



Reconstruction 3D à partir de paires stéréoscopiques en conditions dégradées

Laurent Caraffa

► To cite this version:

Laurent Caraffa. Reconstruction 3D à partir de paires stéréoscopiques en conditions dégradées. Traitement des images [eess.IV]. Université de Paris VI, Pierre et Marie Curie, 2013. Français. NNT : . tel-01365130

HAL Id: tel-01365130

<https://hal.science/tel-01365130>

Submitted on 13 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reconstruction 3D à partir de paires stéréoscopiques en conditions dégradées

THÈSE

soutenue le 29 Novembre 2013

pour l'obtention du

Doctorat de l'Université de Paris 6
(mention informatique)

par

Laurent Caraffa

Composition du jury

<i>Président :</i>	M. Pierre Charbonnier	
<i>Rapporteurs :</i>	M. Fabrice Heitz	Professeur, Université de Strasbourg
	M. Peter Sturm	DR, INRIA Rhône-Alpe
<i>Examineurs :</i>	M. Ryad Benosman	MCF, DR, Paris 6
	M. Pierre Charbonnier	DR, MEDDE, LRPC Strasbourg ERA27 IFSTTAR
<i>Encadrant :</i>	M Jean-Philippe Tarel	CR, IFSTTAR
<i>Directeur :</i>	Mme. Catherine Achard	HDR, Paris 6
	M. Nicolas Paparoditis	DR, IGN

Mis en page avec la classe thesul.

Remerciements

Je remercie l'équipe du LEPSIS de l'IFFSTAR, qui m'a accueilli dans le cadre du doctorat. Durant ces trois années j'ai pu mesurer combien la qualité des relations humaines et la convivialité sont aussi indispensables que les compétences techniques. Ma présence journalière au laboratoire ne s'est pas limitée à une obligation.

J'ai partagé des moments passionnants avec les autres doctorants et post-doctorants qui m'ont soutenu avec une entraide permanente. Dès mon arrivée l'accueil chaleureux et les conseils de Utku et Raouf ont donné le ton. Ces trois années se sont déroulées dans la bonne humeur avec Lancelot, Karin, Gesan. Je souhaite bon courage et réussite à Kevien, Verane, Ferath et Sophie pour leur dernière année de thèse. Je pense aussi à Céline, Emmanuel, Amit. Merci à tous ceux qui étaient présents.

L'aide constante de Sio-Song Ieng et Pierre Charbonnier en mathématiques a été précieuse, je les remercie pour leur écoute. Je remercie Monsieur Didier Aubert, Directeur du LEPSIS, Éric Dumont ainsi que Roland Brémond et Nicolas Hautière pour l'attention régulière qu'ils ont porté à mes travaux.

Je suis très honoré et je remercie les rapporteurs Fabrice Heitz et Peter Sturm pour le temps qu'ils ont bien voulu consacrer au manuscrit avec la prise en compte des contraintes imprévues de calendrier en fin d'année.

Je n'ai jamais eu de contacts directs avec Vladimir Kolmogov et Hiroshi Ishikawa du laboratoire de vision de l'Université de Western Ontario, mais comment ne pas les remercier pour avoir mis leurs outils en libre accès. Sans leurs algorithmes d'optimisation je n'aurais pas pu obtenir certains résultats indispensables pour valider mes travaux.

Je remercie Monsieur Stéphane Zaleski, Directeur de l'Institut Jean Le Rond d'Alembert d'avoir permis que la soutenance se déroule dans les meilleures conditions dans son laboratoire.

Je remercie les directeurs de thèse et professeurs, Madame Catherine Achard, plus particulièrement pour son soutien en fin de thèse, Monsieur Nicolas Paparoditis pour ses conseils avisés, ainsi que Monsieur Jean Devars et les membres du jury, Monsieur Ryad Benosman, Monsieur Pierre Charbonnier président du jury.

Je tiens surtout à remercier Monsieur Jean-Philippe Tarel, qui m'a suivi et guidé de façon permanente pendant ces trois années d'apprentissage. Je réalise les progrès importants que j'ai pu faire grâce à son expertise, ses conseils et sa patience pédagogique.

Je remercie mes parents qui m'ont toujours épaulé.

Sommaire

Introduction

1	Traitement d'image et environnement routier	2
1.1	Modéliser l'environnement	2
1.2	Se déplacer dans un environnement	2
1.3	Aide à la conduite	3
1.4	Domaines de traitement d'image sous-jacents aux applications	3
2	Vision stéréoscopique pour l'environnement routier	3
3	Reconstruction 3D	5
3.1	Reconstruction 3D à partir de paires stéréoscopiques	6
3.1.1	S'inspirer de notre perception 3D	6
3.1.2	Fusionner les indices	7
3.1.3	Apport de connaissances	7
3.2	Spécificité des scènes routières	7
4	Traitement d'image en conditions dégradées	9
4.1	Brouillard	9
5	Modèle bayésien	11
5.1	Modèle bayésien	11
5.2	Optimisation discrète	11
6	Positionnement	12
7	Plan	12

Chapitre 1

Reconstruction 3D à partir de paires stéréoscopiques

15

1.1	Estimation de la carte de disparité	17
1.1.1	Hypothèse photométrique	19
1.1.2	Hypothèse spatiale	20
1.1.2.1	Les méthodes locales	20
1.1.2.2	Les méthodes globales	20

1.1.2.3	Les méthodes coopératives	21
1.1.2.4	Ordre de la régularisation spatiale	21
1.1.3	Prise en compte des variations photométriques et bruit non gaussien . . .	22
1.1.4	Prise en compte des discontinuités	22
1.1.5	Hypothèse du monde segmenté	23
1.1.6	Évaluation comparative	23
1.2	Formulation probabiliste	24
1.2.1	Interprétation probabiliste	25
1.2.1.1	Données du problème	25
1.2.1.2	Maximum <i>a posteriori</i>	26
1.2.1.3	Réseaux bayésiens	27
1.2.1.4	Terme de vraisemblance	28
1.2.1.5	<i>a priori</i>	29
1.2.1.6	Interprétation des valeurs aberrantes	30
1.2.1.7	Pondération photométrique :	31
1.2.2	Optimisation	32
1.2.2.1	Optimisation de fonctions sous-modulaires	33
1.2.2.2	Optimisation de fonctions non sous-modulaires	33
1.2.3	Exemples	33
1.2.4	Conclusion	38

Chapitre 2

Traitement d'image en présence de brouillard

41

2.1	Introduction	41
2.2	Effets du brouillard	44
2.3	Restauration d'image	45
2.3.1	Restauration par rehaussement de couleur et de contraste	45
2.3.1.1	Méthode <i>multiscale retinex</i> (MSR)	45
2.3.1.2	Méthode <i>Contrast-Limited Adaptive Histogram Equalization</i> (CLAHE)	46
2.3.2	Restauration fondée sur la loi de Koschmieder	46
2.3.2.1	Méthode du monde plan (PA)	46
2.3.2.2	Méthode <i>free-space segmentation</i> (FSS)	48
2.3.2.3	La contrainte <i>no-black-pixel</i> (NBPC)	49
2.3.2.4	Méthode <i>Dark Channel Prior</i> (DCP)	51
2.3.2.5	Contrainte <i>no-black-pixel</i> et du monde plan (NBPC+PA)	51
2.3.3	Égalisation locale de l'histogramme (HE+PA)	52
2.3.4	Interprétation bayésienne de la restauration	52

2.3.5	Évaluation	54
2.3.5.1	Brouillard de synthèse	54
2.3.5.2	Comparaison sur des images de synthèse	55
2.3.6	Conclusion	57
2.4	Estimation du coefficient d'extinction	58
2.4.1	L'algorithme du point d'inflexion	59
2.4.2	Conclusion	59

Chapitre 3

Technique d'optimisation d'énergie discrète

61

3.1	Optimisation de fonctions pseudo-bouliennes	63
3.1.1	Définitions et notations	63
3.1.2	Représentation	63
3.1.3	Propriétés des fonctions pseudo-bouliennes	63
3.1.4	Fonction pseudo-boulienne quadratique	65
3.1.4.1	<i>Roof duality</i>	65
3.1.4.2	Modèle de flot maximum / coupe minimale :	66
3.1.5	Cas sous-modulaire	70
3.1.5.1	Recherche du minimum global par coupe de graphe	70
3.1.6	Optimisation de fonctions pseudo-bouliennes de degré supérieur à 2	71
3.1.7	Réduction de fonction de degré quelconque en degré quadratique	72
3.1.7.1	Substitution	72
3.2	Extension au cas multi-labels	74
3.2.1	Stratégie de fusion dans le cas multi-labels	74
3.2.2	Optimisation globale de fonction multi-labels par algorithme itératif	75
3.2.3	α -expansion dans le cas sous-modulaire	75
3.3	Conclusion	76
3.3.1	Récapitulatif	78
3.3.2	Implantation libre	78

Chapitre 4

Restauration des contrastes dans une image

4.1	Restauration de l'image sachant la profondeur	80
4.1.1	Terme d'attache aux données	80
4.1.2	Terme a priori	81
4.1.3	Le choix entre la profondeur et le voile atmosphérique	82
4.2	Inférence du voile atmosphérique	82

4.2.1	Terme d'attache aux données	83
4.2.2	A priori	83
4.3	Restauration d'une seule image	83
4.4	Résultats sur des images de synthèse	85
4.5	Résultat sur des images réelles	86
4.6	Conclusion	87

Chapitre 5

Stéréo reconstruction et restauration simultanées d'images en présence de brouillard 91

5.1	Formulation	92
5.2	Modèle pour la reconstruction et restauration à partir de paires stéréoscopiques .	93
5.2.1	Modèle par champ de Markov	93
5.2.1.1	Terme d'attache aux données	93
5.2.1.2	Contrainte photométrique et <i>a priori</i> sur les pixels blancs	94
5.2.1.3	Terme <i>a priori</i>	94
5.2.1.4	Initialisation de $\tilde{D}$ et de $\tilde{I}_0$	95
5.2.1.5	Modèle complet	95
5.2.2	Optimisation	96
5.2.3	Raffinement avec un <i>a priori</i> du second ordre sur la disparité	97
5.3	Évaluation	98
5.3.1	Initialisation des paramètres	98
5.3.2	Sur images de synthèse	98
5.3.3	Sur image caméra	100
5.4	Conclusion	101

Chapitre 6

Estimation du coefficient d'extinction du brouillard par stéréovision 107

6.1	Estimation du coefficient d'extinction à partir du modèle global	107
6.2	Estimation du coefficient d'extinction par minimisation d'entropie	108
6.3	Problème	109
6.4	Méthode	111
6.4.1	Estimation de I_s	111
6.4.2	Sélection des régions	111
6.4.3	Estimation de β et de I_0	111
6.4.4	Algorithme	112
6.5	Évaluation	113

6.5.1	Capteur statique	113
6.5.2	Capteur en mouvement	114
6.6	Conclusion	115

Chapitre 7

Conclusion et perspectives	117
-----------------------------------	------------

7.1	Résultats et contributions	117
7.1.1	Modèle markovien pour la restauration d'image monoculaire en présence de brouillard	118
7.1.2	Restauration et reconstruction simultanées en présence de brouillard à partir de paires stéréoscopiques	118
7.1.3	Estimation du coefficient d'extinction du brouillard à partir de paires stéréoscopiques	118
7.1.4	Implantation générique pour l'optimisation de champ de Markov	118
7.2	Perspectives	119
7.2.1	Base de tests d'images réelles	119
7.2.2	Accélération graphique	119
7.2.3	Restauration d'image monoculaire	120
7.2.4	Restauration du contraste et reconstruction simultanées à partir de paires stéréoscopiques en présence de brouillard	120
7.2.5	Prendre en compte d'autres conditions dégradées	121

Annexe A

Publications	123
---------------------	------------

Annexe B

Filtre bilatéral guidé	125
-------------------------------	------------

B.1	Introduction	125
B.2	Filtre bilatéral	126
B.2.1	Lien avec l'estimation robuste	126
B.2.2	Filtre bilatéral robuste	127
B.2.3	Estimation linéaire	128
B.3	Le filtre bilatéral guidé	129
B.4	Expérimentations	131
B.4.1	Sans bruit sur le guide	131
B.4.2	Estimation des paramètres à partir des statistiques de l'image	133
B.4.3	Bruit sur le guide	134

B.5 Applications	135
B.5.1 Affinement de la carte de profondeur	135
B.5.2 Restauration d'image en présence de brouillard	136
B.6 Conclusion	137

Bibliographie	139
----------------------	------------

Table des figures

1	Protocole pour la reconstruction 3D en environnement urbain	4
2	<i>Building rome in a day</i>	5
3	Efficacité des différents indices de profondeur en fonction de la distance pour la perception humaine.	7
4	Caractéristiques des scènes en environnement routier	9
	(a) Zones homogènes	9
	(b) Zones planaires non fronto-parallèles	9
	(c) Grande plage de profondeur	9
	(d) Pluie	9
	(e) Neige	9
	(f) Brouillard	9
5	Système d'amélioration de la vision en présence de brouillard	10
1.1	Reconstruction par triangulation	16
1.2	Relation épipolaire	17
1.3	Modèle stéréoscopique rectifié	18
1.4	Paire stéréoscopique et carte de disparité	19
1.5	Résultat avec la prise en compte de l'hypothèse spatiale sur la base de données <i>Middlebury Stereo Evaluation</i>	19
1.6	Principe de l'appariement local par fenêtre multi-échelles	20
1.7	Principe des fenêtres adaptatives	23
1.8	Exemple de structure de champs de Markov classiques avec un voisinage avec des cliques d'ordre 2 et d'ordre 3	28
	(a) Clique d'ordre 2	28
	(b) Clique d'ordre 3	28
1.9	Représentation des occultations et discontinuités par des variables aléatoires . . .	31
1.10	Exemple de la méthode α -expansion sur scène routière	35
1.11	Exemple de construction du graphe pour l'algorithme de fusion	36
	(a) Scène	36
	(b) Graphe	36
1.12	Exemple de la méthode de fusion	39
2.1	Transformation gamma de l'intensité d'une image en présence de brouillard . . .	41
2.2	Effet du coefficient d'extinction d'une scène	44
2.3	Variables liées par l'équation de Koschmieder	45
2.4	Comparaison de différents algorithmes sur des images caméras	47
2.5	Étape de l'algorithme <i>free-space segmentation</i>	48

2.6	Étape de l'algorithme <i>no-black-pixel</i>	49
2.7	Résultat pour différents coefficients de restauration pour la méthode <i>no-black-pixel</i>	51
2.8	Base de données d'images de synthèse en environnement routier	54
2.9	Restauration sur des images de synthèse des algorithmes présentés	56
3.1	Équivalence entre un chemin augmentant dans un graphe et une somme alternée dans une équation pseudo-booléenne	68
	(a) Graphe initial	68
	(b) $\epsilon = 4$, graphe saturé	68
	(c) Graphe résiduel	68
3.2	Le graphe induit 3.2a et le graphe résiduel de l'équation pseudo-booléenne de l'exemple 3.5.	69
	(a) Graphe	69
	(b) Résiduel	69
3.3	Exemple de la minimisation d'une fonction pseudo-booléenne sous-modulaire par coupe de graphe	71
	(a) Graphe	71
	(b) Coupes possibles : $w(c_1) = 5$, $w(c_2) = 2$, $w(c_3) = 12$, $w(c_4) = 7$	71
4.1	Effet du terme décroissant exponentiel ajouté dans le terme de régularisation sur la restauration de la méthode proposée	81
4.2	Effet de l'initialisation de I_0	84
4.3	Comparaison de différentes restaurations	85
4.4	Exemple de la restauration de l'algorithme proposé sur des images de synthèse	86
4.5	Exemple de la restauration de l'algorithme proposé sur des images caméras	88
4.6	Comparaison sur des images caméras de la méthode proposée avec celles de l'état de l'art	89
5.1	Reconstruction à partir de paires stéréoscopiques en présence de brouillard avec différents algorithmes de reconstruction 3D	92
5.2	Exemple de la méthode de reconstruction 3D et de restauration simultanée au cours des itérations	102
5.3	Effet des paramètres sur les résultats de la méthode de reconstruction et restauration	103
5.4	Exemple : Technique de fusion	104
5.5	Résultats sur des images de synthèse	105
5.6	Résultats sur des images caméras	106
6.1	Variation de l'énergie en fonction du coefficient d'extinction	108
	(a) σ_p variables	108
	(b) β_λ variable	108
	(c) β_λ et σ_p variables	108
6.2	Illustration du principe de la méthode d'estimation du coefficient d'extinction β par minimisation d'entropie	110
	(a) Image capteur	110
	(b) Segmentation	110
	(c) Sélection	110
	(d) $I_0 : \beta = 0.0024$	110
	(e) $I_0 : \beta = 0.0086$	110
	(f) $I_0 : \beta = 0.0167$	110

(g)	Histogramme des niveaux de gris de l'image restaurée avec $\beta = 0.0024$	110
(h)	Histogramme des niveaux de gris de l'image restaurée avec $\beta = 0.0086$	110
(i)	Histogramme des niveaux de gris de l'image restaurée avec $\beta = 0.0167$	110
(j)	Entropie en fonction de β pour les trois régions sélectionnées	110
(k)	Somme de l'entropie et l'entropie de la totalité de l'image	110
6.3	Exemple pour différentes valeurs de β de la méthode d'estimation du coefficient d'extinction par minimisation d'entropie	112
(a)	$\beta = 0.0001$	112
(b)	$\beta = 0.0046$	112
(c)	$\beta = 0.0086$	112
(d)	$\beta = 0.0244$	112
(e)	Somme de l'entropie des régions sélectionnées.	112
6.4	Comparaison entre la visibilité estimée et la vérité terrain pendant 1200 minutes pour 3 jeux de paramètres	115
B.1	Estimateur robuste	126
B.2	Comparaison des différents filtres	128
B.3	Exemple : Flash / sans flash	129
B.4	Comparaison filtre bilatéral guidé/croisé	131
B.5	Comparaison de la convergence entre le filtre bilatéral guidé / croisé	132
B.6	Variation du PSNR pour différents types de bruit	132
B.7	Estimation des paramètres	134
B.8	Résultats sur l'image caméra	135
B.9	Cas avec bruit corrélé sur l'image et le guide	136
B.10	Comparatif du PSNR	137
B.11	Application : affinage de la carte de disparité	137
B.12	Application : restauration d'image en condition de brouillard	138

Introduction

Depuis une décennie, l'utilisation des caméras numériques embarquées est devenue de plus en plus courante. Grâce aux progrès constants en terme de matériel, il est possible de nos jours d'assembler une caméra embarquée à moindre coût. L'éventail d'offres actuellement disponibles sur le marché grand public témoigne de l'importance de cette évolution. Tous les *smart-phones* sont équipés d'une caméra de plusieurs millions de pixels. Les caméras polyvalentes comme les Go-pro[®] (environ 300 euros), grâce à leur résistance, leur capteur de plusieurs millions de pixels et à leur grand angle, sont très prisées pour filmer des scènes de sports. L'architecture ARM de l'ordinateur Raspberry Pi est adaptée pour les applications mobiles. Pour un prix de moins de 25 euros on dispose d'un capteur d'une définition de 2592x1944 pixels pour les photos.

Avant de s'imposer sur le marché grand public, les caméras ont été utilisées dans le milieu professionnel pour différentes tâches : la télé surveillance, l'astronomie, la microscopie, la robotique embarquée, etc.

Les caméras sont utilisées dans des domaines très divers, par le grand public pour un usage occasionnel, comme par les professionnels pour des besoins très spécifiques. Un des points communs à toutes ces tâches, est qu'il faut au préalable, traiter et transformer les données acquises par le capteur photo (à savoir l'intensité lumineuse réceptionnée sur chaque photosite du capteur) en une information adéquate afin de l'exploiter au mieux pour une tâche requise. L'utilisateur d'un appareil photo numérique pour une démarche artistique souhaitera que sa photo soit réussie conformément à ses critères de goût. Un chercheur en astronomie voudra obtenir une image la moins perturbée possible afin d'en extraire le plus d'informations. Une fois l'image acquise, il est possible d'extraire des informations de la scène observée, comme la position des visages dans une image, ou la position des panneaux de signalisation d'un environnement routier. Le domaine traitement d'image regroupe donc l'ensemble de ces étapes, depuis la transformation des données brutes du capteur d'image jusqu'à l'extraction de la sémantique de la scène.

Une partie importante des travaux réalisés en traitement d'image consiste à modéliser le monde observé par la caméra. Dans l'absolu, on aimerait pouvoir, à partir d'une image, extraire le maximum d'informations de la scène (le modèle 3D, les sources lumineuses, la sémantique, etc.). Imaginons que cela soit possible pour n'importe quelle image, alors l'éventail des applications serait immense. Mais cette tâche, qui est assez intuitive pour l'Homme, n'est pas simple en informatique.

Grâce à ses connaissances, son apprentissage, l'efficacité de ses sens et sa faculté à en tirer le maximum d'informations, l'Homme reconnaîtra des formes qui lui sont familières indépendamment de la distance ou des configurations lumineuses dans la limite du raisonnable. Cette tâche qui pourtant semble simple, ne l'est pas pour un algorithme ou les traitements informatisés de cas complexes qui sont actuellement bien moins performants que les mécanismes humains.

Néanmoins, grâce à l'augmentation permanente de la puissance de calcul et aux nouveaux modèles, de nombreux algorithmes proposent des résultats exploitables. Dans certains cas, il n'est pas nécessaire d'atteindre un taux de réussite à 100%. Par exemple, la mise en défaut

occasionnelle d'un algorithme peut être tolérée lors du réglage automatique des paramètres d'un appareil photo à partir de la détection d'un visage, dans ce cas l'objectif est que le sujet soit bien exposé et net. Par contre, un taux de réussite proche de 100% est indispensable dans certains cas, par exemple, pour une application de détection d'obstacles qui affecte les commandes d'un véhicule.

1 Traitement d'image et environnement routier

Grâce à l'avancée des nouvelles technologies, le traitement d'image embarqué est de plus en plus courant dans le domaine routier. L'utilisation de caméras et les applications qui en découlent s'avèrent être extrêmement utiles dans de nombreux cas.

1.1 Modéliser l'environnement

La connaissance de la structure (la carte 3D par exemple) et la sémantique de la scène (présence d'espaces navigables, d'obstacles, de positions de panneaux) sont devenues incontournables, pour planifier automatiquement un itinéraire, permettre le déplacement automatique d'un véhicule ou réaliser la visite virtuelle d'une scène.

Les caméras sont des outils puissants pour modéliser l'environnement dans lequel un véhicule évolue. Comme les lasers, il est possible, avec notamment la reconstruction 3D, d'estimer à partir de plusieurs caméras le modèle 3D local d'une scène [Fau93]. De plus, par rapport au laser, on dispose de la texture des objets grâce à l'acquisition de l'intensité lumineuse et la décomposition en 3 canaux rouge, vert et bleu.

Ces informations sont très utiles par exemple pour la détection et l'identification des différents éléments de l'environnement routier. Il est très intéressant de pouvoir segmenter une image pour différencier la chaussée, les trottoirs, les bâtiments [ZWY10, YAC⁺10], les panneaux de signalisation [FCK09] et les marquages au sol [TIC07]. L'intensité lumineuse, pour ces tâches, apporte un plus non négligeable.

Un nouvel aspect, traité plus récemment est la détection des différentes conditions météorologiques [HTLA06, BHT11]. Quand ces conditions sont caractérisées, cela permet d'adapter les traitements aux conditions dégradées. Il est possible de détecter les conditions météorologiques, la présence d'une éventuelle perturbation, ou d'aller jusqu'à l'estimation d'un modèle paramétrique permettant de caractériser la perturbation rencontrée.

1.2 Se déplacer dans un environnement

La modélisation 3D de l'environnement est un point clé pour la navigation autonome d'un véhicule. La structure de l'environnement étant estimée, il est possible de faire évoluer le véhicule dans cet environnement en planifiant préalablement les trajectoires. Néanmoins, dans certains cas, la structure de l'environnement est imprécise ou occupée par d'autres entités dont il est difficile d'anticiper le comportement tels que des piétons. Dans ce cas, des algorithmes spécifiques pour la détection d'obstacles et le calcul dynamique de trajectoires sont nécessaires afin d'avoir un modèle plus complet. L'article [GLU12] récapitule l'ensemble des informations nécessaire pour la navigation autonome d'un véhicule.

1.3 Aide à la conduite

Afin d'assister le conducteur dans la tâche de conduite, et du fait de la présence croissante de l'électronique dans les véhicules motorisés, les aides à la conduite se sont rapidement démocratisées pour assister le conducteur. Citons par exemple l'assistance au freinage (ABS), la direction assistée ou les régulateurs de vitesse. Avec les caméras embarquées de nouvelles possibilités ont vu le jour.

Par exemple, il est possible de corriger la trajectoire en cas de sortie de route en agissant sur le véhicule ou en alertant le conducteur. Un autre exemple est la détection de la présence d'un objet inhabituel sur la chaussée [LAT02].

En plaçant une ou plusieurs caméras au devant du véhicule, il est possible, en présence de brouillard, d'afficher sur le tableau de bord une image restaurée de la scène observée ou augmentée avec des informations complémentaires [THC⁺12].

Nous avons ici énuméré quelques applications possibles basées sur le traitement d'image. Ces applications posent des problèmes en traitement d'image dont certains sont étudiés depuis longtemps comme la reconstruction 3D. D'autres applications posent au traitement d'image de nouvelles questions telle que la restauration d'images en présence de brouillard pour les aides à la conduite.

1.4 Domaines de traitement d'image sous-jacents aux applications

Comme souvent, une application du traitement d'image peut faire appel à plusieurs problématiques, et donc inversement, une même problématique peut intervenir dans de nombreuses applications.

Prenons l'exemple d'un véhicule autonome se déplaçant sur un terrain inconnu. Un travail combiné de reconstruction 3D, de segmentation et de classification devra être réalisé [GLU12].

Comme indiqué précédemment, on souhaiterait pouvoir caractériser entièrement la scène observée à partir d'images. Quand toutes ces informations sont connues, il est alors facile de répondre à beaucoup de questions. Mais cela n'est pas forcément l'approche la plus efficace. Il vaut mieux, dans certains cas, chercher à résoudre chaque question pour une application.

2 Vision stéréoscopique pour l'environnement routier

Nous pouvons remarquer que certaines techniques, comme la reconstruction 3D à partir de paires stéréoscopiques, sont utilisées dans un grand nombre d'applications routières. L'application de la reconstruction 3D la plus importante dans le domaine routier consiste à détecter les obstacles, par exemple, les piétons [BBBDR05]. C'est aussi le principe le plus utilisé pour l'odométrie [How08], la classification de scènes urbaines [ZWY10, YAC⁺10]. Il en découle qu'un algorithme robuste et efficace de reconstruction 3D est primordial pour assurer le bon fonctionnement des applications utilisant la reconstruction par paires stéréoscopiques.

Dans le cadre de l'environnement routier et des applications embarquées, cela se caractérise par l'installation de deux caméras devant le véhicule (d'autres configurations sont bien sur possibles, celle ci étant la plus courante). Avec un tel dispositif, il sera donc possible d'estimer la carte de profondeur à l'avant du véhicule. Cela nécessite de connaître le calibrage extrinsèque des caméras, c'est à dire la position 3D des caméras l'une par rapport à l'autre.

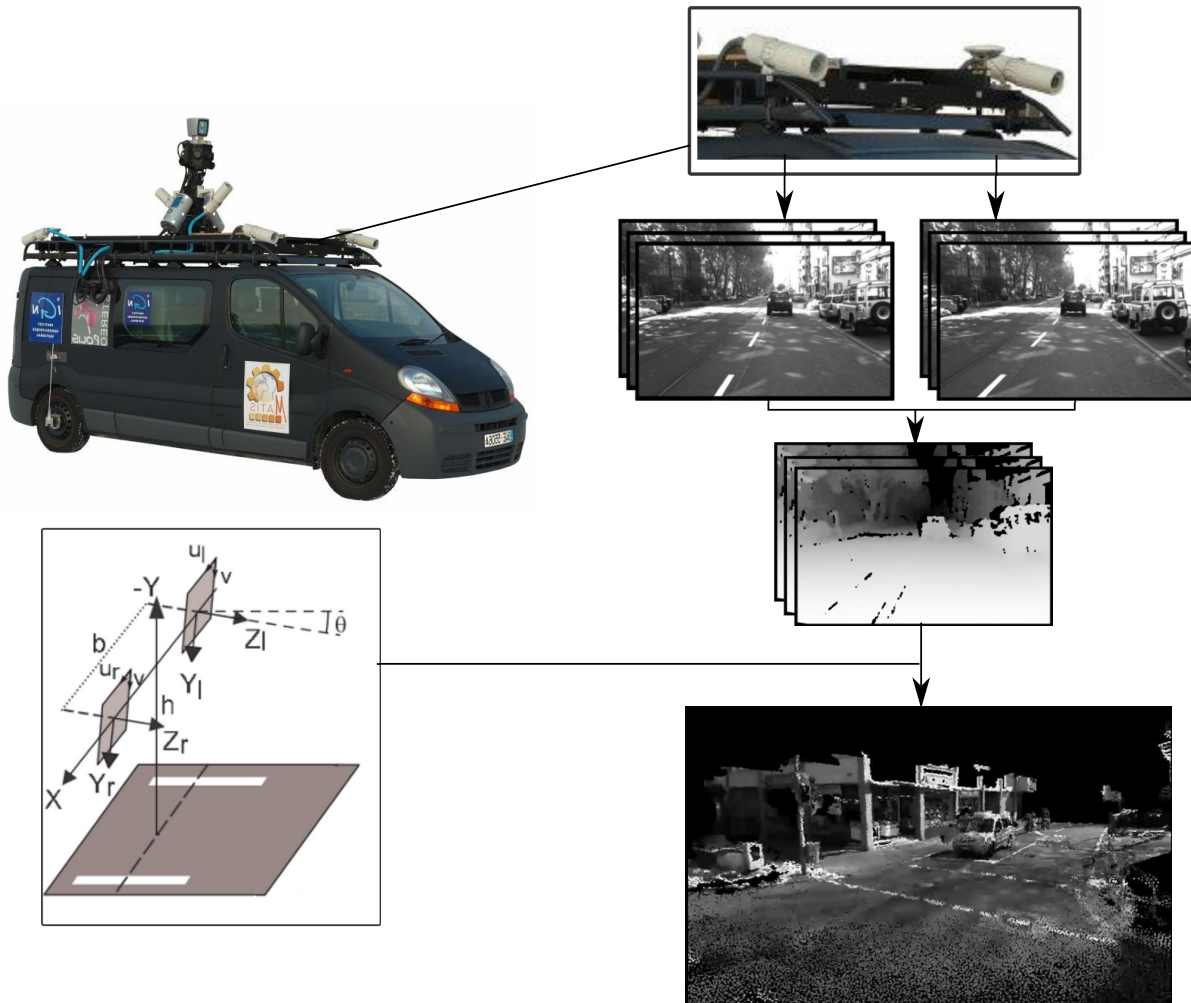


FIGURE 1 – Protocole pour la reconstruction 3D en environnement urbain. Un véhicule¹ est équipé d'une paire stéréoscopique. Une succession de paires stéréoscopiques sont prises. Pour chaque paire la carte de disparité est calculée. Si l'on connaît, pour chaque prise, les coordonnées du véhicule dans la scène, alors il est possible d'estimer à partir de l'ensemble des cartes de disparités le modèle 3D de la scène.



FIGURE 2 – Exemple de reconstruction 3D de l'article *Building rome in a day*, à gauche un ensemble de photos de touristes, à droite le résultat de l'algorithme de reconstruction 3D, la position des caméras (tétraèdre dans l'image) et le modèle 3D de la scène sont retrouvés.

3 Reconstruction 3D

La triangulation est le principe le plus utilisé pour estimer la structure 3D d'une scène à partir de plusieurs vues. Il existe d'autres principes mais cet indice est l'un de ceux qui apporte le plus d'information de profondeur. Le principe est le suivant : si la position et l'orientation relative des caméras sont connues, alors on sait comment projeter un point 3D sur les différentes vues, il sera alors possible de retrouver la position 3D de cette entité par triangulation si l'appariement entre les vues de la projection de l'objet est réalisé.

On arrive actuellement, grâce à plusieurs images de la même scène, à effectuer l'auto-calibration (c'est à dire à estimer automatiquement les paramètres des caméras) et, en simultanée, la reconstruction 3D de la scène observée. L'article *Building rome in a day* [AFS⁺11] présente une méthode complète permettant d'effectuer l'auto-calibration et la reconstruction 3D sur un très grand nombre d'images. La figure 2 montre comme exemple la reconstruction 3D du Colisée de Rome à partir de plusieurs centaines de photos de touristes.

Pour effectuer la reconstruction 3D, plusieurs représentations des données sont possibles. Afin d'effectuer la reconstruction 3D et l'auto-calibration en simultanée, une représentation éparse de points dans l'espace est le plus souvent utilisée pour des questions de rapidité. Dans un premier temps les appariements entre points sont faits entre les images. Ensuite, l'estimation de leur position 3D et des paramètres des caméras peut être effectuée.

Dès lors que le système est calibré, une recherche *dense* d'information peut être faite. Nous appelons recherche dense le fait d'estimer, pour chaque pixel d'une caméra, une profondeur. Ce type de recherche est la plupart du temps utilisée avec un système calibré, et plus particulièrement en reconstruction à partir de paires stéréoscopiques.

La reconstruction 3D à partir de paires stéréoscopiques est un cas particulier de la reconstruction 3D multi-vues. Dans le cas rectifié où seulement deux vues sont considérées, les deux caméras ne diffèrent que par une translation horizontale l'une par rapport à l'autre alignées sur l'axe horizontal des images. Cette configuration présente des avantages : d'une part, d'être facile à calibrer au préalable (contrairement au cas multi-vues), et surtout, les deux projections d'un point sont à la même hauteur sur les deux images.

1. Véhicule de l'équipe MATIS de l'institut géographique national

3.1 Reconstruction 3D à partir de paires stéréoscopiques

La reconstruction 3D à partir de paires stéréoscopiques est un sujet longuement étudié. Il y a actuellement plus d'une centaine d'algorithmes répertoriés et comparés sur la base de tests *Middlebury Stereo Evaluation*² [SS02].

Malgré cette abondance, de nombreux algorithmes sont continuellement proposés. La raison d'un tel engouement autour de la reconstruction 3D s'explique par l'importance en terme applicatif mais aussi au caractère mal posé du problème. Un problème est dit bien posé au sens de Hadamard, si la condition d'unicité de la solution recherchée est garantie et que celle-ci dépend de façon continue des données dans le cadre d'une topologie. Or, pour la reconstruction 3D, il peut exister des ambiguïtés qui permettent plusieurs solutions. En effet, pour un pixel donné, il peut y avoir, dans l'autre image, plusieurs pixels dont l'intensité est proche de celle d'origine. Il n'y a donc pas unicité de la solution. Les principaux obstacles sont les zones homogènes et les structures répétitives.

Il est donc nécessaire de trouver la meilleure façon pour lever ces ambiguïtés. La proposition la plus utilisée est de supposer que la scène est en général spatialement continue et donc que cette hypothèse découle du fait qu'il est plus probable que deux pixels proches dans une image soient à une profondeur proche plutôt qu'à des profondeurs très différentes. Cela conduit à des méthodes dites globales pour minimisation d'une erreur ou d'une énergie.

Au début des recherches, une autre hypothèse plus restrictive a été utilisée. L'idée consiste à apparier non pas pixel à pixel, mais d'une fenêtre autour d'un pixel avec une autre fenêtre. Cela suppose que le monde est constitué de plans parallèles aux capteurs des caméras.

Il est probable qu'il n'existe pas un *a priori* qui permette de lever correctement les ambiguïtés dans toutes les situations, cela explique en partie la multiplicité des méthodes proposées. Une autre explication est la grande variabilité des types de scènes.

En effet, certaines scènes sont plus difficiles à traiter que d'autres. Par exemple, une scène sans discontinuité, avec de nombreuses textures bien éclairées sera plus facile à traiter qu'une scène avec de grandes zones homogènes de pleines ombres. Si les conditions sont satisfaisantes, les algorithmes classiques produisent des résultats acceptables. Dès que les conditions se dégradent, la qualité des reconstructions 3D est fortement diminuée.

3.1.1 S'inspirer de notre perception 3D

En comparaison des méthodes de vision par ordinateur, notre perception dans la 3D se révèle moins précise mais plus robuste. Si l'Homme est si performant, c'est qu'il tire profit de toutes ses facultés sensorielles. Il utilise une multitude d'indices permettant d'inférer la 3D. Dans [Nag91], une énumération des différents indices de profondeur avec lesquels il perçoit la 3D a été faite ainsi que leur utilité en fonction de la distance (figure 3).

Il en ressort que chaque indice apporte plus ou moins de précisions selon la distance. Pour les distances proches le plus efficace est la triangulation. Pour les longues distances c'est le voile atmosphérique. Grâce à l'utilisation combinée de tous les indices de la figure 3, nous obtenons une perception assez complète de l'environnement dans lequel nous évoluons.

Ce mécanisme de fusion est extrêmement complexe car il mélange en plus des indices, les connaissances et l'apprentissage. Au niveau algorithmique, nous sommes encore loin de combiner autant d'indices et *a priori* et de plus, avec une telle rapidité.

2. <http://vision.middlebury.edu/stereo/>

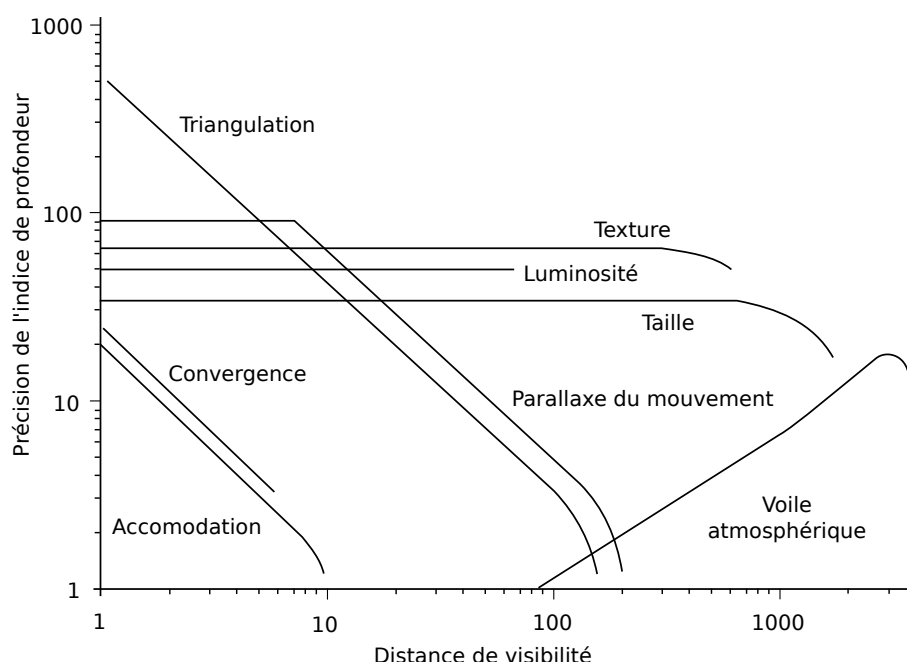


FIGURE 3 – Efficacité des différents indices de profondeur en fonction de la distance pour la perception humaine.

3.1.2 Fusionner les indices

Certains travaux vont dans cette direction, nous pouvons citer par exemple la reconstruction à partir de paires stéréoscopiques combinée avec la segmentation [BRK⁺11] ainsi qu’avec la classification [LSR⁺10]. Le dernier exemple présente un modèle complet permettant de classer les éléments de l’environnement et d’effectuer une reconstruction 3D dense. Ce principe est de plus en plus utilisé comme dans [HB90] où l’estimation du flot optique est combiné avec la segmentation de la scène.

3.1.3 Apport de connaissances

L’autre aspect est de prendre en compte les connaissances sur les particularités de la scène traitée. Par exemple pour l’environnement routier, nous savons que la partie basse de l’image sera occupée principalement par la chaussée, et, la partie haute de l’image aura une forte chance d’être composée de ciel. Ces *a priori* peuvent être utilisés dans les algorithmes. L’*a priori* peut être plus ou moins restrictif, comme supposer que la route suit un modèle polynomial ou un modèle plan. Quand ces *a priori* sont bien utilisés, cela permet de lever des ambiguïtés mais aussi de diminuer l’espace de recherche donc à augmenter la rapidité de traitement.

C’est pour cela qu’il est important de connaître les spécificités des scènes qui nous intéressent : les scènes routières.

3.2 Spécificité des scènes routières

Les algorithmes de reconstruction 3D actuels sont en général peu adaptés aux images routières et aux conditions météo dégradées. Le bon fonctionnement de ces algorithmes en toutes conditions est pourtant crucial dans des applications comme les aides à la conduite, la conduite automatisée, le relevé des caractéristiques de la route.

La base *Middlebury Stereo Evaluation* permet de comparer qualitativement plus d'une centaine d'algorithmes de reconstruction 3D de paires stéréoscopiques. Mais il s'avère que cette base de tests a été réalisée en studio dans des conditions très bien contrôlées. Certes, des scènes avec des zones homogènes ou avec de nombreux sauts de disparités ont été créées. Mais cela reste néanmoins loin des conditions réelles en extérieur où d'autres problèmes surviennent.

C'est à partir de ce constat que la base *KITTI Vision Benchmark Suite* a été créée récemment pour permettre de tester différents algorithmes de reconstruction 3D dans l'environnement routier. [GLU12]. Actuellement, il y a environ 30 algorithmes répertoriés³.

Les scènes routières ont souvent les caractéristiques suivantes :

Zones homogènes : Une grande partie de l'image est constituée par la chaussée qui est en général de couleur grise. Ces zones, posent un problème d'ambiguïté pour la reconstruction 3D (image 3.1a).

Scènes planaires : En plus de la chaussée qui généralement est formée par un plan fortement incliné, l'environnement urbain est constitué de nombreux plans tels que les panneaux de signalisation ou encore les bâtiments (image 3.1b).

Grande plage de profondeur : Une des caractéristiques de la reconstruction 3D des scènes d'extérieur est de pouvoir voir des objets jusqu'à plusieurs kilomètres. Afin de pouvoir prendre en compte un intervalle de profondeur conséquent, la distance entre les caméras doit être grande pour pouvoir faire la triangulation jusqu'à une distance minimale. La conservation de large plage de profondeur pose le problème de temps de calcul (image 3.1c).

Conditions dégradées : S'ajoutant aux caractéristiques géométriques de la scène observée, les conditions dégradées ont des effets non négligeables pouvant altérer de façon imprévisible et significative la qualité des résultats. Les principales conditions dégradées sont :

Pluie : La pluie perturbe de plusieurs façons les algorithmes. Les gouttes entraînent la présence d'eau de pluie sur le pare-brise, mais surtout les surfaces mouillées sont plus sujettes aux spécularités et reflets. Si la réflexion est trop importante, un algorithme de reconstruction interprétera la surface spéculaire comme du vide et reconstruira alors la scène réfléchie. Ce qui dans le cas de la chaussée, entraînera l'apparition de trous dans la carte de profondeur (image 4d).

Neige : La neige a deux conséquences principales, les zones homogènes en cas de neige abondante seront beaucoup plus importantes pour l'ensemble des surfaces recouvertes par la neige. Les flocons de tailles perceptibles feront partie intégrante de la scène et seront donc, si l'algorithme est efficace, reconstruits. S'ils ne sont pas détectés en tant que flocons puis supprimés, cela peut entraîner des défauts importants (image 4e).

Brouillard : L'effet principal du brouillard est la diminution du contraste avec la distance. Les zones de faible distance par rapport à la caméra auront donc un contraste proche d'une scène sans brouillard. Inversement, les objets lointains auront un contraste faible. L'ambiguïté augmente donc fortement avec la distance (image 4f).

