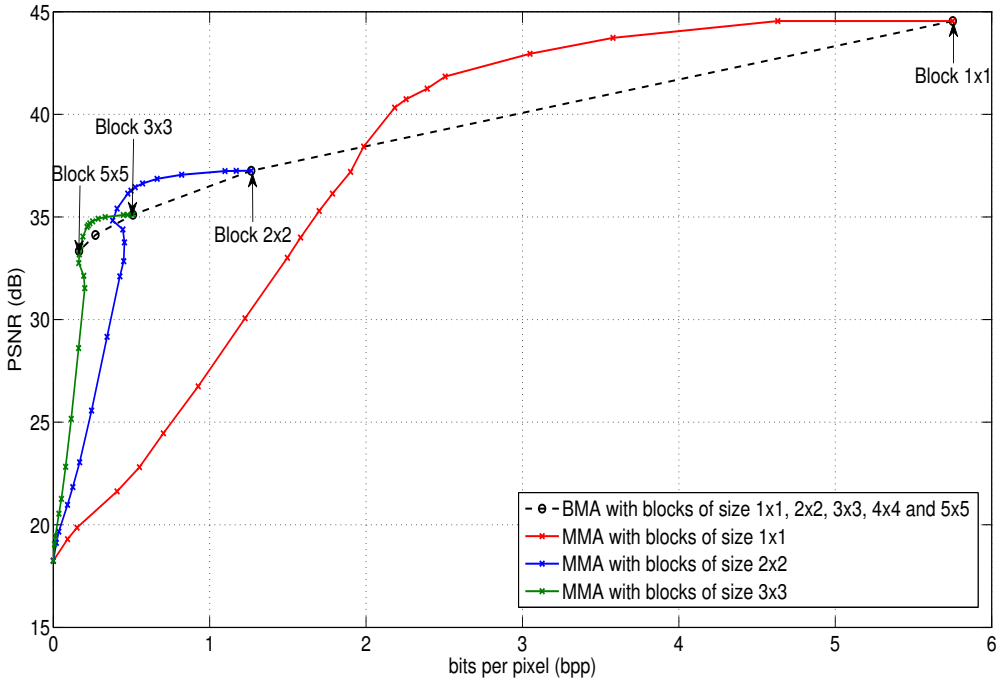

Figure 3.9: Comparison of the MMA and BMA performance on "Tsukuba" using block sizes from 1×1 to 3×3 for MMA and 1×1 to 5×5 for BMA; and bitrates belong to $[0, 5.8]$bpp.

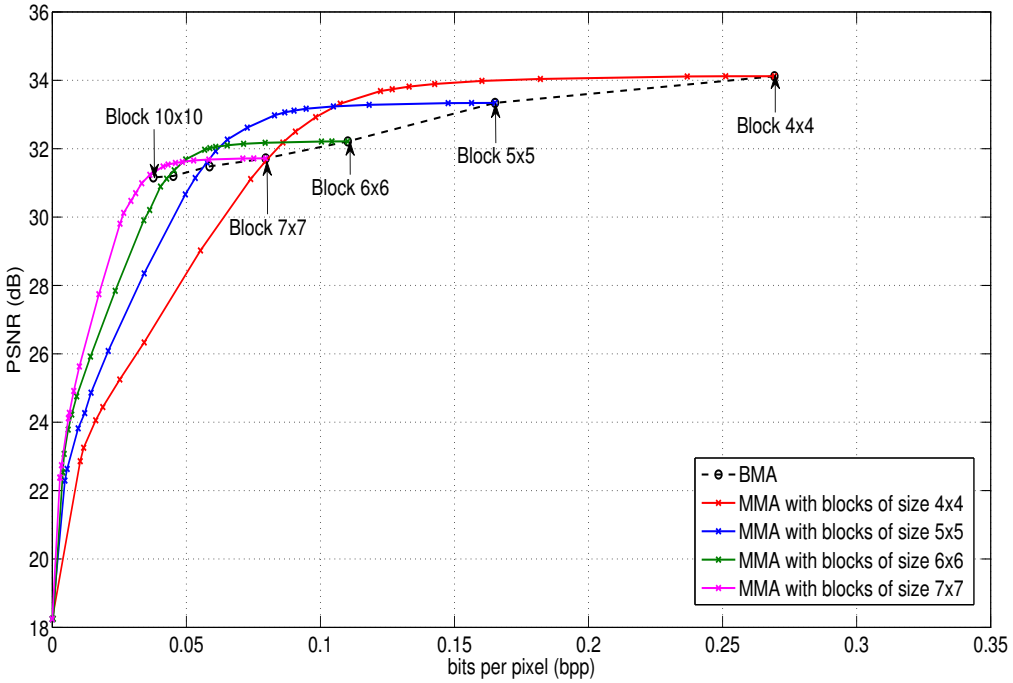

Figure 3.10: Comparison of the MMA and BMA performance on "Tsukuba" using block sizes from 4×4 to 7×7 for MMA and 4×4 to 10×10 for BMA; and bitrates belong to $[0, 0.27]$bpp.

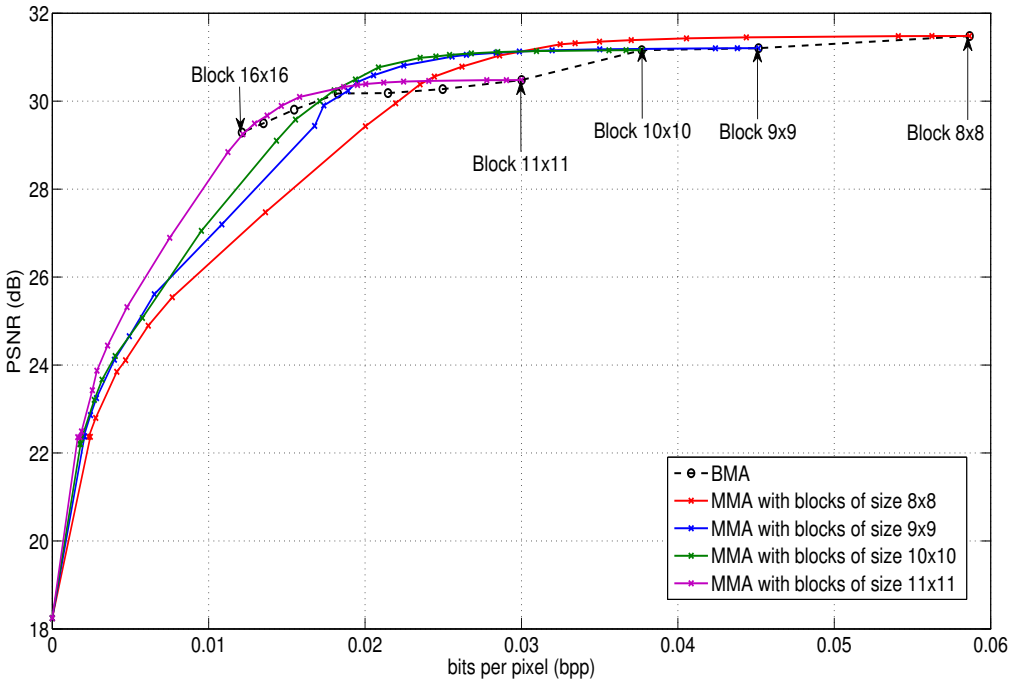

Figure 3.11: Comparison of the MMA and BMA performance on "Tsukuba" using block sizes from 8×8 to 11×11 for MMA and 8×8 to 17×17 for BMA; and bitrates belong to $[0, 0.06]$bpp.

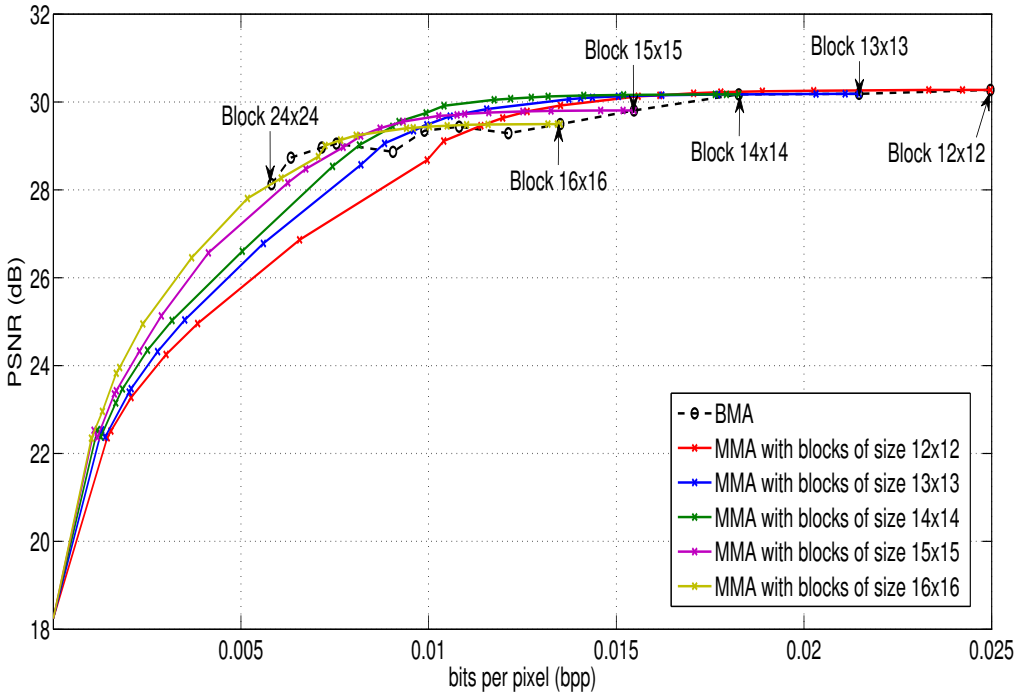

Figure 3.12: Comparison of the MMA and BMA performance on "Tsukuba" using blocksizes from 12×12 to 16×16 for MMA and 12×12 to 24×24 for BMA; and bitrates belong to $[0, 0.025]$bpp.

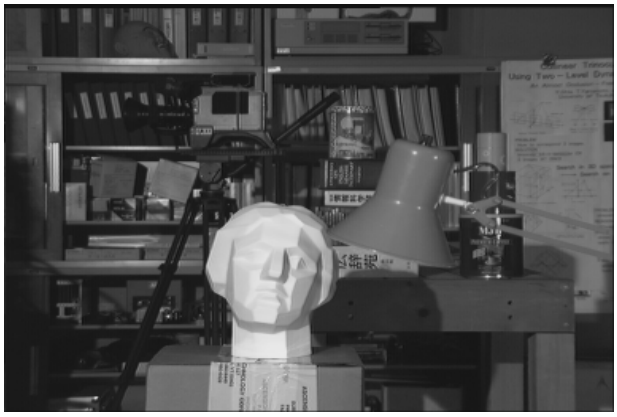

Figure 3.13: Original right image "Tsukuba".

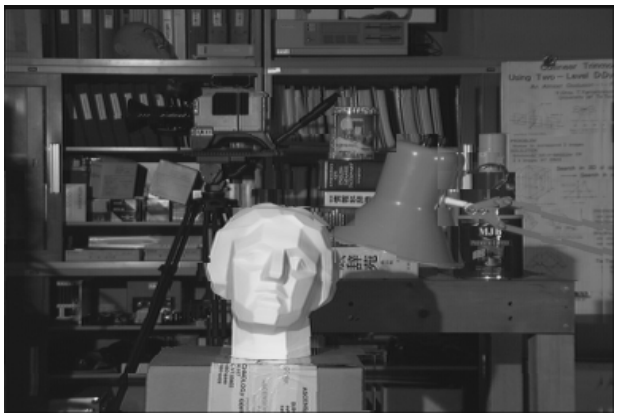

Figure 3.14: Reconstruction using BMA (30.18dB; 0.022bpp).

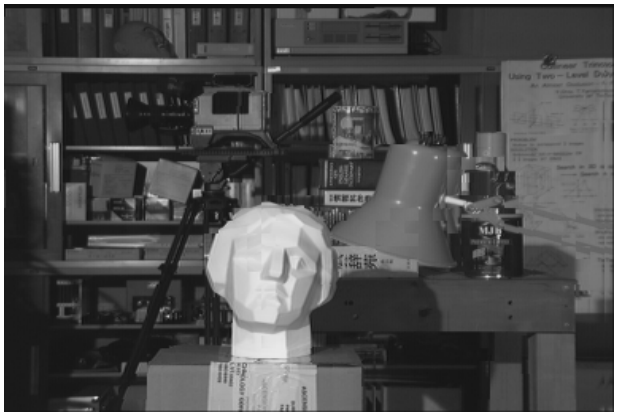

Figure 3.15: Reconstruction using the proposed algorithm (30.81dB; 0.022bpp) with $\lambda = 9.9$.

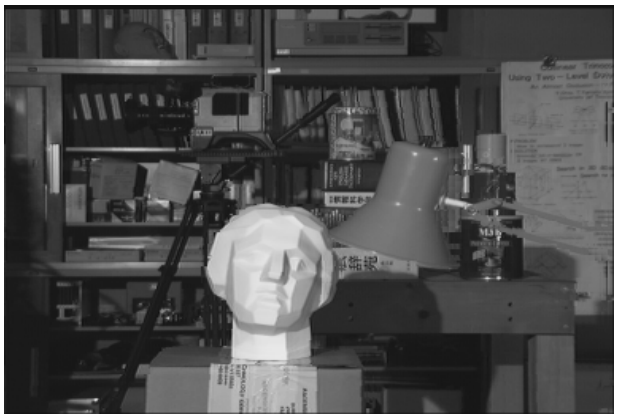

Figure 3.16: Reconstruction using BMA (34.12dB; 0.27bpp).

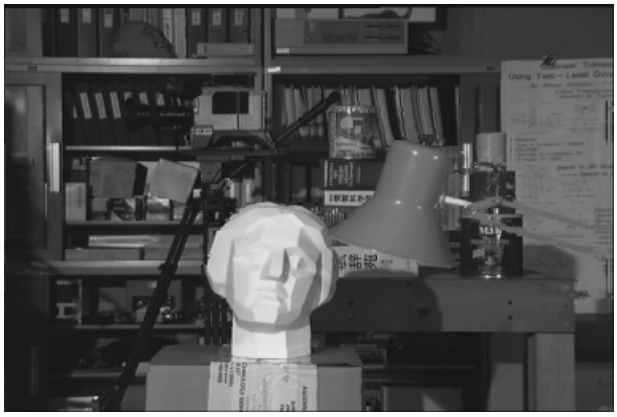

Figure 3.17: Reconstruction using MMA algorithm (33.89dB; 0.14bpp) with $\lambda = 1.8$.

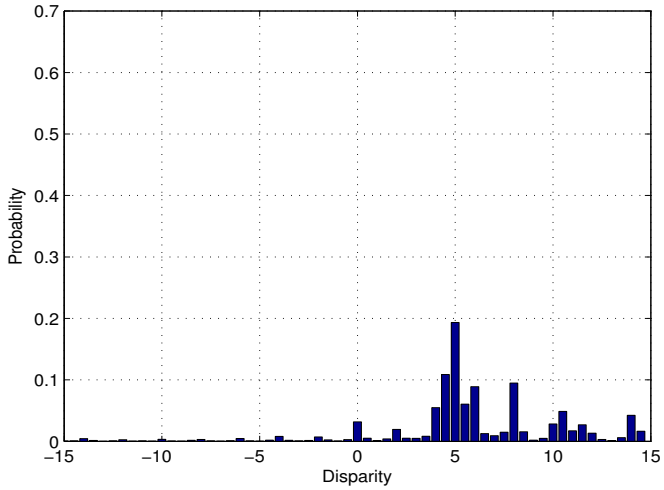

Figure 3.18: Histogram of the disparity map based on the BMA (34.12dB; 0.27bpp).

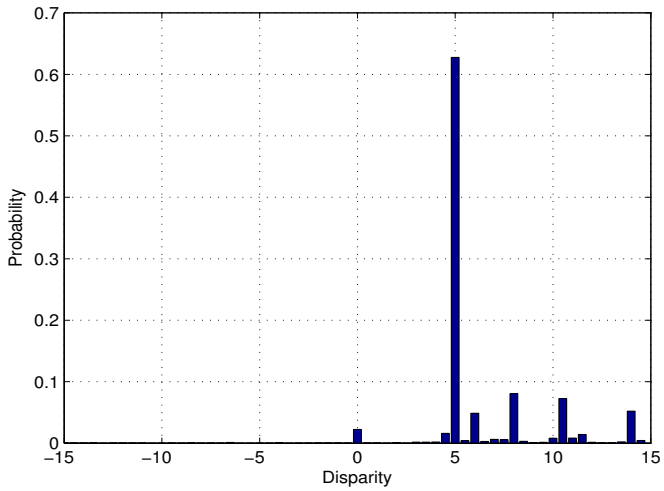

Figure 3.19: Histogram of the disparity map based on the MMA algorithm (33.89dB; 0.14bpp) with $\lambda = 1.8$.

Figure 3.20: Disparity map using BMA (34.12dB; 0.27bpp).

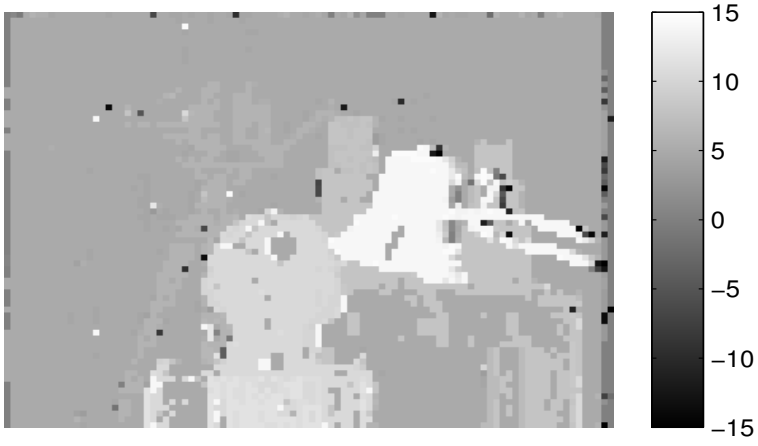

Figure 3.21: Disparity map using the MMA algorithm (33.89dB; 0.14bpp) with $\lambda = 1.8$.

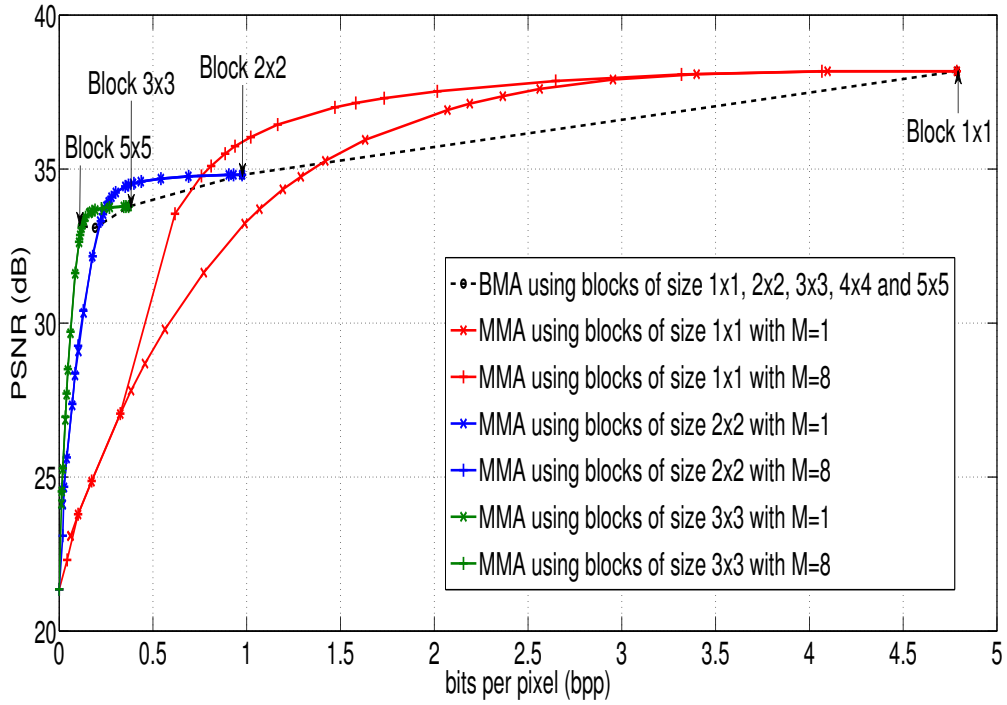

Figure 3.22: Comparison of MMA and BMA performance using "Barn2" with block sizes from 1×1 to 3×3 for MMA and 1×1 to 5×5 for BMA.