3. <http://www.cvlibs.net/datasets/kitti/index.php>

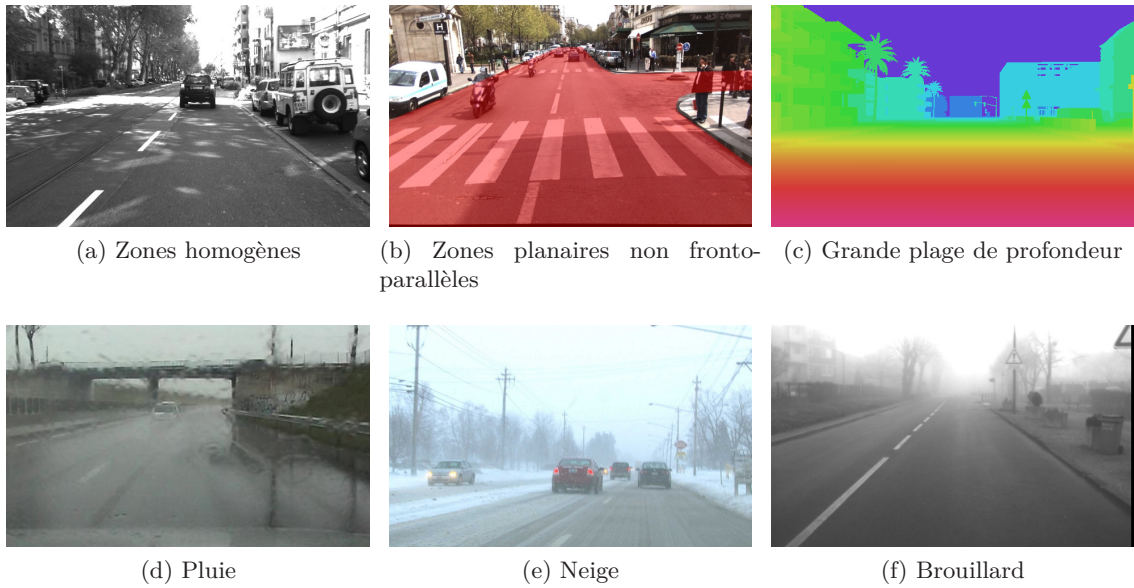


FIGURE 4 – Caractéristiques des scènes en environnement routier

En résumé, nous pouvons constater que l'environnement routier est caractérisé par une géométrie fortement planaire (contrairement à un environnement naturel) mais aussi par la possibilité de conditions dégradées. Plusieurs algorithmes dédiés aux scènes routières ont pris en compte l'aspect géométrique. Etant donné l'importance des conditions dégradées nous nous intéressons maintenant à ce qui concerne ce sujet.

4 Traitement d'image en conditions dégradées

La caractérisation des conditions dégradées peut être faite à plusieurs niveaux, le plus « simple » étant de détecter une perturbation (Est-ce qu'il pleut ? Est-ce qu'il y a du brouillard ?). Une application directe est par exemple l'activation automatique des essuie-glaces ou des feux antibrouillard.

Mais il est intéressant de caractériser plus finement ces conditions dégradées. Comme par exemple la connaissance de la densité du brouillard rencontrée ou la force de la pluie. Dans ces situations, il sera alors par exemple possible de régler la fréquence de balayage des essuie-glaces en fonction de l'importance de la pluie ou encore la puissance des feux antibrouillard en fonction de la densité du brouillard.

Après cette étape de caractérisation des conditions, il sera alors possible d'en tenir compte pleinement dans différents traitements.

4.1 Brouillard

Le brouillard, depuis quelques années, est un sujet d'intérêt en traitement d'image pour des applications embarquées car il est l'une des causes importantes d'accidents. Le brouillard entraîne une augmentation du temps de réaction du conducteur face à un obstacle. Cet effet est principalement dû à la diminution du contraste avec la distance. En plus de la perturbation de la perception humaine, le brouillard entraîne une dégradation de la qualité des résultats obtenus par les algorithmes classiques de traitement d'image embarqués actuels, ce qui les rend



FIGURE 5 – Exemple de système d’amélioration de la vision en présence de brouillard. Une caméra avec l’image de la route captée et restaurée est placée sur le tableau de bord.

inopérants. C’est dans ce contexte que des algorithmes spécifiques ont été développés, d’une part, pour améliorer la visibilité du conducteur [THC⁺12], et d’autre part, pour accroître la fiabilité des algorithmes existants en présence de brouillard.

Afin d’améliorer la visibilité du conducteur, une possibilité est d’afficher une image avec des contrastes restaurés de la scène observée, c’est à dire sans brouillard (image 5). Ce problème n’est pas simple à traiter. En effet, lorsque la profondeur est constante, cela ne pose pas trop de difficultés, car l’intensité de la restauration ne dépend que de la densité du brouillard. Mais quand on est confronté à de grandes différences de profondeurs dans la scène observée, alors l’intensité de la restauration doit dépendre de la densité du brouillard, mais aussi de la profondeur. Il y a donc une ambiguïté entre la profondeur et l’intensité originale de la scène en un pixel donné. Par contre, si l’on connaît la profondeur de la scène et la densité du brouillard, alors il sera possible de retrouver l’intensité originale de la scène. Lorsque seule l’image en présence de brouillard est connue, nous sommes donc en présence d’un problème mal posé.

Malgré cela, des algorithmes ont été proposés pour la restauration du contraste d’une image prise dans le brouillard. La plupart de ces algorithmes sont fondés sur le modèle de Koschmieder qui permet de faire le lien entre l’image en présence de brouillard, l’image restaurée, la profondeur et le coefficient d’extinction du brouillard qui caractérise sa densité. Afin d’inférer l’image restaurée, différentes hypothèses peuvent être faites sur la structure 3D de la scène pour permettre de restaurer au mieux l’intensité. Dans [THC⁺12] de nombreuses méthodes sont comparées pour la restauration d’image en présence de brouillard pour les scènes routières, il s’avère que les méthodes basées sur l’équation de Koschmieder apportent de meilleurs résultats. Certains algorithmes se basent sur le voile atmosphérique. Ce dernier dépend du coefficient d’extinction et de la profondeur. Si le voile est connu, il est alors possible d’effectuer la restauration de l’image sans connaître la profondeur ou le coefficient d’extinction.

La diminution du contraste avec la distance entraîne aussi des dysfonctionnements pour les algorithmes de traitement d’image. Prenons l’exemple de la reconstruction à partir de paires stéréoscopiques, dans les zones proches, l’intensité observée en présence de brouillard est proche de celle des objets d’origine, il n’y a donc pas de très grandes différences avec le cas sans brouillard. Par contre, plus on s’éloigne, et plus l’intensité des objets observés tend vers celle du ciel, la reconstruction par appariement sera donc ambiguë.

Il apparaît judicieux de s'intéresser aux méthodes de reconstruction 3D combinant les indices de la stéréo et du voile atmosphérique.

5 Modèle bayésien

Le cadre bayésien apporte une approche intuitive pour la fusion d'informations. C'est pourquoi, le livre *Computer vision : models, learning and inference* [Pri12] propose de réinterpréter une grande partie des problématiques du traitement d'image en se fondant sur un ensemble restreint de notions liées aux probabilités.

Dans la préface, écrite par *Andrew Fitzgibbon*, cet ouvrage « met en évidence la distinction la plus importante en vision par ordinateur : la différence entre le *modèle* et l'*algorithme* ». Cette distinction est réalisée grâce au langage des statistiques. Ce point est illustré par l'approche d'un problème connu en traitement d'image : la séparation entre les « figures » et l'arrière plan. Deux approches se distinguent : l'une basée sur une succession de méthodes paramétrées par de nombreux « paramètres magiques » complexes à estimer et ne fonctionnant que sur un sous ensemble d'images, et l'autre, l'approche probabiliste, qui permet une distinction nette entre l'objectif et l'algorithme. On formalise d'abord le problème en une énergie, qui, doit être minimisée par la suite.

Le principe des méthodes probabilistes consiste donc à interpréter les variables d'entrée (dans le cas de la reconstruction à partir de paires stéréoscopiques l'image gauche et droite) et les variables à estimer en tant que variables aléatoires, par la suite créer une densité de probabilité modélisant les relations entre elles, et finalement, trouver quelles sont les variables cherchées qui maximisent cette densité de probabilité sachant les variables d'entrée.

Le modèle étant créé, on aura alors un éventail de méthodes d'optimisations permettant de trouver la solution qui minimise le modèle. Bien sur, il n'est pas possible d'optimiser globalement tous les modèles. Il faudra donc bien noter les modèles dont il est facile d'obtenir le minimum global de l'énergie sous-jacente, ou du moins, avec lesquels on peut trouver un minimum acceptable.

5.1 Modèle bayésien

Le formalisme probabiliste permet une interprétation « intuitive » du problème. De nombreux phénomènes physiques découlent de processus aléatoires (celui que nous rencontrons le plus fréquemment est l'acquisition de l'intensité lumineuse par les photosites). Un phénomène peut donc être modélisé par une densité de probabilité. En complément du formalisme « direct » du problème, pour la reconstruction 3D, il est nécessaire de lever l'ambiguïté en ajoutant une connaissance sur la distribution locale des profondeurs car l'appariement pixel à pixel ne suffit pas à lui seul. Pour cela, et grâce aux méthodes bayésiennes, il est possible de créer un estimateur *a posteriori* en ajoutant un *a priori* sur la distribution des profondeurs. Ceci permet de lever les ambiguïtés sur des solutions et de favoriser ainsi celles dont les profondeurs sont proches.

En traitement d'image, une des structures de modèle bayésien la plus rencontrée est celle des champs de Markov. C'est une structure spéciale de modèle bayésien, où chaque variable aléatoire est simplement dépendante (au sens probabiliste du terme) de son voisinage.

5.2 Optimisation discrète

Un modèle étant défini, il faut maximiser une certaine vraisemblance ou minimiser une certaine énergie si l'on passe à la log-vraisemblance.

Les méthodes d'optimisation adaptées sont proposées pour optimiser les énergies très spécifiques découlant des modèles globaux avec un nombre important de variables (ce qui empêche une recherche exhaustive lorsqu'elles sont corrélées). Ces méthodes ont d'abord été créées de façon empirique en traitement d'image telles que les méthodes basées sur les coupes de graphes comme l'algorithme α -expansion [KZ02] ou la programmation dynamique [SSZ02].

Certaines de ces méthodes sont connues dans le domaine de l'optimisation de fonctions discrètes. L'une d'elles, la théorie d'optimisation de fonctions pseudo booléennes, apporte d'une part, un cadre théorique très appréciable pour justifier de façon formelle ces algorithmes, et d'une autre part, un ensemble de nouveaux algorithmes tel que le calcul de la « roof duality » qui permet de minimiser de façon efficace certaines fonctions qui, jusque là, présentaient des difficultés.

Cette approche est très intéressante du fait de la disponibilité de nouveaux algorithmes efficaces. Ce qui est le cas avec l'optimisation de fonctions pseudo booléennes quadratiques. Une grande partie des fonctions relatives aux modèles probabilistes peut être optimisées de cette façon. Avec les processeurs graphiques facilitant les traitements parallèles, de nouveaux algorithmes ont été développés pour résoudre certains problèmes d'optimisation.

Le formalisme probabiliste fournit donc un cadre complet pour la modélisation et l'optimisation de problèmes.

6 Positionnement

Le brouillard au loin diminue le contraste, il entraîne une grande ambiguïté et par conséquent, réduit considérablement la qualité des algorithmes de reconstruction 3D. Étonnamment, très peu d'algorithmes sont spécifiques aux conditions dégradées et plus particulièrement au brouillard. Pourtant, cette problématique est intéressante car le brouillard apporte un indice supplémentaire de profondeur au loin en complément de la stéréovision comme le montre la figure 3.

D'autre part, le traitement d'image en présence de brouillard, sous l'impulsion du développement des aides à la conduite, s'est fortement développé surtout pour la restauration du contraste original de la scène. Ce problème est mal posé à partir d'une seule image, car il y a une ambiguïté entre la profondeur et l'intensité originale de la scène.

Notre idée pour traiter ces deux problèmes est de fusionner les deux indices de profondeur et de lever l'ambiguïté au loin grâce à l'image restaurée, elle même pouvant être calculée grâce à la carte de profondeur.

Pour arriver à cela, l'approche bayésienne semble être la plus appropriée car elle apporte un formalisme intéressant pour modéliser le problème à traiter.

Nous proposons donc de modéliser simultanément le problème de la reconstruction par paires stéréoscopiques et de la restauration d'image en présence de brouillard. Pour cela, nous proposons de créer un modèle probabiliste fusionnant la reconstruction par paires stéréoscopiques et la restauration d'images. Grâce à la stéréovision, un indice de profondeur est disponible pour effectuer la restauration d'images. L'image restaurée permettra donc de lever l'ambiguïté au loin et de favoriser les profondeurs dont l'image restaurée est la plus probable.

7 Plan

Dans un premier temps, un bref état de l'art des méthodes de reconstruction 3D par paires stéréoscopique est présenté dans le chapitre 1. Ensuite, les méthodes fondées sur les modèles probabilistes sont approfondies.

Afin de bien comprendre les problématiques actuelles en traitement d'image en présence de brouillard, dans le chapitre 2, la restauration monoculaire d'image en présence de brouillard est abordée. Le modèle physique permettant de modéliser le brouillard est introduit. Par la suite, différentes méthodes de restaurations fondées sur ce modèle sont détaillées.

Lorsqu'on définit un modèle, les méthodes d'optimisation permettant de l'optimiser sont aussi importantes que le modèle lui-même. Par conséquent, le chapitre 3 introduit l'optimisation de fonctions multi-labels et plus particulièrement l'optimisation de fonctions pseudo-booléennes pour les modèles découlant de la formulation probabiliste.

Le chapitre 4 présente une nouvelle méthode de restauration basée sur le modèle de Koschmieder et l'estimation du voile atmosphérique. Grâce à une formulation énergétique, le voile atmosphérique est dans un premiers temps calculé pour obtenir un indice de profondeur, et par la suite, l'image est restaurée.

Par la suite, le chapitre 5 propose un modèle complet pour la restauration et la reconstruction en présence de brouillard. Pour cela, le modèle précédent est associé à la formulation probabiliste de la reconstruction 3D par paires stéréoscopiques. Ce modèle est optimisé par un algorithme alternant l'optimisation de la carte de profondeur et celle de l'image restaurée.

Malgré les avancées des méthodes de restauration, paradoxalement, la caractérisation du brouillard est encore difficilement réalisable. Dans 6, une nouvelle méthode est introduite, permettant de calculer le coefficient d'extinction à l'aide d'un système stéréoscopique.

Nous avons d'abord envisagé une approche basée sur des descripteurs locaux inspirés du filtre bilatéral croisé du fait des résultats intéressants obtenus dans l'état de l'art sur la reconstruction 3D, mais nous n'avons pas donné suite à cette approche. Nous avons néanmoins exploré une voie pour réinterpréter le filtre bilatéral croisé pour le rendre robuste au bruit non gaussien, ces travaux sont décrits dans l'annexe B.

Chapitre 1

Reconstruction 3D à partir de paires stéréoscopiques

Ce chapitre introduit le problème de la reconstruction 3D à partir de paires stéréoscopiques et les différentes hypothèses faites pour résoudre ce problème. Puis, les méthodes probabilistes sont approfondies.

L'objet de la reconstruction 3D à partir de caméras est d'obtenir la représentation en trois dimensions d'un objet ou d'une scène à partir d'un ensemble d'images prises sous différents points de vue. Si les paramètres propres aux caméras et à leurs positions dans le monde sont connus, et si l'on connaît la projection d'un même objet de la scène dans les différentes images des caméras, alors il est possible, après mise en correspondance entre les vues, d'estimer la position 3D de chaque point de l'objet.

Pour cela, on doit préalablement modéliser les caméras et la relation entre elles. Il existe plusieurs modèles permettant de prendre en compte de nombreux paramètres pour modéliser le plus fidèlement possible une caméra. Le modèle de caméra le plus simple est le modèle sténopé ou « pinhole ». Ce modèle représente la caméra par un plan focal et un centre optique séparé par une distance focale. Les paramètres internes de la caméra sont appelés paramètres intrinsèques. La caméra sera caractérisée dans le repère du monde par un point 3D et une orientation : ce sont les paramètres extrinsèques.

Avec ce modèle, les seuls rayons lumineux atteignant le plan focal seront ceux passant par le centre optique. Ce modèle, ne considère ni les distorsions, ni la profondeur de champ liée à l'ouverture de l'objectif. Il est grandement utilisé, car du fait de sa simplicité, il permet de poser les problèmes plus simplement. Par exemple il permet de poser le problème du calibrage des caméras qui consiste à retrouver les paramètres intrinsèques et extrinsèques. Ce problème qui est très étudié n'est pas ici notre centre d'intérêt.

En effet, nous supposons que les paramètres des caméras sont connus. Il est alors possible, à partir des projections d'un même point dans deux caméras de retrouver sa position 3D (figure 1.1) par triangulation.

La triangulation nécessite des points en correspondance entre les deux images. L'autre avantage important d'un système calibré est la possibilité de réduire l'espace de recherche du correspondant d'un point de la première image dans la deuxième image. L'espace de l'image complète

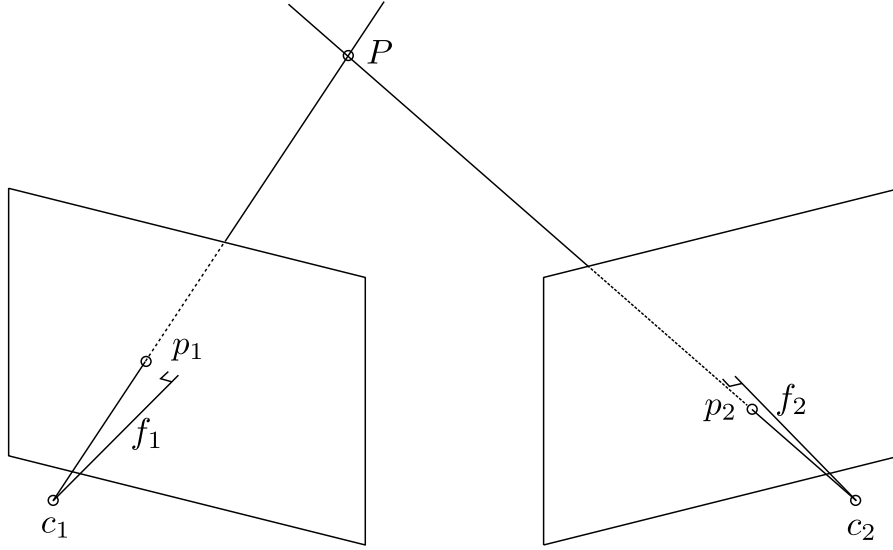


FIGURE 1.1 – Reconstruction 3D par triangulation. P est un point 3D, p_1 sa projection dans une première image de centre optique c_1 et de focale f_1 et p_2 sa projection dans une seconde image de centre optique c_2 et de focale f_2 . Si p_1 et p_2 sont connus, alors il est possible de calculer la position 3D de P en cherchant le point 3D qui minimise la distance entre la droite c_1p_1 et c_2p_2 .

est réduit à une droite. La figure 1.2 met cette propriété en évidence. Soient deux caméras dont c_1 et c_2 sont les centres optiques et un point 3D P . Soit p_1 la projection de P dans la première image, nous cherchons sa projection p_2 dans la seconde caméra. Alors, par construction, p_2 se trouve sur la droite passant par l'épipole e_2 (les épipoles sont la projection du centre optique d'une caméra sur l'image de l'autre caméra) appartenant au plan formé par p_1 et les deux centres optiques. Cette relation est connue sous le nom de géométrie épipolaire.

Lorsque les droites épipolaires sont horizontales et de la même hauteur dans les deux images, on dit que le système est rectifié. L'avantage est évident, la recherche d'appariements ne se fera plus que sur une ligne de même hauteur que celle de p_2 .

Système rectifié : Dans un système rectifié, la différence de position horizontale entre la projection d'un point dans la première image et la seconde image s'appelle la disparité. Grâce à la disparité, il est possible de retrouver la profondeur 3D dans le repère caméra quand le système est calibré. La figure 1.3 représente une image d'un système stéréoscopique rectifié. Les traits gradués représentent les deux plans focaux de la paire stéréoscopique, où chaque graduation est une colonne de l'image avec c_1 et c_2 leurs centres optiques. f est la focale. La base stéréoscopique b représente l'écart entre les deux centres optiques. P est un point 3D de la scène, la projection de centre c_1 sur le plan focal de la première caméra est p_1 , la projection de centre c_2 sur le plan focal de la seconde caméra est p_2 . La différence absolue de l'indice de la colonne du point p_1 et celle du point p_2 est la disparité. Si le ratio pixel/mètre (ici noté λ) est connu, alors il est possible de retrouver la profondeur de la scène p à partir de la disparité avec la relation $p = \frac{\lambda b f}{d}$. Nous pouvons donc constater que la disparité et la profondeur sont inversement proportionnelles. Dans le cas stéréoscopique, δ désignera les paramètres du système des deux caméras tel que $\delta = \lambda b f$.

Dans la partie 1.1 sont introduits les *a priori* nécessaires à la résolution du problème de la reconstruction à partir de paires stéréoscopique ainsi que les différents algorithmes les utilisant. Ensuite, leur interprétation à partir du formalisme probabiliste est approfondie dans la

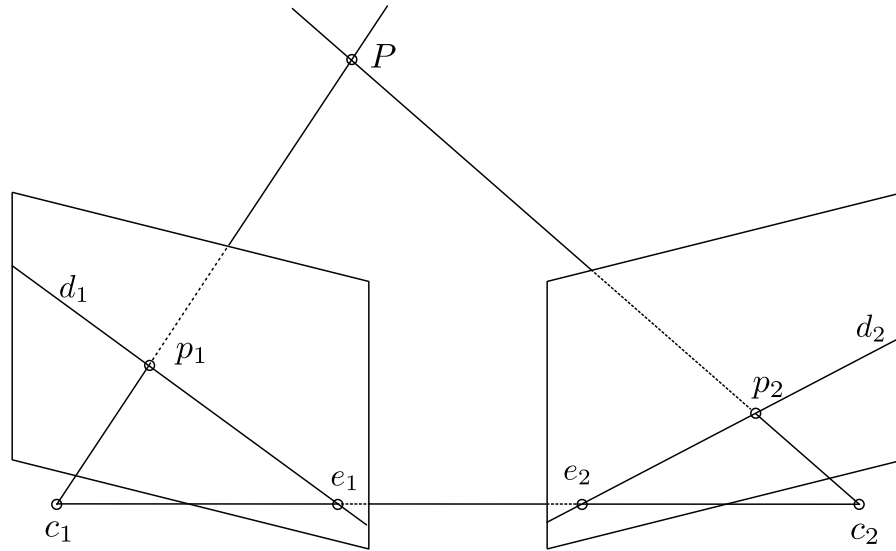


FIGURE 1.2 – Relation épipolaire entre deux caméras. La projection de p_2 de P dans la seconde caméra se trouve toujours sur la droite épipolaire passant par l'épipole c_2 et appartenant au plan formé par les deux centres optiques et la projection du point P projection dans la première caméra.

section 1.2.

1.1 Estimation de la carte de disparité

La reconstruction dense à partir de paires stéréoscopiques consiste à trouver l'ensemble des disparités en chaque pixel d'une image référence. À l'issue de ce processus, le résultat que l'on obtient est une carte de disparité. La figure 1.4 montre un exemple de carte de disparité (dans ce cas la vérité terrain) pour une paire stéréoscopique.

La reconstruction par paires stéréoscopiques dense est un cas particulier qui a été longuement étudié essentiellement du fait de sa plus simple formulation. Pourtant, malgré sa simplicité apparente, il y a des situations où le problème est mal posé. En effet, nous verrons par la suite qu'il existe souvent, pour une paire stéréoscopique, un grand nombre de solutions pour lesquelles l'ensemble des disparités sont probables. Il y a donc des ambiguïtés.

Dans l'article [SS02] (qui introduit la base de tests *Middlebury Stereo Evaluation*) en 2002, un état de l'art des méthodes de reconstructions à partir de paires stéréoscopiques a été proposé. Les algorithmes sont structurés selon quatre étapes :

- Le type d'appariement.
- La méthode d'appariement.
- L'optimisation.
- L'affinement de la solution.

Cette approche est intéressante, car elle permet de décomposer le problème en quatre étapes distinctes, et ainsi, de pouvoir inter-changer et évaluer chaque partie, ce qui est très utile pour faire une évaluation. Néanmoins, cette approche a le défaut de cacher les différents *a priori* faits sur la scène observée. Nous proposons donc, dans ce bref état de l'art, de mentionner les différentes hypothèses souvent faites qui permettent de lever les ambiguïtés et, pour chaque hypothèse, des algorithmes qui en découlent. Ce type de présentation a été inspiré par celle de

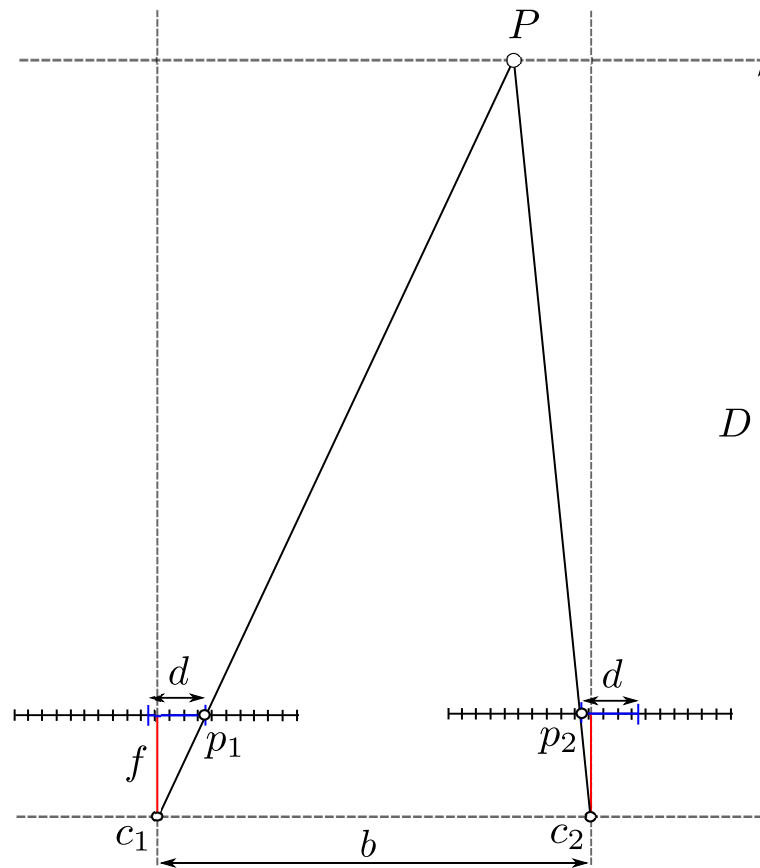


FIGURE 1.3 – Modèle stéréoscopique rectifié vu du dessus. Les deux traits gradués représentent les deux plans focaux de la paire stéréoscopique. c_1 et c_2 sont leurs centres optiques. f est la focale. b la base, P est un point 3D de la scène, p_1 sa projection sur le plan focal de la première caméra, p_2 sa projection sur le plan focal de la deuxième caméra. d est la disparité.



FIGURE 1.4 – L'image gauche et droite d'une paire stéréoscopique et la vraie carte de disparité de la scène.

Michael Bleyer, dans son cours sur la reconstruction à partir de paires stéréoscopiques.



FIGURE 1.5 – Exemple de résultats avec la prise en compte de l'hypothèse spatiale. De gauche à droite : l'hypothèse photométrique seule, avec l'hypothèse spatiale locale [SS02] plus filtrage médian, semi-globale [Hir05], globale avec une régularisation du premier ordre [KZ01] et du second ordre [WTRF09].

1.1.1 Hypothèse photométrique

Afin de pouvoir faire l'appariement entre les deux vues, il a d'abord été supposé que chaque point est vu avec la même couleur et l'intensité lumineuse quelle que soit la direction d'observation. Il est aussi supposé que seul le bruit engendré lors de l'acquisition perturbe les images. La troisième supposition, permettant de simplifier les calculs, est de faire l'hypothèse que l'image gauche est sans bruit, et que seule l'intensité observée dans l'image droite est perturbée par le bruit. Dès lors, la probabilité d'avoir un appariement entre un pixel de l'image droite avec un pixel de l'image gauche est la vraisemblance selon la loi du bruit avec comme moyenne l'intensité du pixel gauche et comme réalisation l'intensité du pixel droit. En terme énergétique, pour une loi de probabilité symétrique décroissante cet appariement pixel à pixel, ne suffit pas, même quand l'hypothèse photométrique est respectée à cause des zones d'intensité uniforme. La première image de la figure 1.5 montre le résultat de l'appariement pixel à pixel. Nous pouvons voir que le résultat est loin de la vérité terrain. Il a donc fallu introduire d'autres hypothèses. C'est en grande partie ce problème d'ambiguïté qui motive les recherches en reconstruction à partir de paires stéréoscopiques. La différence entre les algorithmes est la façon de lever cette ambiguïté. Pour cela, des hypothèses vont être faites sur la solution recherchée pour favoriser une certaine classe de solutions plutôt qu'une autre.



FIGURE 1.6 – Principe des algorithmes fondés sur l'appariement local par fenêtre multi-échelles. Trois gabarits de tailles de voisinage sont utilisés comme descripteur. La plus petite est sujette à une forte ambiguïté étant donné qu'elle ne couvre qu'une zone homogène. La grande fenêtre recouvre un autre objet : l'appariement ne sera donc pas optimal. Une taille de fenêtre intermédiaire est le meilleur choix. Une zone assez conséquente est couverte qui permet de lever les ambiguïtés, et la totalité de cette zone correspond à une même disparité

1.1.2 Hypothèse spatiale

Les algorithmes de reconstructions 3D font l'hypothèse que la profondeur en un pixel, est proche de celles de ses voisins. Cette hypothèse peut être réalisée de différentes façons qui peuvent être classées en 3 grandes catégories : locale, globale et coopérative. La figure 1.5 montre l'impacte de la prise en compte de l'hypothèse spatiale.

1.1.2.1 Les méthodes locales

Ces méthodes proposent de caractériser localement le voisinage du pixel par la valeur d'un descripteur. C'est à dire qu'un pixel ne sera pas représenté uniquement de façon isolée par sa valeur, mais aussi par son voisinage. Les descripteurs étant créés, chaque pixel est apparié avec celui dont le descripteur est le plus semblable dans la seconde image. Nous constatons qu'une nouvelle hypothèse est faite : le monde est supposé continu. Afin de lever l'ambiguïté dans les zones homogènes, ou d'éviter de dépasser au niveau des discontinuités, les descripteurs peuvent être de tailles et de formes variables [KO94] (figure 1.6). Néanmoins, ces méthodes sont limitées du fait de l'étendue importante des zones homogènes telles que la chaussée dans notre application. En effet, pour ces zones, même une grande taille de fenêtre sera insuffisante, dans certains cas, pour lever correctement l'ambiguïté. C'est à la suite de ce constat que les méthodes globales ont été introduites.

1.1.2.2 Les méthodes globales

Les méthodes globales modélisent le problème de la reconstruction à partir de paires stéréoscopiques de façon énergétique. Une énergie est choisie. Le modèle choisi associe une énergie à chaque carte de profondeur de façon à ce que l'énergie soit élevée si la carte de profondeur est éloignée de la vérité terrain. Inversement, si l'énergie est faible, la carte doit être proche de la vérité. Le processus se fait donc en deux étapes. Le calcul de l'énergie en fonction des données, et ensuite, le calcul de la carte de profondeur qui minimise cette énergie.

Le modèle est généralement composé de deux termes. Le terme d'attache aux données qui symbolise l'appariement entre un pixel de l'image de droite et un sur l'image de gauche. Et enfin, un terme de régularisation, qui permet de lever les ambiguïtés en biaisant le résultat vers un certain type de solution. En reconstruction 3D, le terme de régularisation favorisera généralement les solutions dont les disparités sont proches dans un voisinage local [Ter86].

1.1.2.3 Les méthodes coopératives

Les méthodes coopératives (ou aussi appelées semi-globales) ont été introduites pour pallier le fait que les méthodes locales sont limitées par la présence de trop grandes zones homogènes. Ces méthodes combinent les méthodes locales et globales. Les pixels sont d'abord caractérisés par un descripteur et appariés suivant le principe des méthodes locales, et ensuite, l'étape pour minimisation d'énergie est utilisée pour affiner ou compléter le résultat [Hir08, WZ08].

1.1.2.4 Ordre de la régularisation spatiale

L'ordre de la régularisation spatiale est donné par l'ordre de la dérivée de la profondeur ou de la disparité que l'*a priori* régularise. Pour le premier ordre, toutes les différences de disparités dans un voisinage local seront pénalisées. Les cartes de profondeur avec des plans fronto-parallèles seront donc favorisées. Dans le cas du second ordre, alors les plans, quelles que soient leurs directions, seront favorisés. En reconstruction dense, la grande majorité des algorithmes utilise une régularisation soit du premier soit du second ordre, car, ils apportent un bon compromis entre l'ensemble des scènes représentables et la facilité d'optimisation.

Premier ordre : Pour les méthodes locales, l'appariement est fait entre descripteurs modélisant le voisinage. Un des premiers critères utilisés a été la somme des carrés pixel à pixel sur le voisinage. Ce critère suppose donc que la fenêtre recherchée dans l'image gauche n'ait subi qu'une perturbation due au bruit. Il faut donc que le voisinage n'ait été transformé que par une translation. En général pour un objet plan, la transformation est une homographie, ne préservant ni les angles, ni les distances. Seuls les plans fronto-parallèles auront la propriété attendue.

Pour les méthodes globales, la régularisation du premier ordre a été utilisée en premier. Pour cela, on ajoute au terme d'attache aux données, un terme de régularisation pénalisant les différences entre voisins.

Second ordre : Afin de pouvoir prendre l'ensemble des orientations possibles, l'hypothèse du second ordre est meilleure.

Dans le cadre des méthodes locales, une des premières idées est d'estimer localement un plan. De cette façon, chaque pixel pourra être projeté dans la seconde image sachant la distance en supposant un plan. Cela permet par conséquent de prendre en compte les déformations dues à l'homographie.

Dans le cas des méthodes globales, la régularisation du second ordre remplace celle du premier ordre [WTRF09].

Pour les méthodes semi-globales, une des façons d'affiner la solution, consiste, pour l'ensemble des plans voisins, à minimiser la différence de pente entre les plans [YMU13a].

Ce n'est certainement pas la seule solution. De nombreuses autres méthodes ont été proposées permettant la prise en compte du second ordre. Certaines, comme Libelas [GRU10], sont fondées sur une triangulation préliminaire de points d'intérêt. La triangulation est utilisée comme *a priori* lors de l'appariement pour la minimisation d'une énergie.

1.1.3 Prise en compte des variations photométriques et bruit non gaussien

Faire l'hypothèse de bruit gaussien est trop strict, des perturbations altèrent les images. Pour que l'hypothèse photométrique soit valide, il est nécessaire que les objets de la scène aient une surface lambertienne, c'est à dire que la luminance réfléchiée par l'objet soit indépendante de la direction d'observation. Or, il y a beaucoup d'objets spéculaires comme la carrosserie des véhicules ou une surface mouillée qui induiront de nombreuses réflexions.

Une seconde source de perturbation est liée au processus d'acquisition et de pré-traitement de l'image. Cela dépend du calibrage et du mode d'utilisation des caméras, car il est difficile d'obtenir des conditions de réglage identiques. Cela peut entraîner des différences d'intensités. Le bruit de quantification, la transformation Gamma entre les deux caméras souvent réalisée et l'*aliasing* perturbent aussi l'image acquise par le capteur. Cela explique que le bruit est rarement gaussien. Il a donc fallu introduire des estimateurs robustes ou des métriques, lors de l'appariement.

L'article [HS09] fait un état de l'art des différentes métriques réalisées en fonction de leurs propriétés invariantes à des changements d'exposition. Il en ressort que la performance de ces métriques dépend de l'algorithme utilisé. Pour les méthodes locales, la métrique ZNCC (*zero-mean normalized cross-correlation*) apporte de très bons résultats. Dans le cadre des méthodes utilisant l'appariement pixel à pixel, les techniques fondées sur une approche hiérarchique [HS09] donnent de meilleurs résultats, mais elles sont coûteuses en temps de calcul.

Généralement, que ce soit pour les méthodes locales ou globales, les estimateurs robustes permettent d'intégrer dans le bruit certaines perturbations. En pratique, un opérateur tel que la valeur absolue, ou une fonction semi-quadratique, est choisi plutôt que l'erreur quadratique.

1.1.4 Prise en compte des discontinuités

D'autres aspects de la reconstruction 3D sont complexes à prendre en compte. Ce sont les occultations dues aux sauts de disparités. En effet, dans certaines scènes, un point projeté dans la caméra de gauche ne sera pas visible dans l'image droite du fait d'une occultation. Un algorithme ne prenant pas en compte les occultations produira un résultat erroné dans ces zones. Pour permettre de traiter le problème des occultations de nombreux algorithmes ont été proposés.

Méthodes locales : Une des méthodes les plus efficaces proposées est celle des fenêtres adaptatives [HBGR09]. Cette méthode, reprend le principe du filtre bilatéral, et propose de pondérer les coefficients des pixels de la fenêtre en fonction de la différence d'intensité avec celui du centre. En d'autres termes, plus un pixel a une intensité proche de celui du centre, plus il aura de l'importance lors de l'appariement (figure 1.7). Ces descripteurs apportent de bons résultats mais ils sont très coûteux en temps de calcul.

Méthodes globales : L'utilisation d'un estimateur robuste permet de contourner le problème des occultations et d'obtenir un résultat acceptable. Les zones occultées seront normalement caractérisées comme des données aberrantes [CBFAB97, Ter86, SSZ02].

Il existe d'autres approches destinées à gérer les discontinuités et occultations. On constate un usage de plus en plus fréquent de la segmentation en régions pour permettre de mieux prendre en compte les discontinuités.



FIGURE 1.7 – Principe de l'algorithme fondé sur les fenêtres adaptatives. La zone verte dans l'image de droite correspond au pixel ayant une pondération forte dans le calcul du coût. Cette même zone se trouve dans la seconde image pour deux pixels candidats. Nous pouvons observer que l'ensemble des pixels ayant une pondération forte pour le pixel recherché possède une couleur proche.

1.1.5 Hypothèse du monde segmenté

Les forts gradients dans les images sont dus, soit, aux textures d'un objet, ou souvent, aux discontinuités. C'est à partir de cette observation que des méthodes, fondées sur la segmentation d'images ont été proposées. Dans de nombreuses méthodes, la segmentation sert de guide pour que les sauts de discontinuités soient corrélés avec les limites des zones segmentées [WZ08, SSZ02].

1.1.6 Évaluation comparative

Grâce à la base de tests *Middlebury Stereo Evaluation*, et plus récemment à la base de tests *KITTI Vision Benchmark Suite*, une évaluation quantitative de beaucoup d'algorithmes de reconstruction stéréo a été faite. Il en ressort que les méthodes globales ou coopératives dépassent les méthodes locales du fait de la capacité à propager l'information dans les grandes zones homogènes. Cette limite des méthodes locales les rend inapplicables pour la reconstruction 3D de l'environnement routier. En effet, si nous considérons la base de tests *KITTI Vision Benchmark Suite* (courant 2013), les 3 meilleurs algorithmes fondés uniquement sur l'indice stéréoscopique utilisent une méthode globale, et plus particulièrement un formalisme fondé sur les champs de Markov. L'article [YMU13b] obtenant les meilleurs résultats propose une méthode fondée sur l'estimation d'un flot optique le long de la ligne épipolaire. La structure de cette méthode est un champ de Markov où chaque pixel est représenté par un plan, paramétré par trois variables aléatoires. Le flux optique permet d'ajouter une contrainte d'ordre sur les plans et de lever l'ambiguïté spatiale. Le second algorithme [YHMu12] propose lui aussi, un champ de Markov composé de plans orientés permettant de prendre en compte les occultations.

Depuis quelques années, dans de nombreuses applications, telle que la reconstruction à partir de paires stéréoscopiques, les méthodes probabilistes en traitement d'images se situent au niveau général de l'état de l'art.

C'est à partir de ce constat que nous avons décidé de nous intéresser aux méthodes probabilistes.

1.2 Formulation probabiliste

De nombreux autres problèmes en traitement d'images peuvent aussi se formaliser comme une recherche *dense* d'informations, autrement dit nous recherchons en chaque pixel une information de la scène. Par exemple, pour la segmentation, nous chercherons la classe de l'objet visualisé en chaque pixel, pour la restauration l'image corrigée. Ce type de problème peut être formulé de façon globale comme nous l'avons vu pour la reconstruction à partir de paires stéréoscopiques. Cela consiste dans un premier temps, à choisir une énergie modélisant le problème où les solutions probables auront un faible coût, et, ensuite, en chercher la solution qui minimise cette énergie. L'énergie est souvent composée de deux termes, le premier est appelé l'*attache aux données*, ce terme est généralement une fonction locale à chaque pixel qui lie les données (les images acquises par le capteur pour le traitement d'image) aux variables recherchées (profondeur, intensité originale de la scène, label, etc.). Le second est le terme de *régularisation*. Celui-ci va donc, dans le cadre de la reconstruction à partir de paires stéréoscopiques, pénaliser deux profondeurs voisines ayant des valeurs trop différentes pour reprendre l'hypothèse du monde continu et favorisera ainsi des valeurs proches.

Il s'avère que cette structure de problème est bien connue dans d'autres disciplines comme la physique. Dans ces disciplines, ils sont traités par un modèle probabiliste nommé *champs de Markov*. En traitement d'images, les champs de Markov ont apporté une interprétation probabiliste aux méthodes dites *globales*.

Impact des champs de Markov en traitement d'images : Depuis leur introduction en 1984 [GG84], les techniques fondées sur les champs de Markov ont connu un intérêt croissant en traitement d'images. En effet, la majeure partie des problèmes courants en images peut être modélisée de manière probabiliste, dont une grande part directement par les champs de Markov (Par exemple : Segmentation [Bar98], restauration d'images [GG84], flot optique [HB93], reconstruction à partir de paires stéréoscopiques, etc.).

Une des grandes forces de cette approche vient du fait que les méthodes d'optimisation d'un champ de Markov sont indépendantes de la nature du problème posé. En effet, l'approche probabiliste consiste à interpréter les données comme des variables aléatoires et de créer ensuite un modèle représentant le problème à traiter. Que ce soit pour un problème de segmentation ou pour un problème de reconstruction 3D, le modèle final sera une densité de probabilités modélisant le problème où chaque réalisation possible aura une certaine probabilité. Considérons l'exemple de la reconstruction à partir de paires stéréoscopiques. Pour une réalisation des deux images d'entrée, quelle est la réalisation de la carte de disparité qui est la plus probable ? Pour cela, après avoir effectué la réduction du problème de base en un modèle probabiliste, on peut appliquer toutes les méthodes existantes d'optimisation de l'énergie déduite du modèle.

Ce choix de séparer le modèle de la méthode d'optimisation fait la force des modèles probabilistes. En effet, les modèles résultants du formalisme probabiliste pour des classes de problèmes (par exemple les méthodes denses) sont souvent très proches, quel que soit le problème rencontré (reconstruction, segmentation, restauration). Il en découle que s'il existe une méthode efficace d'optimisation pour ce type de modèle, alors toutes ces applications pourront être optimisées de façon efficace.

Dans le cadre des méthodes denses, de nouvelles méthodes d'optimisation, fondées sur la théorie des graphes et l'optimisation de fonctions pseudo-bouliennes ont été introduites en image ces dix dernières années et ont rendu plus facilement utilisables les champs de Markov en image. Cela a eu un tel impact que de nouveaux apports en mathématiques appliquées, plus particulièrement dans le domaine de l'optimisation discrète, sont issus des recherches en

traitement d'images. Ainsi, le problème de réduction de tailles des cliques dans les graphes qui permet d'optimiser des modèles complexes a fait l'objet de nombreuses contributions dernièrement [FGBZ11, Ish09a, GBP11, Ish11]. Cet usage intensif est en partie dû à la structure intéressante des graphes générés par les problèmes en traitement d'images et la capacité des algorithmes à les résoudre.

Dans cette partie, nous allons nous intéresser aux modèles probabilistes dans le cadre du problème de la reconstruction dense (1.2.1) et aux méthodes d'optimisation pouvant être utilisées (1.2.2).

1.2.1 Interprétation probabiliste

L'interprétation probabiliste consiste à modéliser le problème par une densité de probabilités. Pour cela il faut d'abord, interpréter les données comme des variables aléatoires, cf. partie 1.2.1.1. Lorsque les paramètres sont aussi interprétés comme des variables aléatoires on parle de modélisation bayésienne. Ensuite, il faut définir une densité de probabilité liant les variables en paramètres recherchés et les données connues cf. partie 1.2.1.2. Le principe du *maximum a posteriori* est généralement utilisé pour dériver un estimateur des paramètres sachant les données d'entrée. Grâce au théorème de Bayes, la densité de probabilités *a posteriori* peut être réécrite en fonction du terme de *vraisemblance* et de l'*a priori*, sur les paramètres recherchés.

L'*a priori* est une connaissance préalable sur les paramètres à estimer avant même la connaissance de la réalisation des données d'entrée. L'*a priori* peut être introduit comme une corrélation entre une variable en un site, et celles de son voisinage. C'est ce qui caractérise les champs de Markov. La notion de corrélation est introduite plus généralement grâce aux réseaux Bayésiens qui permettent de modéliser les interactions entre les variables corrélées, cf. partie 1.2.1.3.

La vraisemblance modélise la probabilité d'avoir les données d'entrée en fonction des paramètres variables recherchés en chaque pixel. Pour la reconstruction, c'est la probabilité d'avoir une réalisation de l'image gauche, et de l'image droite sachant l'intensité originale de la scène et la disparité en un pixel, cf. partie 1.2.1.4. L'*a priori* ajoute une connaissance sur les variables à estimer cf. partie 1.2.1.5. Dans le cadre de la reconstruction 3D, nous savons qu'en général, quelle que soit la scène rencontrée, la réalisation d'une profondeur en un pixel est corrélée à celle de son voisinage. La reconstruction à partir de paires stéréoscopiques est un problème mal posé, à cause des zones uniformes et des occultations. Ces aspects peuvent être considérés et pris en compte par les modèles probabilistes, cf. partie 1.2.1.6.

1.2.1.1 Données du problème

Il faut tout d'abord interpréter les données et les paramètres comme des variables aléatoires. Les images d'entrée sont interprétées comme des variables aléatoires. On notera $\{I_L, I_R\}$ les images d'entrée gauche et droite de la paire stéréoscopique. Une réalisation de I_L et I_R correspond à une scène observée à un certain moment. On note S un ensemble fini de sites s , chaque site correspondant à un pixel de l'image où $s = \{i, j\}$ est un vecteur 2D représentant la position du pixel de ligne i , et colonne j . On notera $I_x(s)$ la variable aléatoire correspondant à l'intensité d'une image d'indice x en un site s . On notera $I(\{i, j\}, d) = I(i, j - d)$ la variable aléatoire correspondant à l'intensité d'une image sachant un décalage d'une disparité d . Le problème est composé de la scène à estimer $\{D, I_0\}$, où D est un ensemble de variables aléatoires représentant les disparités recherchées avec $D(s)$ la variable du site s et I_0 un ensemble de variables aléatoires représentant l'intensité originale observée dans la même vue que D avec $I_0(s)$ la variable du site s . Par convention, on prendra la vue de gauche comme référence.

L'intensité originale de la scène est perturbée par plusieurs phénomènes, dont le bruit électronique, les repliements de spectre (en anglais *aliasing*). Lorsque les conditions d'acquisitions sont satisfaisantes (condition météo dégagée, luminosité importante afin de limiter le bruit de la caméra), l'intensité originale de la scène peut être supposée connue et égale à l'image acquise à la sortie du capteur. Généralement, et afin de simplifier la modélisation, cette simplification est utilisée sur l'image de référence. Dans un premier temps, le problème sera énoncé dans le cas le plus général.

1.2.1.2 Maximum *a posteriori*

Le modèle probabiliste consiste à donner la loi de probabilité qui lie les variables connues à celles inconnues à estimer. Le principe du maximum *a posteriori* nous indique comment en déduire une estimation des variables inconnues. On cherchera donc quelle est la réalisation de l'ensemble des disparités et des intensités originales qui maximise la densité de probabilité du modèle, sachant une réalisation de l'image gauche et droite. Cette probabilité *a posteriori* est notée $P(D, I_0 | I_L, I_R)$.

En appliquant le théorème de Bayes, la distribution *a posteriori* peut être écrite de la façon suivante :

$$P(D, I_0 | I_L, I_R) = \frac{P(I_L, I_R | D, I_0) P(D, I_0)}{P(I_L, I_R)} \quad (1.1)$$

$P(I_L, I_R | D, I_0)$ est la vraisemblance des données sachant les paramètres. $P(D, I_0)$ est l'*a priori* sur les paramètres et $P(I_L, I_R)$ est l'*a priori* sur les données d'entrée.

L'*a priori* sur les variables à estimer permet d'ajouter des connaissances. Si nous considérons la distribution des disparités dans un voisinage restreint, on voit qu'il y a une forte probabilité d'avoir des valeurs proches, ce qui correspond à l'hypothèse spatiale. L'*a priori* va donc permettre de modéliser cela. En effet, comme nous l'avons vu, si le modèle n'était composé que de la vraisemblance, et donc que chaque site était indépendant de son voisin, la solution sera alors la disparité qui maximise le terme de vraisemblance. Ce genre de solution n'est souvent pas bon. En ajoutant un *a priori* modélisant la distribution de la profondeur dans le voisinage par paires, une meilleure solution sera obtenue.

Simplifications : Plusieurs approximations peuvent être faites pour simplifier l'énergie à optimiser. Premièrement, l'*a priori* sur les données d'entrée n'étant pas fonction de la solution, il peut être ignoré. En conséquence, la densité de probabilité $P(D, I_0 | I_R, I_L)$ sera maximisée à un facteur d'échelle près. Deuxièmement, on suppose que l'image gauche représente l'intensité originale de la scène en posant $I_L = I_0$ et que seule l'image droite est perturbée par du bruit. En prenant en compte ces deux simplifications, nous pouvons réécrire la probabilité *a posteriori* de la façon suivante :

$$P(D | I_L, I_R) \propto P(I_R | D, I_L) P(D | I_L) \quad (1.2)$$

Dans certains travaux, cette dernière simplification n'est pas faite. En particulier, cette approche est très utile pour la reconstruction multi-vues afin de générer un modèle 3D texturé dense [PKG99]. Dans [GS05], un modèle probabiliste complet pour la stéréo multi-vues proposant d'estimer l'intensité sans bruit et la profondeur en simultanée est proposée. Une descente de gradient sur l'énergie relative au modèle est effectuée en alternant l'optimisation de l'intensité de l'image sans bruit et de la profondeur.

1.2.1.3 Réseaux bayésiens

Les réseaux bayésiens permettent de modéliser la notion de dépendance entre deux variables. Cette notion est importante, car une approche naïve ne prenant en compte que la probabilité conjointe de l'ensemble des variables produira un modèle de grandes dimensions difficile à optimiser. En effet, le nombre de paramètres pour caractériser la probabilité conjointe portant sur l'ensemble des variables est très important. En supposant que la dépendance spatiale ne provient que des voisins immédiats, il sera plus facile d'estimer les paramètres.

Afin de modéliser l'interaction entre deux variables, on introduit la notion de voisinage. Nous notons N le voisinage d'un site s en un pixel :

$$N_s = \{t\} \quad \text{tels que} \quad \begin{cases} \cdot & s \notin N_s \\ \cdot & t \in N_s \Rightarrow s \in N_t \end{cases} \quad (1.3)$$

Le voisinage de s est donc l'ensemble des sites possédant une dépendance directe avec le site s . Si un site est vu comme un sommet, et une interaction entre deux sites comme une arête, alors le voisinage N_s est un sous graphe. Un graphe peut être décomposé en cliques. Une clique est soit un singleton, soit un ensemble de sites inter-connectés (et dans le cadre probabiliste, un ensemble de variables toutes inter connectées). On note C l'ensemble des cliques du graphe et C_k l'ensemble des cliques de cardinal k .

Nous pouvons, pour chaque clique, introduire la notion de potentiel. Pour une clique c , le potentiel associé est noté U_c . Cette valeur dépend de la fonction liant les variables de la clique. L'énergie globale U du graphe peut être définie comme la somme des potentiels de toutes les cliques :

$$U = \sum_{c \in C} U_c \quad (1.4)$$

Nous pouvons aussi définir l'énergie locale U_s en un site qui est la somme des potentiels de toutes les cliques voisines où s intervient :

$$U_s = \sum_{c \in C/s \in c} U_c \quad (1.5)$$

Champ de Markov : La structure de voisinage permet donc de modéliser l'interaction entre plusieurs variables. Dans le cadre probabiliste, cela se traduit par la probabilité d'avoir une réalisation d'une variable en un site ne dépend que de la configuration de son voisinage local. C'est ce qui caractérise un *Champ de Markov*. Cette propriété est importante pour l'échantillonnage d'un champ de Markov, car il est possible de calculer la probabilité conditionnelle locale d'une variable (ce qui est impossible quand toutes les variables sont dépendantes du fait de la constante de normalisation devenant rapidement incalculable).

En traitement d'images, les champs de Markov permettent de modéliser l'interaction d'un ensemble de variables autour de chaque pixel. L'exemple 1.1 montre différentes structures de champs de Markov classiques.

Exemple 1.1. La figure 1.8 montre deux configurations de graphes avec différentes structures de voisinages sur une image composée de 12 pixels. Le graphe de la figure 1.8a représente un champ de Markov *4 connexités* (chaque site est connecté à ses voisins directs sur l'axe horizontal et vertical). Un champ *8 connexités* inclura aussi les diagonales. Les cliques sont d'ordre 2. Nous pouvons constater que pour le site $s = \{2, 2\}$, le potentiel local U de l'équation (1.5) correspond

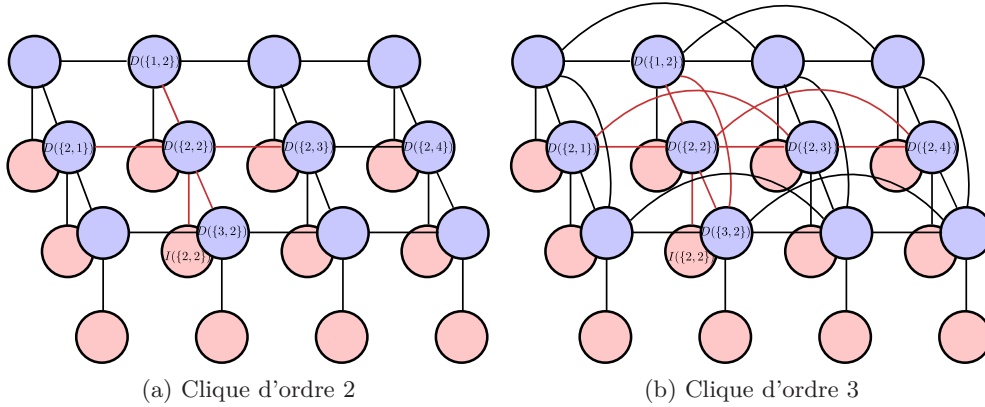


FIGURE 1.8 – Exemple de différentes structures de champs de Markov pour une image composée de 12 pixels. Chaque cercle est une variable aléatoire. Pour chaque pixel, il existe deux variables aléatoires. Les cercles rouges correspondent aux variables observées I et les bleus aux variables à estimer D . La figure 1.8a représente un champ de Markov avec un voisinage représenté par des cliques d'ordre 2, la figure 1.8b avec un voisinage composé de cliques d'ordre 3. Les arêtes qui lient les variables I aux variables D représentant le terme de *vraisemblance*, les arêtes entre les variables D les *a priori*. Les cliques correspondant aux voisinages du site $s = \{2, 2\}$ sont en rouge.

à la somme des potentiels de l'ensemble des cliques où s apparaît. C'est à dire aux couples :

$$\begin{aligned} & \{\{D(\{2, 2\}), I(\{2, 2\})\}, \\ & \{D(\{2, 2\}), D(\{1, 2\})\}, \{D(\{2, 2\}), D(\{2, 1\})\}, \{D(\{2, 2\}), D(\{2, 3\})\}, \{D(\{2, 2\}), D(\{3, 2\})\}\} \end{aligned}$$

Le graphe 1.8b montre une structure composée de cliques d'ordre 3 liant un site avec son voisin direct et un second voisin. L'ensemble des cliques où s apparaît dans ce cas est :

$$\begin{aligned} & \{\{D(\{2, 2\}), I(\{2, 2\})\}, \\ & \{D(\{1, 2\}), D(\{2, 2\}), D(\{3, 2\})\}, \{D(\{2, 1\}), D(\{2, 2\}), D(\{2, 3\})\}, \{D(\{2, 2\}), D(\{2, 3\}), D(\{2, 4\})\}\} \end{aligned}$$

1.2.1.4 Terme de vraisemblance

Le terme de vraisemblance modélise la probabilité des données d'entrée sachant les variables recherchées. En reconstruction 3D, il modélise la probabilité d'une réalisation de I_L et de I_R sachant la réalisation de I_0 et de D . Pour construire $P(I_R, I_L | D, I_0)$, il faut choisir la forme de la distribution sur l'image gauche et droite I_R et I_L avec les paramètres de la distribution comme fonction de D et I_0 . Considérons l'exemple où les images gauches et droites ne seraient perturbées que par un bruit blanc gaussien, alors une loi normale sera choisie pour modéliser la distribution des images acquises avec I_0 comme moyenne.

Dans le cas d'une loi normale, une réalisation aura une forte probabilité si l'intensité des deux images, l'une au site $I_L(s)$, et l'autre au site $I_R(s, D(s))$, sont proches de la moyenne $I_0(s)$. Inversement, si les intensités $I_L(s)$ et $I_R(s)$ sont différentes de celles de $I_0(s)$, la probabilité sera faible.

Dans le cadre général, la vraisemblance des données sachant les paramètres, dans l'hypothèse où chaque site est indépendant, est le produit des probabilités en chaque site. Ceci se traduit par :

$$P(I_R, I_L | D, I_0) = \frac{1}{Z} \exp(-U(I_R, I_L | D, I_0)) \quad (1.6)$$

où $U(I_R, I_L | D, I_0)$ est le potentiel total de l'image et Z la constante de normalisation. Ce terme énergétique est souvent appelé le terme d'attache aux données. Dans le cadre de la reconstruction 3D, il représente donc l'erreur entre l'intensité de l'image gauche et l'intensité originale de la scène et l'erreur entre l'intensité de l'image droite et l'intensité originale de la scène dans l'image sachant une disparité :

$$U(I_R, I_L | D, I_0) = \sum_{s \in S} \rho_p\left(\frac{|I_L(s) - I_0(s)|}{\sigma_p}\right) + \rho_p\left(\frac{|I_R(s, D(s)) - I_0(s)|}{\sigma_p}\right) \quad (1.7)$$

où ρ_p est une fonction liée à la distribution du bruit photométrique avec comme échelle σ_p . Cette formulation générale permet de modéliser des distributions autres que la distribution normale.

Si l'on suppose que I_L représente en fait l'intensité originale de la scène, alors le terme d'attache aux données est seulement l'erreur entre l'intensité d'un pixel de l'image gauche avec celui de l'image droite, sachant une disparité. C'est sous cette forme que le terme d'attache aux données est le plus souvent rencontré :

$$U(I_R | D, I_L) = \sum_{s \in S} \rho_s\left(\frac{|I_R(s, D(s)) - I_L(s)|}{\sigma_s}\right) \quad (1.8)$$

Ce terme d'attache aux données correspond à l'appariement pixel à pixel. Si nous optimisons l'énergie (1.8) directement, plusieurs disparités peuvent minimiser l'énergie en un pixel. Si de surcroît on ajoute le bruit, il y a une forte chance que la disparité minimisant le terme d'attache aux données ne soit pas la bonne.

1.2.1.5 *a priori*

L'*a priori*, dans la plupart des cas, favorisera deux valeurs proches dans un voisinage. Pour cela, on modélise la distribution des variables locales par un ensemble de cliques.

Comme montré dans [Bes74], cette structure connectée est un modèle markovien, car la probabilité d'une variable ne dépend que des variables voisines. La probabilité globale peut donc être décomposée par la multiplication des probabilités locales ce qui se traduit par :

$$P(D | I_L) = \frac{1}{Z} \exp(-\lambda_D U(D | I_L)) \quad (1.9)$$

où $U(D | I_L)$ est un potentiel dépendant des valeurs de disparités entre voisins et Z la constante de normalisation. λ_D est un coefficient permettant de pondérer l'effet de l'*a priori* par rapport au terme de vraisemblance. L'énergie globale sera la somme des potentiels de toutes les cliques du graphe ne dépendant que des variables D . Ce qui se traduit par :

$$U(D | I_L) = \sum_{c \in C_D} \rho_d(\varphi(c, D)) \quad (1.10)$$

où $\varphi(c, D)$ et ρ_d sont donc des fonctions qui vont, dans le cadre énergétique, pénaliser les configurations non désirées.

Choix de l'a priori : Le terme d'a priori le plus utilisé est la norme du gradient. C'est un terme du premier ordre. L'ensemble des cliques C_D est alors l'ensemble de paires de pixels voisins. Pour la norme L1, ce terme est défini par :

$$\varphi(\{p, q\}, D) = D(p) - D(q) \quad (1.11)$$

et $\rho_d(t) = |t|$. Dans ce cas, $|\varphi(\{p, q\}, D)|$ est égal à 0 quand p et q sont identiques. Par contre, le score est d'autant plus élevé que la différence entre p et q est importante. Ce score favorise donc les plans fronto-parallèles au plan image de la caméra. En effet, toutes scènes possédant des plans non fronto-parallèles auront localement, une pénalité non nulle.

Cela pose problème pour les zones homogènes inclinées qui ont besoin de l'a priori. Ce problème est particulièrement visible dans la reconstruction 3D de l'environnement routier, où la chaussée, est de couleur constante mais sur un plan non fronto-parallèle au plan focal. La solution minimisant l'énergie localement sera donc un plan fronto-parallèle, ce qui ne correspond pas à la solution recherchée.

Pour pallier ce problème, il nous faut une régularisation sur n'importe quel plan dans l'espace. C'est pour cela que le terme de la dérivée à l'ordre deux a été introduit :

$$\varphi(\{p, q, r\}, D) = D(p) - 2D(q) + D(r) \quad (1.12)$$

alors, quelle que soit l'orientation d'un plan dans l'espace, ce score, si les disparités aux sites $\{p, q, r\}$ sont sur le plan sera toujours égal à 0. L'utilisation de l'ordre second, contrairement à l'ordre premier, fait intervenir une clique triple. L'ensemble des cliques C_D sera par exemple l'ensemble des patches 3×1 et 1×3 pixels.

Pondération du terme de régularisation en fonction de la vraisemblance : Le paramètre λ_D donnera dans (1.9) plus ou moins d'importance au terme de régularisation. Quand $\lambda_D = 0$ l'a priori n'a plus d'effet et seul le terme de vraisemblance apparaît dans l'énergie finale. Quand λ_D tend vers l'infini, l'image va devenir de plus en plus lisse, jusqu'à tendre vers la solution homogène car le terme de vraisemblance aura un effet négligeable dans l'énergie. Une solution minimisant l'énergie sera donc un plan fronto-parallèle dans le cas d'un terme de régularisation du premier ordre et un plan d'orientation diverse dans le cas du second ordre.

1.2.1.6 Interprétation des valeurs aberrantes

Un des principaux intérêts de la formulation probabiliste est qu'elle permet d'interpréter les phénomènes comme résultants de processus aléatoires, et en conséquence, donne des indications sur comment choisir les fonctions d'erreurs. Considérons l'exemple du terme d'attache aux données (1.7). Il est possible par exemple d'estimer la distribution du bruit dans une zone homogène afin de choisir la fonction ρ_p . Cette façon de procéder néglige les occultations. Il ne faut donc pas être trop strict dans le choix de ρ_p pour tenir compte des occultations. Afin de pallier les problèmes des occultations, comme il a été mentionné dans l'état de l'art, des alternatives ont été proposées. Dans le cadre probabiliste, une façon de procéder est d'interpréter les occultations comme variables aléatoires. Dans [SSZ02], une interprétation des valeurs aberrantes par des variables aléatoires a été spécialement proposée dans le cadre de la reconstruction à partir de paires stéréoscopiques. Il est proposé d'ajouter, pour chaque pixel, 2 variables aléatoires binaires correspondantes aux zones occultées et aux discontinuités. Le problème d'inférence se reformule alors de la façon suivante :

$$P(D, L, O | I_L, I_R) \propto P(I_R | D, L, O, I_L) P(D, L, O | I_L)$$

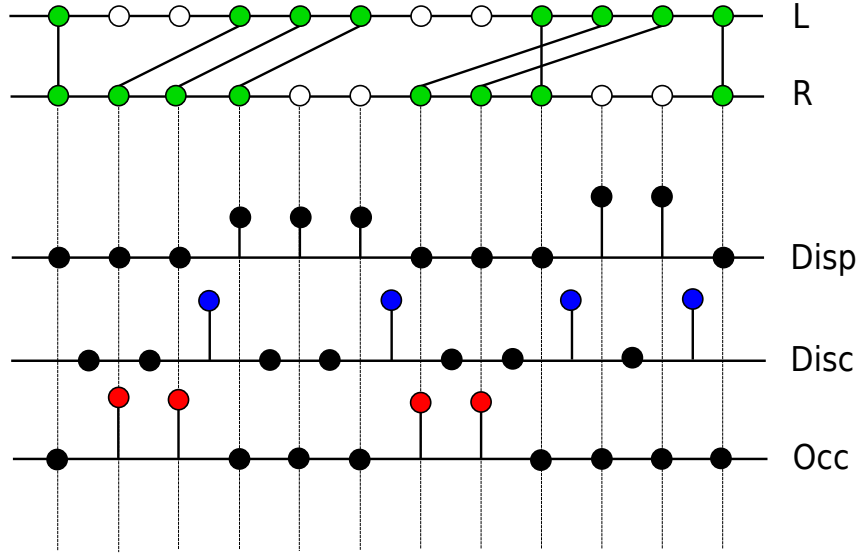


FIGURE 1.9 – Exemple de la représentation par 3 champs de Markov d’une scène avec des discontinuités et occultations. Les lignes L et R représentent 12 pixels horizontaux pour une même position de l’image gauche et droite. Les traits représentent l’appariement d’un pixel de l’image gauche avec celle de l’image droite. Lorsque les sites sont coloriés en vert, cela signifie qu’il sont visibles dans leurs images respectives, inversement, la couleur blanche indique une occultation. La troisième ligne correspond aux disparités. La quatrième représente les discontinuités. La cinquième les occultations.

où L est l’ensemble des variables aléatoires représentant les discontinuités et O est l’ensemble des variables représentant les occultations, et où L et O sont binaires. La figure 1.9 montre un exemple d’une scène et les variables à estimer qui en découlent.

Mais, il s’avère difficile de trouver le maximum *a posteriori* de cette nouvelle formulation. En effet, chaque terme d’attache aux données passe d’une clique d’ordre 1, à une clique d’ordre 3. Ceci est beaucoup plus complexe à optimiser. Il est alors proposé dans un premier temps, au lieu d’estimer l’ensemble des 3 champs, d’interpréter directement les occultations et les sauts de disparités comme des valeurs aberrantes dans les estimateurs. En conséquence, l’estimation ou le choix des paramètres des fonctions ρ_p et ρ_D dans le terme de vraisemblance et dans le terme de régularisation doivent permettre la prise en compte des occultations et des discontinuités. Pour cela, un estimateur robuste est utilisé permettant de mieux modéliser la distribution des appariements en prenant en compte les occultations, la distribution du voisinage et les sauts de disparités.

1.2.1.7 Pondération photométrique :

Nous avons vu qu’il est nécessaire d’introduire des fonctions robustes pour prendre en compte les sauts de disparités. À un saut de disparité, les couleurs de deux objets, le plus souvent, sont différentes. Dans ce cas, on peut donc supposer que les sauts de disparités sont liés aux gradients d’intensité de l’image. Pour prendre en compte ce lien, une pondération photométrique peut être ajoutée au terme de régularisation en fonction de l’image d’intensité :

$$U(D|I_L) = \sum_{c \in C_D} W(c, I_L) \rho_d(\varphi(c, D)) \quad (1.13)$$

où la fonction $W(c, I_L)$ permet de relâcher les contraintes sur la présence de contours de disparité, là où il y a des contours dans I_L . Pour définir la présence d'un contour, plusieurs caractéristiques peuvent être utilisées comme la magnitude locale du gradient [BVZ01a]. La fonction W est alors décroissante. Cela implique que plus le gradient est important, plus l'importance de la valeur du potentiel diminue dans l'*a priori* comparativement au terme d'attache aux données. La fonction $W(c, I_L) = e^{-\frac{|\varphi(c, I_L)|}{\sigma}}$ peut être utilisée. Un autre opérateur de détection de contours ou de segmentation de la scène [SSZ02] peut être aussi utilisé.

1.2.2 Optimisation

Une fois le modèle choisi, il faut trouver la réalisation de D qui maximise au mieux l'estimateur *a posteriori*.

Lorsque que la probabilité *a posteriori* se résume au terme de vraisemblance, une recherche exhaustive suffit pour trouver quelles sont les configurations qui maximisent la densité de probabilités. En effet chaque variable étant indépendante de ses voisines, il suffit, pour chaque variable, de trouver celle qui maximise le terme de vraisemblance. Dans ce cas, la complexité est linéaire en fonction de la taille des données. Dès lors que l'ordre du la plus grande clique est supérieure à 1, le nombre de solutions possible devient trop important pour permettre d'effectuer une recherche exhaustive.

Il existe de nombreuses méthodes d'optimisation permettant de trouver quelles sont les combinaisons qui maximisent la probabilité *a posteriori*. Les premières méthodes qui sont apparues ont été celles fondées sur une approche probabiliste nécessitant un échantillonnage du champ de Markov [GG84]. Ces méthodes mettent un temps important à converger et ne garantissent en aucun cas d'atteindre la solution qui minimise globalement l'énergie. Par la suite, d'autres méthodes ont été proposées [Ter86] fondées sur la minimisation énergétique des potentiels plutôt que sur la maximisation de la densité de probabilité. En effet, la fonction $f(x) = -\exp(x)$ étant monotone décroissante, minimiser le potentiel $U(x)$ équivaut à maximiser la densité de probabilité de la forme $\exp(-U(x))$. Le problème à minimiser sera donc :

$$\min_D [U(I_R, |D, I_L) + U(D|I_L)] \quad (1.14)$$

Il s'avère que ce type de problème a déjà été très étudié, notamment lorsque les labels sont binaires. Actuellement des méthodes efficaces existent pour d'une part, optimiser ces fonctions, mais aussi, les caractériser pour trouver l'algorithme le plus adapté. En effet, certaines énergies peuvent être optimisées globalement en un temps polynomial, il est donc important de pouvoir caractériser ces énergies afin de pouvoir les optimiser au mieux. Dans, [KZ02] il est proposé de savoir quelle énergie peut être minimisée par coupe de graphe.

A partir d'un ordre de clique égale à 2, la notion de sous-modularité intervient. Cette notion est importante, car malgré le caractère *NP-Difficile* de trouver le minimum d'une énergie multi-label avec une clique d'ordre 2 dans le cas général, si cette dernière est sous-modulaire, alors il sera possible de trouver le minimum en un temps polynomial dans le cas binaire.

Une énergie multi-label de degré deux est sous-modulaire si la fonction de coût de chaque terme de degré deux est sous-modulaire. Une fonction de coût de degré deux est sous-modulaire si elle respecte la propriété (3.21). Par exemple, dans le cas multi-label, $f(t) = |t|$ est sous-modulaire, par contre $f(t) = t^2$ ne l'est pas. Le caractère de sous-modularité d'une énergie ne dépend donc pas des termes unitaires, ce qui implique que n'importe quelle fonction ρ_p peut être utilisée pour le terme d'attache aux données (1.7). Par contre, la forme de terme de régulation est importante pour le caractère sous-modulaire de l'énergie.

Quand l'énergie est composée de cliques d'ordre supérieur à deux, il est possible, soit de l'optimiser directement, soit, de la transformer en une énergie de degré deux en ajoutant des variables supplémentaires ayant le même minimum global. Dans ce cas, la réduction en une énergie de degré deux crée forcément des termes non sous-modulaires. C'est le cas de l'*a priori* avec la dérivée du second ordre (1.12).

1.2.2.1 Optimisation de fonctions sous-modulaires

Plusieurs algorithmes existent afin d'optimiser une fonction sous-modulaire multi-label. L'un des premiers utilisé en traitement d'image a été l'algorithme α -expansion (détaillé ici 3.2.3) qui décompose le problème d'optimisation globale en une succession de problèmes binaires. Le principal avantage de cet algorithme est d'apporter un bon compromis entre la rapidité et la qualité d'optimisation [BVZ01b]. En effet, même si le minimum global de la fonction n'est pas atteint, l'énergie après convergence, est assurée d'être inférieure à l'énergie du minimum global multipliée par un facteur.

Plus récemment, des algorithmes fondés sur des méthodes de relaxation ont été proposés afin de converger vers un minimum global de l'énergie [GBO12]. Ces méthodes sont très longues à converger et l'apport qualitatif pour la reconstruction 3D reste souvent négligeable comparativement aux techniques fondées sur les coupes de graphe.

1.2.2.2 Optimisation de fonctions non sous-modulaires

Lorsque l'énergie n'est pas sous-modulaire, des méthodes dites *heuristiques* peuvent néanmoins être appliquées afin de converger vers un minimum local.

Ainsi, les premières méthodes efficaces pour la reconstruction 3D ont été celles fondées sur l'algorithme *belief propagation* [SSZ02]. En pratique, les énergies non sous-modulaires avec des cliques d'ordre 2 (Comme celles proposées dans la partie 1.2.1.6 avec une fonction robuste sur le terme de régularisation) sont efficacement optimisées avec l'algorithme de *belief propagation* [SSZ02]. Mais quand l'ordre des cliques augmente, et par conséquent le nombre de solutions, l'algorithme ne converge plus aussi bien.

Une méthode de fusion (dont l' α -expansion est un cas particulier, décrite plus en détail dans la partie 3.2.1) permet de diminuer l'énergie, mais sans garantie de pouvoir atteindre un minimum global. La difficulté, dans le cas non sous-modulaire, des algorithmes de fusion est de proposer un jeu de labels minimisant l'énergie afin de converger vers un minimum intéressant.

Pour cela, plusieurs types de propositions peuvent être réalisées. Par exemple fusionner le résultat de deux algorithmes indépendants, générer aléatoirement une proposition (suivant une distribution uniforme, ou par exemple, une loi normale avec comme moyenne la solution courante), ou comme proposé dans [Ish09b], en optant pour une solution dans la direction opposée au gradient de l'énergie. Cette approche par fusion est intéressante car elle permet d'obtenir une bonne configuration entre plusieurs cartes de profondeur différentes. Dans [UWSI12], une autre méthode de fusion, fondée sur les méthodes probabilistes, est proposée pour la reconstruction 3D multi-vues appliquée à la reconstruction 3D de scène urbaine. Les résultats montrent que cette approche est efficace.

1.2.3 Exemples

Afin d'illustrer le cas sous-modulaire et le cas non sous-modulaire, nous allons détailler deux exemples de modèles avec des indicateurs sur la méthode d'optimisation appropriée :

Exemple 1.2 (Reconstruction 2 vues avec une régularisation du premier ordre). C'est l'un des premiers modèles utilisés. Ce modèle correspond à un champ de Markov avec des cliques d'ordre 2 (Exemple 1.8a). L'énergie du modèle est la suivante :

$$\sum_{s \in S} \rho_s \left(\frac{|I_L(s) - I_R(s, D(s))|}{\sigma_s} \right) + \lambda_D \sum_{c \in C_D} |\varphi(c, D)| \quad (1.15)$$

où S est l'ensemble des pixels, C_D l'ensemble des patches 2×1 et 1×2 pixels de l'image et σ_s l'échelle. L'avantage de cette énergie est son caractère sous-modulaire. ρ_s est une fonction liée à la distribution du bruit photométrique. La famille des exponentielles lissées peut être choisie : $\rho_\alpha(t) = \frac{1}{2\alpha}((1+t^2)^\alpha - 1)$. Cette famille permet de modéliser correctement plusieurs distributions grâce au paramètre α . Dans cet exemple, $\alpha = 0.7$ est choisi. Une valeur entre 0.5 et 1 pour α est un bon compromis et permet de mieux prendre en compte les discontinuités. Cette énergie, sous-modulaire, peut donc être optimisée par α -expansion (section 3.2.3).

La figure 1.10 montre l'optimisation du modèle (1.15) avec l'algorithme α -expansion pour différentes valeurs de λ_D sur une paire stéréoscopique de la base *KITTI Vision Benchmark Suite*. Nous pouvons voir que lorsque $\lambda_D = 0$ (L'énergie est donc composée seulement du terme d'attache aux données), la reconstruction n'est pas bonne du fait de l'ambiguïté. Avec $\lambda_D = 1$, grâce à l'*a priori*, l'image est plus lisse. Il reste cependant quelques zones comme le bas de la chaussée mal reconstruites. Avec $\lambda_D = 4$, la chaussée est plus lisse, mieux reconstruite. Avec $\lambda_D = 16$, la chaussée au loin est encore mieux reconstruite, cependant, un plan fronto-parallèle s'est formé au niveau de la chaussée à cause du poids important du terme de régularisation. Ce phénomène est encore plus important avec $\lambda_D = 32$ et $\lambda_D = 64$. Pour cette scène, $\lambda_D = 4$ semble être un bon compromis.

Exemple 1.3 (Reconstruction multi-vues avec la prise en compte des occultations et une régularisation du second ordre). Afin d'illustrer les possibilités des méthodes globales, une partie du modèle synthétisé dans [WTRF09] proposant un cadre général pour l'optimisation de méthodes globales avec une régularisation du second ordre est présenté. Dans le cas multi-vues, si l'on connaît les paramètres extrinsèques d'un ensemble de caméras, il est possible de savoir, pour une profondeur donnée dans la vue principale, quelle est la projection dans les autres caméras. Dans cet exemple, on se restreint au cas tri-focal. Soit I_g, I_m et I_d , 3 caméras (respectivement situées à gauche, au milieu et à droite). Dans l'hypothèse où $I_m = I_0$, l'énergie globale est :

$$U(D|I_g, I_m, I_d) = U(I_g, I_d|D, I_m) + U(D|I_m) \quad (1.16)$$

où $U(I_g, I_d|D, I_m)$ est toujours le potentiel correspondant au terme d'attache aux données photométrique. Le modèle proposé dans [WTRF09] permet de prendre en compte les occultations. Pour cela, une variable de visibilité annexe $\mathcal{V}$ est ajoutée. Le terme de vraisemblance aura donc comme coût :

$$f(\Delta I, \mathcal{V}) = \begin{cases} \rho_s(\Delta I) & \text{si } \mathcal{V} = 1 \\ \nu & \text{sinon.} \end{cases} \quad (1.17)$$

Enfin, $U(D|I_m)$ est le potentiel de régularisation du second ordre (1.12). Ce choix implique donc des cliques d'ordre 3.

Pour minimiser (1.16), il est possible d'utiliser la technique de fusion (approfondie dans la partie 3.2.1). Cette technique permet, à partir d'une carte de profondeur existante D_t , et d'une proposition D^p , de trouver la carte fusion D^* des deux configurations qui minimisent l'énergie (1.16). Pour cela, en chaque site est définie une variable booléenne B . Le résultat de la

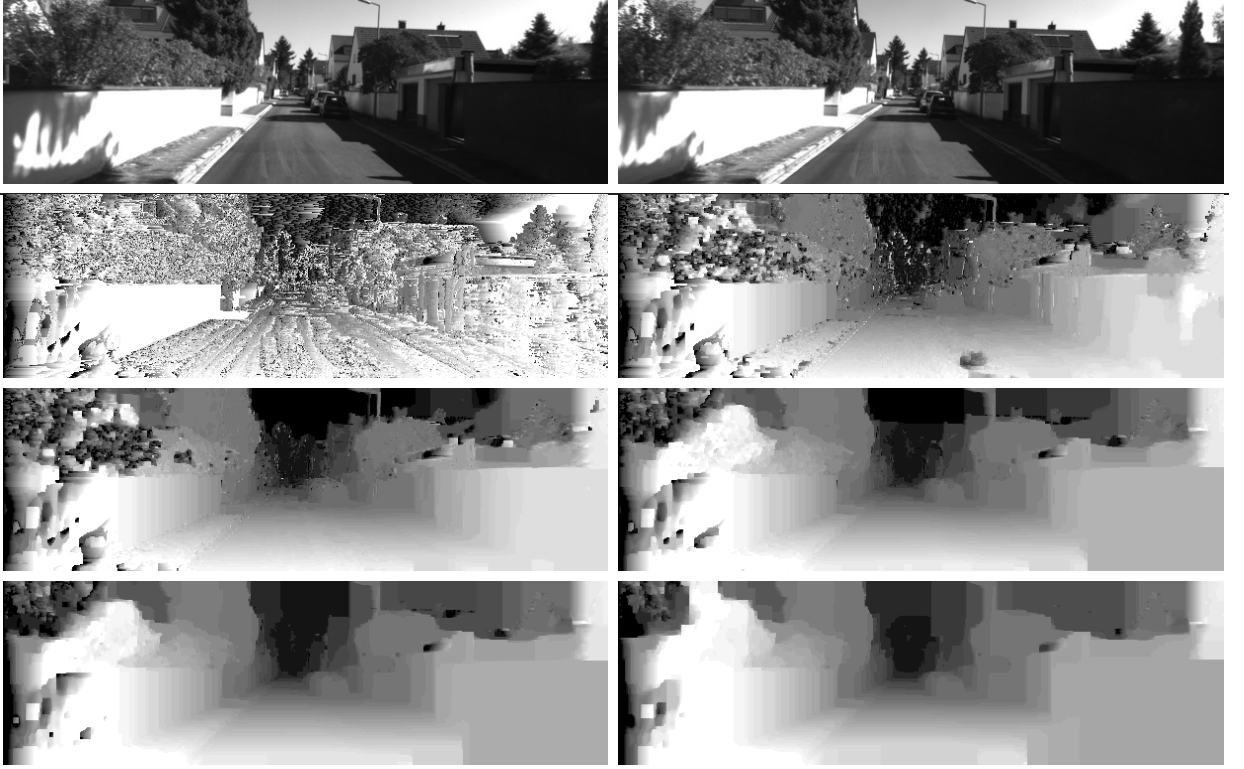


FIGURE 1.10 – Résultat de la minimisation de (1.15) par α -expansion avec une régularisation du premier ordre (1.11) avec plus ou moins d'importance sur l'*a-priori*. Première ligne : l'image gauche et droite de la scène. Seconde ligne : $\lambda_D = 0$ et $\lambda_D = 1$. Troisième ligne : $\lambda_D = 4$ et $\lambda_D = 16$, Quatrième ligne : $\lambda_D = 32$ et $\lambda_D = 64$.

fusion sera une combinaison linéaire de la configuration existante et de la proposition suivant la variable booléenne B comme décrit dans la partie 3.31. En d'autres termes, quand la variable booléenne est égale à 0, la solution initiale est retenue, quand elle est égale à 1, la proposition est retenue.

Création du graphe : L'énergie (1.16) va être représentée sous la forme d'un graphe pour pouvoir la minimiser efficacement. Un graphe est construit, avec comme sommets les variables binaires à optimiser, et comme arêtes les cliques entre les variables. Un terme booléen sera donc l'ensemble des variables composées par la clique avec comme coefficient le coefficient de la clique.

Pour une meilleure compréhension, la figure 1.11 reprend l'exemple utilisé dans [WTRF09]. La Figure 1.11a montre 4 pixels de la caméra centrale alignés horizontalement dans l'image (représentés par les droites noires passant par le point m). La caméra de gauche g et de droite d sont modélisées par 3 pixels. Les solutions D^t et D^p sont montrées pour chaque pixel de la caméra m . La Figure 1.11b montre le graphe créé pour la fusion de D^t et D^p . Pour chaque pixel de l'image référence m_i est associée une variable booléenne B_i . Dans cet exemple, seule la relation spatiale horizontale est représentée.