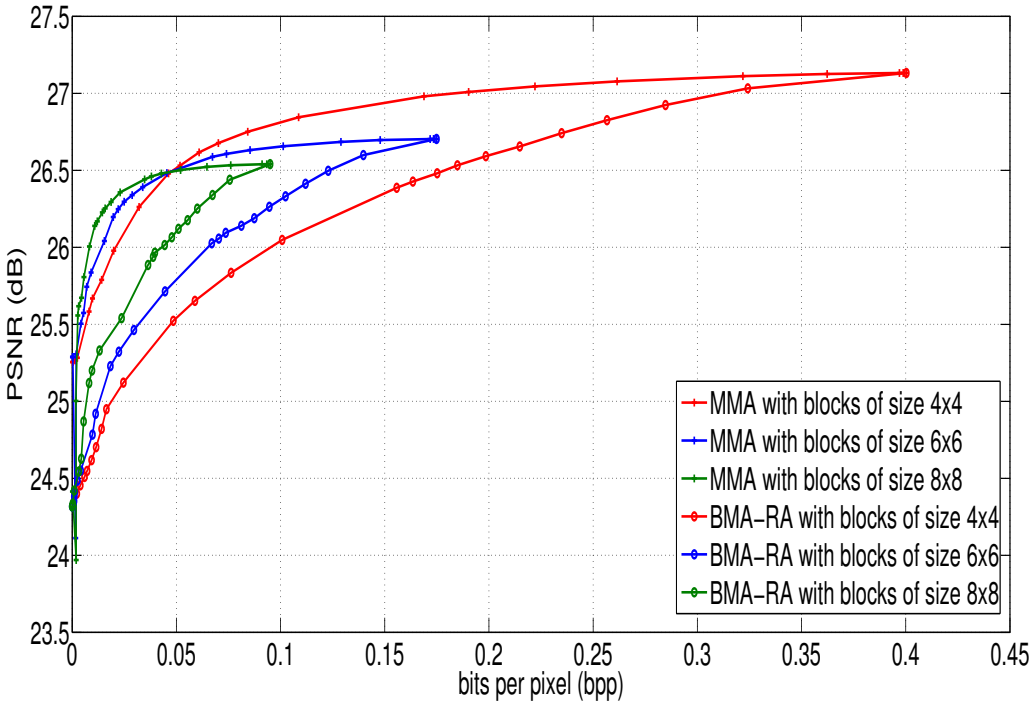

Figure 3.23: Comparison of MMA and BMA-RA performance on "Wood2".

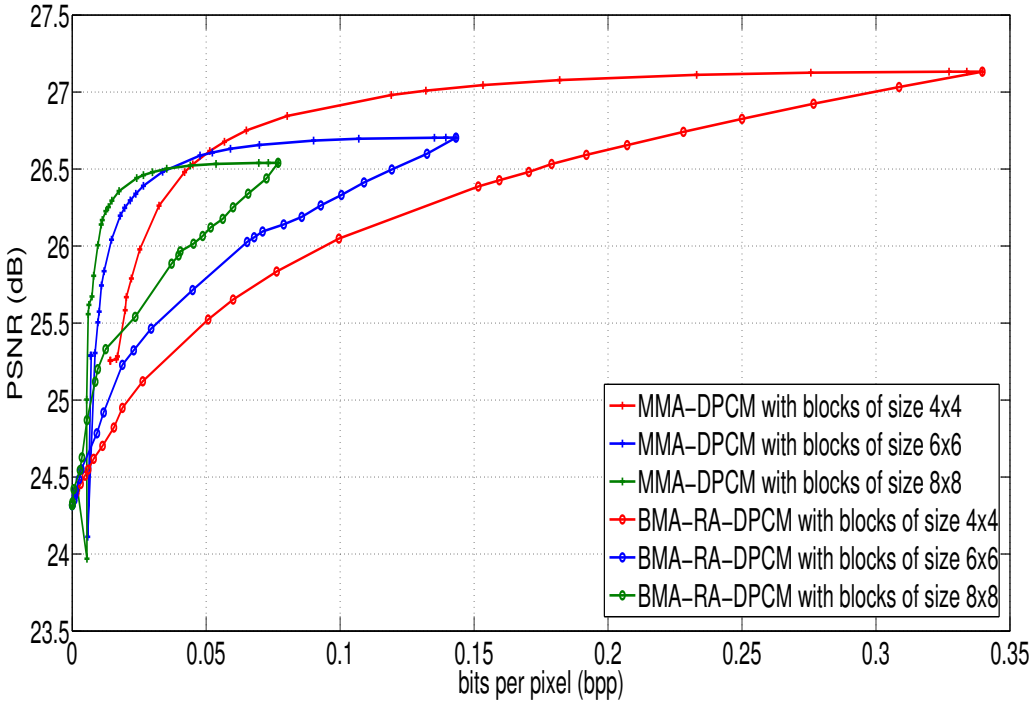

Figure 3.24: Comparison of the MMA-DPCM, BMA-DPCM and BMA-RA-DPCM performance on Wood2.

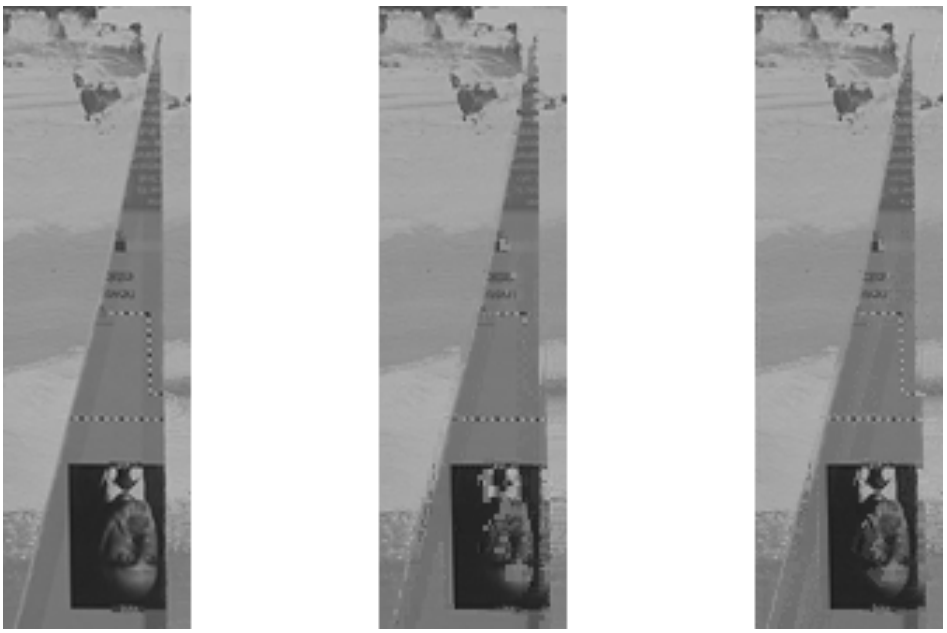

Figure 3.25: From left to right close-up views of the original right image "Barn2", reconstructed image using the BMA; and the reconstructed image using MMA.

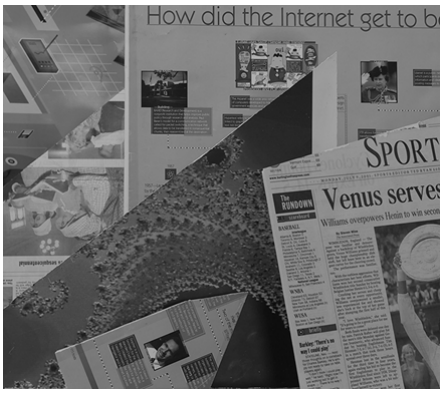

Figure 3.26: Original right image.

Figure 3.27: Reconstructed image using dynamic programming (23.15dB; 3.05bpp).

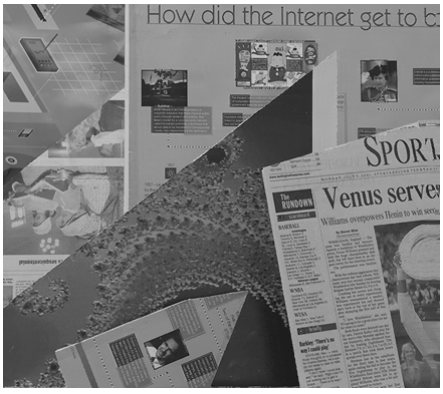

Figure 3.28: Reconstructed image using the MMA (33.97dB; 2.88bpp).

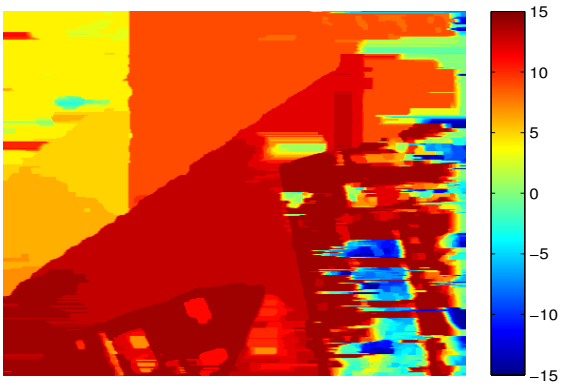

Figure 3.29: Disparity map estimated using dynamic programming (23.15dB; 3.05bpp).

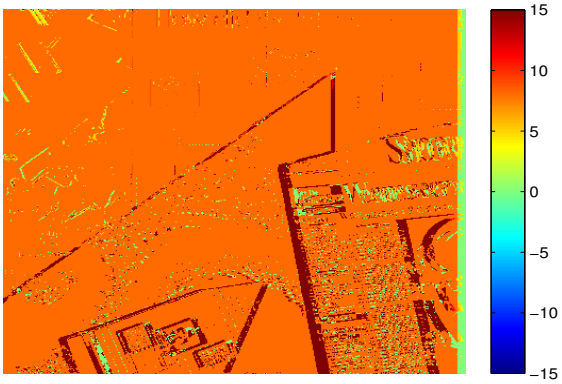

Figure 3.30: Disparity distribution using the MMA (23.81dB; 0.87bpp).

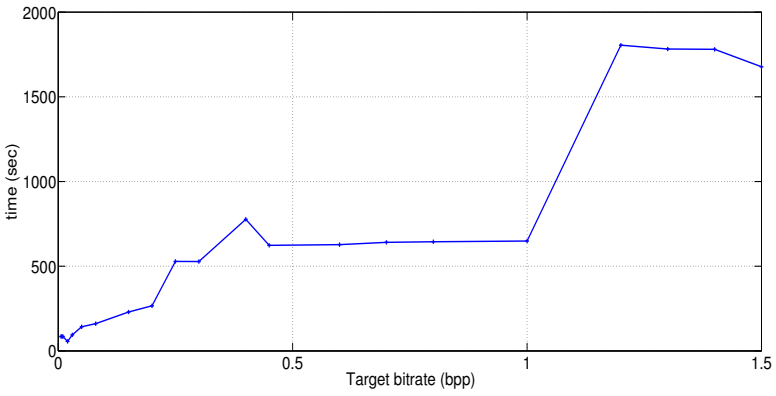

Figure 3.31: MMA processing time according to target bitrates.

Chapter 4

Extension of the MMA to non-rectified stereoscopic images: EMMA

This chapter presents an extension of the MMA described in the previous chapter, called EMMA. The extended algorithm can also be applied to non-rectified stereoscopic images while taking advantage of the MMA performance. We first introduce some modifications in the notations. The adaptation of the stereo-matching problem and the MMA to the case of non-rectified stereoscopic images is then presented. This work has also been presented in [10].

4.1 Notations and assumptions

In the case of non-rectified stereoscopic images, corresponding blocks do not necessarily belongs to the same scan lines. Thus, disparities have two components, one horizontal and one vertical. The disparity $d(i_b, j_b)$ associated to the block at position (i_b, j_b) in the right image (see Section 3.1.1) is redefined as $d(i_b, j_b) = (d_x, d_y)$, with d_x the vertical component of the disparity and d_y its horizontal component as represented in Figure 4.1.

The EMMA uses an enlarge sliding matching window W to perform the block-matching, compared to the MMA to consider also vertical displacements. As shown in Figure 4.1., it is centered on the origin of the block to be match. Its size becomes $N' = W_x \times W_y$ where W_x and W_y are respectively the vertical and horizontal size of the window defined as $W_x = w_{x\_max} - w_{x\_min} + 1$ and $W_y = w_{y\_max} - w_{y\_min} + 1$. In the same way as previously, w now designates a two dimensions vector $w = (w_x, w_y)$ such as $w_x = \{w_{x\_min}, \ldots, w_{x\_max}\}$ and $w_y = \{w_{y\_min}, \ldots, w_{y\_max}\}$. Moreover, the disparity denoted d_l^k representing the disparity of the k-th path at l-th depth (see Section 3.2.1) is now redefined as $d_l^k = ({d_{xl}}^k, {d_{yl}}^k)$.

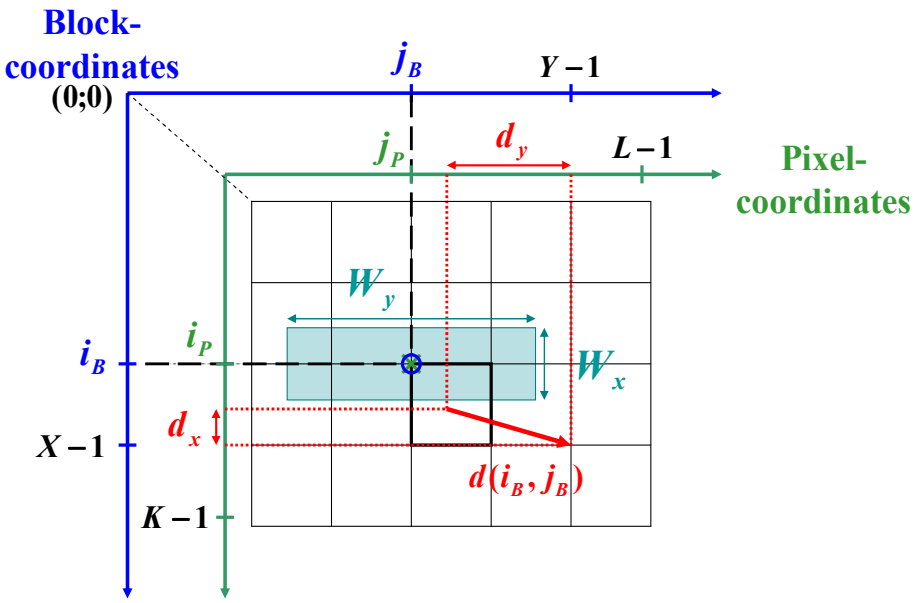

Figure 4.1: EMMA notations.

4.2 Presentation of the extended MMA

Considering the case of non-rectified stereoscopic images, the rate-distortion optimization problem remains unchanged and is still expressed as the following Lagrangian minimization:

$$\widehat{\mathbf{d}} = \text{argmin} J(\lambda, \mathbf{d}) = \text{argmin}(E_G(\mathbf{d}) + \lambda H(\mathbf{d})). \tag{4.1}$$

In this equation, the global distortion of the predicted image $E_G(\mathbf{d})$ given in equation (3.1.1) is modified as follows to consider the two components of the disparity vectors as follows:

$$E_G(\mathbf{d}) = \sum_{i_b=0}^{X-1} \sum_{j_b=0}^{Y-1} E_B(i_b, j_b) \tag{4.2}$$

$$\text{with} \quad E_B(i_b, j_b) = \sum_{u=0}^{x-1} \sum_{v=0}^{y-1} (I_l(i_p + u + d_x, j_p + v + d_y) - I_r(i_p + u, j_p + v))^2. \tag{4.3}$$

The EMMA brings a sub-optimal solution to the rate-distortion optimization problem by following exactly the same steps as the MMA summarized in Table 3.2 where the computation of the local distortion and the entropy of each branch has been updated as explained below.

At each depth of the tree, the EMMA retains the M-best paths and extends each of them by N' new branches. Each new branch corresponds to the choice of a disparity $w = (w_x, w_y)$ inside the sliding matching window W. A local distortion E_{bt}^w is associated to each new branch and its computation is updated as follows:

$$E_{bt}^w = \sum_{u=0}^{x-1} \sum_{v=0}^{y-1} (\widehat{I}_r(i_p + u, j_p + v) - I_r(i_p + u, j_p + v))^2 \tag{4.4}$$

where $\widehat{I}_r(i_p+u, j_p+v) = I_l(i_p+u+w_x, j_p+v+w_y))$ as in MMA.

In the same way, a bitrate is associated to each new branch and is approximated by the computation of the entropy of the corresponding path updated as:

$$H_t^k = -\sum_{w_x=w_{x\_min}}^{w_{x\_max}} \sum_{w_y=w_{y\_min}}^{w_{y\_max}} p_t^k(d=w)\log_2(p_t^k(d=w)) \quad \text{for} \quad k=1,...M\times N'. \tag{4.5}$$

The estimation of the disparity distribution probability include the computation of the uniform distribution probability (see equation (3.13)) which is updated to consider the larger searching window:

$$p_U(d=(w_x,w_y)) = \frac{1}{N'}. \tag{4.6}$$

4.3 Simulation results and discussions

This section provides simulation results to evaluate the performance of the EMMA. Comparisons are carried out with the BMA. As previously, the left image is taken as the reference and the right image as the one to be predicted. The quality of the predicted image is evaluated with the Peak Signal-to-Noise Ratio (PSNR) computed between the original right view and the predicted one. The bitrate of the disparity map is approximated by the entropy in bits per pixels (bpp) calculated using the empirical disparity distribution. Simulations have been conducted on non-rectified and also rectified stereoscopic images taken from the "Middlebury" and "CMU" datasets [47, 51]. The first subsection provides results performed on non-rectified stereoscopic images and the second evaluates the performance of the EMMA in the case of rectified stereoscopic images.

4.3.1 Simulation results on non-rectified stereoscopic images

This section provides an analysis of the performance of the EMMA on a non-rectified stereoscopic image ("sand") taken from the "Middlebury" dataset. We first analyse the impact of the choice of the sliding matching window size dimensions on the predicted image quality. The second part shows the performance achieved by both the EMMA and the BMA on this stereoscopic image considering the results of the first analysis.

4.3.1.1 Impact of the window size choice

In what follows, we have first studied the impact of the sliding matching window dimensions $W_x \times W_y$ on the quality of the predicted image using the stereoscopic image "sand". This evaluation is performed by testing several sizes of symmetrical sliding matching window $W_x \times W_y = (w_{x\_max}-w_{x\_min}+1)\times(w_{y\_max}-w_{y\_min}+1)$ centered on the block to be matched where

$w_{x\_max} = -w_{x\_min} = \{0, 1, 2, \ldots, 20\}$ and $w_{y\_max} = -w_{y\_min} = \{0, 1, 2, \ldots, 150\}$. For each sliding matching window, the BMA estimates the disparity map achieving the best reconstruction of the right view, which quality is measured by the PSNR. A value of PSNR is thus associated to each different size of sliding matching window. Simulations have been carried out with blocks of size 8×8 pixels for the block-matching.

Figure 4.2 shows the performance achieved by the BMA using each different window size: the pixel located at position (i, j) (where $i = \{0, 1, 2, \ldots, 20\}$ and $j = \{0, 1, 2, \ldots, 150\}$) represents the performance achieved by the BMA using a window which bouderies are set as $w_{x\_max} = -w_{x\_min} = i$ and $w_{y\_max} = -w_{y\_min} = j$. Therefore, the window used has a size of $W_x \times W_y = (i \times 2 + 1) \times (j \times 2 + 1)$. For example, the pixel located at position $(i, j) = (10, 60)$ represents the performance of the BMA using a window of size $W_x \times W_y = 21 \times 121$ centered on the blocks to be matched. The PSNR values have been normalized to be represented using a grayscale. Thus, the lowest PSNR value obtained using the BMA (14 dB) is represented in black color and the highest value (24 dB) is represented using white color.

One can distinguish in this figure an horizontal line at line $i = 10$ beyond which almost all the pixels have almost the same white color. This represents the fact that taking a window size where w_{x_max} is greater that 10 vertically does not improve significantly the gains in the quality of the predicted view. This results conducted us to restrict the window size in the following tests performed using the EMMA as well as the BMA.

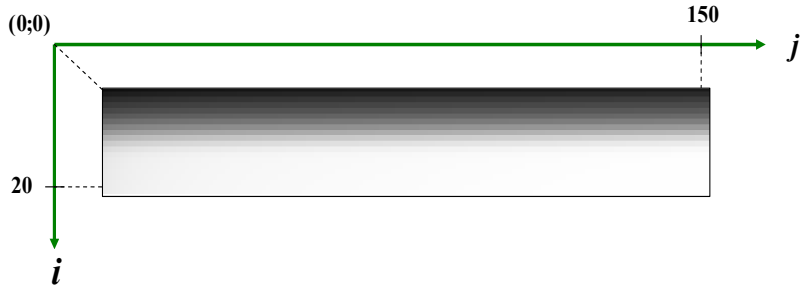

Figure 4.2: BMA performance on "Sand" using different sliding matching window sizes.