Pour chaque pixel, il y a 4 arêtes relatives au terme d'attache aux données : l'appariement de la solution actuelle dans l'image gauche et droite et l'appariement de la solution proposée dans l'image gauche et droite. Une première constatation est que certaines combinaisons des deux configurations apportent des occultations. Par exemple si au site t la solution actuelle est

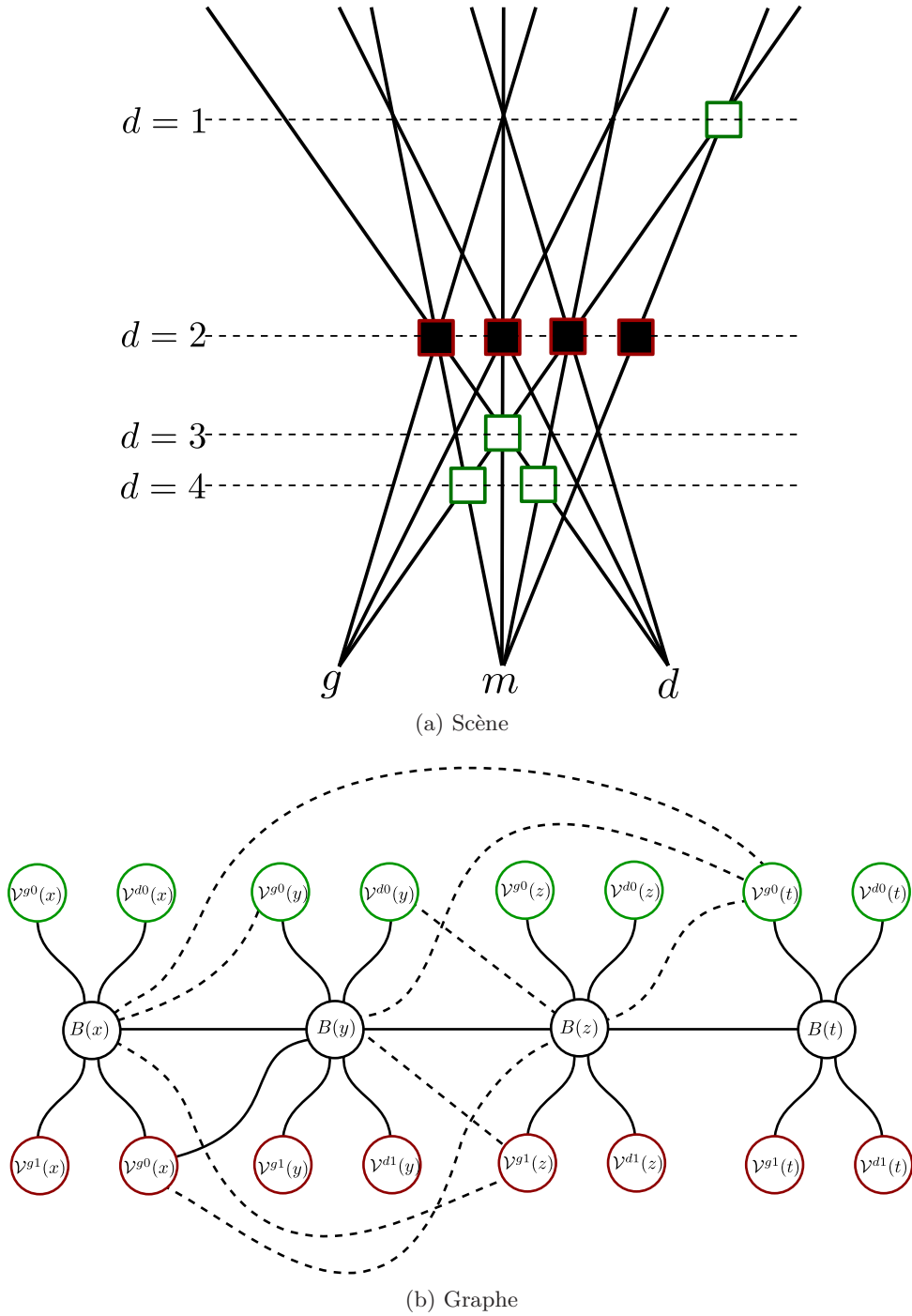


FIGURE 1.11 – Exemple de construction d'un graphe pour l'étape de fusion. La figure 1.11a représente une scène de 4 pixels où m est la caméra référence. Les carrés verts représentent la solution courante, les rouges la solution proposée. La figure 1.11b représente le graphe construit pour l'étape de fusion. Les cercles noirs sont les variables booléennes pour chaque pixel de l'image référence m , les cercles verts sont les variables de visibilité de l'image référence m par rapport à la solution actuelle, et les cercles rouges par rapport à la solution proposée. Les traits liant les variables booléennes aux variables de visibilité représentent le terme d'attache aux données, les traits liant les variables booléennes entre elles sont les cliques d'ordre 3 représentant le terme de régularisation. Les traits en pointillé correspondent aux pénalités d'occultations.

la solution recherchée et au site z la solution proposée est la solution recherchée, alors le modèle devra prendre en compte cette occultation. En effet, le terme de vraisemblance représentant l'appariement au site z de l'image référence m sur l'image gauche g sera faussé (surtout si l'intensité aux sites z et t ont une intensité lumineuse très différente), si le score du potentiel est trop important du fait de la différence d'intensité, cette solution ne minimisera donc pas l'énergie. Pour gérer ce genre de configuration, on introduit des variables de visibilité $\mathcal{V}^{\{g/d\}i}(x)$. Il y a deux variables par vue auxiliaire pour un site x : une pour représenter les occultations de la variable actuelle ($i = 0$) et l'autre pour la variable proposée en plus de la vue de référence ($i = 1$), deux variables de visibilité sont ajoutées par vue supplémentaire.

La construction du terme de vraisemblance sera donc la suivante : chaque variable booléenne représentant un pixel sera connectée aux variables de visibilité correspondantes à la vue. La pondération des arêtes correspondra au terme de vraisemblance.

Prenons l'exemple du site y , quand le label est visible, les termes correspondant aux quatre arêtes du terme d'attache aux données sont : $\rho_s(|I_m(y) - I_g(y, D^t)|)\overline{B}(y)\mathcal{V}^{g0}(y)$ pour le terme de vraisemblance de l'image référence avec l'image gauche pour la profondeur existante, $\rho_s(|I_m(y) - I_d(y, D^t)|)\overline{B}(y)\mathcal{V}^{d0}(y)$ pour le terme de vraisemblance de l'image référence avec l'image droite pour la profondeur existante, $\rho_s(|I_m(y) - I_g(y, D^p)|)B(y)\mathcal{V}^{g1}(y)$ pour le terme de vraisemblance de l'image référence avec l'image gauche pour la profondeur proposée et $\rho_s(|I_m(y) - I_d(y, D^p)|)B(y)\mathcal{V}^{d1}(y)$ pour le terme de vraisemblance de l'image référence avec l'image droite pour la profondeur proposée. Le coefficient de chaque terme correspond au potentiel de l'appariement en fonction du label (existant ou proposé) et de l'image en question (gauche ou droite) de l'équation (1.17) quand $\mathcal{V} = 1$. Inversement, quand $\mathcal{V} = 0$, le score des arêtes sera donc ν , ce qui donne : $\nu\overline{B}(y)\overline{\mathcal{V}}^{g0}(y)$, $\nu\overline{B}(y)\overline{\mathcal{V}}^{d0}(y)$, $\nu B(y)\overline{\mathcal{V}}^{g1}(y)$ et $\nu B(y)\overline{\mathcal{V}}^{d1}(y)$.

Pour permettre une occultation entre deux variables, une arête est connectée entre la variable booléenne correspondant à la variable au premier plan, avec le noeud de visibilité correspondant à la variable occultée. L'idée est d'ajouter une pénalisation importante quand le label au premier plan est égal à 1 et que la variable de visibilité du label occulté est aussi égale à 1, cette pénalisation empêchera donc qu'une variable au premier plan et qu'une variable occultée par celle-ci minimise l'énergie (cette pénalisation doit être assez importante pour que cette possibilité ne se produise jamais et ainsi que la contrainte d'unicité soit respectée). Prenons l'exemple du site x , la variable courante en ce site occulte 3 variables : deux variables proposées (celle en y et en t), et une variable courante en z . Toutes sont occultées dans la vue de gauche. 3 arêtes seront donc créées, deux liant les variables d'occultation représentant les variables proposées des noeuds y et t pour la vue gauche et une liant la variable d'occultation représentant la variable courante du noeud z , pour la vue gauche. Les termes correspondant à ces arêtes seront donc $\infty\overline{B}(x)\mathcal{V}^{g0}(y)$, $\infty\overline{B}(x)\mathcal{V}^{g1}(z)$ et $\infty\overline{B}(x)\mathcal{V}^{g0}(t)$.

Inversement, la variable proposée en x est occultée par deux variables dans la vue droite par les solutions courantes des sites y et z , il y aura donc deux arêtes liant la variable d'occultation au site x correspondant à la vue gauche de la variable proposée, c'est à dire $\mathcal{V}^{g1}(x)$, aux deux variables $\overline{B}(y)$ et $\overline{B}(z)$, ce qui donnera les termes $\infty\overline{B}(y)\mathcal{V}^{g1}(x)$ et $\infty\overline{B}(z)\mathcal{V}^{g1}(x)$. Le coût des arêtes du terme de vraisemblance modélisant l'équation (1.17) se résume donc à :

$$\phi(B(x), \mathcal{V}^{i\alpha}(x)) = \begin{cases} 0 & \text{si } B(x) = 1 \\ \rho_s(\Delta I) & \text{si } \mathcal{V}^{i\alpha}(x) = 1 \\ \nu & \text{sinon.} \end{cases} \quad (1.18)$$

Afin de modéliser la pénalisation du second ordre, les 4 variables booléennes $B(i)$ sont connectées par des cliques d'ordre 3. Prenons l'exemple des sites x , y et z . Ces trois sites, avec un terme

de régularisation du second ordre, sont liés par une clique d'ordre 3 centrée en y . Pour chaque clique d'ordre 3, il y aura donc 8 combinaisons possibles de labels. Si nous choisissons $\rho_p(t) = |t|$, le potentiel local correspondant à la somme des 8 termes de la clique $\{x, y, z\}$ sera donc :

$$\begin{aligned} & |\varphi(\{x, y, z\}, \{D^t, D^t, D^t\})| \overline{B}(x) \overline{B}(y) \overline{B}(z) + |\varphi(\{x, y, z\}, \{D^t, D^t, D^p\})| \overline{B}(x) \overline{B}(y) B(z) + \\ & |\varphi(\{x, y, z\}, \{D^t, D^p, D^t\})| \overline{B}(x) B(y) \overline{B}(z) + |\varphi(\{x, y, z\}, \{D^t, D^p, D^p\})| \overline{B}(x) B(y) B(z) + \\ & |\varphi(\{x, y, z\}, \{D^p, D^t, D^t\})| B(x) \overline{B}(y) \overline{B}(z) + |\varphi(\{x, y, z\}, \{D^p, D^t, D^p\})| B(x) \overline{B}(y) B(z) + \\ & |\varphi(\{x, y, z\}, \{D^p, D^p, D^t\})| B(x) B(y) \overline{B}(z) + |\varphi(\{x, y, z\}, \{D^p, D^p, D^p\})| B(x) B(y) B(z) \end{aligned}$$

où la fonction $\varphi(\{p, q, r\}, \{D^p, D^q, D^r\})$ est la fonction de régularisation du second ordre (1.12) tel que $\varphi(\{p, q, r\}, \{D^s, D^t, D^u\}) = D^s(p) - 2D^t(q) + D^u(r)$.

Prenons l'exemple du terme $|\varphi(\{x, y, z\}, \{D^t, D^t, D^t\})| \overline{B}(x) \overline{B}(y) \overline{B}(z)$, ce terme correspond au coût de l'*a priori* avec les 3 solutions existantes, ce qui donne $|4 - 2 * 3 + 4| \overline{B}(x) \overline{B}(y) \overline{B}(z)$. Si nous prenons le terme, $|\varphi(\{x, y, z\}, \{D^t, D^t, D^p\})| \overline{B}(x) \overline{B}(y) B(z)$, il correspond au potentiel de l'*a priori* avec les solutions existantes pour les sites x et y , et la solution proposée pour z , ce qui donne $|4 - 2 * 3 + 2| \overline{B}(x) \overline{B}(y) B(z)$. Nous pouvons voir avec ces deux exemples que dans le premier cas, la pondération est égale à 2, et dans le second cas, où la solution est un plan, la pondération est égale à 0.

Pour récapituler, l'équation pseudo-booléenne quadratique correspondant au graphe 1.11b à minimiser lors de la fusion sera de la forme (Par souci de visibilité, seuls les termes impliquant le site y sont mentionnés) :

$$\begin{aligned} \min_B E(\mathbf{B}) = & \rho_s(|I_m(y) - I_g(y, D^t)|) \overline{B}(y) \mathcal{V}^{g0}(y) + \rho_s(|I_m(y) - I_d(y, D^t)|) \overline{B}(y) \mathcal{V}^{d0}(y) \\ & + \rho_s(|I_m(y) - I_g(y, D^p)|) B(y) \mathcal{V}^{g1}(y) + \rho_s(|I_m(y) - I_d(y, D^p)|) B(y) \mathcal{V}^{d1}(y) \\ & + \nu \overline{B}(y) \overline{\mathcal{V}}^{g0}(y) + \nu \overline{B}(y) \overline{\mathcal{V}}^{d0}(y) + \nu B(y) \overline{\mathcal{V}}^{g1}(y) + \nu B(y) \overline{\mathcal{V}}^{d1}(y) \\ & + \infty \overline{B}(y) \mathcal{V}^{g0}(t) + \infty \overline{B}(y) \mathcal{V}^{g1}(z) + \infty \overline{B}(y) \mathcal{V}^{d1}(x) \\ & + \infty \overline{B}(x) \mathcal{V}^{g0}(y) + \infty \overline{B}(z) \mathcal{V}^{g0}(y) \\ & + |\varphi(\{x, y, z\}, \{D^t, D^t, D^t\})| \overline{B}(x) \overline{B}(y) \overline{B}(z) + |\varphi(\{x, y, z\}, \{D^p, D^t, D^t\})| B(x) \overline{B}(y) \overline{B}(z) + \dots \\ & + |\varphi(\{y, z, t\}, \{D^t, D^t, D^t\})| \overline{B}(y) \overline{B}(z) \overline{B}(t) + |\varphi(\{y, z, t\}, \{D^p, D^t, D^t\})| B(y) \overline{B}(z) \overline{B}(t) + \dots \end{aligned} \quad (1.19)$$

L'exemple de la figure 1.12 illustre concrètement la fusion entre deux ensembles de labels. La carte D^t est le résultat d'un algorithme d'appariements par fenêtres. Les objets fronto-parallèles sont bien reconstruits, par contre, le plan de la route ne l'est pas du fait de l'hypothèse spatiale du premier ordre. Pour éviter cela, un plan de la route peut être estimé avec les profondeurs actuelles et utilisé comme proposition D^p avec un modèle du second ordre. D^* est le résultat de la fusion. Nous constatons que les profondeurs sur les objets fronto-parallèles sont bien préservées. Au niveau de la route, l'ensemble des profondeurs du plan ont été sélectionnées car elles respectent mieux localement le modèle du second ordre.

1.2.4 Conclusion

Nous avons vu dans cet état de l'art que de nombreuses méthodes de reconstruction 3D ont été proposées et que certaines sont plus particulièrement adaptées à l'environnement routier. Nous avons constaté que les modèles markoviens permettent de modéliser de nombreux aspects des problèmes posés. Par ailleurs, ce sont celles qui actuellement apportent les meilleurs résultats sur les bases de tests en environnement routier, tels que la base *KITTI Vision Benchmark Suite*.

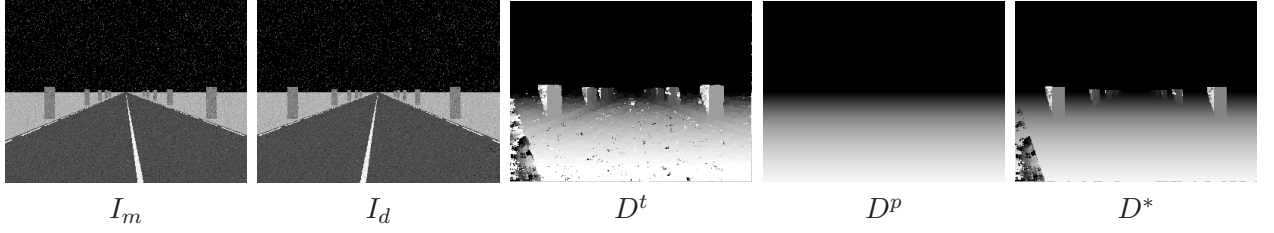


FIGURE 1.12 – Résultat de la fusion avec un modèle du second ordre (1.19). De gauche à droite : l'image gauche et droite, la carte de disparité actuelle D^t , la carte de disparité proposée D^p . Le résultat de la fusion D^* (Notons que dans cet exemple, les occultations ne sont pas prises en compte)

Ainsi, les ambiguïtés posées par le problème de la reconstruction 3D à partir de paires stéréoscopiques peuvent être levées en introduisant plusieurs *a priori* sur les variables recherchées. Dans le cas de l'environnement routier qui est constitué en grande partie de zones planaires, un *a priori* du second ordre est donc fortement conseillé pour l'*a priori* spatial afin de ne pas pénaliser les plans non fronto-parallèles à la caméra.

Une fois un modèle défini, il faut optimiser les modèles pour savoir quel ensemble de labels minimise l'énergie. La complexité d'optimisation change suivant les modèles. L'optimisation de certaines énergies obtenues a une complexité polynomiale : ce sont les énergies dites *sous-modulaires*. Les autres énergies sont dites *non sous-modulaires*, et leurs complexité est *NP-difficile*.

Il s'avère que l'optimisation d'une énergie avec un terme de régularisation du premier ordre est sous-modulaire si la fonction de coût sur le terme de régularisation l'est aussi. Quand le terme de régularisation implique une clique d'ordre 3, alors l'énergie n'est plus sous-modulaire. Il faut donc, lors de la création d'un modèle probabiliste, prendre en compte ce facteur car la capacité à pouvoir trouver le minimum d'un modèle en un temps polynomial est une caractéristique importante. Quand l'énergie n'est pas sous-modulaire, il existe néanmoins des méthodes permettant d'atteindre un minimum local intéressant. Grâce aux méthodes d'optimisation de fonctions pseudo-boléennes appliquées en traitement d'images, de nouveaux algorithmes tels que la technique de fusion ont permis d'optimiser efficacement le type d'énergie, même celles non sous-modulaires.

Dans les deux exemples abordés, le modèle avec une régularisation du premier ordre, bien qu'il ne soit pas parfaitement adapté aux scènes routières, permet de produire un résultat intéressant dans l'exemple 1.2. L'autre exemple montre qu'il est possible d'utiliser un *a priori* du second ordre permettant d'affiner la carte de profondeur ou de fusionner le résultat de plusieurs algorithmes notamment grâce à la méthode de fusion.

Chapitre 2

Traitement d'image en présence de brouillard

Ce chapitre fait l'état de l'art des différentes techniques de restauration en présence de brouillard et des méthodes d'estimation du coefficient d'extinction.

2.1 Introduction

S'il est peu fréquent, le brouillard peut entraîner de graves conséquences. D'une part il diminue la capacité de perception du conducteur, par exemple, la présence de brouillard augmente son temps de réaction. D'autre part il réduit l'efficacité de nombreux algorithmes de traitement d'images en extérieur.

Considérons l'exemple de la stéréo-vision vu dans le chapitre précédent. Les algorithmes de reconstruction 3D en présence de brouillard, dans les zones proches, produisent une carte de profondeur de bonne qualité, mais à partir d'une certaine distance, la qualité de la carte reconstruite se dégrade drastiquement. En effet, le contraste dans la scène diminue avec la distance. De nombreux autres algorithmes comme la détection d'obstacle, le calcul de trajectoire,

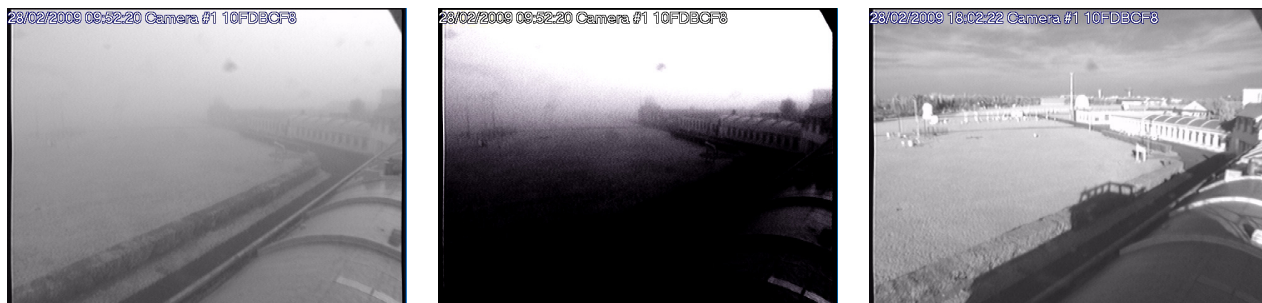


FIGURE 2.1 – Une image en présence de brouillard (image de gauche). La même image avec une transformation gamma permet de faire ressortir les objets au loin. La même scène avec un temps dégagé.

subissent cet effet car ils ne sont pas adaptés et ils supposent une distribution homogène du contraste dans l'image.

Comme pour les algorithmes, les capacités de perception se dégradent avec la distance qui diminue les contrastes. Il s'avère qu'une image acquise en présence de brouillard contient des informations même si elles ne sont pas aisément perceptibles. Par exemple, si l'on observe l'image 2.1, nous pouvons voir qu'il y a des objets au loin à gauche qui ne sont pas perceptibles dans l'image avec brouillard. Si l'on effectue au préalable une transformation adéquate des contrastes de l'image sur la zone en question, ils sont visibles.

Faire ressortir les faibles contrastes a donc été l'objectif des premiers travaux de traitement d'image en conditions de brouillard. Une première idée pour venir en aide à la perception humaine ou pour améliorer l'efficacité des algorithmes est de restaurer le contraste initial de l'image en pré traitement

Ceci est utilisé dans les systèmes d'aide avancé à la conduite (ADAS). Deux classes d'ADAS peuvent être considérées. La première consiste à afficher l'image d'une caméra frontale après la restauration de l'image. Ce type d'ADAS est appelé *système d'amélioration de la vision avec brouillard*, en anglais *Fog Vision Enhancement System* (FVES). La seconde est d'appliquer après la restauration la détection de véhicules arrêtés / voiture en mouvement / piétons / véhicules deux roues, afin de transmettre un avertissement au conducteur en cas de danger. Un exemple est d'avertir lorsque la distance avec le véhicule en mouvement précédant est trop courte par rapport à la vitesse du conducteur. Il est montré dans [HTA10], pour plusieurs types d'algorithmes de détection, que l'amélioration de la visibilité en pré-traitement permet d'améliorer les performances de détection en présence de brouillard. Cela est dû à une meilleure vérification de l'hypothèse selon laquelle les objets à détecter ont un contraste uniforme sur l'ensemble de l'image. Lorsqu'une seule caméra dans le véhicule est utilisée, l'algorithme de restauration doit pouvoir traiter de façon fiable chaque image d'une séquence en temps réel.

A partir d'une seule image en présence de brouillard, la principale difficulté est que la restauration du contraste d'une image en présence de brouillard est un problème mal posé. En effet, en raison des caractéristiques physiques du brouillard, la restauration de la visibilité nécessite d'estimer à la fois la luminance de la scène sans brouillard et la carte de profondeur de la scène. Ce qui implique d'estimer deux paramètres inconnus par pixel à partir d'une seule image.

Restauration d'image monoculaire : La première approche proposée pour traiter le problème de restauration à partir d'une seule image est décrite dans [NN03]. L'idée principale est d'obtenir une carte de profondeur approximative de la géométrie de la scène permettant de déduire l'image sans brouillard. L'inconvénient de cette approche dans le cas d'une ADAS est clair : il n'est pas facile de fournir la carte de profondeur approximative de la géométrie de la scène à partir du point de vue du conducteur tout au long d'un trajet. Dans [HTA07], cette idée de carte de profondeur approximative a été affinée en proposant plusieurs modèles paramétriques géométriques simples dédiés à des scènes de routes observées à partir d'un véhicule. Pour chaque type de modèle, les paramètres sont obtenus à chaque vue en essayant de maximiser la profondeur de la scène sans pour autant produire de pixels noirs pendant la restauration de l'image. La limite de cette approche est le manque de souplesse des modèles géométriques proposés.

Parallèlement, une autre approche a été proposée dans [TPP07] basée sur l'utilisation d'une image en couleur en désaturant les couleurs. Néanmoins, cette approche ne fonctionne pas de façon satisfaisante pour les applications routières. En effet, une grande partie de l'image qui correspond à la route est de couleur grise ou blanche. De plus, dans de nombreuses applications d'aide à la conduite pour véhicules intelligents, seules les images en niveaux de gris sont traitées.

Plus récemment, trois algorithmes de restauration ont été proposés [Tan08, HST10b, TH09], ils fonctionnent à partir d'une seule image en niveaux de gris ou en couleur sans utiliser une autre source d'information externe. Ces trois algorithmes se basent sur une régularisation spatiale locale. De ce fait, ces algorithmes peuvent être appliqués aux situations de brouillards homogènes et hétérogènes. Le principal inconvénient des algorithmes dans [Tan08] et [HST10b] est leur temps de traitement : 5 à 7 minutes et 10 à 20 secondes sur une image de 600×400 pixels respectivement. L'algorithme proposé dans [TH09] est beaucoup plus rapide avec un temps de traitement proche du dixième de seconde pour une taille d'image similaire. Une variante plus rapide de [HST10b] a été récemment proposée dans [HST10a]. L'inconvénient de ces trois méthodes et des variantes associées est qu'elles ne sont pas dédiées aux images de routes. Par conséquent, la partie de la route de l'image qui est grise peut être trop restaurée. Cela est dû à l'ambiguïté entre les objets de couleur claire et la présence de brouillard (comme illustré sur les images de la figure 2.4).

Pour pallier cela, des algorithmes adaptés à l'environnement routier ont été proposés. La propriété importante d'une image de route est qu'une grande partie de l'image, qui correspond à la chaussée, peut être supposée plane. Une méthode de restauration dédiée aux surfaces planes a d'abord été proposée dans [HA05], mais cet algorithme n'est pas en mesure d'améliorer correctement la visibilité des objets hors de la surface plane de la route. Récemment, dans [HTA10], un algorithme dédié aux images routières a été proposé avec également un contraste amélioré des objets hors du plan de la route. Cet algorithme permet une bonne utilisation de l'hypothèse de route plane, mais il repose aussi sur une hypothèse de brouillard homogène.

Dans [THC⁺12], le problème de la restauration est formulé comme l'inférence du voile atmosphérique à partir de trois contraintes. La première contrainte repose sur les propriétés photométriques de la scène en présence de brouillard. La deuxième contrainte, nommée la contrainte *no-black-pixel* (L'algorithme décrit en [TH09] correspond au cas particulier où deux contraintes sont utilisées avec le filtre médian). Finalement, pour tenir compte du fait qu'une grande partie de l'image est constituée par une route plane, comme mentionné initialement dans [HTA10], une troisième contrainte qui repose sur l'hypothèse de route plane est ajoutée. Le nouvel algorithme peut donc être considéré comme la généralisation de l'algorithme fondé sur la contrainte *no-black-pixel* présentée dans [TH09] combinée avec l'algorithme spécifique à la route [HA05] faisant l'hypothèse du monde plan. Cet algorithme est capable de traiter des niveaux de gris ainsi que des images en couleurs avec des temps de traitement compatibles avec du temps réel.

Plus récemment, une méthode par égalisation locale de l'histogramme basée sur la loi de Koschmieder a été proposée dans [HCH11]. Cette méthode, couplée avec l'hypothèse du monde plan, présente l'avantage d'apporter une restauration homogène avec la distance ainsi qu'avec un brouillard hétérogène comparativement à la méthode fondée sur la contrainte *no-black-pixel*.

Dans [NKL12], une interprétation bayésienne du problème de restauration a aussi été proposée. Cette méthode fonctionne sur des images couleurs. L'idée est de linéariser l'équation de Koschmieder et de lier l'image restaurée avec l'image acquise par le capteur et une profondeur relative. Comme le problème est mal posé, afin de contraindre l'espace des solutions, deux *a priori* sont ajoutés. Le premier porte sur la distribution locale de l'intensité de l'image restaurée et l'autre, sur la distribution locale des profondeurs. Une énergie est optimisée par une optimisation alternée de l'image restaurée et de la carte relative des profondeurs.

La partie 2.2 présente le modèle de brouillard que nous utilisons. Ensuite, les différents algorithmes de restauration sont décrits dans la partie 2.3. La partie 2.3.1, résume les algorithmes qui ne sont pas directement liés au brouillard comme l'algorithme *multiscale retinex*



FIGURE 2.2 – Atténuation de contraste de la même scène en fonction de différentes valeurs du coefficient d'extinction β .

(MSR) [JRW97] et l'algorithme *contrast-limited adaptive histogram equalization* (CLAHE). Dans la partie 2.3.2, différentes approches de restaurations d'image en présence de brouillard sont décrites : basées sur l'hypothèse route plane (PA) [HA05], sur la segmentation de la chaussée (FSS) [HTA10], sur la contrainte *no-black-pixel* (NBPC) [TH09], sur le principe du *Dark Chanel Prior* (DCP) [HST10b], la combinaison entre celle basée sur la contrainte *no-black-pixel* et sur l'hypothèse route plane (NBPC+PA) proposée dans [THC⁺12] ainsi que celle basée sur l'égalisation de l'histogramme (HE+PA) proposée dans [HCH11]. Enfin, l'interprétation probabiliste présentée dans [NKL12] est décrite. Dans la partie 2.3.5, la comparaison réalisée dans [THC⁺12] entre les algorithmes MSR, CLAHE, DCP, FSS, NBPC, NBPC+PA et HE+PA est présentée. Elle est basée sur une évaluation quantitative sur deux ensembles de 66×4 et 10×4 images en présence de brouillard, mettant en évidence les propriétés de chaque algorithme.

2.2 Effets du brouillard

Étant donnée la luminance intrinsèque d'un objet $L_0(s)$, sa luminance apparente $L(s)$ avec un brouillard de coefficient d'extinction β est modélisée par la loi de Koschmieder introduite en 1924 [Kos24] :

$$L(s) = L_0(s)e^{-\beta p(s)} + L_s(1 - e^{-\beta p(s)}) \quad (2.1)$$

où $p(s)$ est la profondeur de l'objet au pixel s et L_s est la luminance du ciel. Le brouillard a deux effets : d'abord une décroissance exponentielle $e^{-\beta p(s)}$ de la luminance intrinsèque $L_0(s)$, puis l'ajout de la luminance du voile atmosphérique $V(s) = L_s(1 - e^{-\beta p(s)})$ qui est croissante en fonction de la profondeur de l'objet $p(s)$. Ces deux effets peuvent être constatés sur la même scène dans la figure 2.2 pour différentes valeurs de β . La distance de visibilité météorologique est définie comme $d_m = -\frac{\ln(0.05)}{\beta}$ [HTLA06].



FIGURE 2.3 – Variables liées par l'équation de Koschmieder sur une image de synthèse. De gauche à droite, l'image I correspondant à la luminance de la scène observée en présence de brouillard, l'image I_0 correspondant à la luminance sans brouillard, la carte de profondeur p de la scène (avec une transformation gamma), le voile atmosphérique V pour un β et un L_s donné.

A partir de cet instant, nous supposons que la réponse de la caméra est linéaire, et donc que l'intensité de l'image I se substitue à la luminance L . La figure 2.3 montre la relation entre les différentes variables liées par l'équation de Koschmieder sur une image de synthèse. Nous pouvons voir que plus la profondeur est importante, plus l'intensité lumineuse de l'image en présence de brouillard tend vers celle du ciel.

2.3 Restauration d'image

Dans cette partie, nous allons décrire différentes méthodes de restauration. Premièrement, dans la partie 2.3.1, les algorithmes non fondés sur la loi de Koschmieder, et par la suite, dans la partie 2.3.2, les algorithmes de restauration fondés sur la loi de Koschmieder.

2.3.1 Restauration par rehaussement de couleur et de contraste

Nous abordons maintenant la méthode *multiscale retinex* (MSR) ainsi que les algorithmes *Contrast-Limited Adaptive Histogram Equalization* (CLAHE). Ces deux algorithmes ne sont pas basés sur la loi de Koschmieder (2.1) et sont donc seulement capables, en théorie, d'éliminer une épaisseur constante de brouillard.

2.3.1.1 Méthode *multiscale retinex* (MSR)

L'algorithme *multiscale retinex* (MSR) proposé par [JRW97] permet une restauration non-linéaire de l'image. L'effet principal est d'augmenter le contraste dans les zones de faible contraste / luminosité, sans affecter les zones où, le contraste et la luminosité sont supposés bonnes. La sortie de la méthode MSR est simplement la somme pondérée des sorties de plusieurs échelles *retinex* simple (SSR). Chaque composante de couleur étant traitée indépendamment, la forme de base de la méthode pour une échelle donnée sachant l'image d'entrée $I(s)$ est :

$$R_k(s) = \log I(s) - \log[F_k(s) * I(s)] \quad (2.2)$$

où $R_k(s)$ est le résultat de la méthode SSR, F_k représente la k -ème fonction de contour, et $*$ est l'opérateur convolution. Les fonctions de contour F_k sont des gaussiennes normalisées :

$$F_k(\{v, u\}) = \kappa_k e^{-(u^2+v^2)/\sigma_k^2} \quad (2.3)$$

où σ_k est l'échelle qui contrôle la taille du filtrage, κ_k est une constante de normalisation. Finalement la sortie de l'algorithme MSR est :

$$R(s) = \sum_{k=1}^{k=K} W_k R_k(s) \quad (2.4)$$

où W_k est un poids associé à F_k .

Le nombre d'échelles utilisées pour le MSR est dépendant de l'application. Les résultats obtenus en utilisant le *multiscale retinex* sur trois images avec brouillard sont présentés dans la deuxième colonne de la figure 2.4 avec les paramètres utilisés dans [JRW97]. Ils sont répartis en trois ensembles représentant les contours étroits, moyens, et larges. Ce qui donne : $K = 3$, $\sigma_1 = 15$, $\sigma_2 = 80$, $\sigma_3 = 250$ et $W_k = 1/3$ pour $k = 1, 2, 3$.

2.3.1.2 Méthode *Contrast-Limited Adaptive Histogram Equalization* (CLAHE)

Cette méthode améliore localement le contraste d'image. Comme proposé dans [Zui94], l'algorithme traite une fenêtre de 8×8 , appelée « tuile », plutôt que l'image entière. Le contraste de chaque « tuile » est amélioré, de telle sorte que l'histogramme soit le plus uniforme possible. Les « tuiles » avoisinantes sont ensuite combinées par interpolation bilinéaire afin d'éliminer les artefacts produits par les contours des « tuiles ».

La restauration, en particulier dans les zones homogènes est limitée pour éviter de rehausser le bruit ou des structures indésirables telles celles dues au codage jpg. Les résultats obtenus avec l'algorithme CLAHE sont présentés dans la colonne trois de la figure 2.4.

2.3.2 Restauration fondée sur la loi de Koschmieder

Quatre algorithmes de restauration sont maintenant présentés : La première méthode fait l'hypothèse d'une scène plane (PA), la deuxième se nomme *free-space segmentation* (FSS), la troisième méthode de restauration est basée sur la contrainte *no-black-pixel* (NBPC), la quatrième se nomme *dark channel prior* (DCP) et enfin la dernière utilise la contrainte *no-black-pixel* combinée avec la contrainte de scène plane se nomme (NBPC+PA).

2.3.2.1 Méthode du monde plan (PA)

Afin de pallier les problèmes dus aux grandes zones homogènes grises de la chaussées pour les applications embarquées, l'algorithme (breveté) proposé dans [LTA02] prend en compte la surface plane de la route.

Cette méthode, basée sur la loi de Koschmieder, suppose que le coefficient d'extinction β et les paramètres intrinsèques de la caméra δ par rapport au monde supposé plan sont connus. En effet, il est alors possible d'associer une distance p à chaque ligne de l'image connaissant la position de la ligne d'horizon v_h avec la relation suivante :

$$p = \frac{\delta}{v - v_h} \quad \text{si } v > v_h \quad (2.5)$$

p peut être alors substitué dans l'équation de Koschmieder (2.1) :

$$I(\{v, u\}) = I_0 e^{-\beta \frac{\delta}{v - v_h}} + I_s (1 - e^{-\beta \frac{\delta}{v - v_h}}) \quad (2.6)$$

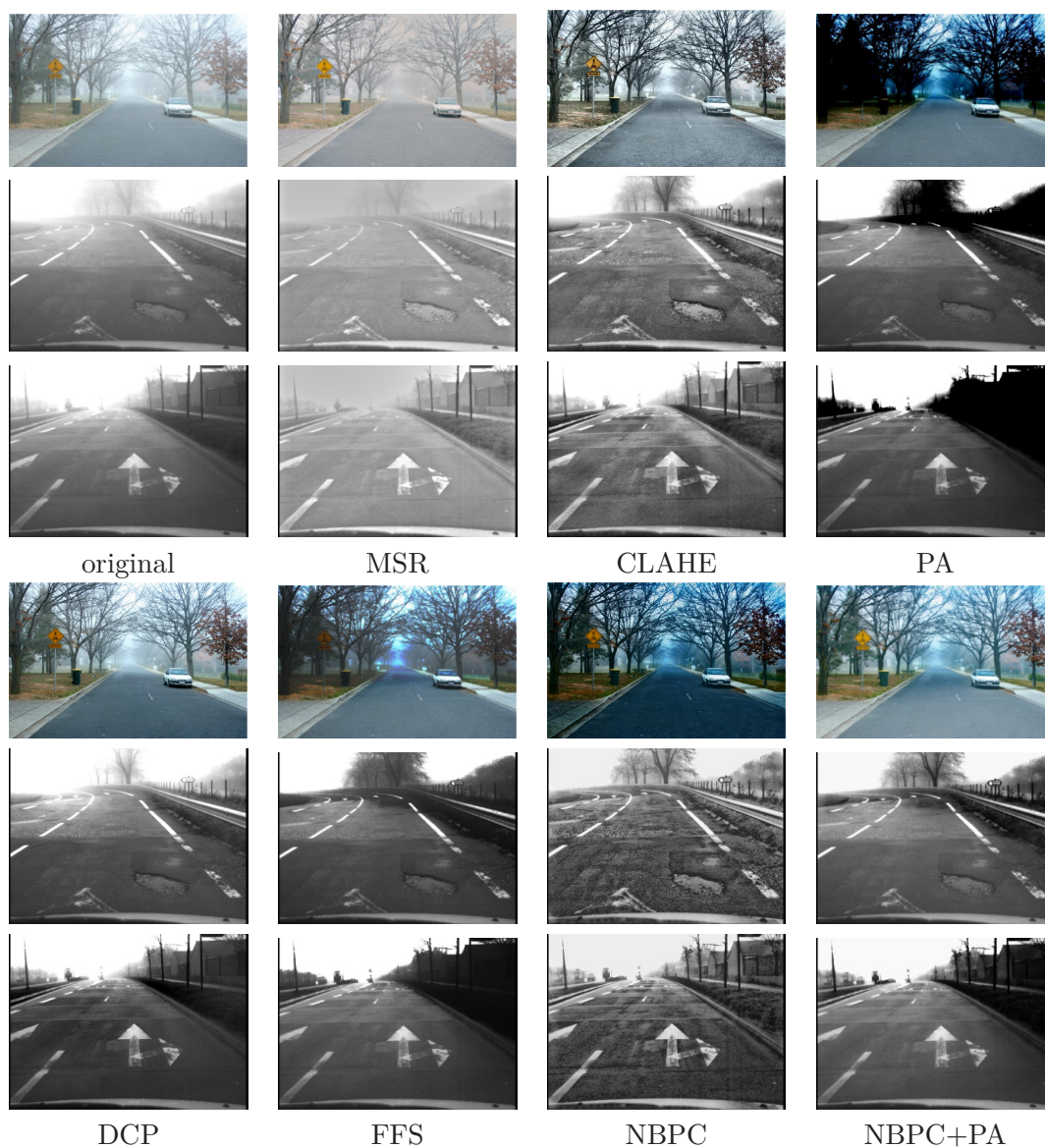


FIGURE 2.4 – De la gauche vers la droite, l'image originale avec brouillard, les images améliorées avec les algorithmes : *multiscale retinex* (MSR), *Contrast-Limited Adaptive Histogram Equalization* (CLAHE), la supposition de la route plane (PA), la méthode *dark channel prior* (DCP), la méthode *free-space segmentation* (FFS), la contrainte *no-black-pixel* (NBPC) et la contrainte *no-black-pixel* associée avec l'hypothèse planéité (NBPC+PA).

Dès lors, les pixels appartenant au plan de la route peuvent être restaurés comme $R(\{v, u\})$ en inversant l'équation de Koschmieder :

$$R(\{v, u\}) = I(\{v, u\})e^{\beta \frac{\delta}{v-v_h}} + I_s(1 - e^{\beta \frac{\delta}{v-v_h}}) \quad (2.7)$$

Afin de permettre la restauration dans toute l'image, un plan de coupe est introduit comme décrit dans [HTA07]. L'idée est de séparer la scène en deux, d'une part on utilisera les profondeurs correspondant au plan de la route dans la partie basse de l'image, les profondeurs dans la partie haute seront prises dans un plan vertical. La ligne séparant la route au plan de coupe est notée c . En conséquence, seulement les distances lointaines sont coupées. Finalement, le modèle géométrique $p_c(\{v, u\})$ d'un pixel à la position $\{v, u\}$ est :

$$p_c(\{v, u\}) = \begin{cases} \frac{\delta}{(v - v_h)} & \text{si } v > c \\ \frac{\delta}{(c - v_h)} & \text{si } v \leq c \end{cases} \quad (2.8)$$

Un exemple de restauration avec l'hypothèse monde plan est montré dans la quatrième colonne de la figure 2.4. Nous pouvons voir, d'après ces résultats, que seule la partie correspondant à la route est bien restaurée, l'intensité dans le reste de l'image tend vers le noir dès lors que la profondeur est surestimée.

2.3.2.2 Méthode *free-space segmentation* (FSS)

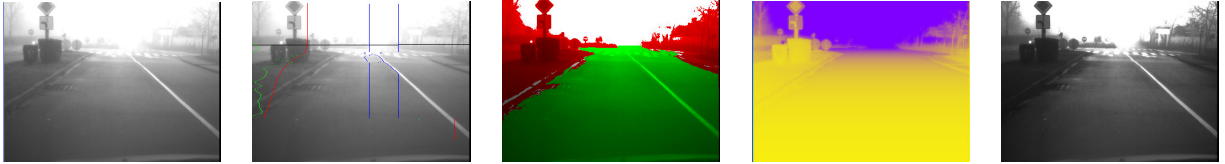


FIGURE 2.5 – De gauche à droite, les étapes de la restauration avec l'algorithme FSS : l'image d'origine, la détection de brouillard avec la méthode du point d'inflexion décrite dans la partie 2.4.1, la segmentation d'objets verticaux (en rouge) et la partie d'espace libre (en vert), estimation relative de la carte de profondeur de scène, la restauration obtenue.

Pour pouvoir améliorer la visibilité dans le reste de la scène, il est nécessaire d'estimer la profondeur $p(s)$ de chaque pixel. Dans [HTA07], un modèle 3D paramétré de la route a été proposé avec une réduction du nombre de paramètres géométriques. Même si ces modèles sont pertinents pour la plupart des scènes de route et même si les paramètres du modèle sélectionné sont optimisés pour obtenir la meilleure restauration, le modèle proposé n'est pas suffisamment générique pour prendre en compte toutes les configurations de l'environnement routier. Dans [HTA10], un schéma différent est proposé. De nouveau, on suppose que la route est plane avec un plan de coupe comme dans la méthode précédente. Quand la contrainte planaire est supposée (2.8), le contraste des objets appartenant au plan de la route est correctement restauré. En revanche, le contraste vertical des objets de la scène (véhicules, arbres, ...) n'est pas correctement restauré du fait que leur profondeur dans la scène est largement surestimée. En conséquence, les intensités restaurées ayant une valeur négative en utilisant la méthode du monde plan décrite précédemment dans la partie 2.3.2.1 sont mises à zéro dans l'image améliorée. Ceux-ci sont nommés les

pixels noirs. L'ensemble des pixels noirs donne une segmentation de l'image en deux régions, l'une à l'intérieur du plan de la route l'autre à l'extérieur. Cela permet donc d'en déduire une zone segmentée sans obstacle au niveau du plan de la route, comme illustré en vert et rouge dans la figure 2.5.

Pour chaque pixel dans la partie correspondant à la route, le modèle de route plane (2.5) s'applique correctement. Pour les pixels en dehors de la route plane (région rouge dans la troisième image de la figure. 2.5), il est proposé, dans [HTA10], d'utiliser le modèle géométrique (2.8) et, pour chaque pixel, de rechercher la plus petite valeur de c qui conduit à une intensité positive dans l'image restaurée. Les valeurs obtenues sont notées $c_{min}(s)$. En effet, lorsque c est proche de v_h , le plan de coupe est éloigné de la caméra et la restauration est seulement légèrement améliorée. Plus la valeur de c est importante, plus le plan de coupe est proche de la caméra, et donc accentue la restauration.

Chaque valeur $c_{min}(s)$ peut être associée à une distance $d_{min}(s)$ en utilisant (2.8). La carte de profondeur résultante de l'image avec brouillard est affichée dans la figure 2.5. Puis, une estimation approximative de la profondeur $p(s)$ est obtenue en prenant un facteur τ de la distance $d_{min}(s)$.

Le pourcentage τ spécifie l'importance de la restauration qui est généralement de 95% avec cette méthode. La carte de profondeur est utilisée pour améliorer le contraste de l'image entière en inversant la loi de Koschmieder comme le montre la quatrième image de la figure 2.5. L'algorithme est détaillé dans [HTA10, HTA09] et des résultats complémentaires sont présentés dans la sixième colonne de la figure 2.4.

2.3.2.3 La contrainte *no-black-pixel* (NBPC)

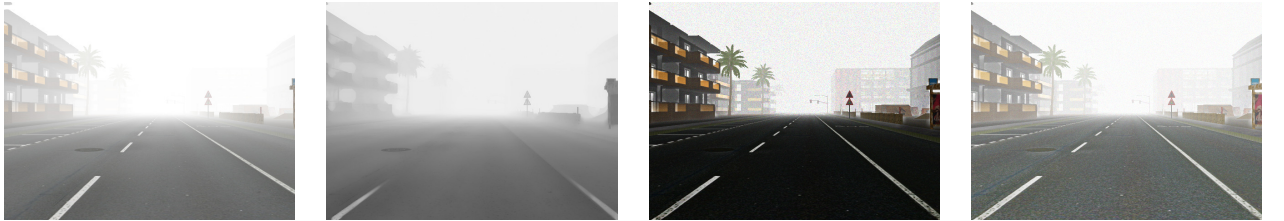


FIGURE 2.6 – Étape de l'algorithme avec la contrainte *no-black-pixel*. De gauche à droite, l'image originale, le voile estimé avec un coefficient $\tau = 0.95$, l'image restaurée intermédiaire grâce au voile estimé, l'image restaurée après l'opération de « tone mapping ».

Dans [TH09], un algorithme basé sur une régularisation locale de l'image est proposé. La distance $p(s)$ étant inconnue, le problème de la restauration monoculaire d'une seule image peut être alors assimilé à celui de l'estimation du voile atmosphérique $V(s) = I_s(1 - e^{-\beta p(s)})$.

La plupart du temps, l'intensité du ciel I_s correspond au maximum de l'intensité de l'image, donc I_s peut être affectée à 1 en supposant que l'image d'entrée est normalisée. Après substitution de V dans (2.1) et avec $I_s = 1$, la loi de Koschmieder est réécrite comme suit :

$$I(s) = I_0(s)(1 - V(s)) + V(s) \quad (2.9)$$

L'intensité de l'image restaurée $R(s)$ est estimée en inversant (2.9) :

$$R(s) = \frac{I(s) - V(s)}{1 - V(s)} \quad (2.10)$$

Le voile atmosphérique $V(s)$ étant inconnu, nous pouvons établir plusieurs contraintes s'appliquant à $V(s)$. Premièrement, $V(s)$ doit être supérieur ou égal à zéro et inférieur à $I(s)$:

$$0 \leq V(s) \leq I(s) \quad (2.11)$$

Ce sont les contraintes photométriques introduites dans [Tan08]. Une nouvelle contrainte qui a pour but de réduire le nombre de pixels noirs dans l'image restaurée R est introduite dans [TH09]. Cette contrainte est appelée la contrainte *no-black-pixel* et stipule que l'écart type autour d'une position de pixel donné doit être inférieur à sa moyenne locale :

$$f \, std(R) \leq \bar{R} \quad (2.12)$$

où f est un facteur de valeur généralement positionné à 1. Dans le cas d'une distribution gaussienne de l'intensité et $f = 1$, ces critères impliquent que 15.8% des intensités deviennent noires. L'utilisation de $f = 2$ conduit à renforcer le critère avec seulement 2.2% des intensités qui deviennent noires.

La difficulté de cette dernière contrainte est qu'elle est fonction du résultat inconnu R . Grâce à la linéarité de (2.10), la contrainte *no-black-pixel* peut être transformée en une contrainte impliquant seulement V et I . Pour cela, une régularisation spatiale locale est appliquée en supposant qu'autour de la position s d'un pixel, la profondeur de la scène est constante et le brouillard est homogène, ce qui équivaut à dire que le voile atmosphérique est localement égal à $V(s)$ à la position s . Compte tenu de cette hypothèse, en utilisant (2.10) nous obtenons que les moyennes locales $\bar{I}$ et $\bar{R}$ sont liées par $\bar{R} = \frac{\bar{I} - V(s)}{1 - V(s)}$ et les écarts types $std(I)$ et $std(R)$ sont liés par $std(R) = \frac{std(I)}{1 - V(s)}$. On obtient donc, après substitution des deux résultats précédents, dans (2.12), la contrainte *no-black-pixel* réécrite comme une fonction de $V(s)$ et de I :

$$V(s) \leq \bar{I} - f \, std(I) \quad (2.13)$$

Le voile atmosphérique $V(s)$ est défini comme un pourcentage τ du minimum entre les deux bornes supérieures (2.11) et (2.13) :

$$V(s) = \tau \, \min(I(s), \bar{I} - f \, std(I)) \quad (2.14)$$

Le pourcentage τ spécifie la force de la restauration. Pour cette méthode il est généralement fixé aux environs de 95%. L'image améliorée est obtenue en appliquant la formule (2.10) en utilisant le voile V précédent. La figure 2.7 montre une image restaurée avec différentes valeurs de τ .

L'algorithme déduit à partir des contraintes photométriques et *no-black-pixel* s'avère être celui décrit dans [TH09] où $\bar{I}$ est obtenue en tant que valeur médiane des intensités locales dans une fenêtre de taille s_v et l'écart type comme la médiane des différences absolues entre les intensités avec $\bar{I}$ utilisant la même taille de fenêtre. D'autres filtres qui respectent les contours peuvent aussi être utilisés tels que la médiane de la médiane le long des lignes [TH09] ou le filtre bilatéral. Du fait du lissage des contours composés de structures assez complexes, de petits artefacts sont produits dans l'image restaurée autour des discontinuités de profondeurs tels que des silhouettes complexes dont les arbres. Un post-traitement avec le filtre bilatéral croisé / joint sur V utilisant I comme un guide peut être appliqué pour nettoyer ces artefacts comme proposé dans [YXL10].

Cet algorithme de restauration est présenté avec une image en niveaux de gris, mais il peut être étendu facilement à des images en couleur ($r(s), g(s), b(s)$) en substituant $I(s)$ dans l'équation précédente par $I(s) = \min(r(s), g(s), b(s))$ après une balance correcte des blancs.

La figure 2.4 montre la restauration obtenue avec l'algorithme NBPC dans la septième colonne. Notons que le contraste de la texture de la partie de la route de l'image améliorée est trop accentué. Ceci est dû au fait que le voile atmosphérique $V(s)$ de la partie route de l'image est surestimé. Ceci est une conséquence de la propriété de localité de l'algorithme NBPC. Comme détaillé dans [TH09], une correction finale peut être utilisée pour atténuer ce problème.



FIGURE 2.7 – Restauration intermédiaire de la méthode *no-black-pixel* avec différentes valeurs de τ .

2.3.2.4 Méthode *Dark Channel Prior* (DCP)

Un algorithme pour la restauration locale appelé *Dark Channel Prior* (DCP) a été proposé dans [HST10b]. Pour les images en niveaux de gris, l'algorithme DCP consiste d'abord à appliquer une érosion morphologique par un élément de taille s_v , qui supprime tous les objets blancs avec une taille plus petite que s_v . Ensuite, le voile atmosphérique $V(s)$ est défini par un pourcentage τ du résultat de l'érosion. Ainsi, cette première étape peut être considérée comme un cas particulier de l'algorithme NBPC utilisant des opérateurs morphologiques comme filtres avec $f = 0$. De façon identique à ce qui a été expliqué dans la partie précédente, une érosion ou une ouverture ne conservent pas les frontières complexes précises le long des discontinuités dues aux profondeurs. Dans [HST10b], un affinement est utilisé pour restaurer les délimitations complexes V . Une alternative plus rapide consiste à utiliser des itérations du filtre guidé, comme proposé dans [HST10a]. Le filtre bilatéral croisé / joint est une autre alternative. L'image améliorée est obtenue en appliquant l'inverse de la loi de Koschmieder (2.10) en utilisant le voile V précédemment calculé.

La figure 2.4 montre la restauration obtenue par l'algorithme DCP dans la cinquième colonne. Comme dans l'algorithme NBPC, les images en couleurs sont traitées en prenant l'intensité minimale sur les 3 canaux $I(s) = \min(r(s), g(s), b(s))$.

2.3.2.5 Contrainte *no-black-pixel* et du monde plan (NBPC+PA)

D'une part, la restauration obtenue avec la méthode FSS, comme expliqué dans la partie 2.3.2.2, effectue une segmentation pour diviser l'image en trois parties : le ciel, les objets hors du plan de la route et l'espace dans le plan de la route. Différents procédés de restauration sont effectués en fonction de la partie considérée. La difficulté avec une approche basée sur la segmentation est de gérer correctement la transition entre les parties.

D'autre part, la restauration avec NBPC et DCP sont des méthodes locales non dédiées aux images de route et qui présentent des difficultés pour la grande zone grise de route [THC⁺10]. En effet, le voile atmosphérique dans la partie inférieure de l'image est surestimé.

Pour combiner les avantages des deux approches, la contrainte planaire PA est ajoutée à la contrainte *no-black-pixel* ce qui empêche la surestimation du voile dans la partie inférieure de l'image en tenant compte de la distance réduite entre la caméra et la route. En pratique, il est très rare d'observer le brouillard avec une distance de visibilité météorologique d_m inférieure à $60m$. En supposant que la distance minimale de visibilité météorologique est de soixante mètres, soit $d_m \geq 60$, nous déduisons $\beta \leq -\frac{\ln(0.05)}{60}$. La supposons que la route est un plan jusqu'à une certaine distance peut être faite, et que le calibrage de la caméra est connu par rapport à la route, c'est à dire que δ et v_h sont connus. Ainsi, en utilisant le dernier terme de l'équation (2.6), le voile atmosphérique est soumis à la troisième contrainte suivante :

$$V(\{v, u\}) \leq I_s(1 - e^{\frac{\ln(0.05)\delta}{d_{min}(v-v_h)}}) \quad (2.15)$$

où d_{min} peut être réglée par exemple à la distance minimale $60m$. La contrainte (2.15) est appelée la contrainte planaire. Comme dans l'algorithme NBPC, le voile atmosphérique $V(s)$ est fixé à un pourcentage τ du minimum par rapport aux trois bornes supérieures :

$$V(\{v, u\}) = \tau \min(I(\{v, u\}), \bar{I} - f \text{std}(I), I_s(1 - e^{\frac{\ln(0.05)\delta}{d_{min}(v-v_h)}})) \quad (2.16)$$

L'image restaurée résulte dans l'application de (2.10). En présence de brouillard avec une distance de visibilité météorologique inférieure à $d_{min} = 60m$, cette troisième contrainte limite les possibilités d'amélioration qui seront partielles pour de courtes distances, même avec $\tau = 100\%$. Une conséquence intéressante de l'introduction de la troisième contrainte est que la correction gamma finale utilisée dans les algorithmes NBPC et DCP ne sont plus nécessaires pour atténuer l'assombrissement de l'image, comme illustré dans la huitième colonne de la figure 2.4.

Plutôt que de fixer $d_{min} = 60m$, une autre approche consisterait à exécuter un algorithme pouvant calculer la densité de brouillard et ainsi estimer β comme il est expliqué dans la partie 2.4.1. β peut alors être utilisé dans (2.15) au lieu de $-\frac{\ln(0.05)}{60}$.

2.3.3 Égalisation locale de l'histogramme (HE+PA)

La méthode présentée dans [HCH11] propose d'étendre localement l'histogramme de l'intensité de l'image sur une fenêtre autour de chaque pixel. Une fois cette intensité calculée, la profondeur, grâce à l'équation de Koschmieder, est calculée et fusionnée avec celle de la chaussée. Afin d'éviter les artefacts dus à la représentation par fenêtre locale, la carte de profondeur est lissée, et ensuite, l'intensité finale est retrouvée en inversant une nouvelle fois l'équation de Koschmieder. Cette méthode est aussi compatible avec la contrainte planaire.

2.3.4 Interprétation bayésienne de la restauration

Nous avons vu dans le chapitre précédant que la reconstruction par paires stéréoscopiques peut être interprétée en tant que modèle probabiliste comme de nombreux autres problèmes. Dans [NKL12], une interprétation probabiliste de la restauration d'image couleur a été proposée.

Pour cela, il est proposé tout d'abord d'effectuer une transformation algébrique du modèle défini par l'équation de Koschmieder (2.1) en un site s de la façon suivante :

$$\ln(1 - \frac{I(s)}{I_s}) = \ln(1 - I_0(s)) - \tilde{d}(s) \quad (2.17)$$

$$\tilde{I}(s) = \mathcal{G}(s) + \mathcal{D}(s) \quad (2.18)$$

où $\tilde{d} = \beta d$. Ici, cette forme permet de séparer explicitement le terme avec l'image restaurée $\mathcal{G}(s) = \ln(1 - I_0(s))$ de la profondeur $\mathcal{D}(s) = -d$. Finalement, la probabilité *a posteriori* est écrite en fonction des variables $\mathcal{G}$, $\mathcal{D}$ et $\tilde{I}$.

$$P(\mathcal{G}, \mathcal{D} | \tilde{I}) \propto P(\tilde{I} | \mathcal{G}, \mathcal{D}) P(\mathcal{G}, \mathcal{D}) \quad (2.19)$$

$$\propto P(\tilde{I} | \mathcal{G}, \mathcal{D}) P(\mathcal{G}) P(\mathcal{D}) \quad (2.20)$$

Dans [NKL12], la probabilité de chaque champ $\mathcal{G}$ et $\mathcal{D}$ est supposée statistiquement indépendante, la probabilité *a posteriori* peut donc être décomposée.

Terme de vraisemblance : Afin de lier l'image en présence de brouillard, la profondeur et l'image restaurée, une loi normale est utilisée avec $\tilde{I}$ comme moyenne.

$$P(\tilde{I} | \mathcal{G}, \mathcal{D}) = \prod_{s \in S} \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(\mathcal{G}(s) + \mathcal{D}(s) - \tilde{I}(s))^2}{2\sigma^2}\right) \quad (2.21)$$

A priori : Deux *a priori* sont ensuite introduits sur la distribution locale de la profondeur et de l'image restaurée. Ces deux *a priori* tendent donc à forcer les variables voisines à des valeurs proches comme vu précédemment dans le cas de la reconstruction par paires stéréoscopiques pour la profondeur. Les *a priori* seront donc de la forme :

$$P(\mathcal{D}) = \frac{1}{Z_{\uparrow}} \exp\left(-\sum_{c \in C_{\mathcal{D}}} \rho_{\mathcal{D}}(\varphi(c, \mathcal{D}))\right) \quad (2.22)$$

$$P(\mathcal{G}) = \frac{1}{Z_{\downarrow}} \exp\left(-\sum_{c \in C_{\mathcal{G}}} \rho_{\mathcal{G}}(\varphi(c, \mathcal{G}))\right) \quad (2.23)$$

$C_{\mathcal{G}}$ et $C_{\mathcal{D}}$ sont respectivement les cliques entre les intensités originales de la scène et les profondeurs relatives. Comme dans le chapitre précédent, $\varphi_{\mathcal{D}}$ et $\varphi_{\mathcal{G}}$ sont deux fonctions favorisant les solutions ayant des valeurs proches dans un voisinage et Z_d et Z_g les constantes de normalisation de la densité de probabilité.

Finalement, en prenant le logarithme du négatif de la probabilité *a posteriori*, l'énergie à minimiser est :

$$\sum_{s \in S} \frac{(\mathcal{G}(s) + \mathcal{D}(s) - \tilde{I}(s))^2}{2\sigma^2} + \sum_{c \in C_{\mathcal{D}}} \rho_{\mathcal{D}}(\varphi(c, \mathcal{D})) + \sum_{c \in C_{\mathcal{G}}} \rho_{\mathcal{G}}(\varphi(c, \mathcal{G})) \quad (2.24)$$

Optimisation : Trouver quelles sont les profondeurs $\mathcal{D}$ et les intensités originales $\mathcal{G}$ qui minimisent l'énergie (2.24) n'est pas possible en un temps polynomial du fait de la combinatoire des solutions entre l'image restaurée et la carte de profondeur relative. Une possibilité pour optimiser cette énergie est de fixer un des deux ensembles de labels (la carte de profondeur relative ou l'image restaurée) et d'optimiser par rapport au second. De cette façon, on retombe sur une structure *classique* de champ de Markov où le terme de vraisemblance est une clique de taille 1 avec un terme de régularisation. Afin de permettre d'explorer un nombre important de solutions, l'écart type σ peut aussi être optimisé. L'algorithme d'optimisation est donc le suivant : chaque champ est optimisé en alternance. Une fois que l'énergie ne diminue plus, l'écart type σ est optimisé. Ces deux étapes sont alternées jusqu'à convergence du système. Ce procédé permet d'obtenir des résultats satisfaisants tout en proposant une approche markovienne intéressante. Mais cela ne s'applique qu'à des images en couleur.

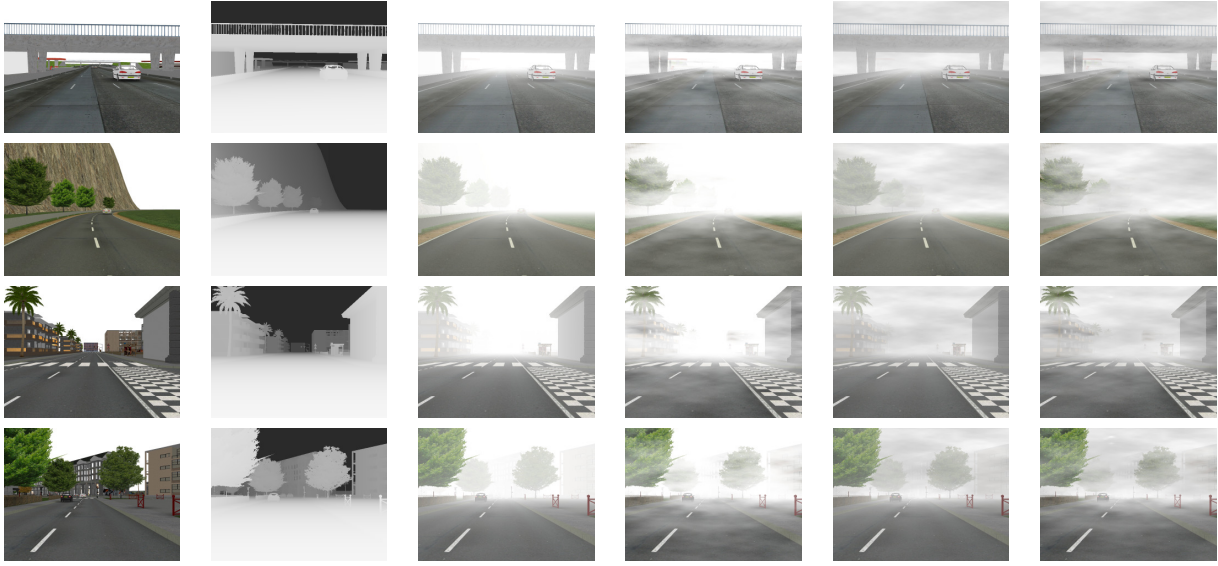


FIGURE 2.8 – Base de données d’images de synthèse en environnement routier. La première colonne représente l’image de synthèse sans brouillard. La deuxième colonne représente la carte de profondeur associée à la scène. À partir de la troisième colonne, l’image originale avec l’ajout de différents types de brouillard artificiels. De gauche à droite : avec un brouillard homogène, avec un brouillard hétérogène où β varie, avec un brouillard hétérogène où L_s varie, avec un brouillard hétérogène où β et L_s varient.

2.3.5 Évaluation

L’évaluation des algorithmes de restauration nécessite des images de la même scène avec et sans brouillard. Il s’avère que l’obtention de ces paires d’images est extrêmement difficile en pratique car elle nécessite de vérifier que les conditions d’éclairage sont identiques dans la scène avec et sans brouillard. En conséquence, pour l’évaluation de l’algorithme de restauration proposé et sa comparaison par rapport aux algorithmes existants, deux ensembles d’images sans brouillard et avec brouillard synthétique ont été créés pour [THC⁺12] à partir de 66 scènes synthétiques.

2.3.5.1 Brouillard de synthèse

Pour [THC⁺12], 66 images de synthèse ont été générées en utilisant SiVICTM, un logiciel qui permet de construire des environnements routiers permettant de générer un véhicule en mouvement avec un modèle de conduite physique axé sur un comportement dynamique [GRdL⁺06], et des capteurs virtuels embarqués. Pour chaque vue, la carte de profondeur effective est également calculée. En effet, la carte de profondeur est nécessaire pour permettre l’ajout du brouillard dans les images. Des images de synthèse ont été produites à partir de la base de données d’images, en utilisant 4 différents types de brouillard :

- *Brouillard uniforme* : La loi de Koschmieder (2.1) est appliquée avec une distance de visibilité météorologique de 80m.
- *Brouillard hétérogène β* : Avec une variation spatiale du coefficient d’extinction.
- *Brouillard hétérogène L_s* : Avec une variation spatiale de l’intensité du ciel.

Algorithmes	Uniforme	Variable β	Variable L_s	Variable $\beta \& L_s$	Tous les types
Image	81.6 ± 12.3	78.7 ± 12.3	69.0 ± 10.9	66.4 ± 10.8	73.9 ± 13.2
MSR [JRW97]	46.7 ± 16.3	86.4 ± 24.7	44.8 ± 17.1	83.7 ± 24.9	65.4 ± 28.9
CLAHE [Zui94]	66.9 ± 10.7	64.5 ± 9.7	54.5 ± 8.5	54.6 ± 7.8	60.1 ± 10.9
DCP [HST10b]	46.3 ± 15.6	46.9 ± 17.0	43.7 ± 16.2	44.1 ± 17.5	45.2 ± 16.7
FSS [HTA10]	34.9 ± 15.1	40.9 ± 13.5	32.5 ± 11.4	36.5 ± 10.3	36.3 ± 13.1
NBPC [TH09]	50.8 ± 11.5	50.5 ± 11.5	38.5 ± 9.0	38.0 ± 8.7	44.5 ± 12.1
NBPC+PA [THC ⁺ 10]	31.1 ± 10.2	36.0 ± 10.3	26.7 ± 5.1	28.4 ± 5.9	30.6 ± 8.9
CM+PA [HCH11]	29.6 ± 11.3	24.0 ± 7.7	25.7 ± 4.4	25.2 ± 6.2	26.1 ± 8.1

TABLE 2.1 – Résultats dans [THC⁺12] des différences absolues moyennes entre images améliorées et les images cibles sans brouillard, pour les 6 algorithmes comparés, sur les 4 types de brouillard de synthèse (66 images pour chaque type) et pour l'ensemble de la base de données (264 images) dans la dernière colonne.

— *Brouillard hétérogène β et L_s* : Avec un mélange entre les deux précédentes variations. Cette base est appelée FRIDA2. Des exemples d'image générés à partir de la base avec les 4 différents types de brouillard sont montrés dans la figure 2.8.

2.3.5.2 Comparaison sur des images de synthèse

Dans [THC⁺12], chaque algorithme a été appliqué sur les 4 types de brouillard. Les algorithmes testés sont : *multi retinex* (MSR), *contrast-limited adaptive histogram equalization* (CLAHE), *dark-channel-prior* (DCP), *free-space segmentation* (FSS), *no-black-pixel* (NBPC), la restauration avec la contrainte *no-black-pixel* et scène plane (NBPC+PA). Par la suite, l'algorithme basé sur égalisation locale de l'histogramme (HE+PA) a été évalué sur la base de test.

Les résultats sur 11 images de brouillard homogène et hétérogène sont présentés dans la figure 2.9. Notons l'augmentation significative du contraste pour les objets plus éloignés : certains objets qui étaient à peine visibles dans l'image avec brouillard apparaissent clairement dans les images restaurées. Une première analyse confirme que, MSR et CLAHE ne sont pas adaptés pour les images avec brouillard. En effet, pour le premier, l'image est plus sombre mais le contraste n'est pas augmenté. Pour le second, les contrastes sont trop augmentés dans les zones proches mais pas suffisamment au loin. Ensuite, malgré une bonne restauration sur une grande partie de la scène, les objets éloignés sont encore trop brumeux avec l'algorithme DCP. Avec l'algorithme FSS, les objets verticaux apparaissent trop assombris. Pour la méthode NBPC, comme souligné précédemment, la chaussée est trop restaurée. Finalement, NBPC+PA se présente comme un compromis satisfaisant.

La comparaison quantitative consiste à calculer la différence absolue entre l'image sans brouillard et l'image obtenue après restauration. Les résultats, en moyenne sur les 66 images, le nombre de pixels de l'image et le nombre de couleurs composantes de l'image, sont présentés dans la table. 2.1. Dans cette moyenne, des pixels dans le ciel, dans l'image d'origine, sont rejetés pour ne pas fausser les résultats. En effet, l'intensité du ciel ne peut être restaurée comme un blanc constant quand L_s est hétérogène. Cette méthode de mesure est appropriée pour les applications de véhicules intelligents, mais il ne l'est probablement pas dans d'autres domaines tel que le traitement de photographies.

L'algorithme *multiscale retinex* (MSR) n'est pas un algorithme de restauration dédié à

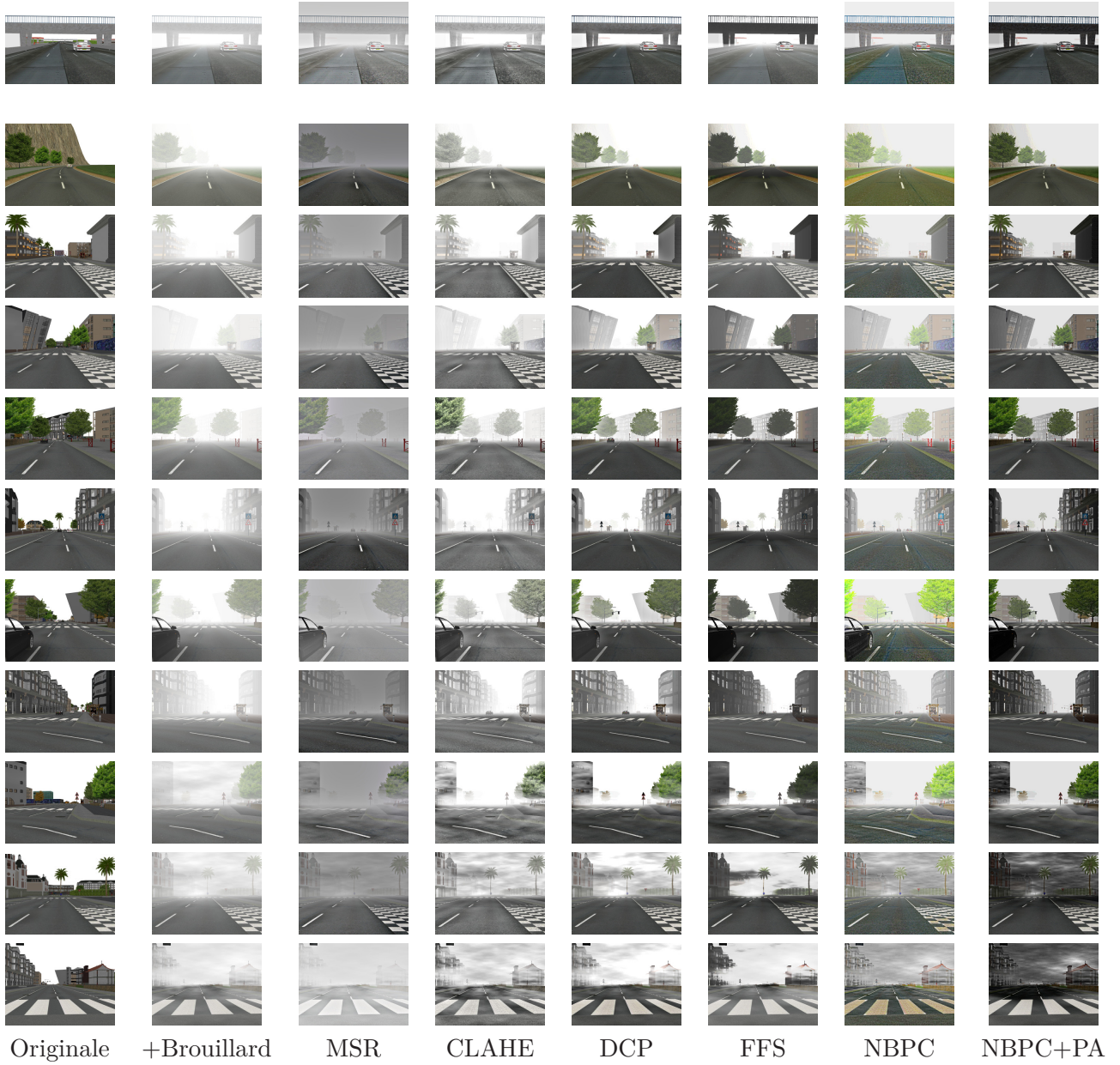


FIGURE 2.9 – Résultats de restauration sur des images de synthèse. De gauche à droite, l'image originale sans brouillard, l'image avec différents types de brouillard ajoutés, l'image restaurée avec l'algorithme *multiscale retinex*, *adaptive histogram equalization*, *dark channel prior*, *free-space segmentation*, *no-black-pixel constraint* et *no-black-pixel constraint* avec le monde plan.

la scène avec différentes profondeurs de l'objet. La différence moyenne est diminuée pour le brouillard homogène et le brouillard hétérogène en L_s , comparativement à l'absence de traitement. Fait intéressant, lorsque β est hétérogène, l'action du *multiscale retinex* empire la situation. Cela est dû au fait que MSR augmente certains contrastes correspondant au brouillard et non à la scène.

Comparativement à l'absence de traitement, la différence moyenne est toujours améliorée lorsqu'on utilise l'égalisation d'histogramme adaptative (CLAHE). Néanmoins, ce n'est pas un algorithme de restauration basé sur la loi de Koschmieder (2.1) et donc l'amélioration est faible. À titre d'illustration, avec CLAHE on obtient des résultats plus mauvais que ceux du *multiscale retinex* avec brouillard homogène et brouillard hétérogène en L_s .

L'algorithme *dark channel prior* (DCP) et la contrainte *no-black-pixel* (NBPC) atteignent des performances similaires en moyenne sur l'ensemble de la base. Néanmoins, nous remarquons qu'avec l'algorithme NBPC, l'image est mieux restaurée pour de grandes distances.

Avec le brouillard homogène, la restauration avec l'algorithme *free-space segmentation* (FSS), celui avec la contrainte *no-black-pixel* (NBPC+PA) et celui basé sur égalisation locale de l'histogramme (HE+PA) tous deux combinés avec l'hypothèse de la scène plane donnent les meilleurs résultats. Un deuxième groupe d'algorithmes avec des performances similaires pour les images de brouillard homogène sont les méthodes *dark channel prior* (DCP), avec la contrainte *no-black-pixel* (NBPC) et l'algorithme *multiscale retinex* (MSR). Ces trois derniers algorithmes sont moins efficaces que les deux premiers en raison de la difficulté à rétablir l'intensité moyenne correcte de la partie route de l'image. NBPC+PA garde les bons résultats de la méthode NBPC pour de grandes distances sans pour autant ajouter des distorsions de contraste sur la partie de la route de l'image grâce à la combinaison avec l'hypothèse plane.

Pour les trois types de brouillard hétérogène, la méthode basée sur l'égalisation locale de l'histogramme (HE+PA) avec NBPC+PA conduit à de meilleurs résultats par rapport à FSS. Cela peut s'expliquer par le fait que l'algorithme FSS s'appuie fortement sur l'hypothèse que β et L_s sont constants sur toute l'image, alors que l'algorithme NBPC+PA et HE+PA supposent uniquement que β et L_s sont localement constants dans l'image. Donc, la plupart du temps, il est plus efficace avec le brouillard hétérogène comparativement aux autres algorithmes.

2.3.6 Conclusion

Nous avons vu dans cette première partie plusieurs algorithmes proposés afin de restaurer les contrastes de l'image sans brouillard. Nous avons constaté que la restauration monoculaire est un problème mal posé du fait de l'ambiguïté entre la distance et l'intensité sans brouillard de la scène pour un même pixel.

Une première idée a donc été d'utiliser des méthodes d'amélioration de contraste d'image déjà existantes pour des applications tierces. Nous avons vu que les résultats ne sont pas satisfaisants dans les scènes en environnement routier du fait des grands écarts de profondeurs.

Pour prendre en compte la profondeur, le modèle de Koschmieder a été utilisé afin de modéliser le lien entre l'image en présence de brouillard, l'image sans brouillard et la profondeur. Il a fallu ajouter des *a priori* et des contraintes sur les deux variables à estimer. Une des premières idées est de choisir une structure 3D approximative de la scène afin qu'il n'y ait plus que l'image restaurée à estimer. Cette méthode est très limitée dans le cadre des applications embarquées pour lesquelles il est difficile d'avoir une bonne estimation tout au long d'un parcours.

Afin de ne pas biaiser le résultat en cas de mauvaise approximation de la profondeur, et plutôt que de fixer une profondeur *a priori*, des contraintes ont été ajoutées sur l'espace des solutions recherchées. C'est dans ce cadre que la méthode estimant le voile atmosphérique a

été proposée. Cette méthode calcule un premier voile soumis à deux contraintes, et par la suite calcule l'image restaurée. Si l'image est en couleur, une meilleure estimation du voile est obtenue.

Néanmoins, la route grise est source d'ambiguïté. La contrainte d'un monde plan a été introduite pour pallier ce problème. Finalement, l'évaluation des différentes méthodes sur des images de synthèse dans [THC⁺12] montre que les méthodes avec la prise en compte de la profondeur avec le modèle de Koschmieder apportent les meilleurs résultats.

2.4 Estimation du coefficient d'extinction

Nous avons vu que le problème de la restauration est mal posé. La restauration d'image peut donc dégrader le contraste dans certains cas. Il est donc important de n'utiliser ces algorithmes qu'en présence de brouillard. En effet, un algorithme utilisant les images produites par un algorithme de restauration peut donner de moins bons résultats en conditions de temps dégagé du fait de l'apparition de certains artefacts par les méthodes de restaurations. Cela implique qu'il est nécessaire, avant d'effectuer un traitement, de caractériser la scène, et de savoir s'il y a du brouillard, et en cas de présence, d'en connaître sa densité.

De plus, l'estimation du coefficient d'extinction peut être utile pour les algorithmes des restaurations. Nous avons vu que la méthode [HTA10] nécessite de connaître la densité du brouillard afin de pouvoir calculer le voile relatif à la route, il est aussi nécessaire dans la troisième contrainte pour les méthodes NBPC+PA. La connaissance du coefficient d'extinction est également utile pour d'autres applications. Par exemple, la puissance et la direction avec laquelle les feux de brouillard sont activés dépendent directement de la densité de brouillard rencontré, et donc du coefficient d'extinction β (comme montré dans le projet ICADAC⁴)

Étonnamment, seulement un nombre réduit de méthodes a été proposé pour détecter le brouillard à partir de caméras et pour estimer le coefficient d'extinction β . La plupart d'entre elles s'appliquent quand la caméra est statique. La première approche a été d'estimer la diminution du contraste d'un objet lointain [BD98, HBD⁺11]. Dans [KNC⁺08], un modèle d'atténuation est estimé en supposant que le modèle 3D de la scène est connu. Cette méthode est intéressante car elle montre que la loi de Koschmieder n'est pas toujours valide sur une grande plage de distances du fait du brouillard hétérogène. La difficulté avec cette méthode est la nécessité de disposer d'un modèle 3D précis de la scène. Quand la caméra est en mouvement, seulement trois approches ont été proposées. La première est basée sur l'utilisation des marquages au sol, la seconde, en faisant l'hypothèse que la route est plane et de couleur homogène, et la dernière, en utilisant la stéréovision. Dans [Pom97], β est obtenue en regardant l'atténuation du contraste des marquages au sol à plusieurs distances. Le problème de cette méthode est qu'elle requière la présence de marquages au sol. Dans [LTA02, HTLA06], une méthode a été proposée appliquée à l'environnement routier, cette méthode fait l'hypothèse que la route est plane avec une couleur constante, et que les paramètres intrinsèques et extrinsèques de la caméra en fonction de la route sont connus. En présence de brouillard, le profil d'intensité de la route dans la direction verticale de l'image dérivée de l'équation (2.1) montre un point d'inflexion. La position de ce point d'inflexion peut être utilisée pour l'estimation de β . La faiblesse de cette méthode est que de nombreuses suppositions sont faites sur la structure de la scène. Un autre problème vient du fait que le point d'inflexion peut être caché par un objet proche, ou tout simplement que le point d'inflexion est trop loin dans le cas d'un brouillard très léger. Dans [HLA06], une méthode basée sur la stéréovision est proposée. Elle consiste en la détection de l'objet le plus lointain avec un

4. Projet ICADAC (6866C0210)

contraste supérieur à 5%. Cette méthode est simple mais elle requiert la présence d'un objet, avec un contraste important à chaque distance, ce qui n'est pas courant en pratique.

2.4.1 L'algorithme du point d'inflexion

L'algorithme proposé dans [LTA02] et [HTLA06] est capable de détecter la présence de brouillard et d'estimer la distance de visibilité directement liée au β de la loi Koschmieder (2.1).

Cet algorithme, connu aussi comme l'algorithme du point d'inflexion, repose principalement sur trois hypothèses : le brouillard est homogène, la majeure partie de l'image montre la surface de la route qui est plane et d'intensité homogène.

En faisant l'hypothèse d'une route avec des propriétés photométriques homogènes (I_0 est constant), le brouillard peut être détecté et le coefficient d'extinction de l'atmosphère β peut être estimé avec la loi de Koschmieder (2.1) si le plan de la route est connu.

En dérivant deux fois I dans l'équation (2.6) par rapport à v , on obtient :

$$\frac{d^2 I}{dv^2}(\{v, u\}) = \beta \frac{\delta(I_0 - I_s)}{(v - v_h)^3} e^{-\beta \frac{\delta}{v - v_h}} \left(\frac{\beta \delta}{v - v_h} - 2 \right) \quad (2.25)$$

L'équation $\frac{d^2 I}{dv^2} = 0$ admet deux solutions. La solution $\beta = 0$ est sans intérêt. La seconde solution qui permet d'estimer le coefficient d'extinction est $\beta = \frac{2(v_i - v_h)}{\delta}$, où v_i désigne la position du point d'inflexion de $I(v)$.

2.4.2 Conclusion

Nous avons vu que les méthodes basées sur la loi de Koschmieder apportent des résultats satisfaisants. Certaines pour être exploitées pleinement nécessitent la connaissance du coefficient d'extinction. De plus, afin de pouvoir caractériser la présence de brouillard ou non, connaître le coefficient d'extinction s'avère nécessaire. Pour détecter le brouillard et estimer β , il existe plusieurs méthodes avec une caméra fixe, mais elles ne sont pas utilisables pour les applications embarquées. Une méthode fondée sur l'hypothèse du monde plan a été décrite. Elle donne des résultats satisfaisants quand une large portion de la route est visible.

Chapitre 3

Technique d'optimisation d'énergie discrète

La théorie d'optimisation de fonctions pseudo-bouliennes est présentée dans le cadre de l'optimisation de fonctions multi-labels.

En traitement d'images, de nombreux problèmes conduisent à la recherche d'un optimum global d'une énergie multi-labels. La solution recherchée est alors représentée par un ensemble de labels (entiers). Soit $L = \{0, \dots, m\}$ l'ensemble des labels (par exemple $\{0, 1\}$ pour une segmentation binaire, $\{0, \dots, 255\}$ pour la restauration d'une image 8 bits, $\{0, \dots, m\}$ pour un problème de reconstruction 3D à partir de paires stéréoscopiques avec m la disparité maximale admissible).

Soit n le nombre de variables, $f : L^n \rightarrow \mathbb{R}$ est une fonction déduite du modèle du problème à traiter, où chaque configuration $\mathbf{x}$ de L^n correspond un score :

$$f(\mathbf{x}) = \sum_{u \in C_1} \vartheta_u(L_u) + \sum_{u,v \in C_2} \vartheta_{uv}(L_u, L_v) + \sum_{u,v,w \in C_3} \vartheta_{uvw}(L_u, L_v, L_w) + \dots \quad (3.1)$$

où C_k est l'ensemble des cliques d'ordre k et ϑ la fonction de coût de la clique. Lorsque le modèle est bien construit, les solutions les plus probables ont les scores les plus faibles, on essayera donc de résoudre le problème de minimisation suivant :

$$\min_{\mathbf{x} \in L^n} f(\mathbf{x}) \quad (3.2)$$

La recherche des labels qui minimisent (3.1) est généralement complexe. Le développement de modèles nécessite la connaissance des techniques d'optimisation.

En optimisation il est primordial de savoir s'il est possible de trouver, et dans quels délais, le minimum global d'une fonction. Si on peut trouver le minimum global d'une énergie, il n'y a pas d'ambiguïté sur le résultat de l'optimisation, le modèle proposé est bien en adéquation avec le problème à résoudre. Inversement, lorsque seulement un minimum local est obtenu, on ne sait pas si c'est une conséquence de ne pas avoir atteint le minimum global de la fonction, ou si c'est le modèle qui est incorrect.

Lorsque le nombre de solutions n'est pas important, il est assez simple de trouver le minimum d'une fonction. Une recherche exhaustive sur l'ensemble des configurations est suffisante pour permettre de trouver celle qui minimise l'énergie. En traitement d'image, certaines méthodes

denses font l'hypothèse qu'il existe une variable par pixel et qu'il existe une interaction entre ces variables. Dans ce cas, la combinatoire des solutions devient très importante même sur une petite image. Par exemple, considérons le cas de l'optimisation d'un champ de Markov d'une image 512x512 avec une labélisation binaire et une cardinalité de clique 2 par voisinage, il existe 2^{262144} configurations possibles. Sachant que les images actuelles sont constituées de plusieurs millions de pixels et que la cardinalité de l'espace de recherche peut être beaucoup plus importante (256 pour du débruitage, ou encore plus pour la stéréo-vision avec une grande base), il est donc d'évidence impossible de parcourir l'ensemble des solutions. Des algorithmes spécialisés ont donc dû être conçus pour optimiser ce type de problème avec un grand espace de recherche.

En optimisation continue, lorsqu'une fonction est convexe, tous les minima locaux sont globaux. De façon semblable à l'optimisation continue, il existe en optimisation discrète une classe de fonctions qui permet, malgré une croissance exponentielle du nombre de solutions, de trouver le minimum global en un temps polynomial. Ce sont les fonctions dites *sous-modulaires* qu'il est important de distinguer des fonctions non sous-modulaires.

Toutefois, l'espace des fonctions sous-modulaires est assez restreint. De plus, au delà d'une certaine cardinalité des cliques, pouvoir dire d'une fonction qu'elle est sous-modulaire s'avère être un problème *NP-Difficile*. Il arrive aussi fréquemment que l'on veuille optimiser des fonctions non sous-modulaires, il est donc utile dans ce cas, d'avoir des méthodes spécifiques d'optimisation. Compte tenu de l'importance de l'optimisation globale dans différents champs d'application, de nombreux algorithmes ont vu le jour.

L'une des approches les plus efficaces pour optimiser (3.2) est celle fondée sur l'optimisation de fonctions pseudo-booléennes.

Optimisation de fonctions pseudo-booléennes : L'étude de l'optimisation des fonctions pseudo-booléennes date des années 50. Observées initialement dans la théorie des jeux, ce fut l'une des principales motivations de leur étude en recherche opérationnelle. Par la suite, la découverte de la présence récurrente de fonctions pseudo-booléennes dans un spectre très large d'applications et de domaines a fortement influé sur l'évolution et le gain d'intérêt de ce domaine.

Dans le cas binaire, on nomme $\mathbf{B} = \{0, 1\}$ l'ensemble des labels. Une fonction $f : \mathbf{B}^n \rightarrow \mathbb{R}$ est dite pseudo-booléenne. Elle peut être écrite de façon unique comme un polynôme de n variables de la façon suivante :

$$f(x_1, \dots, x_n) = c_0 + \sum_{i=1}^n c_i x_i + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j + \sum_{1 \leq i < j < k \leq n} c_{ijk} x_i x_j x_k + \dots \quad (3.3)$$

Dans le cas binaire, l'énergie (3.1) peut être alors réécrite comme une fonction pseudo-booléenne (Cette réduction est expliquée dans la partie 3.2.1).