4.3.1.2 Performance evaluation of the EMMA

The EMMA and the BMA performance for the stereoscopic image "sand" are given in Figure 4.3. The results are given as the PSNR (in dB) of the predicted view versus the bitrate of the disparity map (in bpp). Simulations have been conducted using two sizes of blocks for both algorithms: 8×8 illustrated using red color and 10×10 using blue color. Three sizes of sliding matching window have also been tested for the two algorithms: 1×31 represented by '+' symbol; 11×31 by '$\times$' symbol; and 21×31 by 'o' symbol. The BMA performance is given by the first point at the top right of each curve, which also corresponds to the EMMA performance using λ=0, as both

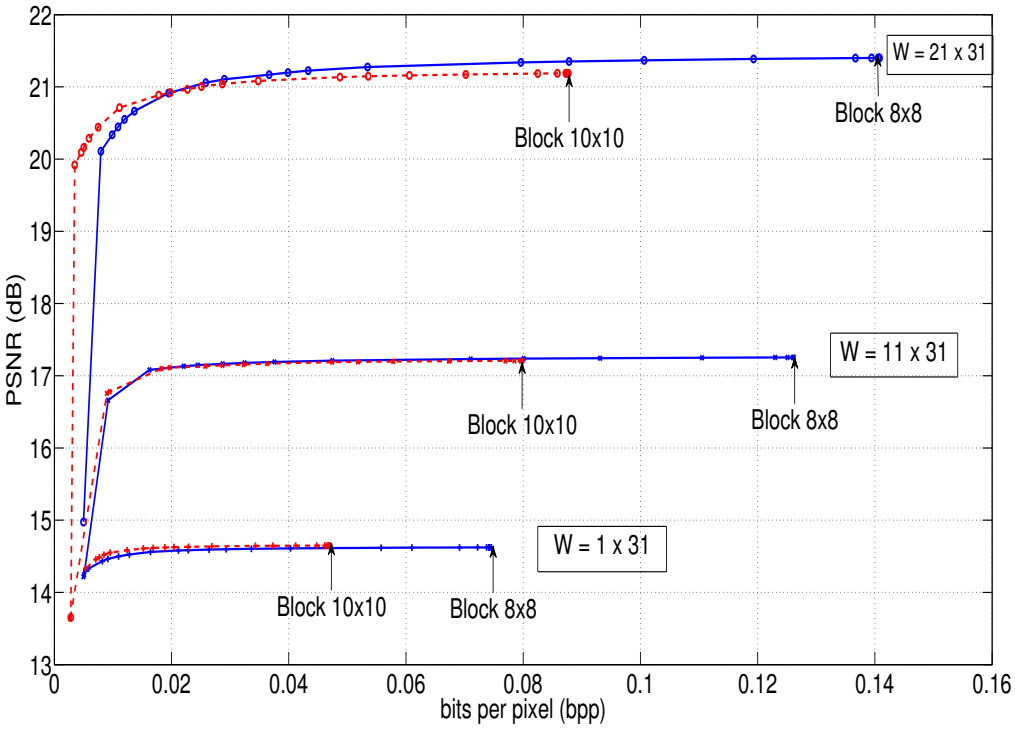

Figure 4.3: Rate-distortion optimization on the non-rectified stereo image "Sand".

algorithm only intend to minimize the distortion of the predicted image in this configuration. From each of these points, starts a curve corresponding to the EMMA performance obtained for different values of λ.

For blocks of size 8×8 and a window of size 1×31, the BMA leads to the R-D point (14.6 dB, 0.07 bpp), while the EMMA achieves the same PSNR but with a much lower bitrate of 0.02 bpp using the same parameter for a given value of λ. The EMMA thus allows a gain of 71% in comparison with the BMA. For a larger window of size 21×31, the BMA leads to the R-D point (21.4 dB, 0.14 bpp), while the EMMA reduces the bitrate down to 0.03 bpp by allowing a little decrease in the prediction quality (21.1 dB). The EMMA achieves a gain of 79% in terms of bitrate allowing this little decrease. These curves also show that for blocks of size 10×10, the BMA yields to the R-D point (21.2 dB, 0.09 bpp), while the EMMA achieves a gain of 0.2 dB over the BMA at the same bitrate using smaller blocks of size 8×8.

4.3.2 Simulation results on rectified stereoscopic images

In what follows, the impact of the choice of the sliding matching window dimensions on the performance when the algorithm is performed on the rectified stereoscopic image "tsukuba" is discussed. Then the performance of the EMMA and the BMA on this image is discussed.

4.3.2.1 Impact of the window size choice

In this part, we present simulation results that have been obtained testing several sizes of the sliding matching window taking the same parameter as in Section 4.3.1.1. Just as previously, for each window size, the BMA estimates a disparity map and the PSNR of the resulting predicted view is computed. The PSNR values corresponding to each size of window have been normalized to be represented using a grayscale. More precisely, the PSNR values have been divided by the maximum PSNR value obtained by the BMA such as all normalized values ranges between 0 and 1 and can be represented using a grayscale. Figure 4.4 shows the performance of the BMA where the black color is associated to the lowest PSNR (17 dB) obtained by the BMA for a given size of window, and the white color is associated to the highest value of PSNR (32 dB).

We notice that the pixels located at a position (i, j) such as $i \geq 1$ and $j \geq 14$ have very close intensity values. It reflects the fact that taking window boundaries $w_{x\_max} \geq 1$ and $w_{y\_max} \geq 14$ does not improve significantly the quality of the predicted view.

In what follows, we restrict the window vertical boundary to $w_{x\_max} > 3$ (i.e. $W_x > 7$) and the horizontal boundary to $w_{y\_max} > 15$ (i.e. $W_y > 31$) for the next simulations performed on the stereoscopic image "tsukuba".

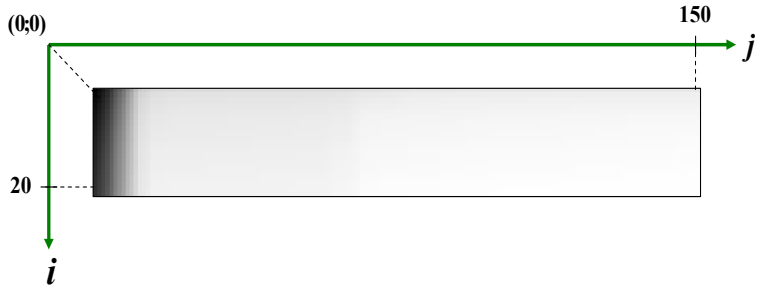

Figure 4.4: BMA performance on "Tsukuba" using different sliding matching window sizes.

4.3.2.2 Performance evaluation of the EMMA

Figure 4.5 shows the performance of the EMMA and the BMA performed on the rectified image "tsukuba". For both algorithms, the same sizes of blocks (4×4 to 6×6) and sizes of the sliding matching window (1×31 and 7×31) have been tested. Results obtained using a window of size 1×31 are represented using blue color while the results obtained using a window of size 7×31 are represented using red color for both algorithms.

The BMA performance is given by the very first points at the the top right of each curve which joints a dashed lines. They are represented using '$\diamond$' symbols with blue color for a window of size 1×31, 'o' symbols in red color for a window of size 7×31. From each of these points, starts a curve showing the EMMA performance using the same parameter as the BMA (block size and window size) and several values of λ. The first point of the EMMA coincides with the

BMA performance as it is obtained using $\lambda = 0$ and in this condition, both algorithms aim at minimizing only the distortion of the predicted image.

When observing the set of blue RD curves obtained using a window of size 1×31 for both algorithms, we note that the BMA leads to the RD point (31.7 dB, 0.09 bpp) using blocks of size 6×6. For the same bitrate, the EMMA achieves a gain of 0.9 dB on the quality of the predicted view using smaller blocks of size 5×5. Although this is a rectified stereoscopic image, when increasing vertically the boundaries of the window such as its dimensions become 7×31, the EMMA still achieves better performance than the BMA using the same window size and also than the EMMA itself using a horizontal window (i.e. 1×31). For example, at the bitrate of 0.13 bpp, the BMA yields to a predicted view with a PSNR of 32.6 dB using blocks of size 6×6 while the EMMA leads to 34.1 dB using smaller blocks of size 4×4, resulting in a gain of 1.5 dB.

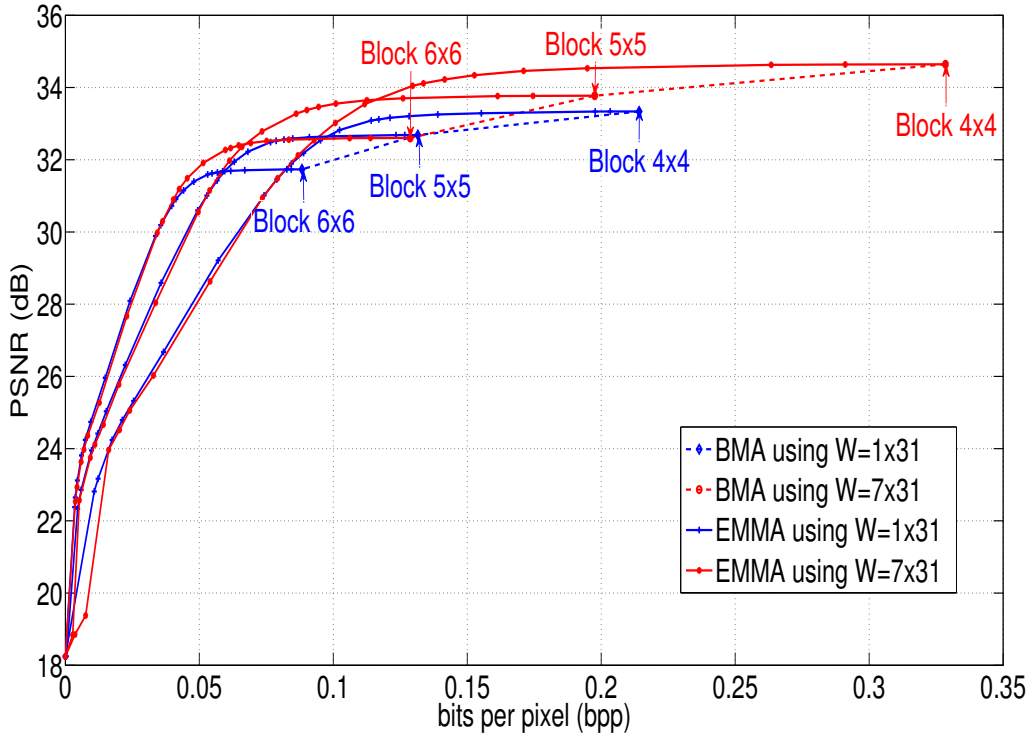

Figure 4.5: Rate-distortion optimization on the rectified stereo image "Tsukuba".

4.4 Conclusion

The block-based disparity map estimation algorithm developed in the previous chapter have been extended to non-rectified stereoscopic images. This algorithm follows exactly the same step as the MMA but some modifications have been made to consider the two dimensions of the disparity vectors in the particular case of non-rectified images. Simulations results still show the advantage of the proposed method in comparison with the BMA.

Benefits of the EMMA in terms of rate-distortion performance have also been observed for rectified stereoscopic images.

Moreover, enlarging the size of the sliding matching window increases the performance of the algorithm but at the expense of an increase in the computational complexity, and thus the processing time of the algorithm. Finding the optimal solution to the rate-distortion optimization problem involves to test a gigantic number of disparity maps inside a search area which size rapidly increases with the size of the sliding matching window. In order to reduce the size of the search area in which the sub-optimal solution is to be found, we propose to start the algorithm with an initial solution and then to explore other disparity maps obtained from the initial one by minimizing the rate-distortion cost. Therefore, the search area and thus the computational complexity of the algorithm can significantly be reduced. This strategy will be explored in the next chapter for rectified stereoscopic images.

Chapter 5

Reference-based block matching algorithm: R-algorithm

In this chapter, we present the reference-based block-matching algorithm, called R-algorithm (for Reference), which estimates the disparity map considering the disparity map computed by the BMA as an initial solution to the rate-distortion optimization problem. Indeed, this map is taken as a reference and is then successively modified as long as improvements are observed in terms of rate-distortion performance. The map yielded by the BMA is the best solution minimizing the distortion of the predicted view. By modifying it successively when improvements are observed allows to explore disparity maps close to the initial one, requiring a smaller bitrate. This algorithm finds a sub-optimal solution to the Rate-Distortion optimization problem with a smaller computational complexity in comparison with the M-algorithm.

The rate-distortion optimization problem is first recalled, then the principles of the R-algorithm are presented before discussing the experimental results. The contributions of this chapter have been presented in [12, 13].

5.1 Rate-distortion optimization problem

This section briefly recall some notations and assumptions to formulate the optimization problem.

5.1.1 Assumptions and notations

In this chapter, we consider the stereoscopic images as rectified. The left view is taken as the reference view and the right view is the one to be predicted.

The notations introduced in 3.1.1 summarized in Fig 3.1 are kept identical (concerning the left and right images, the interpolated left image, the sizes of the images, the block coordinates, the pixel coordinates, the sliding matching window and the disparities).

5.1.2 Formulation of the Rate-Distortion optimization problem

The purpose of the work is to achieve the best compromise between the minimization of the bitrate allocated to the disparity map and the minimization of the distortion of the predicted image. This issue is still formalized with a Lagrangian minimization as in section 3.1.2 but some small modifications have been made on the notations for more accuracy and are introduced below.

The global distortion of the predicted image is expressed as:

$$E_G(\mathbf{d}) = \sum_{i_b=0}^{X-1} \sum_{j_b=0}^{Y-1} E_B((i_b, j_b), d(i_b, j_b)), \tag{5.1}$$

with $E_B((i_b, j_b), d(i_b, j_b))$ the local distortion induced by the disparity $d(i_b, j_b)$ for the block at position (i_b, j_b) in the original view I_r. It is defined as:

$$E_B((i_b, j_b), d(i_b, j_b)) = \tag{5.2}$$
$$\sum_{u=0}^{N_X-1} \sum_{v=0}^{N_Y-1} (\widehat{I}_r(i_p + u, j_p + v) - I_r(i_p + u, j_p + v))^2.$$

Note that the definition of the global distortion (see Equation (5.1)) is similar to the definition given in the chapter 3 by Equation (3.1.1). The local distortion which was denoted $E_B(d(i_B, j_B))$ has simply been renamed $E_B((i_b, j_b), d(i_b, j_b))$. This more accurate notation is used later in this chapter also.

The bitrate of the disparity map is still approximated by its entropy which is computed on the empirical disparity distribution probability as follows:

$$H(\mathbf{d}) = -\sum_{w} p_E(d = w|\mathbf{d}) \log_2 (p_E(d = w|\mathbf{d})). \tag{5.3}$$

where the empirical probability of the disparity w within the disparity map $\mathbf{d}$ is expressed as:

$$p_E(d = w|\mathbf{d}) = \frac{V(\mathbf{d}, w)}{T}. \tag{5.4}$$

with $V(\mathbf{d}, w)$ the number of occurrence of the disparity w within the disparity map and T the total number of blocks. Thus, the entropy of the disparity map can finally be expressed as:

$$H(\mathbf{d}) = -\sum_{w \in W} \frac{V(\mathbf{d}, w)}{T} \log_2 \left(\frac{V(\mathbf{d}, w)}{T} \right). \tag{5.5}$$

The rate-distortion optimization problem is defined in the same manner as in Equation (7.6) as a Lagrangian minimization:

$$\widehat{\mathbf{d}} = \text{argmin} J(\lambda, \mathbf{d}) = \text{argmin}(E_G(\mathbf{d}) + \lambda H(\mathbf{d})), \tag{5.6}$$

where λ is the Lagrangian multiplier.

Note that the R-algorithm presented in this chapter brings a suboptimal solution to this metric minimization.

5.2 Description of the proposed R-algorithm

In this section, we present the sub-optimal solution of the joint entropy-distortion metric minimization problem expressed by Equation (5.6). The developed optimization algorithm, i.e. the R-algorithm, starts with the disparity map yielded by the traditional BMA as an initial Reference disparity map. At each stage of the algorithm, this reference map is modified gradually as long as improvements are observed. The complete process is explained below.

5.2.1 Raster scanning notations

Before describing the proposed algorithm, introduce some notations. Blocks of each view are processed in a raster scanning order. For the sake of simplicity, block-coordinates (i_b, j_b) are now replaced by a 1-D coordinate, denoted t:

$$t = i_b \times Y + j_b \quad \text{with} \quad i_b = 0, ..., X-1 \quad \text{and} \quad j_b = 0, ..., Y-1. \tag{5.7}$$

As a result, t ranges from 0 to $T-1$ where T is the total number of blocks in the right image (i.e. $X \times Y$). The block disparity of coordinate t is now denoted d_t. The Local distortion induced by this disparity is $E_L(t, d_t)$.

The initial reference blockwise disparity map provided by the traditional BMA is defined as follows:

$$\mathbf{d}^R = \{d_0^R, d_1^R, ..., d_t^R, ..., d_{T-1}^R\}, \tag{5.8}$$

where d_t^R is the reference disparity (i.e. the initial disparity) of the t-th block.

5.2.2 Blockwise disparity map based on the joint entropy-distortion criterion

Assume that the optimization R-algorithm has already modified the reference disparity map by processing blocks ranging from 0 to $t-1$. Therefore the updated reference blockwise disparity map $\mathbf{d}^R$ becomes:

$$\mathbf{d}^{t-1} = \{d_0, d_1, \ldots, d_{t-1}, d_t^R, \ldots, d_{T-1}^R\}, \tag{5.9}$$

where $d_0, d_1, \ldots, d_{t-1}$ are the disparities that may have been changed when the algorithm has processed blocks ranging from 0 to $t-1$ and $d_t^R, \ldots, d_{T-1}^R$ are the unchanged disparities of the initial reference map $\mathbf{d}^R$. The selected disparity map $\mathbf{d}^{t-1}$ generates the best global cost expressed as:

$$J(\lambda, \mathbf{d}^{t-1}) = E_G(\mathbf{d}^{t-1}) + \lambda H(\mathbf{d}^{t-1}), \tag{5.10}$$

where $E_G(\mathbf{d}^{t-1})$ is the global distortion of the predicted image and $H(\mathbf{d}^{t-1})$ is the entropy, both related to the disparity map $\mathbf{d}^{t-1}$.

Note that the proposed algorithm changes the disparity of at most one block at the time. Consider the next block, i.e. the t-th block, to be matched. The disparity d_t^R associated to this

block is replaced by each of the $(N-1)\times\alpha$ other candidate disparities $w \in W \setminus \{d_t^R\}$ thus generating $(N-1)\times\alpha$ different disparity maps $\mathbf{d}^t(w)$:

$$\mathbf{d}^t(w) = \{d_0, d_2, ..., d_{t-1}, w, d_{t+1}^R, ..., d_{T-1}^R\}. \tag{5.11}$$

For each modified disparity map, a global cost $J(\mathbf{d}^t(w))$ is computed as follows:

$$J(\lambda, \mathbf{d}^t(w)) = E_G(\mathbf{d}^t(w)) + \lambda H(\mathbf{d}^t(w)), \tag{5.12}$$

where $E_G(\mathbf{d}^t(w))$ and $H(\mathbf{d}^t(w))$ represent respectively the updated global distortion and entropy related to the choice of $\mathbf{d}^t(w)$.

The global costs $J(\lambda, \mathbf{d}^t(w))$ are then sorted in an increasing order. The disparity $w = d_t$ which is associated to the smallest $J(\lambda, \mathbf{d}^t(w))$ is then retained and the disparity map according to the process of the t-th block becomes:

$$\mathbf{d}^t = \{d_0, d_1, ..., d_{t-1}, d_t, d_{t+1}^R, ..., d_{T-1}^R\}. \tag{5.13}$$

Based on this principle, the R-algorithm continues until processing the last block (i.e. $(T-1)$-th block). A new disparity map $\mathbf{d}^{T-1}$ is then estimated introducing a minimal global cost $J(\lambda, \mathbf{d}^{T-1})$.

To further improve the rate-distortion performance, the R-algorithm iterates the described process where the disparity map $\mathbf{d}^{T-1}$ is now considered as a reference disparity map. This process is repeated as long as improvements in terms of rate-distortion are observed. Figure 5.1 summarizes the different steps of the proposed optimization R-algorithm.

5.2.3 Entropy and distortion recursive equations

To avoid being faced with a heavy computational load due to the calculation of $E_G(\mathbf{d}^t(w)$ and $H(\mathbf{d}^t(w)$, the R-algorithm proposes to reuse the previous results obtained after processing the $(t-1)$-th block.

The global distortion is thus updated according to the following recursive equation:

$$E_G(\mathbf{d}^t(w)) = E_G(\mathbf{d}^{t-1}) - E_B(t, d_t^R) + E_B(t, w), \tag{5.14}$$

where $E_B(t, d_t^R)$ and $E_B(t, w)$ qre the local distortion induced respectively by the disparities d_t^R and w for the t-th block:

$$E_B(t, d_t^R) = \sum_{u=0}^{N_X-1}\sum_{v=0}^{N_Y-1} (\widehat{I}_r(i_p+u, j_p+v) - I_{\tilde{L}}(i_p+u, \alpha(j_p+v+d_t^R)))^2, \tag{5.15}$$

$$\text{and} \quad E_B(t, w) = \sum_{u=0}^{N_X-1}\sum_{v=0}^{N_Y-1} (\widehat{I}_r(i_p+u, j_p+v) - I_{\tilde{L}}(i_p+u, \alpha(j_p+v+w)))^2. \tag{5.16}$$

Input: Left image $I_{\tilde{L}}$ and right image I_r

Output: Estimated blockwise disparity map associated with I_r

1. Estimate the disparity map using the traditional BMA;

2. Improve the reference blockwise disparity map:

a. Set initial values: λ; w_{min}; w_{max}; $i_b = -1$; $j_b = -1$; α;

b. Increment by 1 the row index i_b;

c. Increment by 1 the column index j_b;

d. Set the sliding matching window on the pixel $I_{\tilde{L}}(i_p, j_p)$;

e. Replace the disparity of the current block by each of the other disparities w;

f. Update the global distortion induced by the predicted image for each choice of w;

g. Update the entropy for each choice of w;

h. Compute the global entropy-distortion cost for each w;

i. Sort the costs in a decreasing order.

j. Select the disparity w minimizing the global cost.

k. Update the disparity map;

l. Start again from step **c** if $j_b < Y$ otherwise continue;

m. Start again from step **b** if $i_b < X$ otherwise continue;

n. Start again from step 2 if the final global cost is smaller than the initial cost, otherwise continue;

3. End

Figure 5.1: Proposed optimization R-algorithm.