La partie 3.1 introduit l'optimisation de fonctions pseudo-booléennes. Cette partie est une synthèse partielle de l'article [BH02] très complet sur l'état de l'art de l'optimisation de fonctions pseudo-booléennes. L'ensemble des théorèmes que nous présentons est limité à une partie de cet article. Certaines notions sont détaillées par des exemples complémentaires à ceux de l'article

La partie 3.2 présente des algorithmes pour l'optimisation de fonctions multi-labels fondées sur la théorie de l'optimisation de fonctions pseudo-booléennes utilisées dans cette thèse.

3.1 Optimisation de fonctions pseudo-booléennes

3.1.1 Définitions et notations

Dans la suite, on nomme $\mathbf{V} = \{1, 2, \dots, n\}$ l'ensemble des indices de 1 à n , où n est le nombre de variables. $\mathbf{x} = (x_1, \dots, x_n)$ est un vecteur binaire et par définition $\bar{x}_i = 1 - x_i$ le complément pour $i \in \mathbf{V}$. On note $\mathbf{L} = \{x_1, \bar{x}_1, \dots, x_n, \bar{x}_n\}$ l'ensemble de ces symboles.

3.1.2 Représentation

On dispose de deux représentations pour la même fonction pseudo-booléenne :

Forme multilinéaire polynomiale unique :

$$f(x_1, \dots, x_n) = \sum_{S \subseteq \mathbf{V}} c_S \prod_{j \in S} x_j \quad (3.4)$$

On appelle degré de f la taille du plus grand ensemble $S \subseteq \mathbf{V}$ pour lequel $c_S \neq 0$. Cette représentation est utile pour certaines propriétés de f .

Posiforme : Une fonction pseudo-booléenne peut être toujours représentée par une posiforme de la forme suivante :

$$\phi(x_1, \dots, x_n) = \sum_{T \subseteq \mathbf{L}} a_T \prod_{u \in T} u \quad (3.5)$$

où $a_T \geq 0$. Notons qu'on peut toujours représenter une fonction pseudo-booléenne par une posiforme. Une façon de procéder consiste à remplacer le premier élément x par $1 - \bar{x}$ dans chaque terme où $a_T < 0$. Il existe plusieurs posiformes pour une même fonction pseudo-booléenne. On notera $\mathcal{P}(f)$ la famille des posiformes représentant la même fonction f .

Exemple 3.1. ces deux posiformes :

$$\psi_1 = 5x_1 + 4\bar{x}_1\bar{x}_2x_3 + 7x_1x_2x_4 + 9x_3\bar{x}_4$$

$$\psi_2 = x_1 + 4x_1x_2 + 4x_1\bar{x}_2\bar{x}_3 + 7x_1x_2x_4 + 4\bar{x}_2x_3 + 9x_3\bar{x}_4$$

ont la même forme multilinéaire polynomiale unique :

$$g(x) = 5x_1 + 13x_3 - 4x_1x_3 - 4x_2x_3 - 9x_3x_4 + 4x_1x_2x_3 + 7x_1x_2x_4$$

Ces trois fonctions pseudo-booléennes vérifient la même table de vérité (voir table 3.1)

3.1.3 Propriétés des fonctions pseudo-booléennes

La structure de posiforme permet de vérifier certaines propriétés. En effet, pour une posiforme ne possédant aucun terme négatif, leur somme ne peut être inférieure à 0. Dans un sens, minimiser une posiforme équivaut à essayer d'annuler le plus de termes possible. Il en découle que minimiser une posiforme est équivalent au problème d'optimisation de satisfiabilité booléenne maximale *MAX-SAT* (Maximum Satisfiability problem), qui est *NP-complet*. Cette idée permet de détecter si une affectation partielle des variables binaires est « optimale » ([BH02], p.18).

$\mathbf{x}$				$g(\mathbf{x})$
0	0	0	0	0
0	0	0	1	0
0	0	1	0	13
0	0	1	1	4
0	1	0	0	0
0	1	0	1	0
0	1	1	0	9
0	1	1	1	0
1	0	0	0	5
1	0	0	1	5
1	0	1	0	14
1	0	1	1	5
1	1	0	0	5
1	1	0	1	12
1	1	1	0	14
1	1	1	1	12

TABLE 3.1 – Table de vérité de la fonction pseudo-booléenne de l'exemple 3.1

On appelle un vecteur binaire $\mathbf{y} \in \mathbf{B}^S$ correspondant à un sous ensemble $S \subseteq \mathbf{V}$ une affectation partielle. De plus, pour un sous ensemble $S \subseteq \mathbf{V}$ d'indices et un vecteur $\mathbf{y} \in \mathbf{B}^S$, on note par $\mathbf{x}[S] \in \mathbf{B}^S$ le vecteur correspondant aux indices dans S , i.e. $\mathbf{x}[S] = (x_i | i \in S)$. Pour une affectation partielle $\mathbf{y} \in \mathbf{B}^S$ et pour un vecteur $\mathbf{x} \in \mathbf{B}^n$, On définit l'échange de $\mathbf{x}$ comme le vecteur binaire $\mathbf{z}$ par :

$$z_j = \begin{cases} x_j & \text{si } j \notin S \\ y_j & \text{si } j \in S \end{cases} \quad (3.6)$$

et on le note par $\mathbf{z} = \mathbf{x}[S \leftarrow \mathbf{y}]$. Par exemple, si $n = 5$, $S = \{1, 2, 5\}$ et $\mathbf{y}$ est l'affectation partielle $y_1 = 1$, $y_2 = 0$ et $y_5 = 1$ alors l'affectation de $\mathbf{x} = (1, 1, 1, 0, 0)$ par $\mathbf{y}$ va être le vecteur $\mathbf{z} = (1, 0, 1, 0, 1)$.

Soit une fonction pseudo-booléenne f , un vecteur $\mathbf{x} \in \mathbf{B}^n$ et une affectation partielle $\mathbf{y} \in \mathbf{B}^n$ pour un sous ensemble $S \subseteq \mathbf{V}$, alors

Persistance forte : Il y a persistance forte pour f aux valeurs de $\mathbf{y}$ si pour tout $\mathbf{x} \in \text{Argmin}_{\mathbf{B}^n}(f)$ on a $\mathbf{x}[S] = \mathbf{y}$.

En d'autres termes, on dit qu'une affectation est persistante forte si l'ensemble des valeurs de l'affectation $\mathbf{y}$ appartient à tous les minima globaux de la fonction.

Persistance faible : Il y a persistance faible pour f aux valeurs de $\mathbf{y}$ si $\mathbf{x}[S \leftarrow \mathbf{y}] \in \text{Argmin}_{\mathbf{B}^n}(f)$ est vérifiée pour tout $\mathbf{x} \in \text{Argmin}_{\mathbf{B}^n}(f)$.

On dira qu'une affectation est persistante faible si toutes les configurations de $\mathbf{x}$ appartenant à tous les minima globaux de la fonction f le sont encore après affectation.

Exemple 3.2. Pour illustrer les deux notions précédentes, reprenons la fonction g de l'exemple 3.1 et sa posiforme ψ_1 . On considère l'affectation partielle $\mathbf{y}^* = (0) \in \mathbf{B}^{\{1\}}$. On peut consta-

ter pour tous les minima globaux, $\mathbf{B}^{\{1\}} = (0)$, que l'affectation $\mathbf{y}$ est persistante forte. Soit $\mathbf{z}^* = (0, 1) \in \mathbf{B}^{\{1,2\}}$ une affectation partielle. Les solutions $\mathbf{x} = (0, 0, 0, 0)$ et $\mathbf{x} = (0, 0, 0, 1)$ appartiennent au minimum global de la fonction, par contre la deuxième variable $\neq 1$, donc cette affectation n'est pas persistante forte. Néanmoins, l'affectation de $\mathbf{x} = (0, 0, 0, 0)$ et de $\mathbf{x} = (0, 0, 0, 1)$ par $\mathbf{z}^*$ donne $\mathbf{x} = (0, 1, 0, 0)$ et $\mathbf{x} = (0, 1, 0, 1)$. Ces deux vecteurs appartiennent bien au minimum global de la fonction donc cette affectation est persistante faible.

3.1.4 Fonction pseudo-boulienne quadratique

L'optimisation d'une fonction pseudo-boulienne quadratique est un problème abondamment étudié. L'une des principales propriétés de ce type d'optimisation est qu'il est possible de trouver en un temps polynomial une borne inférieure aux valeurs de la fonction appelée la *roof duality*. Une fois cette borne inférieure atteinte, il est possible d'en déduire une affectation partielle ayant comme propriété d'être persistante forte.

On nomme $\mathcal{F}_2$ la famille des fonctions pseudo-bouliennes quadratiques. Une fonction pseudo-boulienne quadratique peut être représentée par sa multilinéaire polynomiale unique

$$f(x_1, \dots, x_n) = c_0 + \sum_{i=1}^n c_i x_i + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j \quad (3.7)$$

ou par une posiforme quadratique de la forme suivante :

$$\phi(x) = a_0 + \sum_{u \in \mathbf{L}} a_u u + \sum_{u, v \in \mathbf{L}, u \neq v} a_{uv} uv \quad (3.8)$$

où $a_u \geq 0$ et $a_{uv} \geq 0$. Généralement, a_0 est nommé *terme constant*, les termes de degré 1 *terme linéaire*, et ceux de degré 2 *terme quadratique*. On note $C(\phi)$ le terme constant a_c de la posiforme ϕ ([BH02], p.34).

3.1.4.1 Roof duality

Trouver le minimum d'une fonction pseudo-boulienne quadratique quelconque est un problème *NP-complet*, mais on peut calculer une borne inférieure aux valeurs de la fonction appelée *roof duality* en un temps polynomial. Une façon intuitive d'interpréter la *roof duality* et de trouver quelle est la posiforme ϕ dans un ensemble de fonctions pseudo-bouliennes vérifiant la même table de vérité, dont le terme constant $C(\phi)$ est le plus grand. Ce problème est en effet connu en optimisation de fonctions pseudo-bouliennes quadratiques. Il existe pour cela plusieurs techniques permettant de calculer cette borne inférieure à la fonction.

- Majorisation notée $M_2(f)$.
- Complémentation notée $C_2(f)$.
- Linéarisation notée $L_2(f)$.
- Autres : Lagrangien, *paved duality*, etc...

Un résultat important est que toutes ces bornes sont équivalentes du fait qu'elles retournent toutes les mêmes valeurs.

Grâce à la structure spécifique de la posiforme (tous les coefficients sont positifs) le terme constant est donc une borne inférieure. En effet, chaque terme étant positif, la valeur minimale d'une posiforme est atteinte quand tous les termes sont annulés.

Théorème 3.1. Pour toutes fonctions pseudo-booléennes quadratiques $f \in \mathcal{F}_2$, on a ([BH02], p.39)

$$M_2(f) = C_2(f) = L_2(f) \leq \min_{\mathbf{x} \in \mathbf{B}^n} f(\mathbf{x})$$

Une fois la *roof duality* atteinte (c'est à dire après avoir transformé la posiforme d'origine en une posiforme avec la plus grande constante possible) la propriété suivante permet d'extraire un sous ensemble persistant fort :

Théorème 3.2. (*Persistence forte*) Soit une fonction pseudo-booléenne quadratique $f \in \mathcal{F}_2$, soit $\phi \in \mathcal{P}_2(f)$ une posiforme la représentant tel que $C(\phi) = C_2(\phi)$, alors si $a_u > 0$ pour des labels $u \in \mathbf{L}$, alors $u = 0$ dans tous les vecteurs binaires $\mathbf{x} \in \text{Argmin}(f)$ minimisant f . ([BH02], p.40)

Autrement dit, si il reste un terme linéaire dans la formule obtenue après avoir atteint la *roof duality*, donc $C(\phi) = C_2(\phi)$, alors chaque affectation annulant le terme linéaire appartiendra à tous les minima globaux.

3.1.4.2 Modèle de flot maximum / coupe minimale :

Trouver la *roof duality* est un problème de classe polynomiale. En effet, elle peut être obtenue en résolvant un problème de programmation linéaire.

Comme précédemment indiqué, plusieurs méthodes existent pour calculer cette borne. Dans cette partie, nous allons détailler une méthode fondée sur la théorie des graphes. Cette méthode est très efficace, car elle est fondée sur des algorithmes de flots de la théorie des graphes, très étudiés et optimisés. Ces algorithmes sont optimaux pour les problèmes discrets

L'algorithme est décomposé en plusieurs étapes : tout d'abord, il faut effectuer la réduction d'une équation pseudo-booléenne en un graphe induit. Par la suite, la *roof duality* peut être calculée grâce à un algorithme de poussage de flot sur le graphe induit, et enfin, les affectations persistantes fortes sont faites à partir d'un parcours du graphe résiduel.

Construction du graphe induit : Soit une fonction pseudo-booléenne quadratique $f \in \mathcal{F}_2$ donnée par la posiforme $\phi \in \mathcal{P}_2(f)$ de forme (3.8). On associe à cette posiforme quadratique un graphe orienté $G_\phi = (N, A)$ où l'ensemble de noeuds est défini par $N = \mathbf{L} \cup \{x_0, \bar{x}_0\}$. Il y a deux sommets par variable : la variable elle même et son complément, plus deux sommets x_0 et $\bar{x}_0$ représentant les constantes $x_0 = 1$ et $\bar{x}_0 = 0$. L'ensemble des arêtes est défini par l'ensemble des termes de l'équation pseudo-booléenne. À chaque terme quadratique correspond deux arcs $(\overrightarrow{u, v})$ et $(\overrightarrow{v, u})$ avec chacun comme poids $\frac{1}{2}c_{uv}$. À chaque terme linéaire correspond les arcs $(\overrightarrow{u, \bar{x}_0})$ et $(\overrightarrow{x_0, u})$ avec chacun comme poids $\frac{1}{2}c_u$ ([BH02], p.42).

Calcul de la borne inférieure par flot maximum : Une fois la réduction faite, Nous allons voir qu'il a été montré qu'il y a équivalence entre une somme alternée dans une équation pseudo-booléenne et un chemin augmentant dans le graphe induit. La caractéristique d'une somme augmentante est qu'il est possible d'extraire par une transformation algébrique une constante de la chaîne, et par conséquent, augmenter le terme constant $C(\phi)$ de la posiforme pour atteindre la *roof duality* $C_2(\phi)$. Cette étape peut être réalisée en effectuant un algorithme de poussage de flot dans le graphe induit précédent grâce à l'équivalence somme alternée / chemin augmentant.

Équivalence somme alternée / chemin augmentant : Un flot possible dans un graphe pondéré $G = (N, A)$ avec comme source x_0 et comme puits $\bar{x}_0$ est une application $\zeta : A \rightarrow \mathbb{R}_+$ respectant les contraintes :

$$\zeta(u, v) \leq c_{u,v} \quad \text{pour tous les arcs } (\overrightarrow{u, v}) \in A, \text{ et} \quad (3.9)$$

$$\sum_{(\overrightarrow{u, v}) \in A} \zeta(u, v) = \sum_{(\overrightarrow{v, w}) \in A} \zeta(v, w) \quad \text{pour tous les noeuds } v \in \mathbf{L} \quad (3.10)$$

Pour un graphe donné $G = (N, A)$ et un flot possible ζ dans ce graphe, on appelle graphe résiduel $G[\zeta] = (N, A^\zeta)$ avec comme capacités résiduelles

$$c_{uv}^\zeta = \begin{cases} c_{uv} - \zeta(u, v) & \text{pour } (\overrightarrow{u, v}) \in A \\ \zeta(u, v) & \text{pour } (\overrightarrow{v, u}) \in A \end{cases} \quad (3.11)$$

Si $u_1, u_2, \dots, u_k \in \mathbf{L}$, alors on appelle l'expression de la forme :

$$u_1 + \bar{u}_1 u_2 + \bar{u}_2 u_3 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k \quad (3.12)$$

une somme augmentante. Une posiforme quadratique ϕ possède la somme augmentante (3.12) de poids ω si nous avons $a_{u_1} \geq \omega$, $a_{\bar{u}_j u_{j+1}} \geq \omega$ pour $j = 1, \dots, k-1$ et $a_{\bar{u}_k} \geq \omega$ pour tous les coefficients correspondants de ϕ ([BH02], p.42).

Proposition 3.1. L'identité suivante est vérifiée pour les sommes alternées :

$$u_1 + \bar{u}_1 u_2 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k = 1 + u_1 \bar{u}_2 + \dots + u_{k-1} \bar{u}_k \quad (3.13)$$

Une posiforme quadratique ϕ qui contient une somme alternée (3.12) de poids ω peut être transformée en une posiforme équivalente avec un terme constant plus important. Pour cela il faut dans un premier temps transformer le terme ϕ :

$$\phi = \omega[u_1 + \bar{u}_1 u_2 + \bar{u}_2 u_3 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k] + \phi'$$

Dans ce cas, ϕ' est aussi une posiforme quadratique. Par conséquent, en appliquant l'identité (3.13), on obtient

$$\phi = \omega + \omega[u_1 + \bar{u}_1 u_2 + \bar{u}_2 u_3 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k] + \phi'$$

Dans cette formulation, il y a une correspondance terme à terme avec une somme alternée contenue dans une posiforme et un chemin augmentant dans le graphe induit correspondant. La proposition suivante peut être énoncée :

Proposition 3.2. On considère une posiforme $\phi \in P_2(f)$ et un flot possible ζ dans le graphe $G = G_\phi$. Alors $x_0, u_1, \dots, u_k, \bar{x}_0$ est un chemin augmentant de capacité $\epsilon > 0$ dans $G[\zeta]$ si et seulement si $u_1 + \bar{u}_1 u_2 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k$ est une somme alternée de poids ϵ dans la posiforme correspondant $\phi_{G[\zeta]}$.

Exemple 3.3. Dans cet exemple, on montre l'équivalence entre la capacité d'un chemin augmentant et le poids d'une somme alternée.

$$10x_1 + 8\bar{x}_1 x_2 + 6\bar{x}_2 x_3 + 4\bar{x}_3 \quad (3.14a)$$

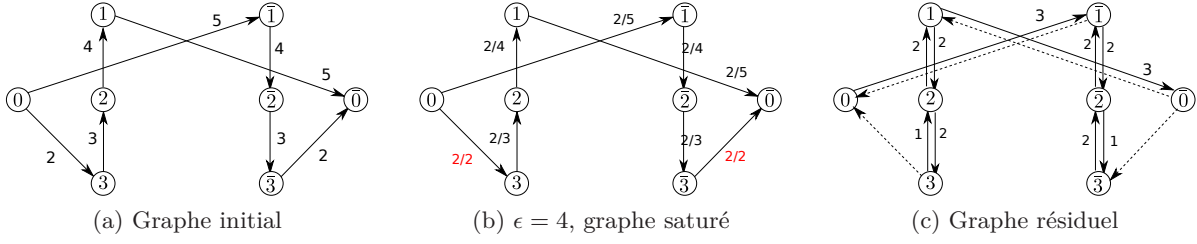


FIGURE 3.1 – Équivalence chemin / somme alternée de l'exemple 3.3. La figure 3.1a montre le graphe induit de la fonction pseudo-booléenne initiale. La figure 3.1b montre les capacités du graphe après poussage de flot. La figure 3.1c représente le graphe résiduel.

$$\underbrace{4}_{\omega} \underbrace{(x_1 + \bar{x}_1 x_2 + \bar{x}_2 x_3 + \bar{x}_3)}_{\text{somme alternée}} + \underbrace{6x_1 + 4\bar{x}_1 x_2 + 2\bar{x}_2 x_3}_{\phi'} \quad (3.14b)$$

$$4(1 + x_1 \bar{x}_2 + x_2 \bar{x}_3) + 6x_1 + 4\bar{x}_1 x_2 + 2\bar{x}_2 x_3 \quad (3.14c)$$

$$4 + 6x_1 + 4x_1 \bar{x}_2 + 4\bar{x}_1 x_2 + 2\bar{x}_2 x_3 + 4x_2 \bar{x}_3 \quad (3.14d)$$

En factorisant par la plus grande constante, on isole la somme alternée (3.14b). En appliquant l'identité (3.13), un terme constant apparaît dans l'équation (3.14c) ce qui a comme conséquence, une fois développé, d'augmenter le terme constant et d'atteindre la *roof duality*.

L'équation (3.14a) correspond au graphe 3.1a. Calculer le flot maximum jusqu'à saturation (graphe 3.1b) correspond à trouver le plus haut terme constant que l'on peut mettre en facteur sur la somme alternée (équation 3.14b). Dans cet exemple on peut constater que $\epsilon = 4$. En prenant l'équation du graphe résiduel une fois les chemins saturés 3.1c, on retrouve l'équation (3.14d).

Grâce à cette équivalence, la propriété suivante peut être énoncée :

Théorème 3.3. Soit une fonction pseudo-booléenne quadratique f , et une posiforme possible la représentant $\phi \in P_2(f)$. Soit ν^* la valeur du flot maximum dans G_ϕ , alors :

$$C_2(f) = C(\phi) + \nu^*$$

La propriété fondamentale de cette réduction en graphe, est que la *roof duality* $C_2(f)$ est atteinte en sommant le terme constant de la posiforme correspondant au graphe induit avec la valeur du flot maximum atteinte par l'algorithme de poussage de flot.

Après avoir calculé le flot maximum, et donc la *roof duality*, il est possible de trouver les affectations persistantes fortes. Une façon efficace est d'utiliser le graphe résiduel.

Théorème 3.4 (Persistance forte et flot maximum). Soit $\phi \in P_2(f)$ pour une fonction pseudo-booléenne quadratique f , soit ζ^* un flot maximum dans $G = G_\phi$, soit $S \in L$ l'ensemble des noeuds dans G qui peuvent être atteints en partant de la source x_0 par un chemin avec une capacité résiduelle positive, alors $u(\mathbf{x}^*) = 1$ pour tout $u \in S$ et pour tous les vecteurs $\mathbf{x}^* \in \text{Argmin}(f)$.

En d'autres termes, pour toutes variables pouvant être atteintes dans le graphe résiduel en partant de la source par des capacités positives, l'affectation qui vérifie $u(\mathbf{x}^*) = 1$ est persistante forte.

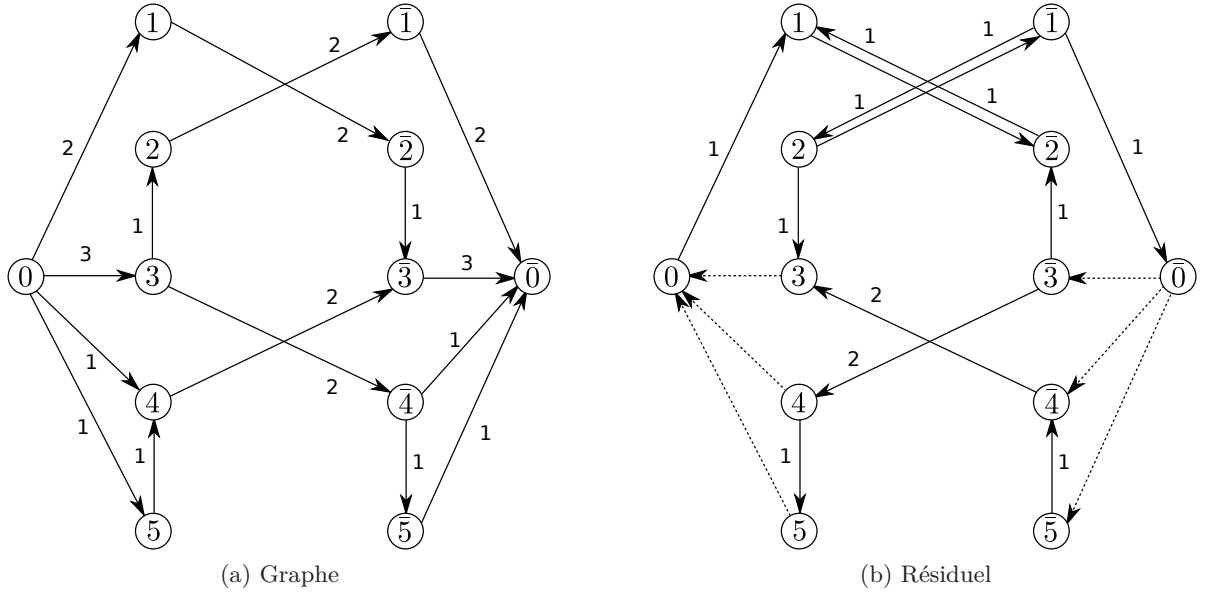


FIGURE 3.2 – Le graphe induit 3.2a et le graphe résiduel de l'équation pseudo-boulienne de l'exemple 3.5.

Récapitulatif : Pour résumer, les affectations persistantes fortes pour une fonction pseudo-boulienne quadratique f sont calculées de la façon suivante :

- Construire le graphe induit G de f .
- Calculer le flot maximum ν de G .
- Prendre le graphe résiduel de $G[\nu]$.
- À partir de la source de $G[\nu]$, pour toutes les variables pouvant être atteintes à partir de la source, affecter $u(x^*) = 1$.

Exemple 3.4. Soit la fonction pseudo-boulienne quadratique suivante :

$$f(x) = 10 - 4x_1 - 4x_3 - 2x_4 + 4x_4x_2 - 2x_2x_3 + 4x_3x_4 - 2x_4x_5$$

En substituant avec l'identité $x = 1 - \bar{x}$ chaque terme quadratique ayant un coefficient négatif, on a la posiforme suivante :

$$\phi = -4 + 4\bar{x}_1 + 6\bar{x}_3 + 2\bar{x}_4 + 2\bar{x}_5 + 4x_1x_2 + 2\bar{x}_2x_3 + 4x_3x_4 + 2\bar{x}_4x_5$$

tel que $\phi \in \mathcal{P}_2(f)$. La figure 3.2a montre le réseau G_ϕ correspondant. En observant la figure 3.2a, on peut constater que dans le réseau G_ϕ , les flots suivants peuvent être poussés séquentiellement pour les chemins augmentants suivants :

$$x_0 \rightarrow x_1 \rightarrow \bar{x}_2 \rightarrow \bar{x}_3 \rightarrow \bar{x}_0 \text{ et } x_0 \rightarrow x_3 \rightarrow x_2 \rightarrow \bar{x}_1 \rightarrow \bar{x}_0, \quad (3.15)$$

$$x_0 \rightarrow x_3 \rightarrow \bar{x}_4 \rightarrow \bar{x}_0 \text{ et } x_0 \rightarrow x_4 \rightarrow \bar{x}_3 \rightarrow \bar{x}_0, \quad (3.16)$$

$$x_0 \rightarrow x_3 \rightarrow \bar{x}_4 \rightarrow \bar{x}_5 \rightarrow \bar{x}_0 \text{ et } x_0 \rightarrow x_5 \rightarrow x_4 \rightarrow \bar{x}_3 \rightarrow \bar{x}_0. \quad (3.17)$$

Ces chemins augmentants correspondent respectivement aux sommes alternées :

$$\bar{x}_1 + x_1x_2 + \bar{x}_2x_3 + \bar{x}_3 \text{ et } \bar{x}_3 + x_3\bar{x}_2 + x_2x_1 + \bar{x}_1, \quad (3.18)$$

$$\bar{x}_3 + x_3x_4 + \bar{x}_4 \text{ et } \bar{x}_4 + x_4x_3 + \bar{x}_3, \quad (3.19)$$

$$\bar{x}_3 + x_3x_4 + \bar{x}_4x_5 + \bar{x}_5 \text{ et } \bar{x}_5 + x_5\bar{x}_4 + x_4x_3 + \bar{x}_3. \quad (3.20)$$

À cet instant, il n'y a plus de chemin augmentant dans le graphe de la figure 3.2a. Le nombre de chemins augmentants maximum a été atteint, la valeur du flot maximum est donc de $\nu = 6$. Le réseau final correspond donc à la forme quadratique suivante :

$$\psi = 2\bar{x}_1 + 2x_1x_2 + 2\bar{x}_1\bar{x}_2 + 2x_2\bar{x}_3 + 4\bar{x}_3\bar{x}_4 + 2x_4\bar{x}_5$$

Avec un flot maximum de capacité $\nu = 6$, on a : $\phi = C(\phi) + \nu + \psi = 2 + \psi$. Donc $C_2(f) = 2$. D'après (3.4), on peut voir que le graphe résiduel a deux sommets directement atteignables à partir de la source : le sommet 1 et $\bar{2}$. D'après (3.4), les affectations $x_1 = 1$ et $x_2 = 0$ sont vérifiées dans chaque minimum global. On peut donc affecter les valeurs à l'équation ψ . Minimiser f équivaut maintenant à minimiser $\psi^* = 4\bar{x}_3\bar{x}_4 + 2x_4\bar{x}_5$. Les affectations qui annulent cette fonction sont persistantes faibles.

3.1.5 Cas sous-modulaire

Une fonction f est sous-modulaire si :

$$f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y) \quad \forall X, Y \quad (3.21)$$

est respecté pour tous les sous ensembles X, Y . Plus généralement, on peut caractériser une fonction pseudo-booléenne $f : \mathbf{B}^n \rightarrow \mathbb{R}$ sous-modulaire si :

$$\Delta_{ij}(X) \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i \partial x_j}(x) \leq 0 \quad (3.22)$$

Autrement dit, une fonction quadratique pseudo-booléenne est sous-modulaire si, sous sa forme multilinéaire polynomiale unique (3.7), sa dérivée seconde est toujours négative ou nulle, ce qui implique que tous ses termes quadratiques ont un signe négatif. Trouver le minimum d'une fonction sous-modulaire peut aussi être réduit en un problème spécifique de coupe de graphe.

3.1.5.1 Recherche du minimum global par coupe de graphe

Quand une fonction est sous-modulaire, il est toujours possible de l'écrire comme une posiforme ϕ sous la forme :

$$\phi(x_1, \dots, x_n) = \sum_{i \in P} a_i x_i \sum_{j \in N} a_j \bar{x}_j + \sum_{1 \leq i \leq j \leq n} a_{ij} x_i \bar{x}_j \quad (3.23)$$

où $P, N \subseteq \mathbf{V}$ et tous les coefficients a_i ($i \in P \cup N$) et a_{ij} ($1 \leq i \leq j \leq n$) sont positifs. On peut alors associer à ϕ le graphe N_ϕ construit de la façon suivante :

- $V(N_\phi) = \{s, t\} \cup \mathbf{V}$.
- $A(N_\phi) = \{(s, j) | j \in N\} \cup \{(i, t) | i \in P\} \cup \{(i, j) | 1 \leq i \leq j \leq n\}$.

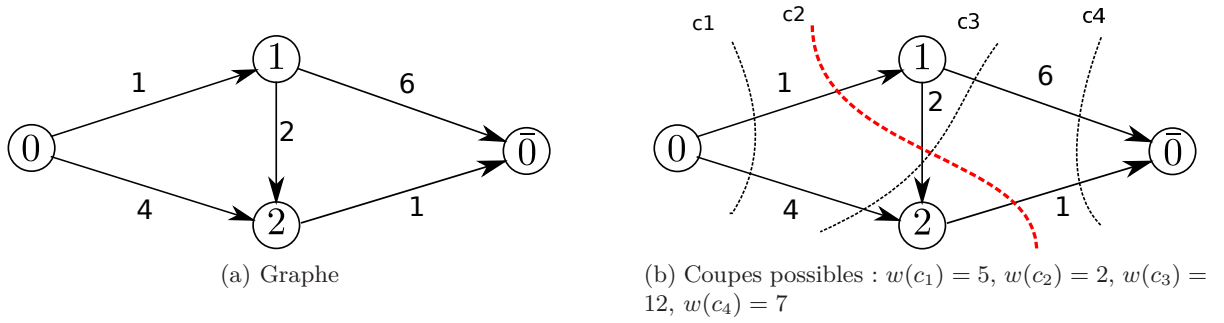


FIGURE 3.3 – Exemple de la minimisation d’une fonction pseudo-booléenne sous-modulaire de degré 2 par coupe de graphe (Exemple 3.5).

où les capacités des arcs sont : $c_{sj} = a_j$ pour $j \in N$, $c_{i,t} = a_i$ pour $i \in P$ et $c_{ij} = a_{ij}$ pour $1 \leq i < j \leq n$.

En d’autres termes, l’ensemble des sommets $V(N_\phi)$ est composé de l’ensemble des indices $\mathbf{V}$ ainsi qu’une source s et un puits t . On construit, pour chaque terme unitaire, une arête. Quand le terme est une variable x_i , une arête du sommet ayant comme indice i est créée, avec comme poids le coefficient du terme a_i . Quand le terme est le complément d’une variable $\bar{x}_j$, une arête de la source s au sommet ayant comme indice j est créée, avec comme poids le coefficient du terme. On construit pour chaque terme $a_{ij}x_i\bar{x}_j$ une arête de la variable i à j .

Il y a donc une correspondance entre la coupe s-t de ce graphe et le vecteur binaire $\mathbf{x} \in \mathbf{B}^n \leftrightarrow S_x \stackrel{\text{def}}{=} \{s\} \cup \{j | x_j = 1\}$. Il est facile de voir qu’avec cette définition on a :

$$\phi(\mathbf{x}) = \sum_{\substack{u \in S_x \\ v \notin S_x}} c_{u,v} \quad (3.24)$$

pour tout $\mathbf{x} \in \mathbf{B}^n$, le minimum de ϕ correspond donc à la coupe de capacité minimum de N_ϕ

Exemple 3.5. Minimiser la fonction :

$$\phi(\mathbf{x}) = \bar{x}_1 + 6x_1 + 4\bar{x}_2 + x_2 + 2x_1\bar{x}_2 \quad (3.25)$$

équivalent à trouver la coupe minimale dans le graphe induit (figure 3.3a) de la fonction 3.25. On peut voir que le poids de chaque coupe du graphe 3.3b correspond à la valeur de la fonction 3.25 avec comme affectation les valeurs correspondantes à la constante de la coupe pour chaque noeud (0 pour la source et 1 pour le puits). Par exemple, la coupe c_2 de la figure 3.3b a comme poids 2 (On rappelle que le poids d’une coupe est la somme des arêtes coupées dont le début est du côté de la source et la fin du côté du puits). Le noeud 1 est coupé au niveau de la source s , donc $x_1 = 0$, le noeud 2 au niveau du puits, donc $x_2 = 1$. La valeur de l’équation pseudo-booléenne (3.25) avec comme affectation $x_1 = 0$ et $x_2 = 1$ est bien égale à 2.

3.1.6 Optimisation de fonctions pseudo-booléennes de degré supérieur à 2

Il est difficile d’optimiser une fonction pseudo-booléenne de degré supérieur à 2. En effet, une méthode générique convergerait très vite dans des minima locaux, surtout, dans le cas non sous-modulaire.

Pour pallier cela, deux approches ont été proposées. La première consiste à réduire une fonction de degré quelconque en degré quadratique possédant le même minimum et de l’optimiser

$\mathbf{x}$	$g(\mathbf{x})$
0 0	5
0 1	2
1 0	12
1 1	7

TABLE 3.2 – Table de vérité de la fonction pseudo-booléenne de l'exemple 3.5

avec les techniques précédemment vues dans le cas quadratique (*roof duality*). La seconde est de généraliser directement le principe de *roof duality* au cas de degré supérieur à deux [KS11]. Dans nos travaux, la méthode utilisée [Ish09a] est fondée sur le principe de substitution, nous allons donc introduire les notions de base de cette approche.

3.1.7 Réduction de fonction de degré quelconque en degré quadratique

La première technique pour réduire une fonction pseudo-booléenne d'ordre n à l'ordre 2 à été introduite en 1975, elle consiste à ajouter une variable auxiliaire afin de réduire les cliques de grandes cardinalités en plusieurs cliques d'ordres inférieurs [Ros75]. Cette méthode est itérative et fonctionne pour un degré quelconque. Le principal défaut de cette méthode, est que, pour chaque substitution, un terme non sous-modulaire de grande cardinalité est ajouté. Au fil des itérations, la cardinalité des termes sous-modulaires augmente dans la forme multilinéaire polynomiale unique.

Mais, le nombre d'affectations persistantes fortes lors du calcul de la *roof duality* dépend directement du nombre et du coefficient des termes quadratiques non sous-modulaires dans la forme multilinéaire polynomiale unique. En effet, lorsque tous les coefficients sont négatifs, l'énergie est sous-modulaire, voir (3.22). Lorsqu'au moins un coefficient est positif sur un terme quadratique, l'énergie n'est plus sous-modulaire.

En pratique et comme constaté dans [GBP11], plus il y a de termes positifs, plus leurs coefficients sont importants, et moins il y a d'affectations fortes lors du calcul de la *roof duality*.

Après ce constat, de nouvelles méthodes sont apparues afin de minimiser l'amplitude et le nombre de termes positifs lors de la réduction d'un degré quelconque à un degré quadratique [Ish09a].

Plusieurs types de réductions sont donc possibles pour une même clique, dans [GBP11], il est proposé, pour chaque clique, de choisir entre différents types de décompositions. Pour cela, une technique d'inférence est proposée pour permettre de choisir le type de réduction en minimisant soit, le nombre d'arêtes non sous-modulaires, soit l'amplitude des termes non sous-modulaires dans la fonction pseudo-booléenne finale.

3.1.7.1 Substitution

La méthode introduite dans [Ros75] est une méthode de substitution qui permet, en remplaçant le produit de deux variables par une variable auxiliaire, de réduire le degré maximum d'une fonction pseudo-booléenne. On considère ces deux équivalences :

$$xy = z \iff xy - 2xz - 2yz + 3z = 0 \quad (3.26)$$

$$xy \neq z \iff xy - 2xz - 2yz + 3z > 0 \quad (3.27)$$

Les deux équivalences précédentes peuvent être facilement vérifiées en essayant les 8 combinaisons possibles pour les variables x, y et z .

Il est donc possible, dans une équation pseudo-booléenne, de remplacer le produit xy par z dans une clique de degré supérieur à deux, et d'ajouter l'expression $xy - 2xz - 2yz + 3z$ comme un terme de pénalité (avec un grand terme multiplicatif).

Alors, d'après [Ros75], le vecteur minimisant l'équation réduite est équivalent au vecteur minimisant l'équation originale en omettant les variables supplémentaires introduites, ce qui se traduit par :

$$\min_{w,x,y} wxy = \min_{w,x,y,z} wz + a(xy - 2xz + yz + 3z) \quad (3.28)$$

où w est la variable ajoutée et a le coefficient multiplicatif. a doit être assez grand, pour cela, il peut être initialisé avec $a = 1 + \sum_{S \in \mathbf{V}} |c_S|$. Cette étape de substitution peut être itérée jusqu'à ce que le degré maximum soit égal à 2.

Exemple 3.6. Soit la fonction pseudo-booléenne :

$$f_a(x_1, x_2, x_3) = 2x_1x_2 - 3x_1x_2x_3 \quad (3.29)$$

D'après (3.26), en substituant x_1x_2 par une nouvelle variable x_4 avec $a = 1 + 2 + 3 = 6$, la fonction (3.29) se réécrit :

$$\begin{aligned} f_{ar}(x_1, x_2, x_3, x_4) &= 2x_1x_2 - 3x_1x_2x_3 + 5(x_1x_2 - 2x_1x_4 - 2x_3x_4 - 3x_4) \\ &= 18x_4 + 8x_1x_2 - 12x_1x_4 - 12x_2x_4 - 3x_3x_4 \end{aligned} \quad (3.30)$$

Si on observe la table de vérité 3.3 de la forme originale (3.29) et de la forme réduite (3.30), on constate que le minimum de la fonction réduite est vérifié pour le vecteur $\mathbf{x} = (1111)$, si on omet les variables ajoutées pour la réduction (ici x_4), le vecteur correspondant aux variables d'origine est $\mathbf{x} = (111)$, on peut voir que ce vecteur est bien le minimum global de la fonction d'origine.

Mais, cette technique ajoute des termes de cardinalité de plus en plus grands au fil des itérations, si ces termes ont un coefficient positif dans la forme multilinéaire polynomiale unique (énergie non sous-modulaire) alors le calcul de la *roof duality* ne sera pas optimal. En effet, plus la cardinalité des termes non sous-modulaires est importante et moins il y a d'affectations persistantes fortes.

Pour pallier cela, d'autres méthodes de réductions ont été proposées afin de minimiser l'amplitude et le nombre de termes non sous-modulaires de la forme quadratique. Une première amélioration consiste en un autre changement de variables. Pour cela, plusieurs méthodes ont été proposées afin de réduire le nombre et l'amplitude des termes non sous-modulaires [KZ02, Ros75, Ish09a]. Ces méthodes ont permis d'augmenter considérablement le nombre de persistances fortes lors du calcul de la *roof duality*, ce qui a permis de rendre réellement exploitable les fonctions quadratiques résultantes de la réduction. Néanmoins, le choix de la bonne substitution n'est pas simple, en effet, il y a plusieurs façons d'effectuer une réduction.

Dans [GBP11], il a été proposé de traiter le problème de réduction comme un problème d'optimisation, où l'on cherchera quelles sont les réductions qui minimisent l'amplitude ou le nombre de termes non sous-modulaires, cette méthode a l'avantage de tirer parti de chaque méthode de réduction et d'augmenter le nombre d'affectations persistantes fortes.

$\mathbf{x}$	$f_a(\mathbf{x})$	$f_{ar}(\mathbf{x})$
0 0 0 0	0	0
0 0 0 1	0	18
0 0 1 0	0	0
0 0 1 1	0	15
0 1 0 0	0	0
0 1 0 1	0	6
0 1 1 0	0	0
0 1 1 1	0	3
1 0 0 0	0	0
1 0 0 1	0	6
1 0 1 0	0	0
1 0 1 1	0	3
1 1 0 0	2	8
1 1 0 1	2	2
1 1 1 0	-1	8
1 1 1 1	-1	-1

TABLE 3.3 – Table de vérité de la fonction pseudo-booléenne de l'exemple 3.6

D'autres méthodes ont par la suite été proposées pour améliorer le taux d'affectation et la rapidité de la réduction. En effet, il est possible de réduire directement une fonction de degré quelconque en degré quadratique. Dans [Ish11], la méthode par substitution proposée fonctionne quel que soit le degré de la fonction à réduire. Une autre méthode est proposée dans [FGBZ11] par coupe de graphe.

3.2 Extension au cas multi-labels

L'optimisation multi-labels peut être approchée comme une succession de sous-problèmes binaires. En effet, si nous prenons deux jeux de labels pour une fonction donnée, il est possible de trouver grâce aux méthodes vues précédemment, quels sont les labels, sur l'ensemble des deux configurations, qui minimisent l'énergie. C'est ce que l'on appelle, dans le cadre d'une optimisation multi-labels une *fusion* [LRRB10].

3.2.1 Stratégie de fusion dans le cas multi-labels

Pour chaque variable, un choix booléen est réalisé. Soit L^t un jeu de labels courants, soit L^p un nouvel ensemble. Le but est de fusionner les deux ensembles de façon optimale pour avoir un nouvel ensemble L^{t+1} pour lequel l'énergie (3.1) est plus faible qu'avec l'ensemble L^t . Afin de combiner les deux jeux de labels, une combinaison $L^b(B)$ est définie par la fonction booléenne auxiliaire suivante :

$$L^b(B) = (1 - B)L^t + BL^p \quad (3.31)$$

où B est un ensemble de variables booléennes. Il y en a une par site s . Quand $B(s) = 0$, le label courant est gardé au site s , quand $B(s) = 1$, le label proposé est choisi. Avec cette formulation, nous pouvons donc obtenir une fusion des labels grâce au choix de B , donc :

$$L^{t+1} = L^b(B^*) \quad (3.32)$$

Pour cela, il faut donc trouver quelle est la combinaison de variables booléennes B^* dont les labels suivant l'équation (3.31) minimisent l'énergie (3.1) :

$$B^* = \min_B f(L^b(B)) \quad (3.33)$$

L'équation pseudo-booléenne correspondant à la fusion entre les labels courants L^t et les labels proposés L^p minimisant l'énergie (3.1) en fonction des variables booléennes B est :

$$\begin{aligned} \phi(B) = & \sum_{u \in C_1} \vartheta_u(L_u^p)B(u) + \vartheta_u(L_u^t)\bar{B}(u) \\ & + \sum_{u,v \in C_2} \vartheta_{uv}(L_u^p, L_v^p)B(u)B(v) + \vartheta_{uv}(L_u^p, L_v^t)B(u)\bar{B}(v) \\ & \quad + \vartheta_{uv}(L_u^t, L_v^p)\bar{B}(u)B(v) + \vartheta_{uv}(L_u^t, L_v^t)\bar{B}(u)\bar{B}(v) \\ & + \sum_{u,v,w \in C_3} \dots \end{aligned} \quad (3.34)$$

Si l'énergie (3.1) est sous-modulaire, alors l'équation booléenne (3.34) est aussi sous-modulaire, ainsi, toutes les variables B pourront être fixées et la fusion sera optimale. Dans le cas contraire, un sous ensemble sera fixé dans un premier temps avec le calcul de la *roof duality*, par la suite, les affectations persistantes faibles seront calculées en minimisant l'équation résiduelle. Finalement, les labels restants doivent être fixés, pour cela, plusieurs stratégies peuvent être utilisées, la plus simple étant de garder la solution courante.

Si la fonction (3.34) à un degré supérieur à 2, elle peut être réduite en degré quadratique, et donc être optimisée avec les méthodes vues précédemment.

3.2.2 Optimisation globale de fonction multi-labels par algorithme itératif

Une fonction multi-labels peut être optimisée avec un algorithme itératif grâce à la technique de fusion. En effet, une fusion permet de choisir, parmi un sous-ensemble de labels, les labels qui minimisent l'énergie. En répétant cette étape, il est donc possible de converger vers un minimum d'énergie.

Dans le cas général, une proposition quelconque peut être fournie. Si l'énergie n'est pas sous-modulaire, alors la fusion fournira un sous ensemble persistant fort et le reste des labels peut être fixé grâce à la persistance faible afin de garantir que l'énergie n'augmente pas. Dès lors, on se trouve dans le cadre d'une optimisation non convexe et différentes techniques peuvent être explorées pour optimiser les propositions. Dans le cas sous-modulaire, l'étape de fusion est appelée α -expansion.

3.2.3 α -expansion dans le cas sous-modulaire

Une énergie multi-labels est sous-modulaire si les fonctions d'erreur sur les termes quadratiques vérifient la propriété (3.21).

L'algorithme α -expansion est un cas particulier de la technique de fusion. Il est très efficace dans le cadre d'une énergie pseudo-booléenne quadratique sous-modulaire. Il consiste à effectuer

une succession de fusions, avec comme proposition L^p un ensemble de labels identiques et d'itérer sur les différents labels jusqu'à la stabilisation [BVZ01b].

Grâce au caractère sous-modulaire de la fonction, cette étape peut être réduite en une coupe de graphe qui permet d'assigner 100% des labels (partie 3.1.5.1), le minimum global du problème binaire sera donc trouvé lors de la fusion. Par contre, seulement un minimum local de l'énergie multi-labels sera obtenu [BVZ01b].

L'algorithme est le suivant :

- Initialiser L .
- Tant qu'il y a des changements dans L :
 - Pour chaque label α , effectuer une fusion entre la carte de symbole L et α .

3.3 Conclusion

L'étude de l'optimisation des fonctions pseudo-booléennes nous a permis de pouvoir caractériser une fonction non sous-modulaire par rapport à une fonction sous-modulaire et de proposer, dans ce dernier cas, des algorithmes en temps polynomial permettant de trouver le minimum global de l'énergie (coupe de graphe).

Mais, l'ensemble des fonctions sous-modulaires est assez restreint. En effet, dans le cadre des champs de Markov, seuls les modèles avec des cliques d'ordres maximales égale à 2 avec une métrique sous-modulaire, donnent une fonction sous-modulaire qu'il est possible d'optimiser globalement en un temps polynomial. Les fonctions avec des cliques d'ordres supérieures à 2, ou avec une métrique non sous-modulaire, ne sont pas sous-modulaires. Il existe des techniques qui permettent de transformer une fonction pseudo-booléenne quelconque en fonction pseudo-booléenne d'ordre deux ayant le même minimum global. Mais, même avec une métrique sous-modulaire sur les cliques d'ordres supérieures à 2, la transformation d'une clique d'ordre supérieur à 2 entraînera l'apparition de terme non sous-modulaire, il n'est donc plus possible d'appliquer les algorithmes d'optimisation globale sous-modulaire.

Étonnamment, et malgré le caractère non sous-modulaire, certaines affectations de labels, appartenant au minimum global, peuvent néanmoins être trouvées en un temps polynomial. En effet, la structure éparse de certains graphes permet de détecter, pour un sous ensemble de variables, si l'affectation d'un label pour une variable donnée appartient au minimum global ou non. Cette notion est importante car elle permet, même si le minimum global d'une énergie ne peut pas être atteint du fait de sa non sous-modularité, d'identifier un sous ensemble de variables pour lesquelles on peut trouver l'affectation appartenant au minimum global. Cette notion est introduite par la *roof duality*.

Pour passer à l'optimisation multi-labels, une heuristique est de reformuler en une succession de problèmes binaires où il faut choisir entre une solution courante, et une proposition de label. Chaque étape de fusion sélectionne une solution globale de l'énergie binaire si la fonction est sous-modulaire. Pour autant, en itérant les fusions, il n'y a aucune garantie de convergence vers un minimum local. Nous allons nous intéresser plus particulièrement aux résultats en optimisation de fonctions pseudo-booléennes, car elles ont démontré leur efficacité expérimentale.

En conclusion, nous avons vu que l'optimisation de fonctions pseudo-booléennes quadratiques apporte un cadre complet pour optimiser les énergies rencontrées en traitement d'image.

Ce bref état de l'art, qui a pour but de présenter succinctement les méthodes utilisées dans nos travaux ne représentent qu'une petite partie de l'optimisation de fonctions pseudo-booléennes. Pour aller plus loin, l'article [BHT06] propose de nouvelles méthodes pour l'optimisation de fonctions pseudo-booléennes quadratiques tel que le *probing* qui permet d'améliorer le calcul

de la *roof duality*. La thèse [Str12], soutenue en 2012, propose un état de l'art complet sur l'optimisation discrète appliquée au domaine de la vision pour ordinateur.

3.3.1 Récapitulatif

Pour résumer, les points clés à retenir de l'optimisation de fonctions pseudo-booléennes sont :

- N'importe quelle fonction pseudo-booléenne peut être transformée en une posiforme.
 - Les affectations vérifiant la persistance forte avec la *roof duality* peuvent être trouvées en un temps polynomial.
 - Les affectations persistantes faibles peuvent être trouvées en résolvant la posiforme résiduelle.
- Si la fonction pseudo-booléenne est sous-modulaire, alors on peut trouver le minimum global en un temps polynomial.
- Une fonction pseudo-booléenne quelconque peut toujours être réduite en une posiforme quadratique en ajoutant des variables supplémentaires ayant le même minimum global.
 - Il existe plusieurs réductions possibles.
 - La réduction entraîne l'apparition de termes non sous-modulaires.
 - Le nombre et les coefficients des termes non sous-modulaires dépendent directement de la réduction utilisée.
 - Le nombre et les coefficients des termes non sous-modulaires influent sur la qualité de la minimisation.
- On peut ramener un problème multi-labels en un problème binaire avec les techniques de fusions.

3.3.2 Implantation libre

Plusieurs implantations d'algorithmes sous licence libre sont disponibles pour résoudre différents problèmes abordés dans ce chapitre :

α -expansion :

<http://vision.csd.uwo.ca/code/>.

Optimisation de fonctions pseudo-booléennes quadratique :

<http://pub.ist.ac.at/~vnk/software.html>.

Réduction de fonctions pseudo-booléennes quelconques en degré 2 :

<http://www.f.waseda.jp/hfs/software.html>

Chapitre 4

Restauration des contrastes dans une image

Une interprétation probabiliste de la méthode de restauration d'image en présence de brouillard fondée sur la contrainte *no-black-pixel* est proposée. Cette méthode est décomposée en deux étapes : l'estimation du voile atmosphérique et la restauration du contraste de l'image.

Nous avons vu dans le chapitre 2 que la restauration sur une seule image est un problème mal posé. Pour lever ces ambiguïtés plusieurs *a priori* et contraintes peuvent être ajoutés. Les contraintes sont moins strictes. En effet une contrainte limite l'espace de recherche pour permettre d'obtenir une solution valide. C'est ce qui est fait avec l'algorithme *no-black-pixel* avec la contrainte photométrique sur le voile atmosphérique. Les contraintes sont insuffisantes pour rendre le problème bien posé. Les contraintes étant posées, un *a priori* supplémentaire doit être ajouté pour biaiser la solution vers un certain type de solution. Si l'*a priori* n'est pas correctement choisi ou inadapté à une scène observée, la restauration ne produira pas un résultat de bonne qualité. Le choix de l'*a priori* est donc primordial. Dans le cas de l'algorithme *no-black-pixel*, le voile est maximisé. Un autre *a priori* est introduit avec les contraintes en empêchant avec un facteur entre 0 et 1 le voile d'être trop grand. Mais, cette façon de faire diminue trop fortement la profondeur des zones éloignées et, inversement, de façon insuffisante la profondeur des zones proches. Dans le chapitre 2, le problème de la restauration et reconstruction a été formalisé de façon probabiliste. Dans [NKL12], une transformation algébrique est effectuée sur l'équation de Koschmieder afin d'avoir un lien linéaire entre une variable fonction de l'image de la scène avec brouillard et une variable fonction de l'image restaurée, et la profondeur. Cette transformation non linéaire a pour conséquence que le bruit n'est plus uniforme.

Dans ce chapitre, nous proposons de reformuler la méthode *no-black-pixel* de façon probabiliste, c'est à dire interpréter le voile atmosphérique et l'image restaurée comme des variables aléatoires. Contrairement au modèle déjà proposé dans [NKL12], l'image acquise par le capteur ne subit pas de transformation. L'algorithme d'optimisation est décomposé en deux étapes : le calcul du voile qui minimise l'énergie, et de l'image restaurée sachant le voile.

Cette approche en deux temps apporte une nouvelle façon de biaiser le résultat pour éviter

les zones sombres. L'image restaurée est en effet initialisée à une valeur donnée lors du calcul du voile. Cela présente l'avantage de biaiser uniformément le résultat quelle que soit la profondeur.

Dans la partie 4.1, nous traitons du problème de la restauration connaissant la distance. Dans cette partie, nous cherchons à estimer l'image restaurée sachant la vraie profondeur. Un nouveau terme de régularisation sur l'image restaurée est introduit. Ensuite, le calcul du voile atmosphérique est décrit. Le voile est substitué à la profondeur. Finalement, la partie 4.2 décrit l'inférence du voile atmosphérique. Les différentes contraintes sont rappelées et un modèle est proposé. Des résultats sont montrés sur des images de synthèse et sur des images réelles. L'évaluation de l'algorithme est effectuée sur la base FRIDA2, ce qui permet de comparer qualitativement cet algorithme à ceux de l'état de l'art.

4.1 Restauration de l'image sachant la profondeur

Nous cherchons à estimer l'intensité originale de la scène I_0 pour une image I donnée en supposant que la profondeur D est connue ainsi que le coefficient d'extinction β et l'intensité du ciel I_s . I_s est généralement fixée comme l'intensité maximale de l'image (voir [TPP07]). Pour, β , il peut être estimé avec des méthodes citées dans la partie 2.4.

Quand la carte de profondeur D est connue, ainsi que β et I_s , I_0 peut être estimée directement en résolvant l'équation de Koschmieder (2.1). Cette méthode est valide si I est sans bruit. En présence de bruit, du fait du faible rapport signal à bruit sur les objets lointains, le bruit est fortement amplifié lors de la restauration. La formulation par champ de Markov est donc particulièrement adaptée. Elle permet en effet de mieux prendre en compte le bruit lors de la restauration, sachant la carte de profondeur. La solution proposée dans cette partie donne une restauration lissée. La qualité de l'image restaurée est améliorée sur l'ensemble des distances, surtout en présence de bruit.

Le problème de la restauration du contraste sachant la carte de profondeur peut être formulé comme la maximisation de la probabilité *a posteriori* $P(I_0|D, I)$. En utilisant le théorème de Bayes, la probabilité *a posteriori* peut être réécrite de cette façon :

$$P(I_0|I, D) \propto P(I|I_0, D)P(I_0|D) \quad (4.1)$$

où $P(I|I_0, D)$ est le terme de vraisemblance et $P(I_0|D)$ est l'*a priori* sur I_0 . Comme déjà mentionné, il est préférable de minimiser moins le logarithme de la probabilité plutôt que de maximiser la densité de probabilité. L'énergie dérivée de (4.1) en utilisant le négatif du logarithme est donc :

$$U(I_0|I, D) = U(I|I_0, D) + U(I_0|D) \quad (4.2)$$

Cette énergie est la somme de deux termes, le terme d'attache aux données $U(I|I_0, D)$ et le terme de régularisation $U(I_0|D)$

4.1.1 Terme d'attache aux données

Par définition, le terme d'attache aux données est le logarithme de la probabilité d'avoir l'intensité de l'image observée sachant la profondeur et l'intensité de la scène, en prenant en compte que I_0 est observée à travers le brouillard (2.1). Par conséquent, il peut être réécrit de la façon suivante :

$$U(I|I_0, D) = \sum_{s \in S} \rho_p \left(\frac{|(I_0(s) - I_s)e^{-\beta p(s)} + I_s - I(s)|}{\sigma_p} \right) \quad (4.3)$$

où S est l'ensemble des sites de l'image et ρ_p est une fonction d'erreur liée à la distribution du bruit d'échelle σ_p . Dans l'hypothèse où le modèle de Koschmieder est respecté, une distribution gaussienne est généralement utilisée, par conséquent $\rho_p(t) = t^2$.

4.1.2 Terme a priori

L'*a priori* a pour conséquence de lisser l'image restaurée en pénalisant les grandes différences d'intensités entre les pixels voisins. Il peut être écrit de la façon suivante :

$$U(I_0|D) = \lambda_{I_0} \sum_{\{i,j\} \in C_{I_0}} e^{-\beta p(i)} \rho_{I_0}(\varphi(\{i,j\}, I_0)) \quad (4.4)$$

où C_{I_0} est l'ensemble des cliques binaires entre les variables I_0 (ici 4 *connexités* d'ordre 2) avec $\varphi(c, I_0)$ la fonction pénalisant les grands écarts d'intensité dans le voisinage (1.11), λ_{I_0} est le facteur qui contrôle la force de l'*a priori* comparé au terme d'attache aux données et ρ_{I_0} est une fonction relative à la distribution des gradients d'intensité dans le voisinage en absence de brouillard. En pratique, la fonction valeur absolue pour ρ_{I_0} , qui permet d'avoir une énergie sous-modulaire, donne de bons résultats. Notons que dans [NKL12], une méthode judicieuse est proposée afin d'estimer avec plus de précision les paramètres du terme de régularisation.

En plus de la pénalisation, un terme avec une décroissance exponentielle est introduit dans (4.4). Ce terme est important pour notre modèle. Du fait du coefficient décroissant sur I_0 dans le terme d'attache aux données, l'effet du terme de régularisation diminue avec l'augmentation de la profondeur par rapport à l'effet du terme d'attache aux données. Dans ce cas, le terme d'attache aux données prédomine et l'image devient de plus en plus lisse avec la distance. Pour éviter cet effet, le même terme $e^{-\beta p(s)}$ est introduit dans l'*a priori*. La figure 4.1 montre

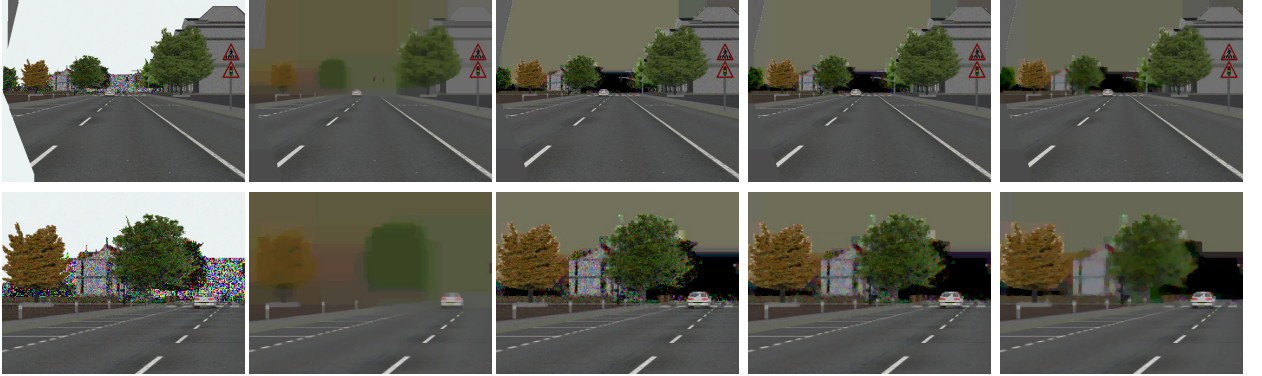


FIGURE 4.1 – Effet des différents paramètres de l'*a priori* sur la restauration à partir de la vérité terrain. De gauche à droite : sans le terme de régularisation i.e $\lambda_{I_0} = 0$, quand $\lambda_{I_0} = 1$ sans le terme décroissant exponentiel. Finalement avec le terme décroissant exponentiel quand $\lambda_{I_0} = 1$, $\lambda_{I_0} = 2$ et $\lambda_{I_0} = 4$.

l'image restaurée obtenue avec différents paramètres à partir de la vérité terrain. La première image correspond à une restauration sans le terme de régularisation, en inversant l'équation de Koschmieder. Puis, avec le terme de régularisation sans la pondération exponentielle, et finalement avec le terme de régularisation ainsi que la pondération pour différentes valeurs de λ_{I_0} . Les résultats montrent que sans le terme de régularisation, la restauration est correcte dans les zones proches, car le signal n'est pas amplifié, mais dès que la distance augmente, le bruit est fortement

amplifié. Quand le terme de régularisation est ajouté, les zones proches sont bien restaurées avec $\lambda_{I_0} = 1$, mais, l'intensité des objets proches s'étale sur l'intensité des objets lointains (nous pouvons observer la disparition des poteaux avec la distance). Avec la pondération exponentielle dans le terme de régularisation, l'effet de lissage est homogène avec la distance et les détails sont bien préservés même sur les objets lointains avec $\lambda_{I_0} = 1$. Quand λ_{I_0} est plus important, le lissage augmente de façon homogène sur la distance. Nous pouvons donc conclure d'après ces tests que le modèle proposé modélise convenablement la distribution de l'image restaurée sachant la distance.

4.1.3 Le choix entre la profondeur et le voile atmosphérique

La carte de profondeur D du modèle markovien peut être remplacée par le voile atmosphérique introduit dans le chapitre 2. Nous rappelons que le lien entre le voile atmosphérique $V(s)$ et la profondeur $p(s)$ en un site s est le suivant :

$$V(s) = I_s(1 - e^{-\beta p(s)}) \quad (4.5)$$

Après substitution de V dans (2.1), l'équation de Koschmieder se réécrit :

$$I = I_0(1 - \frac{V}{I_s}) + V \quad (4.6)$$

Ce qui permet de réécrire le terme d'attache aux données (4.3), comme :

$$U(I|I_0, V) = \sum_{s \in S} \rho_{I_0} \left(\frac{|I_0(s)(1 - \frac{V(s)}{I_s}) + V(s) - I(s)|}{\sigma_p} \right) \quad (4.7)$$

et l'*a priori* (4.4), comme :

$$U(I_0|V) = \lambda_{I_0} \sum_{\{i,j\} \in C_{I_0}} (1 - \frac{V(i)}{I_s}) \rho_{I_0}(\varphi(\{i,j\}, I_0)) \quad (4.8)$$

Finalement, on obtient un modèle de restauration du contraste dans les images en présence de brouillard sachant le voile atmosphérique V . Cette formulation, comme dans le chapitre précédent, présente l'avantage de ne pas nécessiter la connaissance du coefficient d'extinction β .

4.2 Inférence du voile atmosphérique

Dans cette partie, nous proposons un modèle markovien afin d'inférer le voile atmosphérique pour une image en présence de brouillard, l'estimation de V est obtenue comme la minimisation d'une énergie composée d'un terme d'attache aux données et d'un *a priori*.

$$U(V|I) = U(I|V) + U(V) \quad (4.9)$$

Du fait de l'ambiguïté entre l'image restaurée et le voile atmosphérique dans l'équation (4.5), V ne peut pas être estimé à partir d'une seule image. Comme mentionné dans la partie 2.3.2.3, le voile est sujet à plusieurs contraintes, que nous reprenons ici, pour construire l'attache aux données.

4.2.1 Terme d'attache aux données

Afin d'estimer le voile atmosphérique, nous proposons d'utiliser deux contraintes sur le voile présentées dans la partie 2.3.2.3. La contrainte photométrique (2.11) est nécessaire et la contrainte planaire (2.15) qui est uniquement pour les images en environnement routier. Par contre nous n'utilisons pas la contrainte *no-black-pixel* (2.3.2.3).

Le terme d'attache aux données va donc permettre de maximiser la valeur du voile atmosphérique en respectant les bornes des deux précédentes contraintes. En maximisant la valeur de V , on tend indirectement vers un contraste maximum de l'image restaurée. Dans la méthode *no-black-pixel* un coefficient est introduit sur le voile atmosphérique pour le biaiser et empêcher qu'il atteigne le maximum et conduise à des zones noires. En remplacement, nous proposons d'initialiser arbitrairement I_0 par une valeur constante dans (4.6). Cela permet de biaiser la solution vers la valeur de I_0 dans les zones grises plutôt que vers 0.

Le terme d'attache aux données est de la forme :

$$U(I|V) = \sum_{s \in S} \begin{cases} \rho_{p'}\left(\frac{|I_0^b(s)(1 - \frac{V(s)}{I_s}) + V(s) - I(s)|}{\sigma_p'}\right) & \text{si } 0 \leq V(s) \leq \min(V_{route}(s), I(s)) \\ +\infty & \text{sinon.} \end{cases} \quad (4.10)$$

où I_0^b est la valeur initiale de l'intensité, ρ_p' est une fonction croissante arbitraire et σ_p' l'échelle. En pratique, la fonction valeur absolue donne de bons résultats et permet une bonne propagation du voile, ce qui est utile pour les images en noir et blanc. $V_{route}(s)$ est la contrainte planaire (2.15) introduite dans la partie 2.3.2.5.

4.2.2 A priori

Le voile étant lié à la profondeur et comme dans le cadre de la reconstruction 3D, on suppose que les pixels voisins ont une grande chance d'avoir le même voile atmosphérique. Le terme de régularisation pénalisera donc de grands écarts du voile atmosphérique dans chaque voisinage. En conséquence, le terme de régularisation à la forme suivante :

$$U(V) = \lambda_V \sum_{c \in C_V} \rho_V(\varphi(c, V)) \quad (4.11)$$

où C_V est l'ensemble des cliques binaires entre les variables V (ici 4 *connexités* d'ordre 2) avec $\varphi(c, V)$ la fonction pénalisant les grands écarts d'intensité dans le voisinage (1.11), λ_V est la pondération du terme de régularisation par rapport au terme d'attache aux données et ρ_V est une fonction relative à la distribution du voile dans le voisinage.

4.3 Restauration d'une seule image

L'algorithme est décomposé en deux étapes. Le voile atmosphérique V est d'abord obtenu à partir de I en minimisant l'énergie $U(V|I)$ avec I_0^b fixé qui est la somme entre le terme d'attache aux données (4.10) et l'*a priori* (4.11). Puis, quand V est inféré, l'image restaurée est calculée en minimisant l'énergie $U(I_0|I, V)$ qui est la somme entre le terme d'attache aux données (4.7) et l'*a priori* (4.8).

Ces deux énergies ont une structure de Markov « classique », composée d'une part d'une somme de cliques d'ordre 1 pour l'attache aux données, et d'une autre part d'une somme de

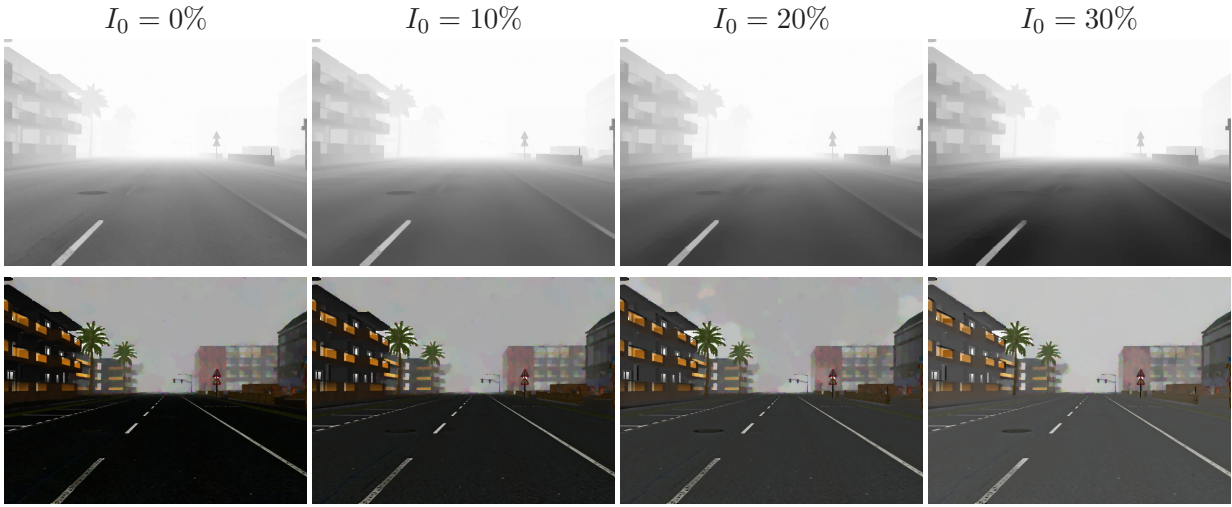


FIGURE 4.2 – Effet de l’initialisation de I_0^b sur la restauration. Première ligne : le pourcentage du blanc du ciel en niveau de gris utilisé pour l’initialisation de I_0^b . Seconde ligne : le résultat de l’estimation du voile atmosphérique. Dernière ligne : Résultat de la restauration.

cliques d’ordre 2 pour le terme de régularisation dans le cas où seuls les voisins directs sont pris en compte. Si les cliques d’ordre 2 ont une fonction d’erreur sous modulaire, alors l’énergie est sous modulaire. Elle peut donc être minimisée avec l’algorithme α -expansion (décrit dans la partie 3.2.3.)

L’algorithme complet est donc le suivant :

- D’abord, le voile V est inféré en minimisant (4.9) avec l’algorithme d’optimisation α -expansion en utilisant la fonction valeur absolue pour ρ_V .
- Ensuite, l’image restaurée est obtenue en minimisant l’énergie (4.2) avec l’algorithme α -expansion en utilisant la fonction valeur absolue pour ρ_{I_0} . Pour les images en couleurs, la restauration est calculée pour chaque canal.

Initialisation de I_0^b : Le choix de I_0^b lors de l’initialisation va permettre de biaiser la solution afin d’éviter un trop grand nombre de zones noires. La figure 4.2 montre différentes restaurations de la même scène avec une initialisation différente. Quand I_0^b est initialisé à 0% de l’intensité en niveau de gris du ciel, l’intensité de l’image dans les zones grises est proche du noir. Néanmoins, nous pouvons voir que les zones de couleurs, grâce à la contrainte photométrique sur le canal le plus clair, sont convenablement restaurées. Plus la valeur initiale de I_0^b augmente et plus l’image restaurée est claire dans les zones grises. Les zones de couleurs, elles, ne varient pas beaucoup. Nous avons trouvé, sur l’ensemble des images en environnement routier testées, que $I_0^b = 0.15 \times I_s$ était un bon compromis.

Images couleurs : Pour plus de concision, l’algorithme est décrit pour une image en noir et blanc. Pour le calcul du voile (4.9), il suffit d’effectuer la somme des potentiels du terme d’attache aux données (4.10) sur chaque canal. Pour le calcul de l’image restaurée, chaque canal est calculé indépendamment.

4.4 Résultats sur des images de synthèse

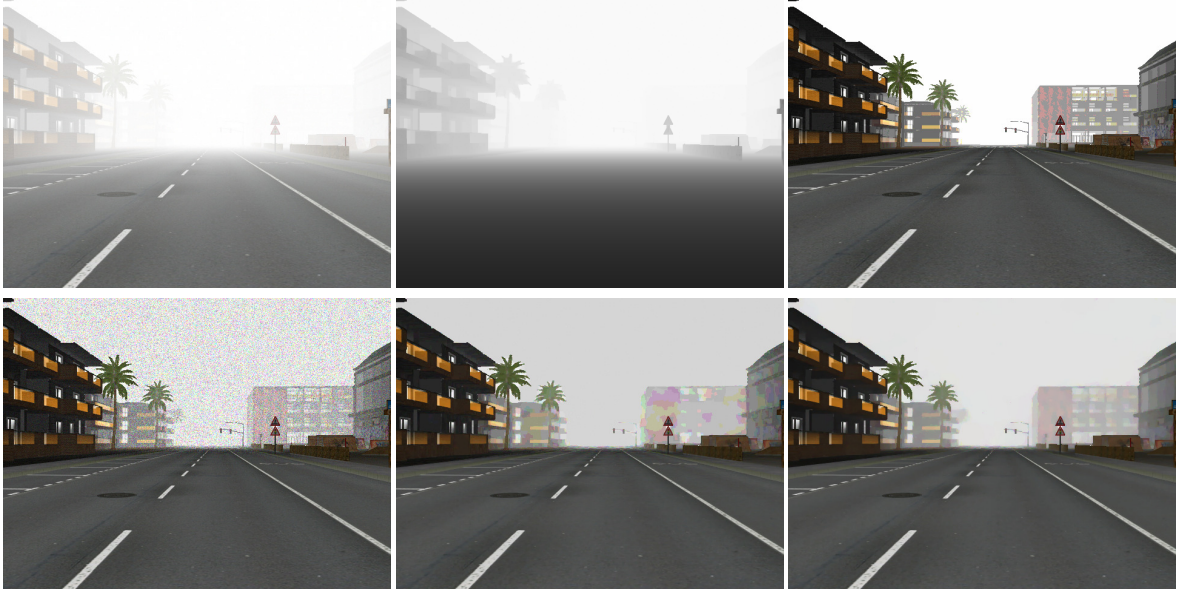


FIGURE 4.3 – Comparaison de différentes restaurations. De gauche à droite, première ligne : l'image en présence de brouillard, le voile atmosphérique obtenu en minimisant (4.9) avec la contrainte planaire, la restauration sans le terme de régularisation et sans bruit ajouté sur l'image. Seconde ligne : la restauration sans régularisation avec un bruit ajouté sur l'image d'entrée, la même méthode mais en lissant l'image d'entrée au préalable, l'algorithme proposé.

La figure 4.3 montre le résultat de l'algorithme pour différents paramètres sur la même image synthétique avec et sans bruit ajouté. Sans bruit, l'image restaurée peut être calculée sans régularisation (avec $\lambda_{I_0} = 0$), c'est à dire en inversant l'équation de Koschmieder. Lorsque l'image est bruitée, le bruit est très amplifié sur les objets à longue distance. L'image peut être lissée avant d'effectuer la restauration, mais nous voyons que le lissage entraîne l'apparition d'artéfacts importants tels que de grandes zones homogènes colorées sur les objets lointains. Quand le terme d'*a priori* est utilisé, l'effet de lissage est assez uniforme, en particulier sur les objets à longue distance où les détails sont les mieux restaurés.

Nous avons évalué et comparé l'algorithme sur la base de données FRIDA2 utilisée dans l'état de l'art (partie 2.3.5). Les paramètres utilisés dans l'estimation du voile sont : $\rho_{p'}(t) = t$, $\sigma_{p'} = 1$ et $\lambda_V = 2$. Les paramètres utilisés dans l'estimation de l'image restaurée sont : $\rho_p(t) = t^2$, $\sigma_p = 1$ et $\lambda_{I_0} = 2$. Les résultats montrent que l'algorithme proposé apporte de meilleurs résultats que les autres. En effet, l'erreur moyenne sur l'ensemble de la base de données de l'algorithme proposé est 19.6 tandis que le meilleur score du précédent algorithme est 19.9 avec la méthode CM+PA qui est aussi adaptée aux scènes routières.

Néanmoins, ces résultats sont obtenus sur des images non bruitées. Or, le bruit, même faible, apparaît de façon importante à longue distance. Afin de tester l'algorithme avec le bruit, un bruit blanc gaussien a été ajouté (avec un écart type de 1) sur toutes les images. L'erreur moyenne est encore plus faible et l'erreur atteinte est de 16.9 en comparaison au cas sans bruit. Cela est probablement dû au fait que le modèle utilisé pour la vraisemblance correspond à la distribution du bruit. Dans ce cas, l'écart entre l'algorithme proposé et ceux de l'état de l'art est encore plus

TABLE 4.1 – Moyenne et erreur absolue entre l'image restaurée et l'image sans brouillard, pour 6 méthodes sur 66 images de synthèse avec un brouillard uniforme et un bruit blanc gaussien d'écart type 1.

(Base d'images FRIDA2)		
Algorithme	erreur moyenne (en niveaux de gris)	erreur moyenne (+ bruit)
Image	81.6 ± 12.3	81.2 ± 12.4
DCP [HST10b]	46.3 ± 15.6	49.2 ± 14.4
FSS [HTA10]	34.9 ± 15.1	non disponible
NBPC [TH09]	50.8 ± 11.5	50.4 ± 11.8
NBPC+PA [THC ⁺ 10]	31.1 ± 10.2	26.4 ± 8.7
CM+PA [HCH11]	19.9 ± 11.1	19.1 ± 6.7
Proposé+PA [CT13]	19.6 ± 10.7	16.9 ± 5.1

important. Grâce au lissage, l'algorithme proposé est aussi plus stable avec un écart type de 5.1 niveau de gris, comparé à 10.7 dans le cas sans bruit. Quelques exemples de restauration sur la base d'images FRIDA2 sont montrés dans la figure 4.4. On peut remarquer que les véhicules au loin sont bien plus visibles sur l'image restaurée, et surtout que le traitement est homogène avec la distance.

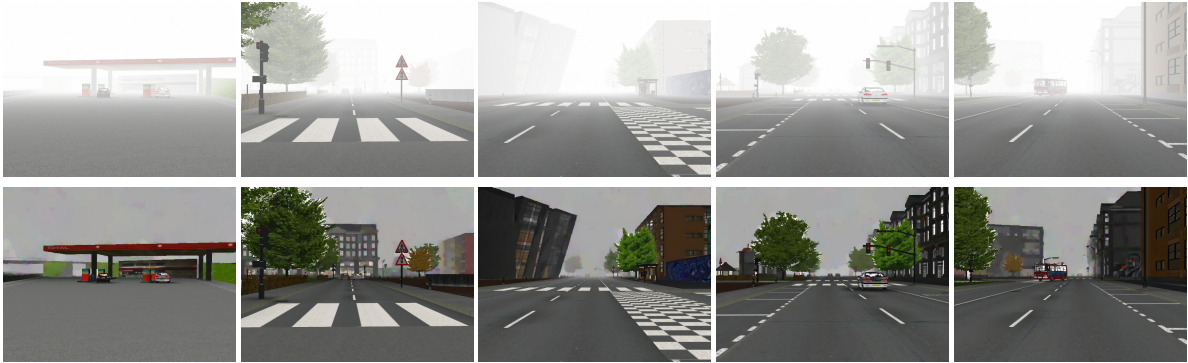


FIGURE 4.4 – Résultats de l'algorithme sur des images de synthèse de la base FRIDA2. Première ligne : l'image avec un brouillard homogène et un bruit blanc gaussien. Seconde ligne : l'image restaurée obtenue avec l'algorithme proposé.

4.5 Résultat sur des images réelles

De bons résultats sont aussi obtenus sur les images réelles comme illustré sur la figure 4.5. Le coefficient du terme de régularisation λ_{I_0} est fixé à la valeur 2 pour ces résultats. Nous pouvons remarquer que même avec le terme de régularisation, les détails à longue distance sont bien préservés.

Dans la figure 4.6, nous comparons les résultats de l'algorithme proposé avec ceux de l'état de l'art : [Tan08], DCP [HST10b], NBPC [TH09], NBPC+PA [THC⁺10]. Seul NBPC+PA est dédié aux images de routes avec la contrainte planaire comme l'algorithme proposé. On remarque que le résultat obtenu par la méthode proposée est proche de NBPC+PA. Par contre, le bruit est

mieux pris en compte.

4.6 Conclusion

Dans ce chapitre, nous proposons un nouveau modèle markovien pour la restauration du contraste dans une image avec du brouillard. Ce modèle conduit à une optimisation en deux étapes par l'algorithme α -expansion dans le cas où l'énergie est sous modulaire. Comme dans [THC⁺10], la contrainte de la route plane peut être incluse dans les images de route. Ceci améliore considérablement les résultats dans les zones homogènes de la chaussée.

La première étape consiste à estimer le voile atmosphérique. Après l'estimation du voile, l'image restaurée peut être calculée en optimisant une autre énergie sachant le voile atmosphérique.

Deux points clés ont été améliorés. Le modèle probabiliste permet de prendre en compte de façon satisfaisante l'effet du bruit dans l'étape de restauration. Afin d'éliminer les zones sombres après la restauration, I_0^b peut être fixé pour biaiser le résultat vers cette valeur. L'avantage est que le biais est homogène avec la distance. L'évaluation sur les images de synthèse montre les améliorations apportées par la méthode proposée par rapport à l'état de l'art en restauration du contraste sur une seule image avec un brouillard homogène.

Dans la méthode proposée, le problème a été décomposé en deux étapes en utilisant la méthode *no-black-pixel*. Néanmoins, il pourrait aussi être posé sous la forme d'une seule énergie, dans le même esprit que [NKL12] en utilisant le même processus d'optimisation. Le problème étant mal posé, il est nécessaire d'introduire un *a priori* sur la solution indiquant quel est le pourcentage de blanc qui appartient à la couleur et l'autre au brouillard.

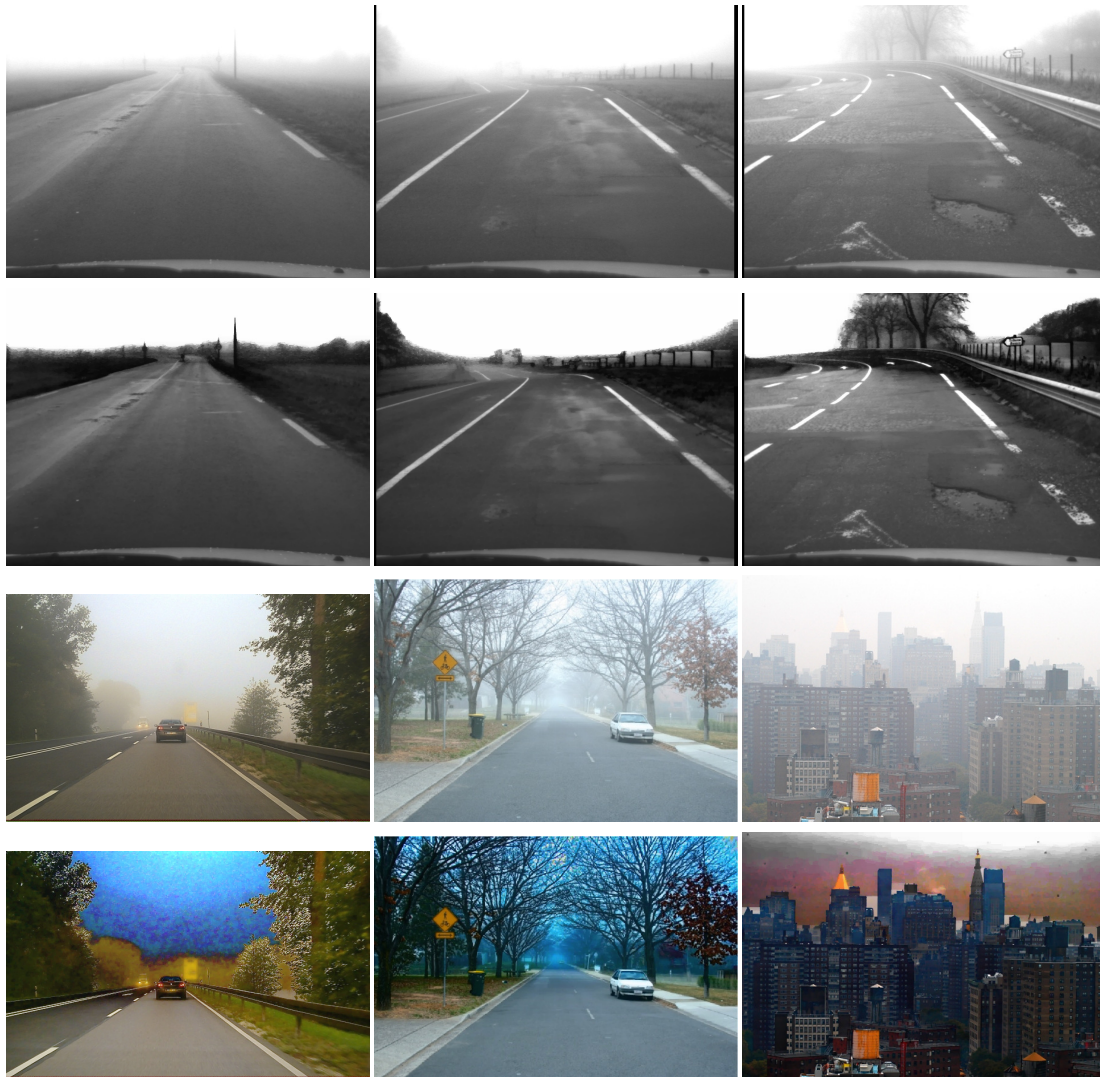


FIGURE 4.5 – Exemple de restauration sur des images réelles. Première et troisième ligne : les images originales. Seconde et quatrième ligne, l'image restaurée avec l'algorithme proposé. La contrainte du monde plan est utilisée pour les cinq premières images. La quatrième image appartient à Audi, la cinquième image est proposée dans [Tan08].