Note that the permutation of disparity w with d_t^R, increases the number of occurrence $V(\mathbf{d}^t, w)$ of one unit as compared to $V(\mathbf{d}^{t-1}, w)$, while $V(\mathbf{d}^t, d_t^R)$ has decreased of one unit as compared to $V(\mathbf{d}^{t-1}, d_t^R)$. Hence, the entropy of the modified disparity map is also updated according to the following equation:

$$\begin{aligned} H(\mathbf{d}^t(w)) = H(\mathbf{d}^{t-1}) \\ + \left(\frac{V(\mathbf{d}^{t-1},d_t^R)}{T}\right)\log_2\left(\frac{V(\mathbf{d}^{t-1},d_t^R)}{T}\right) - \left(\frac{V(\mathbf{d}^{t-1},d_t^R)-1}{T}\right)\log_2\left(\frac{V(\mathbf{d}^{t-1},d_t^R)-1}{T}\right) \\ + \left(\frac{V(\mathbf{d}^{t-1},w)}{T}\right)\log_2\left(\frac{V(\mathbf{d}^{t-1},w)}{T}\right) - \left(\frac{V(\mathbf{d}^{t-1},w)+1}{T}\right)\log_2\left(\frac{V(\mathbf{d}^{t-1},w)+1}{T}\right). \end{aligned} \tag{5.17}$$

5.3 Performance evaluation

5.3.1 Rate-distortion performance of the R-algorithm

This section discusses simulation results conducted on different stereoscopic images, from "Middlebury" and "Deimos" datasets [47, 52], to evaluate the performance of the proposed optimiza-

tion R-algorithm. It is compared first with the traditional BMA. Then some additional results are provided to carry out comparisons with the MMA presented in the chapter 3.

The base image (corresponding to the reference image in previous chapters, which is renamed here to avoid confusion with the reference disparity map) is considered as the left one. The predicted image is the right one which is derived from the uncompressed left image and the estimated blockwise disparity map using the developed R-algorithm. The bitrate associated to the disparity map is estimated by the entropy expressed in bits per pixel (bpp). For all algorithms, estimated disparities have the same quarter-pixel precision ranging from -30 up to $29+3/4$. The interpolated right image is computed using the same filters employed in the H.264 standard [36].

Figure 5.2 provides simulation results conducted on the stereoscopic image "Tsukuba" (Middlebury dataset). The curve with circles (o) illustrates the PSNR involved by the BMA disparity map versus bpp for 13 different block sizes (4×4 up to 16×16). For the sake of visibility, circles are joined with a dashed line.

The curves with red solid line show the performance in terms of PSNR involved by the R-algorithm disparity map versus bpp. Each curve is related to a given block size and plotted for different values of λ. Note that each curve at its right end is connected to a circle recalling that for $\lambda = 0$, the R-algorithm and BMA have the same performance. The analysis of these curves clearly shows the advantage of our approach compared to the traditional BMA. Indeed significant gains in terms of rate-distortion are observed.

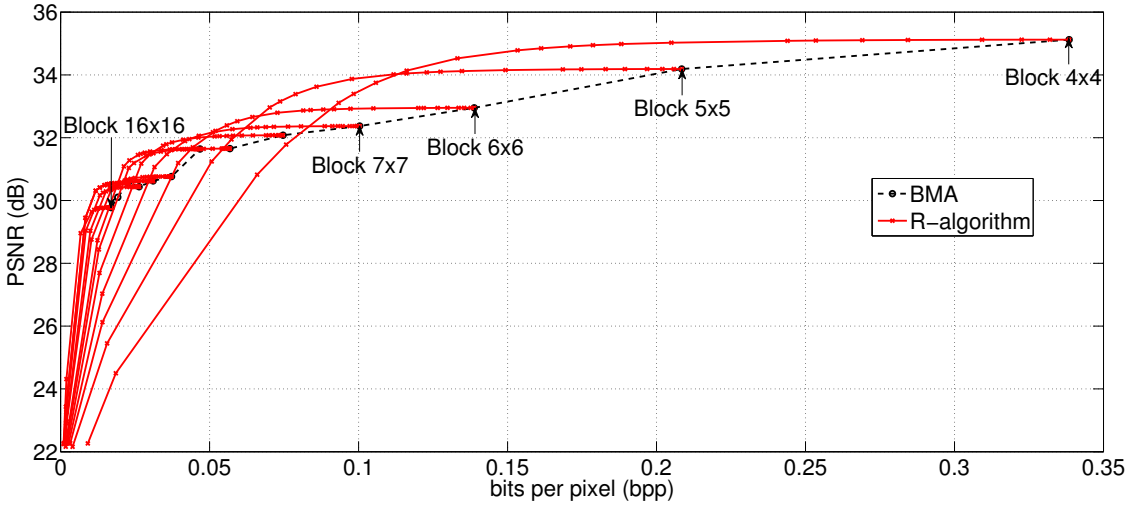

Figure 5.2: Rate-distortion performance on "Tsukuba".

Table 5.1 compares the performance using four stereoscopic images "Tsukuba", "Sawtooth", "Teddy" (from "Middlebury" dataset) and "Stereo_13" (from "Deimos" dataset) for low, medium and high bitrates with 3 different block sizes (4×4, 6×6 and 8×8). Note that the R-algorithm achieves a significant reduction in terms of bpp for a small reduction of the re-

construction quality. Hence, when processing "Tsukuba" stereoscopic image using blocks of size 4×4, the BMA requires 0.338 bpp to ensure a reconstruction quality equal to 35.12 dB, while the R-algorithm requires only 0.188 bpp for a very similar reconstruction quality equal to 34.98 dB. This important bitrate reduction is obtained by a different choice of disparities yielding a lower entropy.

Stereo		**BMA**		**R-algorithm**	
Images	**Block Size**	**PSNR**	**Bitrate**	**PSNR**	**Bitrate**
Tsukuba	4x4	35.12	0.338	34.98	0.188
	6x6	32.95	0.138	32.88	0.084
	8x8	32.08	0.074	32.00	0.046
Stereo_13	4x4	30.57	0.430	30.48	0.317
	6x6	29.35	0.180	29.27	0.124
	8x8	28.88	0.097	28.80	0.065
Sawtooth	4x4	33.37	0.362	33.30	0.285
	6x6	31.79	0.154	31.74	0.126
	8x8	30.69	0.085	30.62	0.069
Teddy	4x4	26.14	0.416	26.06	0.266
	6x6	25.10	0.177	25.03	0.110
	8x8	24.37	0.095	24.31	0.062

Table 5.1: Comparison of rate-distortion performance between the BMA and the R-algorithm.

For equivalent PSNR (see first line of Table 5.1), the BMA disparity map histogram, provided by Figure 5.3, is composed of many bars of medium or small height. While the R-algorithm disparity map histogram, in Figure 5.4, contains less bars of increased height resulting in a reduction of the disparity map entropy.

Figure 5.5 shows the original right image "Tsukuba", it indicates with a white box the corresponding location of the close-up views of Figure 5.6, showing among other items, the upper part of a lampshade. This figure shows two reconstructions of the right image with the BMA on the left and with the R-algorithm on the right. Both require the same bitrate 0.14 bpp, the BMA uses blocks of size 6×6, whereas the R-algorithm uses blocks of size 4×4. When measuring the distortion of the whole image in PSNR, we note that the R-algorithm achieves an improvement on the BMA of 1.5 dB: 34.5 dB as compared to 32.9 dB. With a closer look we see a better reconstruction quality with the R-algorithm as a gray block is lacking on the upper left corner of the lampshade on the left image.

Figure 5.7 is provided to compare the performance of the R-algorithm with the MMA. Parameter settings are the same as for Figure 5.2 except that simulations are conducted on the stereoscopic image "Stereo_13" and that the performance of the R-algorithm and the MMA are

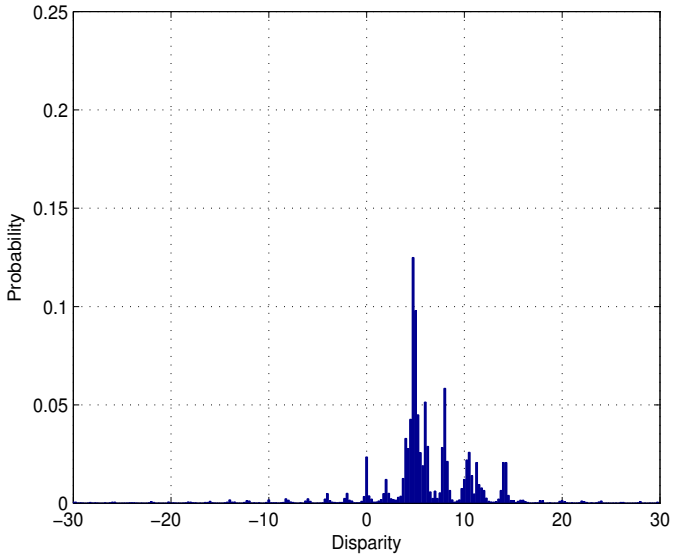

Figure 5.3: BMA blockwise disparity map histogram.

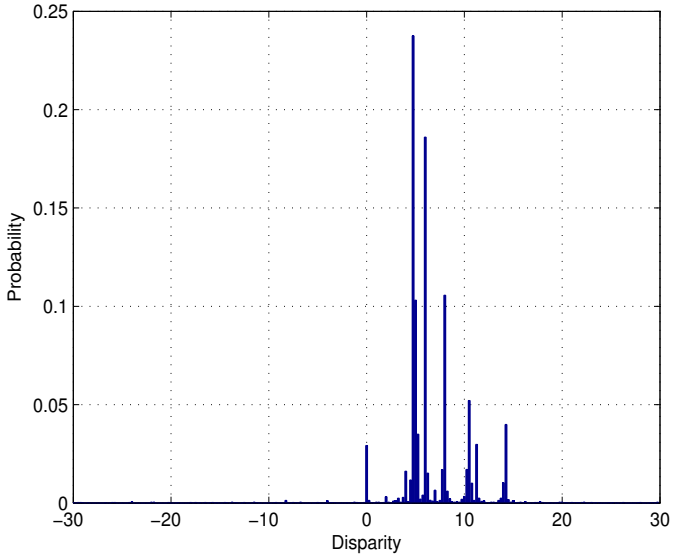

Figure 5.4: R-algorithm blockwise disparity map histogram.

plotted for three block sizes: 4×4, 6×6 and 8×8. Results obtained with the R-algorithm are given by the set of solid lines in red joining the "x" symbols. Those obtained with the MMA are given by the set of solid lines in blue joining the "+" symbols. This figure shows that the R-algorithm and the MMA perform better than the BMA for a given bitrate as well as for a given quality of reconstruction. The R-algorithm performs sligthly better than the MMA, with a gain of 0.2 dB in terms of PSNR at the bitrate of 0.13 bpp.

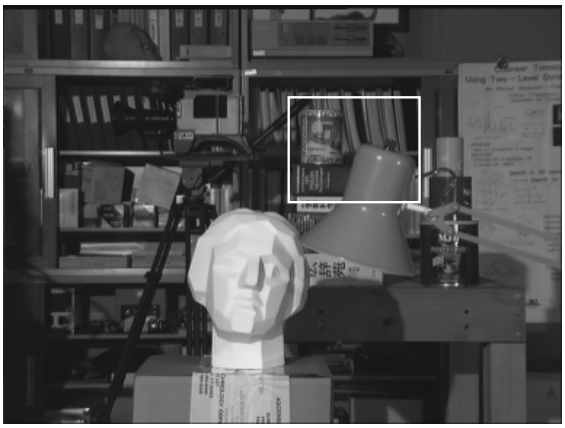

Figure 5.5: "Tsukuba" original right image.

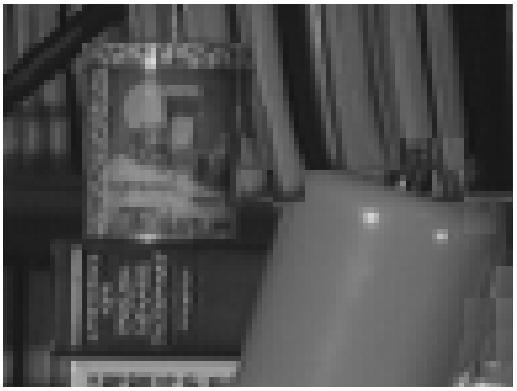

Figure 5.6: Close-up views of the reconstructed image using the BMA (left figure) and the R-algorithm (right figure).

5.3.2 Processing time and complexity of the R-algorithm

The above simulations have been conducted in the same conditions as in section 3.3.7, that is to say under Matlab environment (version R2012a) on a Personal Computer having an Intel Core i5 processor (3.20 GHz) and a RAM of 4 GB. For the stereoscopic image "Stereo_13" of size 202×320 pixels, using disparities amongst the range $[-30, ..., 29]$ and blocks of size 8×8, the R-algorithm takes in average 30 seconds to estimate a disparity map for a given value of the parameter λ, corresponding to one Rate-Distortion point.

The R-algorithm has a complexity reduced compared to that of the MMA. This complexity is computed as the number of elementary operations computed by the R-algorithm to estimate a disparity map for a given value of λ. This complexity is computed in the case of a single iteration of the R-algorithm and is given by:

$$3N_w XY N_X N_Y + 11N_w XY - 7Nw - 1 \tag{5.18}$$

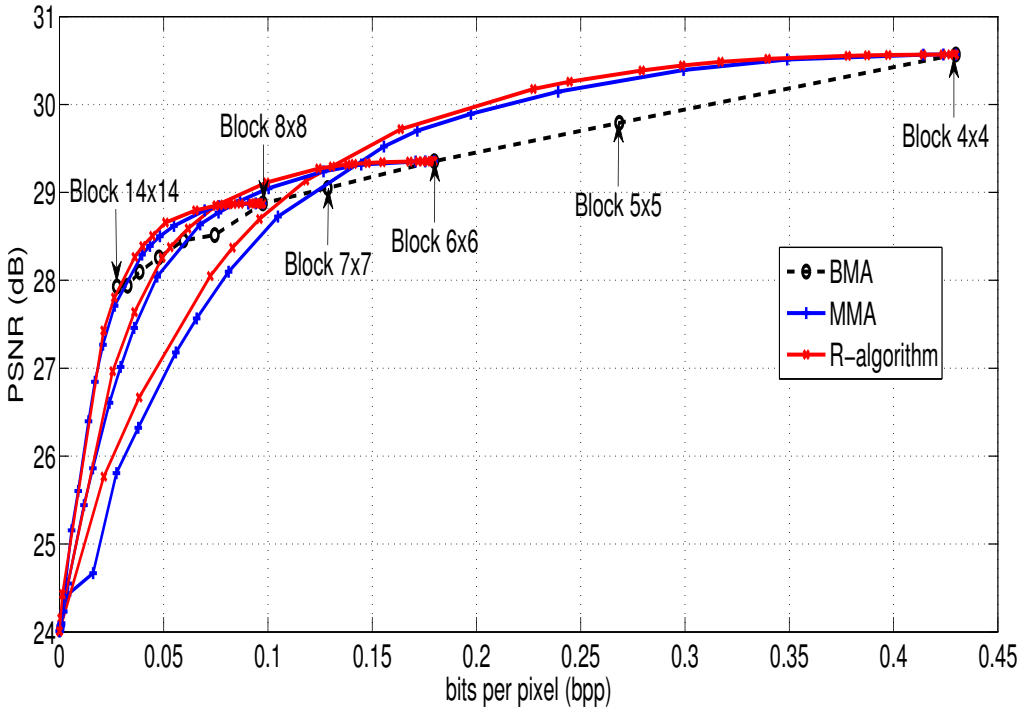

Figure 5.7: Rate-distortion performance on "Stereo_13".

The processing time to compute one RD-point is reduced in comparison to the MMA as the parameter M does not make increase the complexity of the R-algorithm.

5.4 Conclusion

This chapter presented the reference block-matching algorithm (i.e. R-algorithm) which is a sub-optimal algorithm relying on the reference disparity map provided by the traditional BMA. This reference map is modified as long as improvements in terms of rate-distortion are observed. Moreover, this algorithm has been concerned with reducing the computational load when updating the global distortion and entropy. Simulation results conducted on stereoscopic images show that the proposed R-algorithm performs better in terms of rate-distortion than the traditional BMA. To further improve the performance of such an algorithm, one idea is to enlarge the search area in which the disparities are selected. Doing this will increase the complexity of the algorithm and thus its processing time also. In the next chapter, we propose to take advantage of a large search area without increasing the complexity by restraining the search area to relevant sets of disparities.

Chapter 6

Improving the BMA performance based on a selection of disparities sets: BMA_H and BMA_S

During a stereo-matching process, as in the BMA, disparities are traditionally selected amongst a search area by minimizing a local distortion. The larger the search area is, the better the quality of the predicted image is, thanks to more adequate choices of disparities. But enlarging the search area can also result in a more expensive disparity map in terms of bitrate, as the range of the selected disparities may also increase.

The search area is a rectangular window in most cases. Instead of using all possible disparities included in the search area, two approaches relying on the initial set of disparities selected by the BMA are proposed. They allow to consider only a set of disparities belonging to the search area allowing in this chapter to achieve a given bitrate while minimizing the distortion of the predicted image.

Simulation results confirm the benefits of the proposed algorithms compared to the BMA in terms of bitrate-distortion. The rest of the chapter is organized as follows. Section 6.1 raises the issue of selecting a set of disparities in a given large search area. Section 6.2 then proposes two blockwise disparity map estimation algorithms: BMA_H and BMA_S. Simulation results are then discussed in Section 6.3. Finally Section 6.4 concludes this work. The contributions presented in this chapter have also been published in [14].

6.1 Selection of disparities sets in a given search area

This section states the problem of selecting a set of disparities in a given large search area, considered as an optimization problem, in such a way to minimize the distortion of the predicted view for a given disparity map bitrate. Before presenting the problem statement, the following section introduces some assumptions and notations.

6.1.1 Notations and assumptions

The same notations as in section 3.1.1 are used concerning the left and right views (I_l and I_r), the size of the images ($K \times L$ pixels) and the sliding matching window of size $W_x \times W_y = N_W$ pixels resulting in the same number of different choices of possible disparities for each block. in addition, W denotes the initial set composed of these N_W disparities. S_n is a set of n different disparities selected amongst the initial set W. $\mathcal{P}(W)$ represents all possible sets included in W.

Given a set S_n of disparities, the BMA associates to each block of the right view a unique two-coordinates disparity $\underline{d}$ chosen amongst this set. The predicted view is denoted $\widehat{I}_{r,D}$ as it depends on the disparity map denoted D composed of the disparities selected for each block. The set of all different disparities in D is denoted $\mathrm{Co}(D)$ as it can be regarded as the codomain of the mapping function D.

The global distortion of the predicted view, denoted $E(D)$, depends on the estimated disparity map and is computed as:

$$E(D) = \sum_{i=0}^{K-1} \sum_{j=0}^{L-1} \left(\widehat{I}_{r,D}(i,j) - I_r(i,j) \right)^2, \tag{6.1}$$

where $\widehat{I}_{r,D}$ is the right predicted image using the disparity map D.

Given a set S_n of disparities, the BMA computes the best disparity map $D(S_n)$ minimizing the global distortion of the predicted view:

$$D(S_n) = \underset{\{D|\, Co(D) \subset S_n\}}{\arg\min} \; E(D), \tag{6.2}$$

where $\{D|\, \mathrm{Co}(D) \subset S_n\}$ is the collection of all disparity maps whose codomain are included in S_n.

The minimization issue of Equation (6.2) is achieved in an efficient manner as for each block the disparity is selected by minimizing a local distortion metric regardless of the disparities selected for the other blocks.

6.1.2 Problem statement

Considering a disparities set S_n, the BMA estimates a disparity map which requires a certain bitrate to be encoded and induces a global distortion on the predicted image.

Therefore, the objective is to find the best set S_n of disparities in the search area W minimizing the global distortion of the predicted view at a given bitrate b. This objective is formulated as:

$$S_n = \underset{\substack{S'_n \subset \mathcal{P}(W) \\ H(D(S'_n)) \leq b \\ 1 \leq n \leq N_W}}{\arg\min} \; E\left(D(S'_n)\right), \tag{6.3}$$

where $E(D(S_n^{'}))$ is the global distortion associated to the disparity map $D(S_n^{'})$.

The entropy associated to the disparity map $D(S_n^{'})$ approximating the bitrate b of this map, denoted $H(D(S_n^{'}))$, is expressed as:

$$H(D(S_n^{'})) = - \sum_{\underline{s} \in Co(D(S_n^{'}))} P(\underline{d} = \underline{s}) \log_2 \left(P(\underline{d} = \underline{s}) \right), \tag{6.4}$$

with $P(\underline{d} = \underline{s})$ the occurrence frequency of the disparity $\underline{s}$ in the map $D(S_n^{'})$.

Solving the objective formulated by Equation (6.3) for a given bitrate b requires to test all possible disparities sets S_n of all possible sizes that can be made considering the initial set W. For example, considering an initial set composed of $3 \times 121 = 363$ disparities, solving Equation (6.3) for a given bitrate requires to process $\sum_{k=1}^{363} \binom{363}{k} \approx 10^{108}$ different disparities sets to find the best one minimizing the global distortion.

Instead of solving this complex optimization problem, we address the following problem: the new objective is to find the best subset S_N composed of exactly N different disparities minimizing the global distortion of the predicted view. This optimization problem is of less complexity than the previous one: indeed the bitrate constraint is no longer taken into account.

Moreover, the initial set of all possible disparities W is restricted to $W_0 = \mathrm{Co}(D(W))$: this latter corresponds to the set of disparities effectively selected by the BMA when it is processed using the initial set W. The new optimization problem can finally be expressed as:

$$S_N = \underset{\substack{S'_N \subset \mathcal{P}(W_0) \\ Card(S'_N) = N}}{\arg\min} \; E\left(D(S'_N)\right). \tag{6.5}$$

Solving this equation for different sizes N ranging from 1 to $Card(W_0)$ leads to a family of sets S_N having decreasing values of global distortion. Indeed increasing the number of disparities inside a set enables to choose more adequate disparities for the blocks to be matched, thus reducing the global distortion of the predicted image. As the size of S_N increases, it is also expected that the entropy increases (however note that it may not increase). A sub-optimal solution of equation (6.3) is obtained by selecting the highest index N for which the bitrate constraint is respected, i.e. $H(D(S_N)) \leq b$, and considering S_N as the sub-optimal solution. The following section proposes two sub-optimal algorithms solving equation (6.5).

6.2 Proposed sub-optimal BMA_H and BMA_S algorithms

Before presenting the sub-optimal algorithms denoted BMA_H (for BMA based Histogram) and BMA_S (for BMA based Sets), this section discusses the computational complexity of the optimal solution of equation (6.5).

Input: I_l, I_r, initial disparity set W with $Card(W) = N_W$ and N the target size of the disparity set such as $N \leq N_W$.
Output: Disparity set S_N and disparity map $D(S_N)$.

1. Compute the BMA using disparities contained in W, get the estimated disparity map $D(W)$;
2. Sort the disparities contained in the set W in a decreasing order according to their occurrence frequency in $D(W)$;
3. Retain the N first disparities as the selected set S_N;
4. Compute the BMA using disparities contained in S_N, get the estimated disparity map $D(S_N)$

Figure 6.1: BMA_H sub-optimal algorithm.

6.2.1 Computational complexity of the optimal solution

To solve equation (6.5), one method consists in processing the BMA on all possible sets $S_N \in \mathcal{P}(W_0)$ composed of N disparities. The optimal solution is the set S_N for which the global distortion of the predicted image is the smallest one. Depending on the size of W_0 and the value of N, this rapidly becomes a complex combinatorial problem for which this method would require a long processing time. Consider as an example, a rectangular window containing $121 \times 3 = 363$ disparities. Finding the set of 15 disparities minimizing the global distortion of the predicted image leads to process $\binom{363}{15} \approx 10^{26}$ sets. Even if this window is replaced by a small search area W_0 containing 33 disparities, we would still have to process $\binom{33}{15} \approx 10^9$ sets. To overcome this issue, two sub-optimal algorithms are presented in the following.

6.2.2 Selecting the best set of disparities using BMA_H algorithm

The first sub-optimal algorithm, called BMA_H (H for Histogram), consists in processing the BMA on the initial disparity set W so as to estimate the disparity map $D(W)$. The disparities contained in $D(W)$ are then sorted in a decreasing order of their occurrence frequency in $D(W)$ (using the histogram). The set containing the N first disparities is then defined as the set S_N satisfying the constraint $S_N \in \mathcal{P}(W_0)$ and $Card(S_N) = N$. The set S_N is then considered as the sub-optimal solution of equation (6.5). Finally the BMA estimates the disparity map $D(S_N)$ using the sub-optimal set S_N. Figure 6.1 summarizes the BMA_H steps.

This algorithm is interesting in terms of computational complexity as it requires only a little more time processing than that when using the BMA on W. However when N is much lower than $Card(W_0)$, we should not expect good performance in terms of reducing the global distortion. Indeed assessing the utility of a given disparity in a set S_N, depends on the other disparities of S_N. It is likely that some of the disparities of an optimal set for N small may have little occurrence frequency in $D(W)$ and thus be discarded when using BMA_H. In an attempt to overcome this issue, the following disparity set selection strategy is proposed.

6.2.3 Selecting the best set of disparities using BMA_S algorithm

The second disparity map estimation algorithm, denoted BMA_S (S for Set), provides a sub-optimal solution of equation (6.5). The BMA_S is based on a sequential reduction of the size of the set from $Card(W_0)$ down to N by considering first the set $S = W_0$ and at each step by pruning one disparity, the one for which the global distortion is being the less increased.

Assume that the proposed algorithm has already found a set S_m. There are m subsets of size $m-1$ that can be extracted from the set S_m. These subsets are defined as $S_m \backslash \{\underline{s}\}$ for all possible disparities $\underline{s} \in S_m$. Selecting the best set S_{m-1} is solved by

$$S_{m-1} = S_m \backslash \{\underline{s}\}, \qquad (6.6)$$
$$\text{where } \underline{s} = \underset{\underline{s}' \in S_m}{\arg\min}\, E\left(D(S_m \backslash \{\underline{s}'\})\right).$$

The process is then iterated until obtaining the best subset S_N composed of N disparities.

For a given set of size m, equation (6.6) suggests processing the BMA m times to select the best subset S_{m-1}. Hence the BMA_S requires processing the BMA up to $Card(W_0)!$ times to solve equation (6.5) (if $N = 1$). It is actually possible to reduce the computational complexity of our algorithm. Indeed we propose to process the BMA only once and save in memory all local distortions induced by each disparity for each block. When solving equation (6.6), the algorithm just has to read the required information stored in memory.

Note that this algorithm is not optimal since the minimization is computed on a restricted domain instead of $P(W_0)$. The rationale is that if S_m is a good choice it is likely to have a good choice for S_{m-1}. As for BMA_H, it remains possible that a disparity of a set solving optimally equation (6.5) for N small is pruned by the BMA_S while processing sets of a larger size. The steps of the BMA_S are summarized in Fig 6.2.

6.3 Analysis and discussion of the simulation results

This section discusses the simulation results of the proposed blockwise disparity map estimation algorithms (BMA_H, BMA_S) and the traditional BMA algorithm.

Simulations are conducted on stereoscopic images from the CMU-VASC database [51]. The window W contains disparities ranging horizontally from -60 to 60 and vertically from -1 to 1. Indeed simulations have shown that disparities having a small vertical component could clearly improve the performance of the disparity map estimation (see section 4.3.2.1). On one hand it seems that the human visual system is able to cope with such disparities and still perceive depth. On the other hand stereoscopic cameras may not always be that precise.

Input: I_l, I_r, initial disparity set W with $Card(W) = N_W$ and N the target size of the disparity set such as $N \leq N_W$.
Output: Disparity set S_N and disparity map $D(S_N)$.

1. Compute the BMA using disparities contained in the set W, get the estimated disparity map $D(W)$;
2. Extract from $D(W)$ the set of disparities W_0;
3. Set $S_m = W_0$ with $m = Card(W_0)$;
4. If $m = N$, go to step 11. Else if $m > N$, continue;
5. Create m new sets S_{m-1} by removing one disparity $\underline{s}$ at each time from S_m (i.e. $S_{m-1} = S_m \backslash \{\underline{s}\}$);
6. For each set S_{m-1}, compute the BMA using only the disparities contained in this set;
7. Compute the global distortion $E(D(S_{m-1}))$ associated to each set;
8. Retain the set of disparities S_{m-1} introducing the minimal distortion $E(D(S_{m-1}))$;
9. Set $m = m - 1$;
10. Go to step 4;
11. Select the best set S_N in terms of distortion;
12. Compute the BMA using disparities contained in S_N, get the estimated disparity map $D(S_N)$.

Figure 6.2: BMA_S sub-optimal algorithm.

6.3.1 Comparing BMA_H, BMA_S and the optimal solution

This section compares the performance of the sub-optimal algorithms developed in sections 6.2.2 (i.e. BMA_H) and 6.2.3 (i.e. BMA_S) to the optimal disparity map estimation algorithm described in section 6.2.1. For this, a small stereoscopic test image of size 64×64 pixels extracted from the stereo set "sand" is considered [51]. For all algorithms, the BMA is processed with blocks of equal size 10×10 pixels.

Figure 6.3 illustrates the bitrate-distortion performance of both sub-optimal disparity map estimation algorithms. The dashed curve joining black circles concerns the BMA_H. While the solid line joining red circles represents the BMA_S. Both curves start from the same point corresponding to the traditional BMA yielding a disparity map $D(W)$ containing 33 disparities. The circles of each curve provides the performance when the size of the selected disparity set ranges from 33 disparities to 1 disparity. These curves show clearly that the BMA_S achieves better performance than the BMA_H.

Note that the curve of the BMA_H ends with the point corresponding to a set of 3 disparities. In this example, indeed no sets of smaller size could have been tested since some blocks of the image have no adequate choices of disparities in a more reduced set due to the side effects (e.g. blocks located at the extreme left side of the image cannot have a negative disparity).

In Figure 6.3, the blue cross shows the performance achieved with the optimal set of 5 disparities found after processing all possible sets of 5 disparities amongst the 33 disparities of W_0, meaning that $\binom{33}{5} = 237336$ sets have been processed! In this experiment, the optimal set provides exactly the same performance as the the BMA_S sub-optimal algorithm for $N = 5$. Moreover the same disparities have been selected by both algorithms. This result shows that BMA_S algorithm although it is based on a sub-optimal solution remains efficient while reducing the computational complexity of the optimal solution.

6.3.2 Rate-distortion performance of sub-optimal algorithms

This section compares the performance of both sub-optimal disparity map estimation algorithms (BMA_H, BMA_S) with the traditional BMA on the stereoscopic image "house2" [51]. Simulations have been conducted using block sizes from 4×4 to 12×12 pixels.

Figure 6.4 illustrates the simulation results. The performance of the BMA_S is represented by the red curves in solid line joining circles while the one of the BMA_H is represented by the black dashed curves joining circles. Each curve is associated to a given block size where each point is related to a disparity set of different size.

For a given block size, the curves of both algorithms start from the same point. This point is also associated to the BMA rate-distortion using the same block size. The dashed blue line

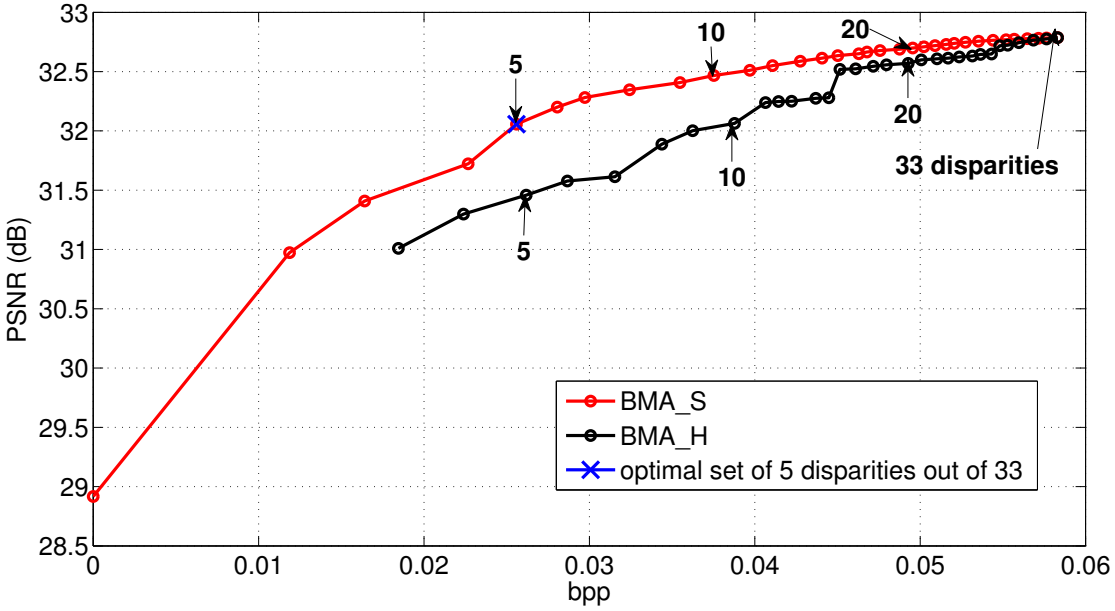

Figure 6.3: RD performance on a small stereoscopic image extracted from "sand".

joining squares illustrates the BMA performance where each square corresponds to a specific block size. An analysis of these RD curves show that the BMA_S performs better than the BMA_H and BMA. For a given bitrate of the disparity map, the BMA_S predicts the right view with a greater precision as compared to the BMA and BMA_H. For example, at a bitrate of 0.09 bpp, the BMA and the BMA_H lead to a predicted image of quality 26.4 dB while the BMA_S yields to 28.6 dB resulting in a gain of 2.2 dB.

6.3.3 Comparing the predicted right view for a given bitrate

This section compares the quality of the "house2" predicted right view computed with the BMA_H, BMA_S and BMA algorithms for a given bitrate equal to 0.09 bpp.

The original "house2" left and right views are represented on Figure 6.5 where the white square on the right view frames a specific area. This area is shown in Figure 6.6 as close-ups of the three predicted views computed respectively with BMA (block size 8×8), BMA_H (block size 8×8) , and BMA_S (block size 6×6) using their respective disparity set (of size 184, 184 and 155 disparities). It is clearly seen that the window of the house is better predicted with the BMA_S. Note that this higher performance is achieved using blocks of smaller size for BMA_S than those used for BMA_H and for BMA as these two algorithms would not have coped with such small blocks under the bitrate constraint of 0.09 bpp.

6.3.4 Performance evaluation using Bjøntegaard metric

An average PSNR difference, using the Bjøntgaard's metric [53], is calculated between: (i) the BMA_S and the BMA denoted ΔBMA; and (ii) the BMA_S and the BMA_H denoted

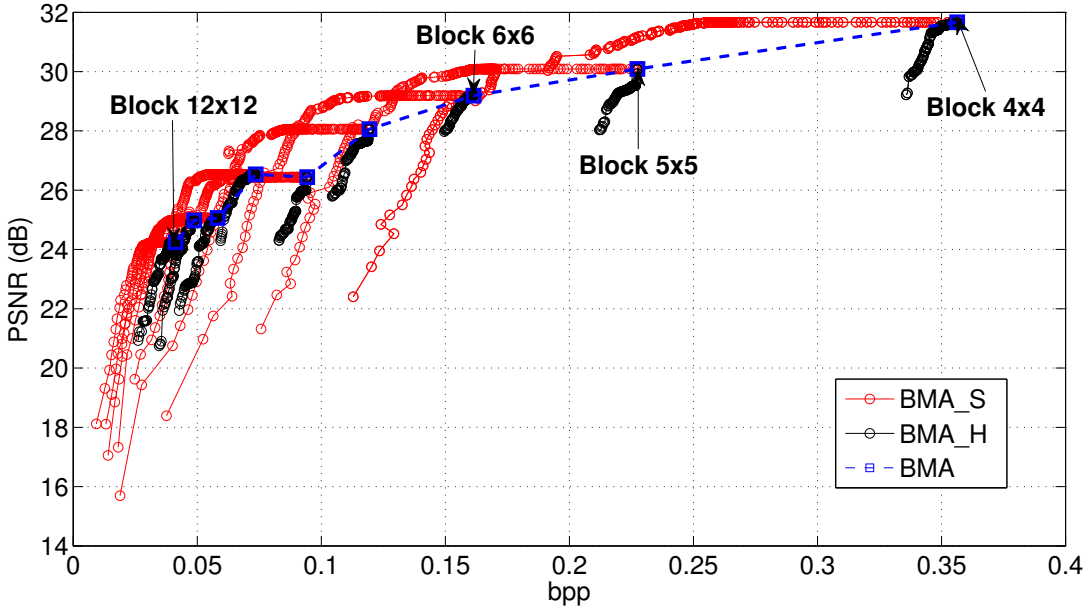

Figure 6.4: RD performance on the stereoscopic image "house2".

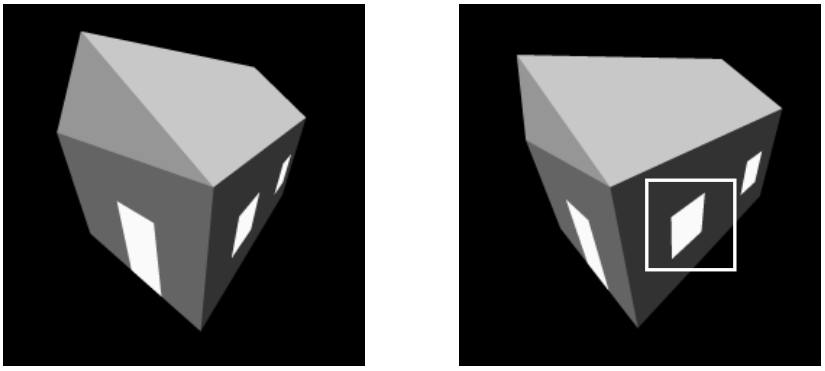

Figure 6.5: "house2" original left view (at left) and original right view (at right).

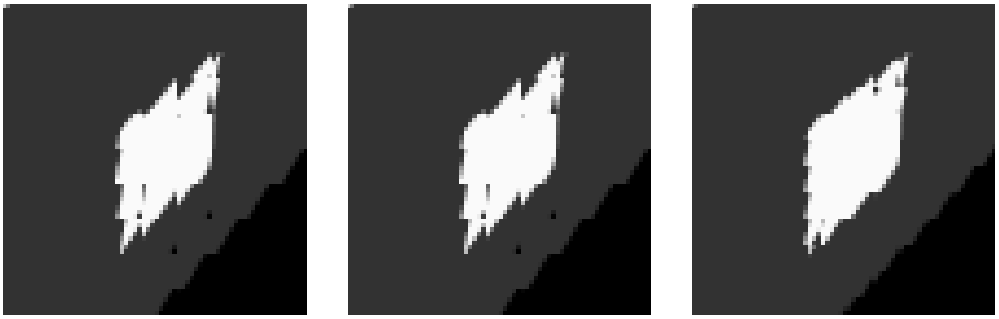

Figure 6.6: Close-up of the predicted right view using the BMA (left figure), the BMA_H (middle figure) and the BMA_S (right figure).

ΔBMA_H. The average PSNR difference is calculated respectively for low and medium bitrates considering the following target bitrates for all algorithms: $[0.06; 0.07; 0.08; 0.09]$ and $[0.1; 0.25; 0.40; 0.55]$ (in bpp). For each target bitrate, the point achieving the best performance in terms of PSNR under the bitrate constraint is retained for the computation of the Bjøntegaard metric for each algorithm.

Simulations have been conducted on nine stereoscopic images from the CMU-VASC database: "whouse", "wdc2r", "toys", 'telephone", "rubik" and "mars1r" which are natural images, and also "sphere", "house2" and "house1" which are synthetic images [51].

Figure 6.7 illustrates an example of the retained points using the bitrate-distortion results provided by Figure 6.4. Consider for example the target bitrate of 0.55 bpp. The best points selected are identical and correspond to the first top right point of all curves as this is the best point in terms of PSNR under this bitrate for all algorithms (see Figure 6.4). This point is also considered as the best one in terms of PSNR for the target bitrate equal to 0.40 bpp as the bitrate of this point (0.36 bpp) is inferior to it.

At low bitrate, the average PSNR difference of the BMA_S as compared to the BMA is of 1.81 dB and the average PSNR difference of the BMA_S as compared to the BMA_H is of 1.21 dB.

At medium bitrate, the average PSNR difference of the BMA_S as compared to the BMA and the BMA_H is respectively of 1.56 dB and 1.54 dB. Table. 6.1 summarizes the relative per-

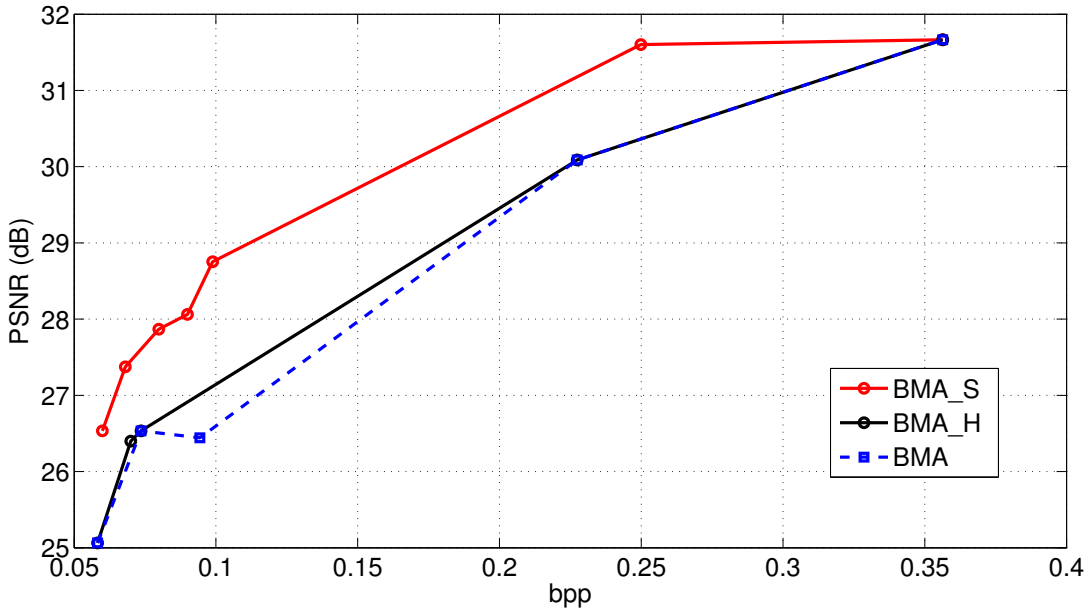

Figure 6.7: Best RD points under bitrate constraints.

formance of the proposed algorithms using the Bjøntegaard metric at low and medium bitrates. The columns ΔBMA, at low and medium bitrates, represent the average PSNR difference of the BMA_S as compared to the BMA. While the columns ΔBMA_H, at low and medium bitrates,

give the average PSNR difference of the BMA_S as compared to the BMA_H.

One clearly sees the improvements of the BMA_S as compared both to the BMA and the BMA_H in terms of bitrate-distortion performance. Note that the BMA_H achieves in most case better performance than the BMA.

Stereo	**Low bitrate**		**Medium bitrate**	
Images	**ΔBMA**	**ΔBMA_H**	**ΔBMA**	**ΔBMA_H**
mars1r	0.30	0.10	0.35	0.20
whouse	0.33	0.21	0.35	0.24
sphere	0.37	0.20	0.47	0.37
wdc2r	0.39	0.12	0.47	0.29
telephone	0.48	0.29	0.46	0.42
rubik	0.52	0.36	0.22	0.22
toys	1.64	0.25	0.56	0.28
house2	1.81	1.21	1.56	1.54
house1	1.40	1.36	0.59	0.59

Table 6.1: Average PSNR gain of the BMA_S over the BMA and the BMA_H

6.3.5 Processing time and complexity of the proposed algorithms

The above simulations have been conducted in the same conditions as mentioned in section 3.3.7.

Both the BMA_H and BMA_S relies on three main steps: the BMA is applied first to estimate a disparity map, on the proposed methods in applied to select a subset of disparities and then finally the BMA is applied again using this subset to estimate a new disparity map. The BMA has a complexity of:

$$N_w(XY)(3N_xN_y - 1) \tag{6.7}$$

Thus, the complexity of the BMA_H can be deduced as:

$$(N_w + N)(XY)(3N_xN_y - 1) \tag{6.8}$$

The complexity of the BMA_S is then computed as:

$$(XY)(3N_xN_y - 1)\left[N_w + \frac{W_0(W_0+1)(2W_0+1)}{6} - \frac{N(N+1)(2N+1)}{6}\right] \tag{6.9}$$

For the stereoscopic image "Stereo_13" of size 202×320 pixels, using disparities amongst the range $[-30, ..., 29]$ and blocks of size 8×8, the BMA_H takes in average 0.4 seconds to compute one Rate-Distortion point (corresponding to a set of size N). The BMA_S takes between 0.4 and 24 seconds to compute one point depending on the size of the set.

6.4 Conclusion

Two blockwise disparity map estimation algorithms have been proposed to take advantage of a large search area, allowing a better prediction of the right view than the traditional BMA, while reducing the required bitrate to encode the disparity map. This is achieved by selecting an appropriate disparity set and by processing the BMA with this specific set. Simulation results conducted on several stereoscopic images have shown the benefits of both proposed algorithms as compared to the BMA with a particular advantage for the second one i.e. the BMA_S.

Chapter 7

Joint disparity and block-length maps optimization algorithm: JDBLMO

In block-based disparity map estimation, usually one disparity is estimated for each block, meaning that pixels in any given block are predicted using the same disparity. This reduces considerably the bitrate needed to encode the disparity map. Using blocks of different sizes can further reduce the bitrate of the disparity map in some cases. Indeed, textured regions or small objects are coded with more precision using small blocks, while in the same image, uniform regions at roughly the same depth are coded using larger block sizes. Using blocks of variable sizes comes at a cost, that of transmitting to the decoder sufficient information to know how the blocks of variable sizes are displayed. In this chapter, a new algorithm jointly optimizing the disparity and the block-length maps is presented. The developed method relies on an entropy-distortion metric taking into account the reduction of the distortion of the predicted view but also an approximation of the bitrate needed to encode both the disparity and the block length maps. At each step of the proposed algorithm, a decision is taken to decide if a given block should be predicted using a single disparity value or if that block should be divided into four subblocks and predicted using four different disparity values. This choice is coupled with a refinement of the disparity map. Simulation results conducted on several stereoscopic images from the CMU-VASC dataset confirm the benefits of this approach as compared to competitive block matching algorithms.

The contributions presented in this chapter have been submitted to [15].

The rest of the chapter is organized as follows. Section 7.1 introduces the stereo-matching optimization problem in the case of variable block sizes. Section 7.2 proposes the suboptimal optimization algorithm. Section 7.3 provides simulation results evaluating the developed algorithm's performance. Finally section 7.4 concludes this chapter.

7.1 Rate-Distortion optimization problem

This section formalizes the rate-distortion trade-off in the case of variable block sizes. In fact, the stereo-matching optimization algorithm is intended to yield a good estimate of the predicted view from a given reference view while requiring a low bitrate to encode the disparity map. But to further improve the performance, blocks of variable sizes are considered. All considered block layouts follow a traditional quad-tree structure, that is square subblocks may only be divided into four identical non-overlapping square blocks. Some notations and assumptions are introduced before describing the rate-distortion optimization problem in the following section.

7.1.1 Notations and assumptions

In what follow, the same notations for the couple of left and right images (I_L and I_R) and their common size ($I \times J$ pixels) are kept. Let consider in addition the following notations.

The right view is first divided into non-overlapping squared blocks of fixed size denoted $L_{max} \times L_{max}$. Each of these blocks may be further partitioned into a quadtree structure. An example of such a layout is shown on Figure 7.1 with blocks denoted as B_k. As it is difficult to keep track of indexes in a quadtree structure, blocks are denoted as B and the map containing all blocks is denoted $\mathbf{B}$.

Each block B is located at position (i_B, j_B) in a coordinates system starting from the top left of the image, with i belonging to $\{0, \ldots, I-1\}$ and j to $\{0, \ldots, J-1\}$. Each block has a size of $l_B \times l_B$ pixels where the length l_B belongs to $L = \{L_{min}, 2L_{min}, \ldots, L_{max}\}$. Each block is also associated to a unique disparity d_B chosen amongst a set of disparities $\{d_{min}, \ldots, d_{max}\}$. Hence, B is identified by (d_B, l_B, i_B, j_B).

The disparity map, the block-length map and the block map are represented by $\mathbf{d}, \mathbf{l}, \mathbf{B}$. All these maps are of the same size denoted T.

7.1.2 Coding of the block layout

Using blocks of variable sizes requires to inform the decoder about the block layout, i.e the size and the location of each block.

The pixel values of the predicted right view $\widehat{I}_R$ on the block B are related to that of the left view I_L using the disparity d_B as follows:

$$\widehat{I}_R(i_B + u, j_B + v) = I_L(i_B + u, j_B + v + d_B) \text{ where } \; u, v \in \{0, \ldots, l_B - 1\}. \tag{7.1}$$

It may seem from Equation (7.1) that in order to compute the predicted right view on block B, the decoder needs not only the disparity d_B and the length l_B, but also the location i_B, j_B. In terms of bitrate this would be quite excessive. However the storage of the location information

Example of a block-length map, split into six subsequences for clarity purpose:
`8,8,4,4,4,4,8,16,16,8,8,8,8,  32,  32,`
`32,  16,16,8,8,8,4,4,4,4,16,  32,`

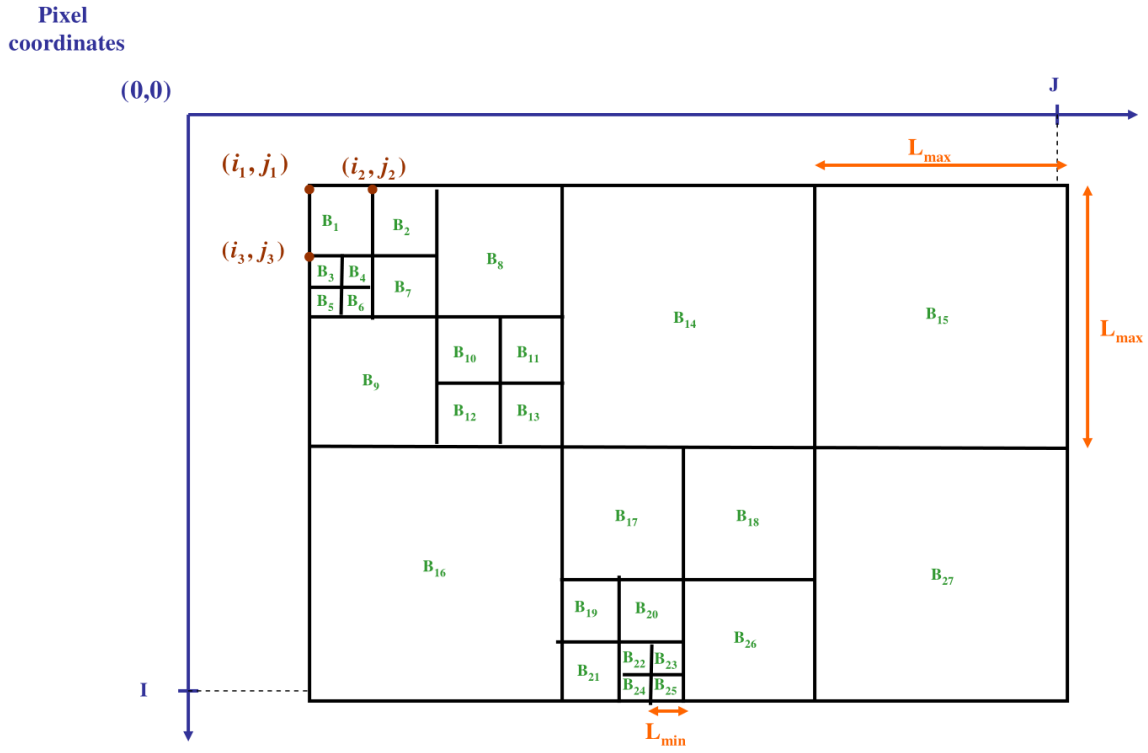

Figure 7.1: Example of a block-length map and its corresponding block layout.

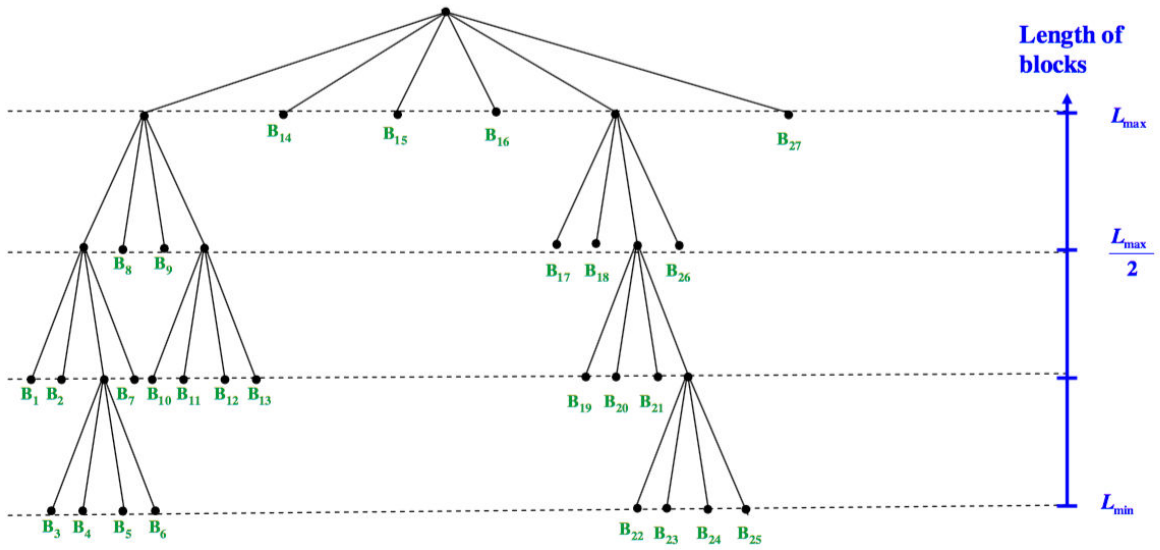

Figure 7.2: Tree structure associated with the block layout shown in Figure 7.1.

of each block is avoided since a specific tree structure is adopted where an example is shown in Figure 7.2. Its first layer has as many nodes as the number of non-overlapping square blocks that fit into an image of size $I \times J$. The other layers are subsequent quadtree decompositions. Each block corresponds to a leaf in this tree structure. A block is located at the first layer if its size is equal to $L_{max} \times L_{max}$. It is located at the second layer if its size is equal to $\frac{L_{max}}{2} \times \frac{L_{max}}{2}$, and so on for the blocks of smaller size. As a result, by assigning to the tree an exploring order, the block-length map $\mathbf{l}$ and the image dimensions (I and J) are sufficient to recover the block layout without sending the locations of each block to the decoder.

The proposed exploring order corresponds to the breadth-first search in the tree. Indeed all neighbor nodes are explored before the children nodes. This order has been chosen as opposed to the depth-first search because breadth-first search is expected to better balance the optimization process on the whole image rather than focusing on the first blocks right at the beginning of the algorithm. The exploring order considered in this chapter takes into account the raster scanning order when ordering neighboring nodes. This is illustrated in Figure 7.1 by showing a block-length map and the corresponding image of size 64×96 pixels with a specific block layout whose size is ranging from 4×4 to 32×32.

As a consequence in terms of notations, it makes sense to claim that all quantities depending on a block B, depend only of $\mathbf{d}$ and $\mathbf{l}$.

7.1.3 Formulation of the rate-distortion problem

In this section, the quality of the predicted right view is measured with the global distortion, $E_G(\mathbf{B})$, defined as the sum of the square differences. Note that applying the remark ending

subsection 7.1.2, $E_G(\mathbf{B})$ can be replaced by $E_G(\mathbf{d}, \mathbf{l})$. As only one disparity is assigned to each block, this global distortion is the sum of all local distortions denoted $E_B(B)$ accounting for the difference between the predicted right view and the true right view on a given block B:

$$E_G(\mathbf{d}, \mathbf{l}) = \sum_{B \in \mathbf{B}} E_B(B) \tag{7.2}$$

$$\text{where } E_B(B) = \sum_{u=0}^{l_B-1} \sum_{v=0}^{l_B-1} \left(\mathbf{I}_L(i_B + u, j_B + v + d_B) - \mathbf{I}_R(i_B + u, j_B + v)\right)^2 . \tag{7.3}$$

The disparity map is encoded using a lossless coder (entropy encoder). Thus the bitrate needed to encode the disparity map is approximated by T times the measure of its entropy computed using the empirical distributions of the disparities in the map as follows:

$$H(\mathbf{d}) = -\sum_{d \in D} p_E(d = \underline{d}|\mathbf{d}) \log_2 \left(p_E(d = \underline{d}|\mathbf{d})\right), \tag{7.4}$$

where $p_E(d = \underline{d}|\mathbf{d})$ is the ratio of the number of blocks having the disparity $\underline{d}$ on the total number of blocks T.

The block-length map is also encoded using a lossless coder. The same notation choices have been made. $TH(\mathbf{l})$ approximates the bitrate needed by the entropy coder. More specifically, it denotes T times the entropy of the block length map computed on the empirical distribution of the lengths in the associated map:

$$H(\mathbf{l}) = -\sum_{l \in L} p_E(l = \underline{l}|\mathbf{l}) \log_2 \left(p_E(l = \underline{l}|\mathbf{l})\right), \tag{7.5}$$

where $p_E(l = \underline{l}|\mathbf{l})$ is computed as the ratio between the number of blocks having the length $\underline{l}$ on the total number of blocks T.

Finally, the bitrate required to encode the disparity map is approximated by $T\left(H(\mathbf{d}) + H(\mathbf{l})\right)$.

To achieve the goal, the optimization problem can be posed in several ways resulting in a low distortion of the predicted view requiring a low bitrate. One might expect the lowest global distortion for a given bitrate or the smallest bitrate for a given global distortion. Solving one issue or the other for every bitrate or every global distortion value leads to the same rate-distortion curve made from all the optimal points. We propose to use once again the Lagrangian formulation to solve the optimization problem.

The Lagrangian cost function $J(\lambda, \mathbf{d}, \mathbf{l}, T)$ is then defined as follows:

$$J(\lambda, \mathbf{d}, \mathbf{l}, T) = E_G(\mathbf{d}, \mathbf{l}) + \lambda T \left(H(\mathbf{d}) + H(\mathbf{l})\right). \tag{7.6}$$

Minimizing the cost function $J(\lambda, \mathbf{d}, \mathbf{l}, T)$ for every value of the Lagrange multiplier λ leads to the same optimal curve. Finding the optimal solution of this cost function is a complex task. Indeed any modification of a given disparity at a given block may modify the appropriateness of a block division or the choice of a disparity at another block.

7.2 Joint disparity and block-length maps optimization algorithm (JDBLMO)

This section focuses on the minimization of the Lagrangian cost function $J(\lambda, \mathbf{d}, \mathbf{l}, T)$ introduced in the previous section, for a given value of λ. Given the complexity of this optimization problem, a sub-optimal optimization algorithm is developed. In the rest of the chapter, this algorithm is called JDBLMO for Joint Disparity and Block-Length Maps Optimization. Figure 7.3 presents the steps of the algorithm.

The proposed sub-optimization algorithm is performed according to three main stages. The first stage, described in Subsection 7.2.1, is related to the initialization process where the initial disparity and block-length maps are computed using the traditional BMA. The second stage, described in Subsection 7.2.2, is an adaption of the R-Algorithm developed in chapter 5 since the this algorithm has been developed only for same size blocks. This algorithm is named ARA (Adaptation of the R-Algorithm). The third stage, described in Subsection 7.2.3, reduces the function cost $J(\lambda, \mathbf{d}, \mathbf{l}, T)$ by means of some Block Divisions and is called Block Divisions Algorithm (BDA). The BDA and ARA are repeated as long as an improvement is being observed.

7.2.1 Initialization process

In this initialization step, the right view is first partitioned into blocks of size $L_{max} \times L_{max}$, setting the block-length map as a constant map. The traditional BMA is then applied on each block. For a given block B, the disparity d_B is selected by minimizing the Sum of Squared Differences (SSD) computed between the pixel-values of the right view and the predicted view considering only pixels in B. Using E_B defined in (7.3), d_B is expressed as:

$$d_B = \arg\min_{d \in D} E_B(d). \tag{7.7}$$

The collection of all d_B forms an initial disparity map $\mathbf{d}$. When associated with the constant map $\mathbf{l} = [L_{max} \dots L_{max}]$, this forms an initial block map $\mathbf{B}$. This block map $\mathbf{B}$ is then converted into a queue whose processing is consistent with the tree-structure exploring order defined in section 7.1.2.

Let us introduce some notations to handle $\mathbf{B}$ as a queue. B_f is the front block of the queue and $\mathbf{B}_t$ is the tail of the queue:

$$\mathbf{B} = [B_f, \mathbf{B}_t]\,. \tag{7.8}$$

Removing the first block and placing it at the queue's end leads to:

$$\mathbf{B} := [\mathbf{B}_t, B_f]\,. \tag{7.9}$$

This action concerns the steps 2.b.; 2.d. and 3.f. in Figure 7.3. For the sake of simplicity each block $B \in \mathbf{B}$ is now considered as a couple (d_B, l_B) and the front block is denoted as (d_f, l_f).

Input: Left view $\mathbf{I}_L$, right view $\mathbf{I}_R$,
minimum and maximum lengths, respectively L_{min} and L_{max} pixels.
Output: Joint estimated disparity and block-length map $(\mathbf{d}, \mathbf{l})$.

1. Initialization.
  1.a Partition the right view into square blocks of lengths L_{max}.
  1.b Compute for each block B a disparity d_B using the BMA.
  1.c Consider this initial block map $\mathbf{B}$ as a queue.
  1.d Set $l_{min} = L_{max}$.
2. Apply the ARA on $\mathbf{B}$:
  2.a. Set $J_{old} := J(\mathbf{B})$ and $c := T$.
  2.b. Remove the queue's front block denoted as (d_f, l_f).
  2.c. Select the best disparity $\widehat{d}$ minimizing $J(\mathbf{B}_t, (d, l_f))$ as $d \in D$.
  2.d. Update the block map: $\mathbf{B} := (\mathbf{B}_t, (\widehat{d}, l_f))$.
  2.e. Decrement c.
  2.f. Go back to step 2.b. if $c > 0$.
  2.g. Go back to step 2. if $J(\mathbf{B}) < J_{old}$.
  2.h. Go to step 3. if $l_{min} > L_{min}$.
  2.i. End here the JDBLMO.
3. Apply the BDA on $\mathbf{B}$:
  3.a. Set $l_{min} = \frac{l_{min}}{2}$.
  3.b. Set $c := T$ and $f := 1$.
  3.c. Go to step 3.f. if $l_f = l_{min}$.
  3.d. Compute $\mathbf{B}_s$ as in (7.13)
  3.e. Set $\mathbf{B} := \mathbf{B}_s$ and $f := 0$ and go to step 3.g. if $J(\mathbf{B}_s) < J(\mathbf{B})$.
  3.f. Set $\mathbf{B} := [\mathbf{B}_t, B_f]$.
  3.g. Decrement c.
  3.h. Go to step 3.c. if $c > 0$.
  3.i. Go to step 3.b. if $f = 0$
  3.j. Go to step 2.

Figure 7.3: Description of the JDBLMO algorithm

A decreasing counter denoted c is introduced, both in the ARA and the BDA. It ensures that each block B is processed exactly once, regardless of the evolving size of the queue. This guarantees that at the end of the processing the queue's first block is located at the top left of the image. This is how it works: c is first set as the initial size of the block map, it is then decremented at each iteration and repetition terminates when c reaches 0. Control of the repetition is achieved using the decreasing counter c (steps 2.a., 2.e. and 2.f. and steps 3.b., 3.g. and 3.h. in Figure 7.3).

7.2.2 Adapation of the R-Algorithm (ARA)

The ARA is an adaptation of the R-algorithm developed in chapter 5. The main steps of this latter are briefly reminded to extend its strategy to blocks of non-equal sizes.

The R-algorithm finds a suboptimal solution of an optimization problem slightly different from the one expressed by Equation (7.6). Indeed the cost function denoted here as J_r that is minimized by the R-algorithm does not take into account the size of the blocks since it is assumed to be of equal-size $l \times l$. Thus the term $H(\mathbf{l})$ related to the block length map is absent from the cost function:

$$J_r(\lambda, \mathbf{d}) = E_{Gr}(\mathbf{d}) + \lambda_r H(\mathbf{d}), \tag{7.10}$$

where $E_{Gr}(\mathbf{d}) = E_G(\mathbf{d}, \mathbf{l})$ with $\mathbf{l}$ the constant block length map only composed of the same repeated value l and $\lambda_r = \lambda T$.

The disparity map $\mathbf{d}$ is first set using the BMA as in (7.7). The R-algorithm processes the disparity map by successively re-evaluating each selected disparity: each of them is being replaced by all the other disparities of the search area D, resulting in as many disparity maps as disparities in the search area. A rate-distortion cost is computed for each disparity map and the one minimizing the cost is set as the new disparity of the current block. The process is then iterated until re-evaluating all the disparities in the map. After that, considering the new updated disparity map, the whole process is re-iterated again until no improvements are observed in terms of rate-distortion.

This R-algorithm is now adapted to cope with not only the variable number of blocks but also the variable sizes of these blocks. The process is achieved through the use of $\mathbf{B}$ as a queue. The disparity of the front block, d_f is replaced by all disparities $d \in D$, resulting in as many block maps $[(d, l_f), \mathbf{B}_t]$. The selected disparity $\widehat{d}$ minimizes J is given by:

$$\widehat{d} = \arg\min_{d \in D} J\left([(d, l_f), \mathbf{B}_t]\right). \tag{7.11}$$

This modified block is removed from the front and placed at the queue's end:

$$\mathbf{B} := \left[\mathbf{B}_t, (\widehat{d}, l_f)\right]. \tag{7.12}$$

Using the decreasing counter c, this process is iterated in such a way as to process all blocks once and to ensure that the front block is the block located at the image's top left corner. It is further iterated as long improvements are made, this is checked by comparing the current value $J(\mathbf{B})$ of the cost function with a stored value denoted J_{old}.

7.2.3 Block Division Algorithm (BDA)

This section deals with the third stage, Block Division Algorithm (BDA), of the optimization algorithm. The BDA consists in processing all blocks. For each block, is decided, whether dividing the block into four subblocks helps reducing the function cost $J(\lambda, \mathbf{d}, \mathbf{l}, T)$ or not. This decision is taken by comparing the cost function when applied to two block maps, one where the front block is not divided and one where the front block is divided and disparities of each subblock are adjusted so as to minimize $J(\lambda, \mathbf{d}, \mathbf{l}, T)$. Note that the BDA modifies the tree-structure only within a specific layer consisting of all blocks of a given size. When it is first applied, it attempts to divide blocks of size $L_{max} \times L_{max}$. When applied a second time, it considers blocks of size $\frac{L_{max}}{2} \times \frac{L_{max}}{2}$ and so on until reaching the desired size $L_{min} \times L_{min}$. The parameter l_{min} at the step 3.c. in Figure 7.3 controls the size of the blocks that might be divided by the BDA.

Let us first describe how to transform a block map $\mathbf{B}$ into a new block map, denoted $\mathbf{B}_s$, containing four new subblocks (corresponding to the step 3.d. in Figure 7.3):

- The front block (d_f, l_f) is divided into four subblocks denoted as $(d_f, \frac{l_f}{2})$,$(d_f, \frac{l_f}{2})$, $(d_f, \frac{l_f}{2})$ and $(d_f, \frac{l_f}{2})$;

- A new block map is obtained by removing the front block and placing at the queue's end these four blocks:

$$\left[\mathbf{B}_t, (d_f, \frac{l_f}{2}), (d_f, \frac{l_f}{2}), (d_f, \frac{l_f}{2}), (d_f, \frac{l_f}{2})\right];$$

- The first subblock disparity d_f is replaced by all disparities d in the search area, and the disparity d_1 is selected so as to minimize $J(\lambda, \mathbf{d}, \mathbf{l}, T)$:

$$d_1 = \arg\min_{d \in D} J\left(\left[\mathbf{B}_t, (d, \frac{l_f}{2}), (d_f, \frac{l_f}{2}), (d_f, \frac{l_f}{2}), (d_f, \frac{l_f}{2})\right]\right);$$

- The selection of the second subblock disparity d_2 takes into account the choice of d_1. It is also computed by testing the cost function on all disparities within the search area:

$$d_2 = \arg\min_{d \in D} J\left(\left[\mathbf{B}_t, (d_1, \frac{l_f}{2}), (d, \frac{l_f}{2}), (d_f, \frac{l_f}{2}), (d_f, \frac{l_f}{2})\right]\right);$$

- The selection of the third and fourth subblock disparities d_3, d_4 are obtained in the same way:

$$d_3 = \arg\min_{d \in D} J\left(\left[\mathbf{B}_t, (d_1, \frac{l_f}{2}), (d_2, \frac{l_f}{2}), (d, \frac{l_f}{2}), (d_f, \frac{l_f}{2})\right]\right),$$
$$d_4 = \arg\min_{d \in D} J\left(\left[\mathbf{B}_t, (d_1, \frac{l_f}{2}), (d_2, \frac{l_f}{2}), (d_3, \frac{l_f}{2}), (d, \frac{l_f}{2})\right]\right),$$

- The proposed block map is then expressed as:

$$\mathbf{B}_s = \left[\mathbf{B}_t, (d_1, \frac{l_f}{2}), (d_2, \frac{l_f}{2}), (d_3, \frac{l_f}{2}), (d_4, \frac{l_f}{2})\right]. \tag{7.13}$$

The decision consists in comparing $J(\mathbf{B})$ with $J(\mathbf{B}_s)$. If $J(\mathbf{B}_s)$ is smaller, then the new block map considered is $\mathbf{B}_s$. If, on the contrary, $J(\mathbf{B})$ is smaller, then the front block is placed at the queue's end indicating that it has been processed and the new block map considered is

$$\mathbf{B} := [\mathbf{B}_t, (d_f, l_f)]$$

The block-division algorithm consists in repeating these steps (computing $\mathbf{B}_s$ and making the comparison between the cost functions) for all the blocks of the same layer in the tree (see Figure 7.2), i.e all the blocks having the same size. Repetition goes on until the first block is back at the front position of the queue. This repetition is further iterated with the blocks of a same layer as long as at least one block division has been decided, this is checked with a flag f that is initially switched on. It is switched off as soon as a block is divided. The control of this flag is done by steps 3.b and 3.i in Figure 7.3.

7.3 Performance evaluation and discussions

This section analyzes the performance of the proposed JDBLMO algorithm. The ability of this algorithm to achieve a good prediction of the right view knowing the exact left view while requiring the least bitrate is discussed. The rationale is that such an algorithm should also yield good performance when it is integrated in a disparity compensated coding scheme.

The Peak Signal-to-Noise Ratio (PSNR) is adopted to measure the prediction quality computed between the original and the predicted right view. The bitrate is expressed in bits per pixel (bpp) and computed according to the entropy approximation applied to both the block-length and the disparity maps.

Simulations are conducted on stereoscopic images taken from the "CMU-VASC" dataset [51]. The JDBLMO performance is compared with the traditional BMA using same size-block for stereoscopic images but also with the modified version of the block-matching algorithm Intra/Inter-frame Block Segmentation Coding (IIBSC) using variable size-blocks developed in [54].

The predicted view is first divided into fixed-size blocks. Each block is encoded either using an intra-frame or an inter-frame coding scheme. The latter scheme makes use of a metric assessing the quality of the predicted block as compared to the corresponding block on the original image using the Sum of Squared Differences (SSD). This scheme consists in two steps: blocks are repeatedly divided so as to lower the SSD below a threshold, then these divided blocks are repeatedly merged as long as the SSD remains below this threshold. This algorithm has been adapted to our context and is called MIIBSC (Modified IIBSC). The modifications made to this algorithm consist in allowing only the inter-frame coding and in adapting this prediction mode to the context of stereoscopic image coding. The inter-frame coding becomes the inter-view coding taking into account the two adjacent views. Furthermore only the first step is retained and is performed as follows. The fixed-size blocks are repeatedly and recursively divided into four subblocks of equal size until the mean squared error of each divided subblocks gets below a threshold.

The performance discussion is organized as follows. JDBLMO, BMA and MIIBSC algorithms are compared first in a more detailed fashion in subsection 7.3.1 and then in terms of average rate-distortion performance in subsection 7.3.2.

7.3.1 JDBLMO performance evaluation compared to BMA and MIIBSC

This section compares the performance of the proposed algorithms to the test algorithms described above using the stereoscopic image "rubik" shown at the top of Figure 7.5.

In the discussed simulations results, the right view is first divided into blocks of size 32×32 pixels. The candidate disparities are selected within the set $[-30, -29, \ldots, 30]$. The MIIBSC is tested according to a large number of thresholds ranging from 0 to 1000 which are applied to the mean SSD. For a given threshold, a predicted view, a PSNR-value, a bit-rate and a global block layout are obtained. While the JDBLMO is performed using a large number of λ-values. The BMA is performed with four fixed set of size-blocks: 4×4, 8×8, 16×16 and 32×32. For each parameter value, each algorithm yields a predicted view, a global block layout, and a rate-distortion point composed of a PSNR value and a bitrate value.

Figure 7.4 illustrates the rate-distortion performance of the algorithms by joining with lines the experimental rate-distortion points. Note that the JDBLMO (black dashed curve joining black squares) performs better than the BMA (red curve joining red pluses) and even better than the MIIBSC (blue curve joining blue circles). For a given bitrate of 0.015 bpp, the JDBLMO yields a PSNR of 32.0 dB improving by 1.8 dB the BMA's performance and by 0.9 dB the MIIBSC's performance.

Figure 7.5 presents above the original right view and below three close-up views of the right part of a "wooden egg cup" extracted from the three predicted views processed with respectively from left to right, the JDBLMO, MIIBSC and BMA. The JDBLMO yields the best visual

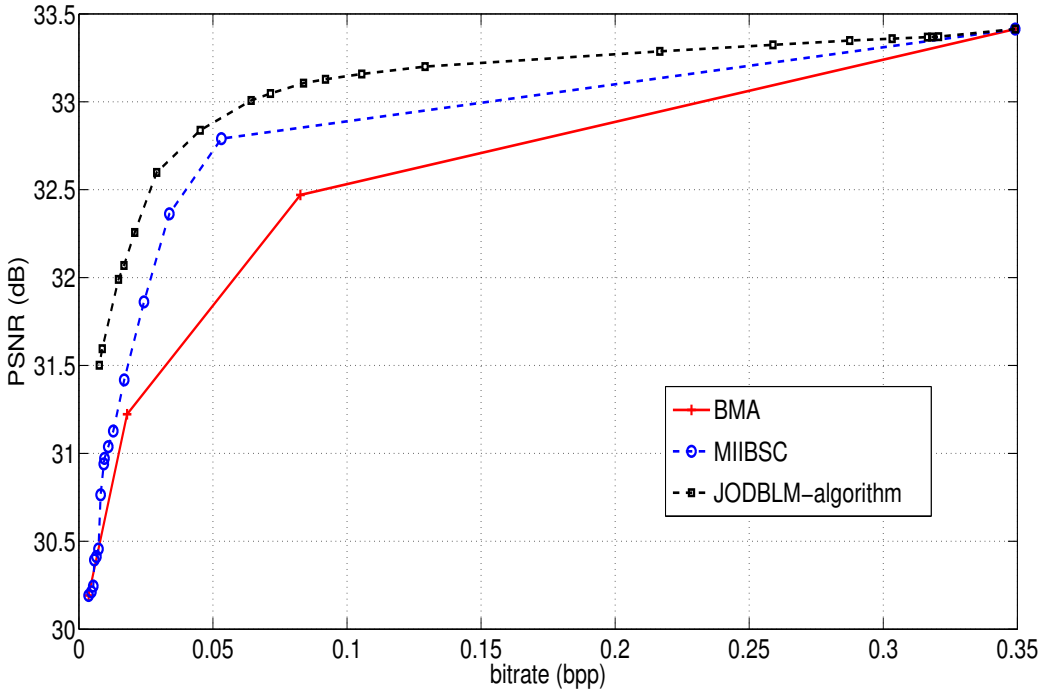

Figure 7.4: Rate-distortion performance of the JDBLMO, MIIBSC and BMA performed on "rubik".

reconstruction as the two others fail to render the triangular shape. Indeed this can be explained by the block layout selected respectively by the JDBLMO and the MIIBSC given by Figure 7.6 and Figure 7.7. A closer look at these figures show that the triangular shape is processed with seven subblocks by the JDBLMO and with only one block by the MIIBSC.

7.3.2 JDBLMO average rate-distortion performance compared to BMA and MIIBSC

This section analyzes the average rate-distortion performance of the described algorithms. Simulations are performed on ten stereoscopic images, namely "book", "books", "sphere", "wdc2r", "whouse", "rubik", "telephone", "toys", "cdc1", "house2". Eight target bitrates in bpp are defined: [0.06 0.07 0.08 0.09] and [0.1 0.25 0.40 0.55], the former corresponding to low bitrate and the latter corresponding to medium bitrate. The parameter values (λ for JDBLMO, threshold for MIIBSC and size-block for BMA) are set so as to comply the target bitrates. The Bjøntegaard metric [53] is then used to compute at low and at medium bitrate, an average PSNR difference between the JDBLMO and the BMA, and between the JDBLMO and the MIIBSC. These average PSNR differences are listed in Table. 7.3.2 for each of the ten stereoscopic images, where Δ`BMA` stands for the former average difference and Δ`MIIBSC` the latter average difference. One can observe that the JDBLMO achieves better results when compared to the BMA and in

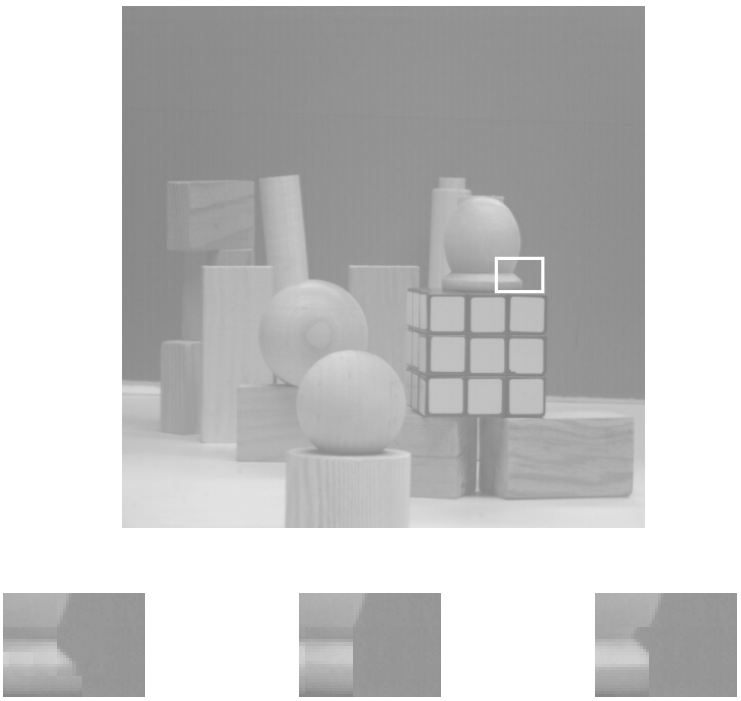

Figure 7.5: Above: original right view "rubik". Below, from left to right close-up views extracted from the three predicted views processed with the JDBLMO, MIIBSC and BMA.

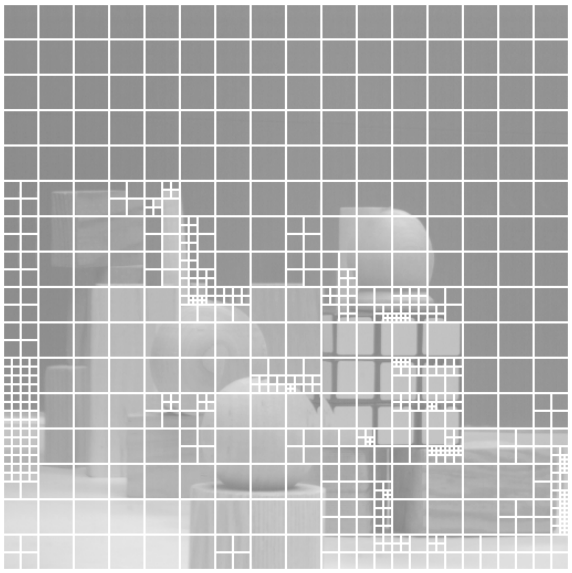

Figure 7.6: Block-length map resulting from the JDBLMO

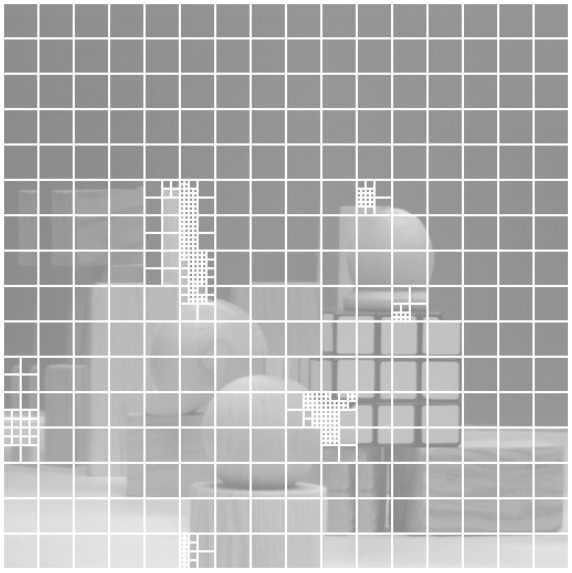

Figure 7.7: Block-length map resulting from the MIIBSC

most cases when it is compared to the MIIBSC. The gain is even higher at low bitrate compared to the BMA.

7.3.3 JDBLMO rate-distortion performance compared to MMA and R-algorithm

This section compares the performance of the JDBLMO with the MMA and the R-algorithm developed in the previous chapters and also with the BMA.

Figure 7.8 shows the results of an experiment conducted on the stereoscopic image "Stereo_13". The disparity map is estimated selecting disparities between the range $[-29, \ldots, 29, 30]$.

The black dashed line joining circles gives the performance of the BMA using of size 4×4 to 14×14 pixels. Each performance achieved with a given size of block is represented by a circle. The set of blue curves joining '+' symbols represents the performance of the MMA achieved using blocks of size 4×4 (for the right blue curve) and 8×8 (for the left blue curve) where each point is obtained using a different value of the parameter λ. In the same way, the set of red curves represents the performance of the R-algorithm using blocks of size 4×4 (for the right blue curve) and 8×8 (for the left blue curve) where each point is obtained using a different value of λ. Finally, the curve in purple joining '+' symbols represents the performance of the JDBLMO algorithm using variable block sizes for 32×32 to 4×4 pixels. Each point is obtained using a different value of λ. One can see that for high bitrate, i.e. form 0.17bpp to 0.3bpp, the JDBLMO is globally less performant than the R-algorithm which is better than the MMA itself better than the BMA. For example, at the bitrate of 0.31, the MMA, the R-algorithm and the BMA leads to same prediction with a quality of 29.6dB. While for the same quality, the JDBLMO needs 0.35bpp. This can be explained by the fact that all algorithms uses small blocks to reconstruct

Stereo	Low bitrate		Medium bitrate	
Images	ΔBMA	ΔMIIBSC	ΔBMA	ΔMIIBSC
book	0.95	0.51	0.67	0.03
books	1.08	0.10	0.90	0.19
sphere	1.14	0.90	1.02	0.25
wdc2r	1.30	0.55	0.45	-0.74
whouse	1.45	0.55	1.46	0.44
rubik	1.64	0.22	0.74	0.44
telephone	2.45	0.81	1.89	0.34
toys	3.09	-0.27	2.51	0.13
cdc1	3.14	0.50	1.80	0.04
house2	3.24	0.12	0.51	-0.01

Table 7.1: Average PSNR gain of the JDBLMO over the BMA and the MIIBSC

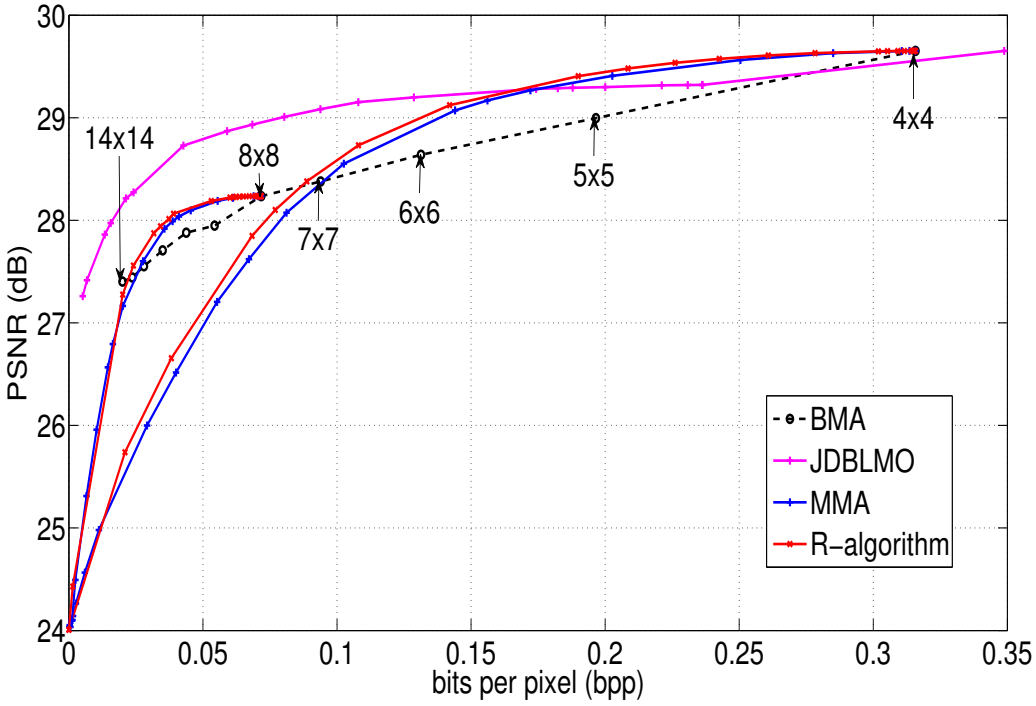

Figure 7.8: Rate-distortion performance of the MMA, R-algorithm, JDBLMO and BMA on "stereo_13".

the predicted view which leads to a good reconstruction quality. Nevertheless, the JDBLMO has to encode also a block length map containing the constant size of the blocks which is not useful in this case.

At lower bitrates, from 0.005bpp to 0.13bpp, the JDBLMO performs better than all the other algorithms. It takes advantage of coding large homogenous areas using big blocks but still ensuring a good prediction quality thanks to smaller blocks.

These results confirms the use of blocks of variable block sizes at low bitrates.

7.3.4 Processing time and complexity of the JDBLMO

The simulations above have been conducted in the same conditions as mentioned in section 3.3.7. For the stereoscopic image "Stereo_13" of size 202×320 pixels, using disparities amongst the range $[-30, ..., 29]$, the JDBLMO takes between 7 and 35 seconds to estimate a disparity map for a specific value of the parameter λ. The complexity of the JDBLMO depends on the depth q of the tree structure. It has been calculated in too cases. IN the first case, the JDBLMO has the minimum complexity at very low bitrate when no blocks are being divided, thus the ARA and the BDA are applied only once and the complexity of the JDBLMO is given by:

$$18N_w L_{max}^2 + 6N_w XY + 5N_w + 12 \tag{7.14}$$

The complexity of the JDBLMO has also been calculated at very high bitrate when all block are being divided until the minimum block size allowed. We have assumed that the ARA and the BDA are applied only once for each block size. Thus the complexity of the JDBLMO is given by:

$$\begin{gathered} N_w(XY)(3L_{max}^2 - 1) \\ + \sum_{k=0}^{q-2} 3N_w XY L_{max}^2 + 11N_w \left(2^{2k} XY\right) - 7N_w - 1 \\ + \sum_{k=0}^{q-2} \left(2^{2k} XY\right) \left(4N_w \left(\frac{L_{max}^2}{2^{2k}} - 1\right) + 12N_w + 13\right) \end{gathered} \tag{7.15}$$

7.4 Conclusion

This chapter has presented a new stereo-matching image algorithm using variable size-block. This algorithm optimized jointly the disparity and block-length maps so as to ensure a good reconstruction of the predicted view while minimizing the bitrate. Simulation results shown that this sub-optimal stereo-matching algorithm achieved better prediction performance when compared to the competitive BMA algorithm using variable size-block.

Chapter 8

Conclusion and future research

The developed work in this thesis has been undertaken in a context where the number of applications using 3D-technology is continuously increasing and subjected to a growing demand from the users in terms of quality and comfort of use. Stereoscopic images require efficient techniques for their storage or their transmission. The contributions of the work concern the compression of stereoscopic images and more particularly, we focus on the disparity map estimation process improving the quality of the predicted view with a constraint of low bitrate.

The disparity map estimation for stereoscopic image compression has been at the center of many works. We have proposed five original algorithms to cope with the Rate-Distortion optimization problem of coding stereo images. The main contributions of this thesis are recalled below and then some clues of possible improvements are discussed.

Our first contribution was to develop a disparity map estimation algorithm (MMA) using a joint entropy distortion metric. Most of the methods developed in the literature include a smoothness constraint assuming that a smoother disparity map requires a bitrate less important. Contrarily to these methods, the MMA includes an estimate of the bitrate needed to encode the disparity map in the metric to minimize as well as a common term measuring the dissimilarity between the blocks to be matched. This allows the MMA to ensure good bitrate-distortion performance at each step of the disparity selection process. Furthermore as the complete disparity map distribution is unknown during the estimation process, the final bitrate cost of the disparity map is estimated at each step of the disparity map computation thanks to a finite mixture probabilities. This mixture is defined such as the disparity distribution estimate becomes more and more close to the true distribution as the image is being processed. To have the ability to explore a large number of paths, the MMA relies in a tree where each path corresponds to a different disparity map. At each depth of the tree, only the M-best paths are extended.

This method has shown better performance on the stereoscopic images that were tested compared to the Block-Matching Algorithm (BMA) with and without regularization, also when coding the estimated map with a DPCM coder.

Improvements in terms of Rate-Distortion performance were mainly obtained setting the

parameter M to 1 (i.e. by just extending only one disparity map at each depth of the tree). Some of the simulations have shown a significant improvements in the bitrate-distortion performance by increasing the values of M. But increasing the value of M is done at the expense of an increase in the processing time also.

No particular attention have been taken in this algorithm to make it faster: programming it in C-language or including parallel programming in the algorithm could make it run faster. We are optimist that making some improvements dedicated to the reduction of the processing time could results in a much less important processing time and allows to increase the value of the parameter M to further improve the bitrate-distortion performance.

Another point to consider is that this algorithm involves many parameters. We have conducted a short study on how to set all the parameters to obtain best performance. A future research could consist in incorporating this algorithm in a complete disparity compensated scheme. Then the bit allocation problem should be investigated and would be more complex as it would involve all the extra parameters of the MMA which all have an influence on the bitrate of the disparity map.

Another contribution of this paper was the proposed R-algorithm which takes initially the disparity map computed by the BMA which minimizes the distortion as a reference. Then this map is successively modified as long as improvements are observed in terms of rate-distortion performance. The disparities are also modified according to the minimization of the proposed joint entropy-distortion metric. This algorithm shows to perform better than the BMA and the previously developed MMA on the stereoscopic images tested. This algorithm has also the advantage of being simpler to use than the MMA as it does not involves any specific parameters, and thus its processing time is significantly reduced in comparison to the MMA.

Then we get interested in improving the performance of the developed algorithms by enlarging the search area in which the disparities are selected. Enlarging the search area can of course improves the performance of the algorithm in terms of distortion. But it can also increase the bitrate needed to encode the disparity map as the range of selected disparities gets larger. Thus, we propose two methods to select relevant sets of disparities inside the search area, each set minimizing the distortion of the predicted view at a given bitrate. Simulation results have shown that both methods perform better that the BMA. In future investigations, this algorithm could be used jointly with the other disparity map estimation algorithm proposed in this thesis. Indeed, one could at first estimate the best set of disparities improving the distortion performance given a search area and then this best set could be used by either the R-algorithm or the JDBLMO as the set of disparities on which they rely to estimate the disparity map. This is expected to improve the Rate-Distortion performance of both algorithms as they will be able to take advantage of a larger search area for a small increase of the computational complexity.

Finally we have investigated the use of variable block sizes for stereoscopic image compression with the JDBLMO. As it is expected to improve the bitrate-distortion performance: larger zones having roughly the same depth can be encoded using a single disparity. But this comes

at the cost of sending also to the decoder information about the block sizes. We have propose a novel algorithm which jointly optimize the disparity map as well as the block sizes while minimizing the distortion of the predicted view. The proposed algorithm performs better than the MIISBC which is a method using variable block sizes inspired from video coding techniques.

In future investigations, all the developed techniques could be integrated in a disparity compensated scheme. In fact, we have already started these works with the R-algorithm. The base view and the residual image were coded using HEVC coder and the disparity map with a lossless entropy coder. The distortion was measured on both views of the stereo images and the bitrate computed as the sum of the bitrate cost of encoding the left view, the residual image and the disparity map. The first simulations results we obtained are very encouraging to pursue in this way as improvements have been observed in terms of global Rate-Distortion performance for very low bitrate in comparison to the use of 3D-HEVC to encode the stereoscopic images tested.

Moreover, the developed algorithms could be included in the 3D-HEVC coder to further improve its performance on stereoscopic video coding.

For example, the BMA_S can be used as a initial step to select the best disparities set in a larger search area, so that this set could be used during the Disparity Compensated Prediction (DCP) of the 3D-HEVC.

Another idea would be to run the R-algorithm on the blocks coded using DCP in 3D-HEVC. The disparities selected by this 3D-HEVC for the blocks in one frame could be further improved by applying the R-algorithm on this partial disparity map (as some blocks are coded using intra prediction or inter-motion prediction). Improvements in terms of Rate-Distortion performance are expected to be more important in the case of the coding of I-frames stereoscopic pairs as these frames are coded independently of the rest of the videos (they are coded using only intra coding and DCP).

Bibliography

[1] Peter Kauff and Oliver Schreer. An immersive 3D video-conferencing system using shared virtual team user environments. In *Proceedings of the 4th international conference on Collaborative virtual environments*, pages 105–112. ACM, 2002.

[2] John C Byrn, Stefanie Schluender, Celia M Divino, John Conrad, Brooke Gurland, Edward Shlasko, and Amir Szold. Three-dimensional imaging improves surgical performance for both novice and experienced operators using the da vinci robot system. *The American Journal of Surgery*, 193(4):519–522, 2007.

[3] Aljoscha Smolic, Karsten Mueller, Nikolce Stefanoski, Joern Ostermann, Atanas Gotchev, Gozde B Akar, Georgios Triantafyllidis, and Alper Koz. Coding algorithms for 3dtv - a survey. *IEEE transactions on circuits and systems for video technology*, 17(11):1606–1621, 2007.

[4] Shiping Li, Mei Yu, Gangyi Jiang, Tae-Young Choi, and Yong-Deak Kim. Approaches to h. 264-based stereoscopic video coding. In *Image and Graphics (ICIG'04), Third International Conference on*, pages 365–368. IEEE, 2004.

[5] Woontack Woo and Antonio Ortega. Stereo image compression with disparity compensation using the mrf model. In *Visual Communications and Image Processing'96*, pages 28–41. International Society for Optics and Photonics, 1996.

[6] Udo Ahlvers, Udo Zoelzer, and Stefan Rechmeier. Fft-based disparity estimation for stereo image coding. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 1, pages I–761. IEEE, 2003.

[7] Kemal Ugur, Alexander Alshin, Elena Alshina, Frank Bossen, Woo-Jin Han, Jeong-Hoon Park, and Jani Lainema. Motion compensated prediction and interpolation filter design in h. 265/hevc. *IEEE Journal of Selected Topics in Signal Processing*, 7(6):946–956, 2013.

[8] Aysha Kadaikar, Gabriel Dauphin, and Anissa Mokraoui. Sequential block-based disparity map estimation algorithm for stereoscopic image coding. *Signal Processing: Image Communication*, 39:159–172, 2015.

[9] Aysha Kadaikar, Anissa Mokraoui, and Gabriel Dauphin. Entropy-constrained dense disparity map estimation algorithm for stereoscopic images. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, pages 241–245. IEEE, 2014.

[10] Aysha Kadaikar, Gabriel Dauphin, and Anissa Mokraoui. Extended map estimation algorithm using joint entropy-distortion metric for non-rectified stereoscopic images. In *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on*, pages 505–510. IEEE, 2015.

[11] Aysha Kadaikar, Gabriel Dauphin, and Anissa Mokraoui. Estimation de cartes de disparité basée sur l'algorithme revisité d'appariement de blocs pour la compression d'images stéréoscopiques. In *Proceedings of the GRETSI Conferences*, 2015.

[12] Aysha Kadaikar, Gabriel Dauphin, and Anissa Mokraoui. Modified block matching algorithm improving rate-distortion performance for stereoscopic image coding. In *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on*, pages 478–483. IEEE, 2015.

[13] Aysha Kadaikar, Gabriel Dauphin, and Anissa Mokraoui. Méthode d'optimisation pour l'appariement de pixels d'images stéréoscopiques basée sur une métrique conjointe entropie-distorsion. In *Compression et Représentation des Signaux Audiovisuels (CORESA)*, 2014.

[14] Aysha Kadaikar, Gabriel Dauphin, and Anissa Mokraoui. Improving block-matching algorithm by selecting disparity sets minimizing distortion for stereoscopic image coding. In *Visual Information Processing (EUVIP), 2016 6th European Workshop on*, pages 1–5. IEEE, 2016.

[15] Aysha Kadaikar, Gabriel Dauphin, and Anissa Mokraoui. Joint disparity and variable size-block optimization algorithm for stereoscopic image compression. *Signal Processing: Image Communication*, 2017 (under revision).

[16] Olivier Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT press, 1993.

[17] Zhengyou Zhang and Gang Xu. A general expression of the fundamental matrix for both perspective and affine cameras. In *Proceedings of the Fifteenth international joint conference on Artifical intelligence-Volume 2*, pages 1502–1507. Morgan Kaufmann Publishers Inc., 1997.

[18] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty: A review. *International journal of computer vision*, 27(2):161–195, 1998.

[19] Richard Hartley and Rajiv Gupta. Computing matched-epipolar projections. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, pages 549–555. IEEE, 1993.

[20] Andrea Fusiello, Emanuele Trucco, and Alessandro Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.

[21] Demetrios V Papadimitriou and Tim J Dennis. Epipolar line estimation and rectification for stereo image pairs. *IEEE transactions on image processing*, 5(4):672–676, 1996.

[22] Yuichi Ohta and Takeo Kanade. Stereo by intra-and inter-scanline search using dynamic programming. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (2):139–154, 1985.

[23] Gerard Medioni and Ramakant Nevatia. Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31(1):2–18, 1985.

[24] Cordelia Schmid and Andrew Zisserman. The geometry and matching of curves in multiple views. In *European Conference on Computer Vision*, pages 394–409. Springer, 1998.

[25] Takeo Kanade. Developement of a video-rate stereo matching. *Proceedings of Image Understanding Workshop*, pages 549–557, 1994.

[26] Woontack Woo and Antonio Ortega. Overlapped block disparity compensation with adaptive windows for stereo image coding. *Circuits and Systems for Video Technology, IEEE Transactions on*, 10(2):194–200, 2000.

[27] Andrea Fusiello, Vito Roberto, and Emanuele Trucco. Symmetric stereo with multiple windowing. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(08):1053–1066, 2000.

[28] Jae Chul Kim, Kyoung Mu Lee, Byoung Tae Choi, and Sang Uk Lee. A dense stereo matching using two-pass dynamic programming with generalized ground control points. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 1075–1082. IEEE, 2005.

[29] Olga Veksler. Stereo correspondence by dynamic programming on a tree. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 384–390. IEEE, 2005.

[30] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001.

[31] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions using graph cuts. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 508–515. IEEE, 2001.

[32] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(7):787–800, 2003.

[33] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient belief propagation for early vision. *International journal of computer vision*, 70(1):41–54, 2006.

[34] Berthold K Horn and Brian G Schunck. Determining optical flow. In *1981 Technical Symposium East*, pages 319–331. International Society for Optics and Photonics, 1981.

[35] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[36] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra. Overview of the H.264/avc video coding standard. *IEEE Transactions on circuits and systems for video technology*, 13(7):560–576, 2003.

[37] Krzysztof Wegner Ying Chen, Gerhard Tech and Sehoon Yea. Test model 10 of 3d-hevc and mv-hevc. Technical report, ISO/IEC JTC1/SC29/WG11, 2014.

[38] Kenneth Zeger et al. Residual image coding for stereo image compression. *Optical Engineering*, 42(1):182–189, 2003.

[39] Haluk Aydinoglu, Faouzi Kossentini, Qin Jiang, and Monson H Hayes. Region-based stereo image coding. In *Image Processing, 1995. Proceedings., International Conference on*, volume 2, pages 57–60. IEEE, 1995.

[40] Jianmin Jiang and Eran A Edirisinghe. A hybrid scheme for low bit-rate coding of stereo images. *IEEE Transactions on Image Processing*, 11(2):123–134, 2002.

[41] Torsten Palfner, Alexander Mali, and Erika Muller. Optimized progressive coding of stereo image using discrete wavelet transform. In *World Multiconference on Systemics, Cybernetics and Informatics*, volume 1, pages 35–39, 2002.

[42] Amir Said and William A Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on circuits and systems for video technology*, 6(3):243–250, 1996.

[43] Carla L Pagliari and Tim J Dennis. Stereo disparity computation in the dct domain using genetic algorithms. In *Image Processing, 1997. Proceedings., International Conference on*, volume 3, pages 256–259. IEEE, 1997.

[44] Aldo Maalouf and Mohamed-Chaker Larabi. Bandelet-based stereo image coding. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 698–701. IEEE, 2010.

[45] Rafik Bensalma and Mohamed-Chaker Larabi. A wavelet-based quadtree driven stereo image coding. In *IS&T/SPIE Electronic Imaging*, pages 72371Z–72371Z. International Society for Optics and Photonics, 2009.

[46] Frederick Jelinek. Fast sequential decoding algorithm using a stack. *IBM Journal of Research and Development*, 13(6):675–685, 1969.

[47] Daniel Scharstein and Richard Szeliski. Middlebury stereo vision page. *Online at http://vision.middlebury.edu/stereo*, 2, 2002.

[48] Miloš Klíma, Karel Fliegel, Petr Páta, Stanislav Vítek, Martin Blažek, Petr Dostal, Lukáš Krasula, Tomáš Kratochvíl, Václav Říčný, Martin Slanina, et al. Deimos-an open source image database. *Radioengineering*, 20(4):1016–1023, 2011.

[49] Iain E Richardson. *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia*. John Wiley & Sons, 2004.

[50] Matlab toolbox: http://www.mathworks.fr/products/computer-vision/code-examples.html?file=%2fproducts%2fdemos %2fshipping%2fvision%2fvideostereo.html.

[51] Cmu vision and autonomous systems center (vasc) image database. http://vasc.ri.cmu.edu/idb/html/stereo/.

[52] M. Klima. Deimos-an open source image database, 2011.

[53] G. Bjontegaard. Calculation of average psnr differences between rd. *ITU-T SC16/Q6, 13th VCEG Meeting, Document VCEG-M33*, April 2001.

[54] MH Chan, YB Yu, and AG Constantinides. Variable size block matching motion compensation with applications to video coding. *IEE Proceedings I-Communications, Speech and Vision*, 137(4):205–212, 1990.