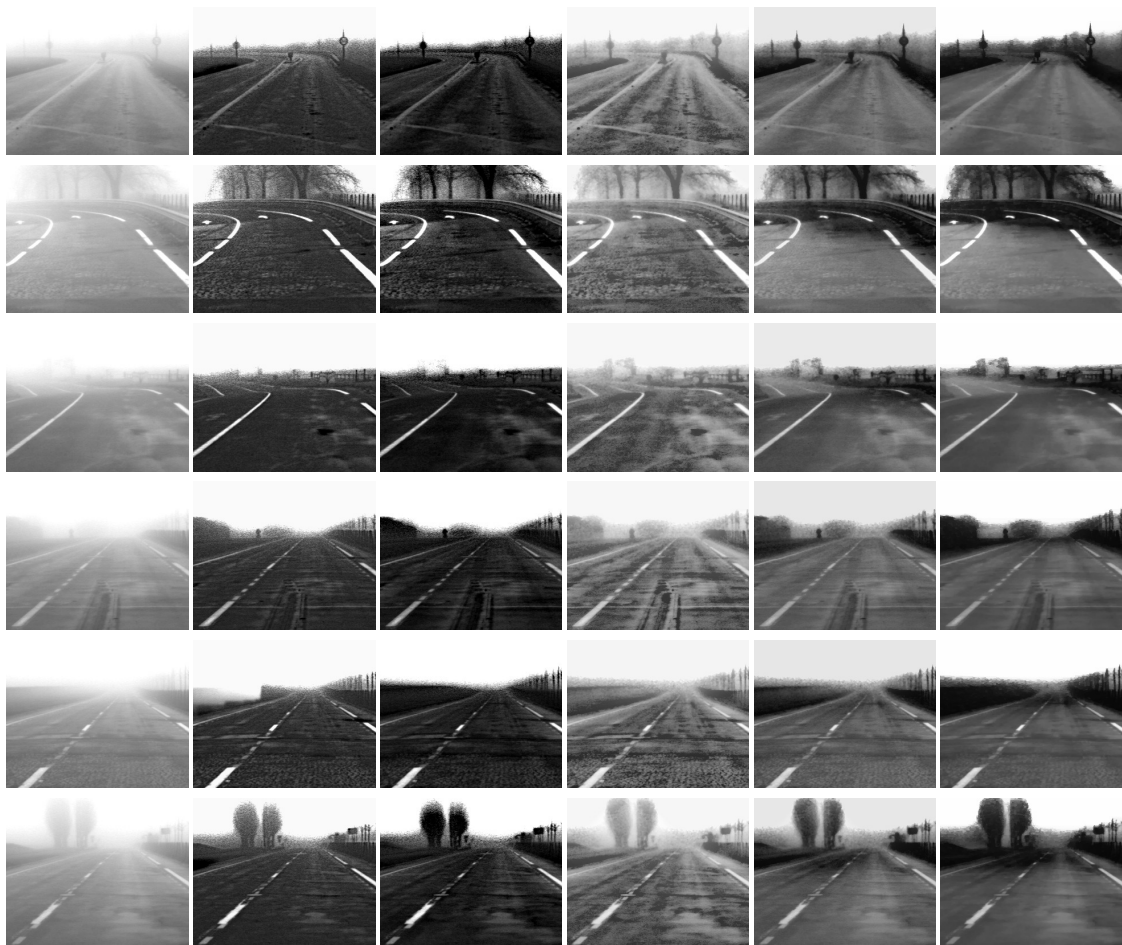


FIGURE 4.6 – Comparaison sur des images réelles. De la colonne gauche à droite : L'image originale avec brouillard, le résultat en utilisant l'algorithme [Tan08], DCP [HST10b], NBPC [TH09], NBPC+PA [THC⁺10] et enfin avec la méthode proposée.

Chapitre 5

Stéréo reconstruction et restauration simultanées d'images en présence de brouillard

Le problème de la reconstruction 3D à partir de paires stéréoscopiques est abordé en présence de brouillard. Un modèle complet est proposé pour estimer simultanément la carte de profondeur et l'image restaurée.

Un des principaux problèmes de la reconstruction 3D en présence de brouillard est l'ambiguïté importante à longue distance. En effet, avec la distance, le contraste diminue fortement. La figure 5.1 montre plusieurs algorithmes testés sur des images synthétiques de la base de données FRIDA3 (introduite dans la partie 5.3.2). Nous pouvons constater qu'à partir d'une certaine distance, la carte de disparité est très mal estimée. Cela est dû à la forte ambiguïté provoquée par le voile atmosphérique qui s'accroît de façon importante à partir d'une certaine distance. Si nous regardons la figure 5.1, les résultats obtenus par la méthode de modélisation markovienne avec l'optimisation par α -expansion, nous pouvons voir qu'avec la distance, le terme de régularisation devient prédominant sur le terme d'attache aux données à cause de cette ambiguïté. Les méthodes comme Elas [GRU10], fondées sur l'appariement de points d'intérêt, produisent de mauvais résultats à cause des mauvais appariements à longue distance.

Pour lever cette ambiguïté, nous proposons d'estimer en plus de la carte de disparité, l'image restaurée. En effet, nous avons vu dans le chapitre précédent que l'intensité d'une scène observée en un pixel est souvent proche de celles de ses voisins. C'est cette hypothèse qui, dans le chapitre précédent, nous a permis de générer des images au contraste restauré avec une atténuation du bruit. En estimant l'image restaurée en plus de la carte de profondeur, on favorisera les profondeurs produisant une image restaurée avec des intensités proches.

Nous avons noté dans l'état de l'art au chapitre 2, que l'estimation de l'intensité originale de la scène n'est pas simple dans le cas monoculaire à cause de l'ambiguïté entre la profondeur et l'intensité originale de la scène en un pixel. Or, la stéréovision apporte un indice de profondeur. Nous allons montrer qu'avec la connaissance du coefficient d'extinction du brouillard et avec un système de caméras calibrées, il est possible de mieux retrouver l'intensité originale de la scène en plus de la reconstruction.

Pour cela, nous proposons un modèle markovien complet liant la carte de disparité avec l'image restaurée et les images observées. Grâce à l'indice de profondeur, l'intensité originale de la scène peut être retrouvée, même aux distances proches. Les images restaurées seront sujettes à la régularisation comme dans le chapitre précédent. Ainsi, grâce au voile atmosphérique et à l'image restaurée, les profondeurs lointaines seront aussi estimées.

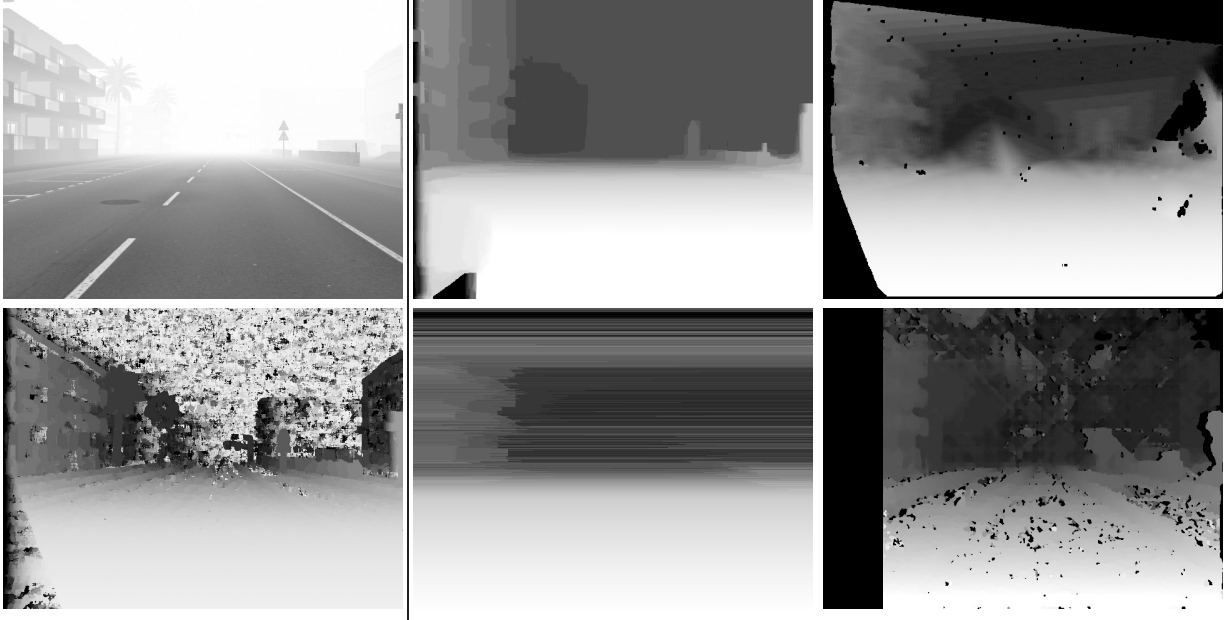


FIGURE 5.1 – Exemple de reconstruction à partir de paires stéréoscopiques sur une scène en présence de brouillard. Première ligne, de gauche à droite : l'image originale, la carte de disparité obtenue en utilisant l'algorithme α -expansion [BVZ01b], Elas [GRU10]. Seconde ligne : l'appariement par fenêtres, la programmation dynamique sur chaque ligne et une méthode co-opérative [Hir05].

Dans la partie 5.1, nous rappelons la formulation probabiliste de l'estimation de la carte de disparité et de l'image restaurée, sachant l'image gauche et droite. Ensuite, un premier modèle est introduit dans la partie 5.2 ainsi qu'un algorithme d'optimisation approprié permettant de converger vers un minimum local intéressant. Puis, la technique de fusion est appliquée au modèle actuel pour pouvoir prendre en compte la régularisation du second ordre. Une évaluation sur images de synthèse est réalisée dans la partie 5.3 ainsi que des tests sur des images caméras.

5.1 Formulation

Dans le chapitre 1, nous avons vu que le problème de la reconstruction à partir de paires stéréoscopiques peut être posé de façon générale dans le cadre probabiliste comme la maximisation de la probabilité *a posteriori* d'avoir la carte de profondeur et l'intensité originale de la scène $\{D, I_0\}$ sachant les images gauche et droite $\{I_L, I_R\}$ acquises par la caméra (1.7).

Dans le cas de la reconstruction à partir de paires stéréoscopiques avec un ciel dégagé, l'intensité observée des objets n'est que très peu perturbée. On peut donc supposer que l'une des deux images correspond à l'intensité recherchée comme cela est fait habituellement en reconstruction. En présence de brouillard, ce n'est pas le cas. Par conséquent, nous proposons de laisser I_0 en

inconnue et d'estimer la carte de profondeur et l'intensité originale de la scène sachant l'image gauche et droite. Cela donne en terme probabiliste l'*a posteriori* suivant à maximiser :

$$P(D, I_0 | I_L, I_R) \propto P(I_L, I_R | D, I_0) P(D, I_0) \quad (5.1)$$

où $P(I_L, I_R | D, I_0)$ est le terme de vraisemblance et $P(D, I_0)$ l'*a priori* sur le couple de variables $\{D, I_0\}$. Comme précédemment, on cherchera à maximiser de façon équivalente moins le logarithme de la probabilité, ce qui donne :

$$U(D, I_0 | I_L, I_R) = U(I_L, I_R | D, I_0) + U(D, I_0) \quad (5.2)$$

5.2 Modèle pour la reconstruction et restauration à partir de paires stéréoscopiques

Afin de créer le modèle complet, nous proposons de fusionner le modèle probabiliste de la reconstruction à partir de paires stéréoscopiques vu dans la partie 1.7 et celui de la restauration d'image en présence de brouillard fondé sur l'équation de Koschmieder qui a été introduite dans le chapitre précédent.

5.2.1 Modèle par champ de Markov

5.2.1.1 Terme d'attache aux données

Pour la reconstruction à partir de paires stéréoscopiques en présence de brouillard, il faut introduire le modèle de Koschmieder (2.1) dans le modèle de la stéréo-vision où I_0 est inconnue (1.7) sur chaque image. Pour cela, la profondeur en un site $p(s)$ est substituée par la disparité $D(s)$ suivant $D(s) = \frac{\delta}{p(s)}$. Ce qui nous donne le terme de vraisemblance suivant :

$$\begin{aligned} U_{stéréo_kosh}(I_L, I_R | D, I_0) = \sum_{s \in S} \rho_p \left(\frac{|I_0(s) e^{\frac{-\beta \delta}{D(s)}} + I_s(1 - e^{\frac{-\beta \delta}{D(s)}}) - I_L(s)|}{\sigma_p} \right) \\ + \rho_p \left(\frac{|I_0(s) e^{\frac{-\beta \delta}{D(s)}} + I_s(1 - e^{\frac{-\beta \delta}{D(s)}}) - I_R(s, D(s))|}{\sigma_p} \right) \end{aligned} \quad (5.3)$$

Lorsque $\beta = 0$, nous retrouvons le terme général de la reconstruction à partir de paires stéréoscopiques quand il n'y a pas de brouillard (1.7). Notons, que quelle que soit la valeur de I_0 , l'appariement sur les images en présence de brouillard doit aussi être respecté. En effet, pendant l'estimation simultanée de I_0 et D , la valeur I_0 peut être éloignée de la valeur recherchée. Pour cela, nous proposons d'ajouter au terme (5.3), une contrainte forçant l'appariement sur les images avec du brouillard à être respecté :

$$U_{stéréo} = U_{stéréo_kosh} + \alpha U_{stéréo_direct} \quad (5.4)$$

où $U_{stéréo_direct}$ est le terme d'attache de la stéréovision appliqué directement sur les deux images d'entrée, c'est à dire quand I_L est utilisé comme approximation de I_0 comme dans l'équation (1.8). α est un coefficient donnant plus ou moins d'importance à cette contrainte. Grâce à cette contrainte, même si l'estimation de I_0 est erronée, les cartes de profondeurs seront biaisées vers celles dont l'appariement sur les images en présence de brouillard sont respectées.

5.2.1.2 Contrainte photométrique et *a priori* sur les pixels blancs

Comme pour la restauration, le voile est sujet à plusieurs contraintes. Nous pouvons donc contraindre l'espace des solutions en ajoutant les contraintes sur le voile (2.11). Cette contrainte doit être vérifiée sur l'image gauche et l'image droite.

Finalement, nous proposons de biaiser vers la disparité minimale les pixels ayant une intensité proche du ciel. Le terme de vraisemblance final sera donc :

$$U(I_L, I_R | D, I_0) = \begin{cases} U_{\text{stéréo}} & \text{si } V(s) \leq I_L(s) + 3\sigma_p \\ & \text{et } V(s) \leq I_R(s, D(s)) + 3\sigma_p \\ & \text{et } I_L(s) \neq I_s \\ 0 & \text{si } I_L(s) = I_s \text{ et } D(s) = 0 \\ +\infty & \text{sinon.} \end{cases} \quad (5.5)$$

5.2.1.3 Terme *a priori*

Dans l'équation (5.1), la probabilité *a priori* $P(D, I_0)$ dépend de deux variables : la disparité D et l'intensité I_0 . Cette structure d'*a priori* rend l'énergie difficile à optimiser. Comme vu précédemment, ces deux variables sont liées. En effet, les sauts de disparité sont souvent corrélés avec les forts gradients.

Pour pallier cela, nous proposons de réécrire le terme *a priori* de la façon suivante $P(D, I_0) = P(D | \check{I}_0) P(I_0 | \check{D})$, où $\check{D}$ et $\check{I}_0$ sont respectivement une approximation fixe de D et de I_0 . De cette façon, les variables sont découplées. Il est donc possible d'écrire l'énergie de l'*a priori* comme une somme de l'*a priori* sur la carte de profondeur (1.10) et de l'*a priori* sur l'image restaurée (4.11). Avec cette approche, nous retrouvons les *a priori* introduits dans les chapitres précédents : $P(D | \check{I}_0)$ correspondant à l'*a priori* sur la carte de disparité introduit dans la partie 1.2.1.5 et $P(I_0 | \check{D})$ correspond à l'*a priori* sur l'image restaurée introduit dans la partie 4.1.2, ce qui nous donne :

$$U(D, I_0) = \sum_{c \in C_{I_0}} \lambda_{I_0} e^{-\frac{\beta \delta}{D(c)}} W_{I_0}(c, \check{D}) \rho_{I_0}(\varphi(c, I_0)) \\ + \sum_{c \in C_D} \lambda_D W_D(c, \check{I}_0) \rho_D(\varphi(c, D)) \quad (5.6)$$

où C_{I_0} est l'ensemble des cliques sur l'image correspondant à l'intensité originale I_0 et C_D l'ensemble des cliques de la carte de disparité. Comme précédemment, les fonctions $\rho_{I_0}(\varphi(s, I_0))$ et $\rho_d(\varphi(c, D))$ poussent, respectivement l'image restaurée et la carte de profondeur, à avoir des valeurs proches dans chaque voisinage. Les facteurs λ_{I_0} et λ_D contrôlent respectivement l'importance donnée à l'*a priori* sur l'image restaurée et sur la carte de disparité.

$W_D(c, \check{I})$ et $W_{I_0}(c, \check{D})$ permettent de faire correspondre les discontinuités de la carte de disparité avec les forts gradients de la scène, et vice versa. En effet, la supposition sous-jacente est que la scène est composée d'objets et qu'il y a plus de chance d'observer un saut de disparité ou un changement d'intensité entre deux objets.

Pour cela, il faut donc choisir pour le terme de régularisation sur la disparité une fonction dont le poids décroît en fonction des gradients de l'image restaurée. Pour le terme de régularisation sur l'image restaurée, il faut une fonction dont le poids décroît en fonction des sauts de disparités.

En conséquence notre choix est : $W_D(c, \check{I}) = e^{-\frac{|\varphi(c, \check{I})|}{\sigma_{I_0}}}$ et $W_{I_0}(c, \check{D}) = e^{-\frac{|\varphi(c, \check{D})|}{\sigma_D}}$.

5.2.1.4 Initialisation de $\ddot{D}$ et de $\ddot{I}_0$

Les variables $\ddot{D}$ et $\ddot{I}_0$ doivent donc être une approximation de la carte de disparité et de l'image restaurée respectivement dans le terme $U(D, I_0)$. Les pondérations W_D et W_{I_0} sont introduites pour atténuer la régularisation au niveau des contours. Nous avons trouvé que le voile atmosphérique V , que ce soit pour la profondeur $\ddot{D}$ ou pour l'image restaurée $\ddot{I}_0$ est un bon substitut. Nous proposons donc de remplacer $\ddot{D}$ et $\ddot{I}_0$ par $\ddot{V}$ qui est une approximation du voile V .

Par définition, à partir de (2.1), $V = 1 - e^{-\frac{\beta\delta}{D}}$. Par conséquent, $\ddot{D}$ peut être obtenu à partir de $\ddot{V}$. Cela implique que le facteur $e^{-\frac{\beta\delta}{D}}$ dans l'équation (5.6) peut être substitué par $1 - \ddot{V}$. Pour estimer $\ddot{V}$, nous pouvons utiliser la méthode proposée dans le chapitre précédent expliquée dans la partie 4.2.

Une autre conséquence est que $\ddot{D}$ dans W_{I_0} est substitué par $\ddot{V}$ ainsi que $\ddot{I}_0$ dans W_D . Ces fonctions doivent relâcher le terme de régularisation au niveau des gradients. Or, le voile V ajoute un gradient sur les zones homogènes de I_0 . Par conséquent, afin de ne pas relâcher le terme de régularisation sur les zones homogènes en présence de voile, le gradient du voile est utilisé plutôt que le voile lui-même. Par conséquent, soit $\ddot{I}_0$ est remplacé par $\nabla\ddot{V}$ dans le poids W_D et $\ddot{D}$ est remplacé par $\nabla\ddot{V}$ dans le poids W_{I_0} .

5.2.1.5 Modèle complet

Pour résumer, le problème de la reconstruction et de la restauration simultanée minimise l'énergie suivante :

$$\min_{D, I_0} [U(I_L, I_R | D, I_0) + U(D, I_0)] \quad (5.7)$$

En pratique, la fonction valeur absolue est choisie pour les fonctions ρ_D et ρ_{I_0} . Le bruit sur l'image est supposé gaussien, par conséquent, ρ_p est la fonction quadratique. ρ_s est une fonction de la *famille des exponentielles lissées*. Pour les pixels qui vérifient la contrainte photométrique et qui ont une intensité différente de celle du ciel, l'énergie finale qui est minimisée est :

$$\begin{aligned} U(D, I_0, \sigma_p) = & \sum_{s \in S} \left\{ \frac{1}{\sigma_p^2} \left(|I_0(s)e^{\frac{-\beta\delta}{D(s)}} + I_s(1 - e^{\frac{-\beta\delta}{D(s)}}) - I_L(s)|^2 \right. \right. \\ & \left. \left. + |I_0(s)e^{\frac{-\beta\delta}{D(s)}} + I_s(1 - e^{\frac{-\beta\delta}{D(s)}}) - I_R(s, D(s))|^2 \right) \right. \\ & \left. + \alpha \rho_s \left(\frac{|I_L(s) - I_R(s, D(s))|}{\sigma_S} \right) \right. \\ & \left. + 4 \log(\sigma_p) \right\} \\ & + \sum_{\{p, q\} \in C_{I_0}} \lambda_{I_0} (1 - \ddot{V}(p)) e^{-\frac{|\varphi(\{p, q\}, \nabla \ddot{V})|}{\sigma_{I_0}}} |I_0(p) - I_0(q)| \\ & + \sum_{\{p, q\} \in C_D} \lambda_D e^{-\frac{|\varphi(\{p, q\}, \nabla \ddot{V})|}{\sigma_D}} |D(p) - D(q)| \end{aligned} \quad (5.8)$$

Cette énergie est définie à un facteur d'échelle près, elle peut être arbitrairement divisée par λ_{I_0} . Pour prendre en compte les images en couleur, comme pour la méthode du chapitre précédent, il suffit de sommer les coûts de chaque canal couleur et par conséquent, d'appliquer la contrainte photométrique sur tous les canaux.

5.2.2 Optimisation

Afin d'optimiser D et I_0 dans (5.8), une optimisation alternée est réalisée : la première étape consiste à minimiser par rapport à D et la suivante à minimiser par rapport à I_0 . Quand un des deux champs est fixé, on retombe dans le cas d'une formulation classique avec un seul champ de Markov. En effet, le terme de vraisemblance devient un terme de degré 1. De plus, sur les deux termes, l'énergie est sous modulaire, et l'algorithme α -expansion peut donc être utilisé pour l'optimisation. Avant d'effectuer la première étape, le modèle est initialisé avec le résultat de l'algorithme de reconstruction stéréo supposant l'absence de brouillard.

Comme souligné dans [NKL12], la distribution d'une image avec brouillard peut changer d'une image à l'autre. Cela implique qu'après la division de (5.8) par λ_{I_0} , le facteur $\sigma_p \sqrt{\lambda_{I_0}}$ peut être différent d'une image à l'autre. Grâce à la fonction quadratique utilisée sur le terme d'attache aux données dans (5.8), le facteur $\sigma_p \sqrt{\lambda_{I_0}}$ qui minimise l'énergie peut être facilement estimé en calculant l'écart type des résidus du terme de vraisemblance (5.3). Avec cette formulation, il est donc possible d'optimiser $\sigma_p \sqrt{\lambda_{I_0}}$, en plus de I_0 et de D .

Pour résumer, l'algorithme d'optimisation est le suivant :

- β est précédemment calculé par une méthode annexe.
- Calculer $\ddot{V}$ en minimisant (4.9) par α -expansion.
- Initialiser D en minimisant l'énergie de la reconstruction classique (1.14) sur les images en présence de brouillard avec les contraintes photométriques (5.5) en fonction de D par α -expansion.
- Jusqu'à convergence, itérer :
 - Jusqu'à convergence, itérer :
 - Minimiser (5.8) en fonction de I_0 avec l'algorithme α -expansion.
 - Minimiser (5.8) en fonction de D en utilisant l'algorithme α -expansion.
 - Mettre à jour $\sigma_p \sqrt{\lambda_{I_0}}$ en calculant l'écart type des résidus du terme de vraisemblance.

La figure 5.2 montre la carte de disparité et l'image restaurée au cours des itérations de l'algorithme d'optimisation ainsi que l'évolution de l'énergie du modèle, l'erreur entre la carte de disparité par rapport à la vérité terrain et de l'image restaurée par rapport à la vérité terrain, au cours des itérations. La seconde ligne montre le résultat de l'algorithme de reconstruction stéréoscopique classique (1.14), c'est à dire en n'utilisant que le terme $U_{stéréo_direct}$ pour la vraisemblance du modèle avec la restauration obtenue relative à la carte de disparité avec une échelle proche de celle du bruit. Nous pouvons constater, qu'après une certaine distance, le terme de régularisation l'emporte sur le terme de vraisemblance. Cela a pour effet de créer un plan vertical au-delà d'une certaine distance. La troisième ligne montre l'initialisation de l'algorithme, le résultat est amélioré d'une part grâce aux contraintes photométriques (la zone noire en bas à gauche de l'image a disparu grâce à la contrainte photométrique car elle ne peut pas se situer à l'infini étant donné que son intensité est trop sombre). D'autre part, grâce aux biais des pixels blancs à l'infini, la carte de profondeur est nettement améliorée au loin. Les lignes suivantes montrent l'évolution au cours des itérations. L'algorithme étant initialisé avec $\sigma_p \sqrt{\lambda_{I_0}} = 20$, la première restauration est très lissée. À la seconde itération, une amélioration notable est visible à courtes et moyennes distances. La profondeur autour des poteaux à gauche de l'image est mieux reconstruite, ainsi que la chaussée. Au fur et à mesure que l'échelle $\sigma_p \sqrt{\lambda_{I_0}}$ diminue, la restauration et la carte de profondeur s'améliorent avec la distance. Notons que même si une

grande valeur de $\sigma_p \sqrt{\lambda_{I_0}}$ a été choisie au départ, grâce au terme stéréo standard, nous constatons que la solution ne diverge pas de façon aberrante.

Afin d'illustrer l'utilité de l'ajout du terme d'attache aux données directement sur les images réelles, la figure 5.3 montre une comparaison des résultats avec différentes valeurs des paramètres. Les trois premières colonnes montrent le résultat quand $\alpha = 0$ pour les valeurs suivantes de l'échelle : $\sigma_p \sqrt{\lambda_{I_0}} = 2$, $\sigma_p \sqrt{\lambda_{I_0}} = 5$ et $\sigma_p \sqrt{\lambda_{I_0}} = 10$. Nous pouvons voir qu'avec une échelle égale à 2, la solution après convergence est relativement proche de l'initialisation. Avec l'échelle égale à 5, nous pouvons voir une amélioration significative au loin, néanmoins, le terme de régularisation de la carte de profondeur prend le dessus. Cela est dû aux grandes échelles, ce qui fait converger la solution vers un résultat non satisfaisant. On note aussi que la carte de profondeurs est biaisée dans certaines zones. Avec une échelle égale à 10, la solution converge vers une solution très loin de la solution recherchée. La dernière colonne correspond à l'algorithme avec $\alpha = 4$ ainsi que $\sigma_p \sqrt{\lambda_{I_0}} = 20$. Dans ce cas, la solution est plus satisfaisante.

5.2.3 Raffinement avec un *a priori* du second ordre sur la disparité

Dans le modèle précédent, grâce au premier ordre, chaque champ de Markov peut être optimisé efficacement par α -expansion car il y a la propriété de la sous modularité de l'énergie minimisée à chaque étape. Par ailleurs, nous savons que l'ordre premier favorise les solutions fronto-parallèles. Cela n'est pas très adapté aux scènes routières. Le second ordre est préférable pour la reconstruction 3D de l'environnement routier. Cependant, nous avons vu dans l'état de l'art qu'il n'est pas possible d'optimiser globalement une énergie avec un terme du second ordre (qui est une fonction non sous modulaire) en un temps polynomial. Néanmoins, une affectation partielle peut être effectuée avec la méthode de fusion.

L'avantage de l'approche par fusion, introduite dans la partie 3.2.1, est de permettre de fusionner deux solutions en choisissant localement celle qui minimise l'énergie totale. Cette caractéristique intéressante nous permet d'utiliser d'une nouvelle façon la contrainte planaire introduite dans le chapitre précédent. Comme le montre l'exemple 1.3, le plan de la chaussée peut être fusionné avec une carte de profondeur actuelle. Ainsi, nous constatons avec cet exemple que le résultat minimisant l'énergie prend en compte les spécificités des deux solutions.

Cette technique peut être appliquée dans le cas avec brouillard avec une régularisation du second ordre. La figure 5.4 montre un exemple de fusion entre deux solutions, d'une part le résultat de l'algorithme précédent avec le premier ordre, et d'une autre part, le plan de la route et l'image restaurée obtenue avec l'algorithme de restauration du chapitre précédent. La figure montre que l'algorithme à l'ordre premier traite difficilement le bas de l'image. Cela est dû au choix du terme de régularisation qui favorise un plan fronto-parallèle. Une contrainte forte comme celle utilisée pour la restauration n'est pas envisageable dans ce cas. En effet, l'un des premiers intérêts de la reconstruction 3D à partir de paires stéréoscopiques est de détecter les obstacles et les trous dans la chaussée. Avec l'approche par fusion, la meilleure des deux propositions est sélectionnée. Si l'énergie est minimale dans une des cartes de profondeur où une discontinuité est observée, localement elle sera préservée dans le résultat de la fusion.

Nous proposons donc d'utiliser l'approche par fusion pour affiner la solution actuelle. Pour cela, le modèle du monde plan introduit dans [THC⁺10] est une proposition D^p pour la carte de disparité D . L'image obtenue avec l'algorithme de restauration peut aussi être utilisée comme la proposition I_0^p pour I_0 .

Étant donné qu'il s'agit de fusionner une paire de labels I_0^p et D^p , nous proposons un algorithme de fusion à partir de paires des variables actuelles $\{D, I_0\}$ avec les solutions $\{D^p, I_0^p\}$. Cette approche a pour conséquence d'avoir une énergie avec des cliques d'ordre 1 lors de la

fusion, et donc, de pouvoir calculer efficacement la *roof duality*.

La figure 5.4 montre le résultat de la fusion avec la régularisation du premier ordre et la restauration avec la contrainte du monde plan. Le résultat de la fusion est très satisfaisant. En effet, en bas de l'image, la carte de profondeur proposée a été conservée, ainsi que les détails proches des jonctions entre les deux cartes tels que les trottoirs. Notons que les bâtiments et les autres objets hors de la chaussée n'ont pas été modifiés.

5.3 Évaluation

5.3.1 Initialisation des paramètres

La méthode proposée est principalement paramétrée par α qui pondère le poids accordé à la régularisation $U_{photo_stéréo}$ de la stéréo classique sans brouillard. Quand α est proche de zéro, la carte de disparité obtenue peut être erronée si I_0 n'est pas bien estimée. Quand la valeur de α est très importante, la disparité obtenue est de bonne qualité dans les zones proches. Néanmoins, le terme de vraisemblance n'a plus beaucoup d'effet sur la carte de disparité. En conséquence, nous avons estimé que $\alpha = 4$ donne un bon compromis. Un autre paramètre important est l'initialisation de $\sigma_p \sqrt{\lambda_{I_0}}$, le choix de cette valeur peut conduire à converger vers un autre minimum local. Plus la valeur de $\sigma_p \sqrt{\lambda_{I_0}}$ est importante à l'initialisation, plus la carte de profondeur sera lissée après convergence.

5.3.2 Sur images de synthèse

Pour évaluer la reconstruction à partir de paires stéréoscopiques en présence de brouillard, nous nous sommes appuyés sur des images de synthèse compte tenu de la difficulté de disposer d'une référence sur la 3D de la scène. Cela permet aussi d'avoir la vérité terrain, l'image sans brouillard, pour faire l'évaluation de la restauration. Nous avons généré 66 images en présence de brouillard en utilisant le logiciel SiVIC. Un bruit blanc gaussien est ajouté sur chaque pixel des deux images avec un écart type de 1. Cette base de données est appelée FRIDA3, elle est disponible en ligne⁵. La figure 5.5 montre des résultats sur la base de données FRIDA3. Notons que l'intensité de l'image restaurée à l'infini a une distribution équiprobable, nous la fixons donc à une valeur arbitraire pour la visualisation. Nous avons choisi I_s .

Pour évaluer la reconstruction à partir de paires stéréoscopiques, nous avons comparé le résultat de l'algorithme proposé avec deux autres algorithmes de reconstruction 3D à partir de paires stéréoscopiques : L'algorithme Elas [GRU10]⁶ (ELAS), l'approche classique par champ de Markov décrite dans l'exemple 1.2 (MRF_STEREO) et l'approche coopérative [Hir05]⁷ (SGBM). Ces algorithmes sont effectués sur les images en présence de brouillard et sur les images restaurées par deux algorithmes de restauration du contraste différents : l'algorithme du chapitre précédent (MRF_RESTAURATION) et l'algorithme fondé sur la contrainte *no-black-pixel* (NBPC). Pour l'algorithme proposé, nous avons évalué la première itération de la méthode (UNE ITERATION), ensuite après convergence de l'algorithme avec comme paramètres $\sigma_p \sqrt{\lambda_{I_0}} = 20$ et $\alpha = 4$ (FINALE) et après fusion avec la proposition du monde plan (FINALE+PA). Les résultats sont montrés dans le tableau 5.1 sur une moyenne de 66 images stéréoscopiques. Le pourcentage ne prend en compte que les pixels visibles dans les deux images avec une disparité supérieure à 1 (c'est à dire sans prendre en compte le ciel). La reconstruction de l'algorithme

5. <http://perso.lcpc.fr/tarel.jean-philippe/visibility/fogstereo.zip>

6. Implantation disponible à l'adresse <http://www.cvlb.net/software/libelas/>

7. Implantation disponible à l'adresse <http://opencv.org>

TABLE 5.1 – Comparaison du pourcentage de disparités correctes en moyenne sur 66 images stéréoscopique de synthèse en utilisant : l’approche par champ de Markov classique avec un terme de régularisation du premier ordre (MRF_STEREO), l’algorithme Elas (ELAS), la méthode semi-globale (SGBM). Ces trois algorithmes sont évalués sur les images en présence de brouillard, sur les images restaurées par l’algorithme avec la contrainte *no-black-pixel* (NBPC) et sur les images restaurées avec l’algorithme du chapitre précédent (MRF_RESTAURATION). L’algorithme proposé est évalué à la première itération (UNE ITERATION), après convergence (FINALE) et finalement après l’algorithme de fusion (FINAL+PA). Les pourcentages sont donnés pour différentes valeurs de l’erreur maximale *err* sur l’écart des disparités (en pixel).

Algorithme	$err \leq 1$	$err \leq 0.66$	$err \leq 0.33$
Images d’entrée : en présence de brouillard			
MRF_STEREO	81.2	75.9	53.8
SGBM	72.2	62.1	41.6
ELAS	82.2	72.9	46.7
Images d’entrée : contraste restauré avec MRF_RESTAURATION [CT13]			
MRF_STEREO	82.5	77.0	53.2
SGBM	73.9	62.5	40.6
ELAS	74.4	61.6	37.5
Images d’entrée : contraste restauré avec NBPC [TH09]			
MRF_STEREO	74.4	61.6	37.5
SGBM	72.7	62.4	41.3
ELAS	77.6	68.6	45.2
Algorithme proposé			
UNE ITERATION	85.8	81.4	58.8
FINALE	86.9	82.6	61.8
FINALE+PA	93.4	89.5	75.7

TABLE 5.2 – Erreur moyenne absolue entre l'image restaurée et l'image sans brouillard pour 9 méthodes de restauration sur 66 images de synthèse avec l'ajout d'un brouillard uniforme et d'un bruit d'écart type $\sigma^2 = 1$.

Algorithme	Erreur moyenne (en niveau de gris)
Sans traitement	81.6 ± 12.3
DCP [HST10b]	46.3 ± 15.6
FSS [HTA10]	34.9 ± 15.1
NBPC [TH09]	50.8 ± 11.5
NBPC+PA [THC ⁺ 10]	31.1 ± 10.2
CM+PA [HCH11]	19.1 ± 6.7
MRF+PA [CT13]	16.9 ± 5.1
Proposé	22.9 ± 18.8
Proposé+PA	18.9 ± 13.0

MRF_STEREO obtient un résultat de 81.2% de disparité correcte et l'algorithme Elas 82.2%. Les scores des deux algorithmes appliqués sur les images restaurées sont généralement moins bons, cela est principalement dû à la perte d'information et à l'amplification du bruit produit par les algorithmes de restauration. Seul l'algorithme MRF_STEREO appliqué sur les images restaurées par l'algorithme proposé dans le chapitre précédent (MRF_RESTAURATION) apporte un meilleur résultat avec un score de 82.5%. L'algorithme SGBM, malgré plusieurs jeux de paramètres testés, ne produit pas une bonne reconstruction.

Pour l'algorithme proposé dans ce chapitre, le pourcentage, après une itération est de 85.8%. Après convergence, le pourcentage atteint est 86.9%. Finalement, avec la fusion du monde plan, on obtient un score de 93.4%. La méthode proposée améliore donc de façon significative les résultats par rapport à l'algorithme classique ne prenant pas en compte la présence de brouillard. L'amélioration en pourcentage ne semble pas très important au premier abord quand on compare le score de la première étape au score après convergence. Cela est dû au fait qu'il y a, en proportion dans l'image, bien moins de zones lointaines que de zones proches dans les scènes proposées. Mais ces itérations sont bien nécessaires pour améliorer au loin les résultats lorsque l'échelle converge vers une valeur proche de la distribution du bruit.

Pour évaluer le résultat de la restauration, nous avons comparé l'image restaurée après convergence aux autres algorithmes de restauration sur la base FRIDA2. Les résultats récapitulatifs sont montrés dans la table 5.2. L'algorithme proposé permet d'obtenir des résultats très proches de ceux de l'algorithme du précédent chapitre. L'algorithme proposé (Proposé) obtient 22.9. En ajoutant la contrainte planaire, la méthode proposée obtient 18.9 (Proposé+PA).

5.3.3 Sur image caméra

Nous avons comparé la méthode proposée à la méthode de reconstruction sans prendre en compte le brouillard et la méthode de restauration décrite dans [TH09]. Dans cette partie, la valeur de β est fixée arbitrairement. Les résultats montrent que la reconstruction et l'image restaurée sont de meilleure qualité qu'avec les deux algorithmes appliqués séparément. La figure 5.6 montre deux exemples : le premier dans une zone urbaine et le second en campagne. Premièrement, l'amélioration la plus significative provient de la contrainte photométrique, en effet nous pouvons voir sur les deux images que certaines zones ambiguës avec l'algorithme classique tendent vers zéro. Grâce à la contrainte photométrique, ces solutions ne sont pas autorisées

et la qualité dans ces zones est bien meilleure avec l'algorithme proposé. Ensuite, nous observons que la carte est plus lisse dans les zones planes et surtout que les contours des immeubles dans la première image au loin sont beaucoup mieux dessinés. La restauration quant à elle, est aussi de meilleure qualité, en effet, les objets proches ne sont pas saturés tandis que les objets à longue distance sont mieux restaurés et le bruit est mieux pris en compte.

5.4 Conclusion

Nous avons proposé une formulation markovienne pour résoudre le problème de la reconstruction à partir de paires stéréoscopiques et la restauration simultanées en présence de brouillard. Cet algorithme correspond à la combinaison de deux modèles : l'approche classique de la stéréo reconstruction sans brouillard et la restauration sachant la profondeur. La méthode proposée inclut la contrainte photométrique et un *a priori* sur les pixels blancs du ciel. Le modèle complet peut être optimisé en alternant l'optimisation de la carte de profondeur et de l'image restaurée par l'algorithme α -expansion. La convergence vers un minimum local est donc garantie. L'algorithme est évalué sur un ensemble d'images de synthèse et testé sur des images réelles pour démontrer l'efficacité du modèle. Grâce à l'information de profondeur, la disparité est correcte dans les zones proches de la caméra. La restauration est améliorée dans les zones proches comparativement aux résultats des algorithmes monoculaires. La restauration simultanée entraîne une amélioration conséquente de la qualité de la carte de profondeur dans les zones ambiguës au loin.

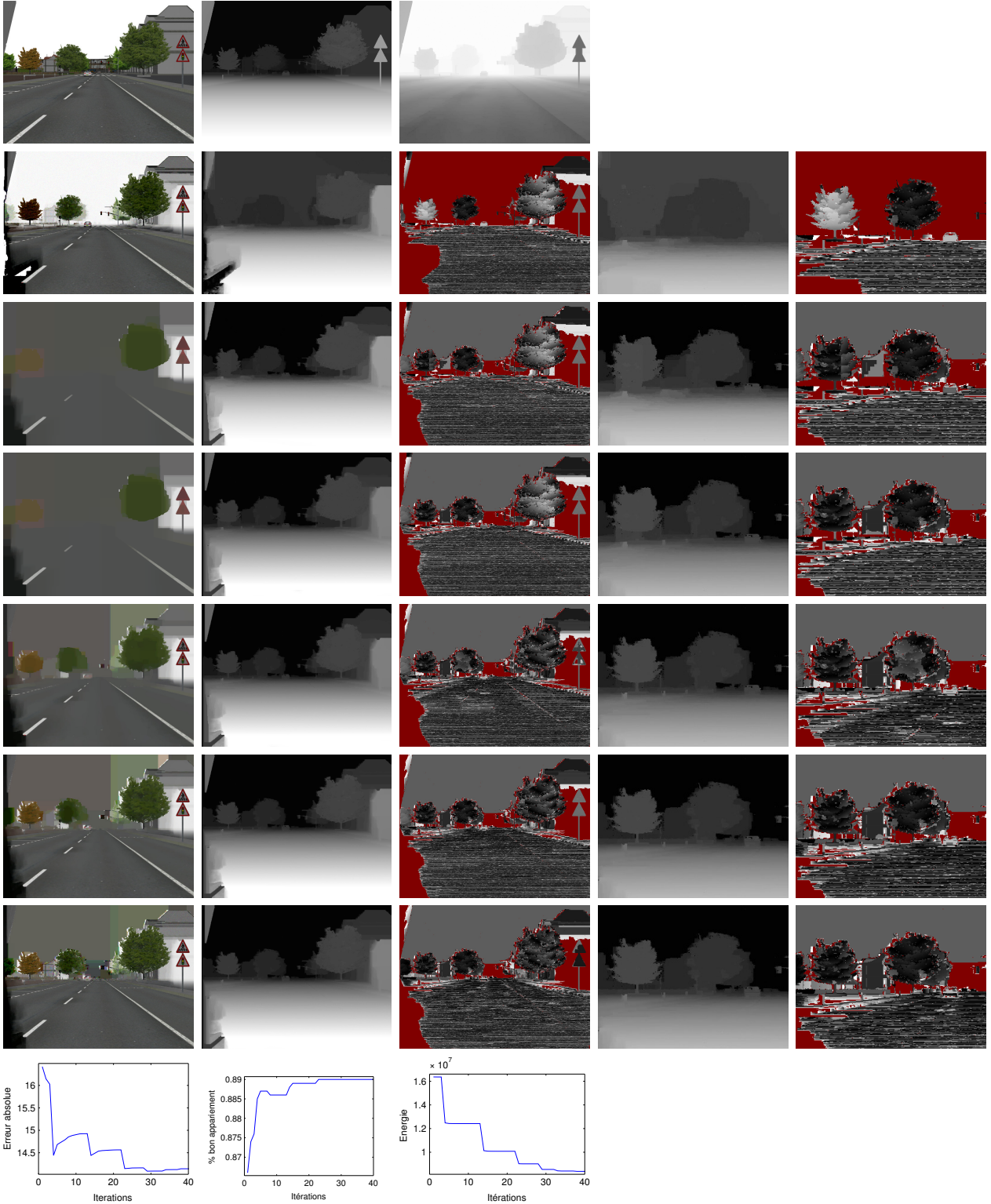


FIGURE 5.2 – Exemple de la méthode proposée au cours des itérations. Première ligne, de gauche à droite : l'image sans brouillard, la vérité terrain de la carte de disparité et l'estimation du voile $\tilde{V}$. Seconde ligne : l'image restaurée avec $\sigma_p \sqrt{\lambda_{I_0}} = 1$ et la carte de disparité correspondant ainsi que la carte d'erreur de la carte de disparité générée avec la vérité terrain. Une sous partie est montrée dans les deux dernières colonnes. Troisième ligne : le résultat de la méthode proposée pour $\sigma_p \sqrt{\lambda_{I_0}} = 20$ à l'initialisation. Puis, la seconde, dixième et vingtième itérations et après convergence. Dernière ligne, le nombre de bons appariements, l'erreur absolue entre l'image restaurée avec l'image sans brouillard et l'énergie au cours des itérations.

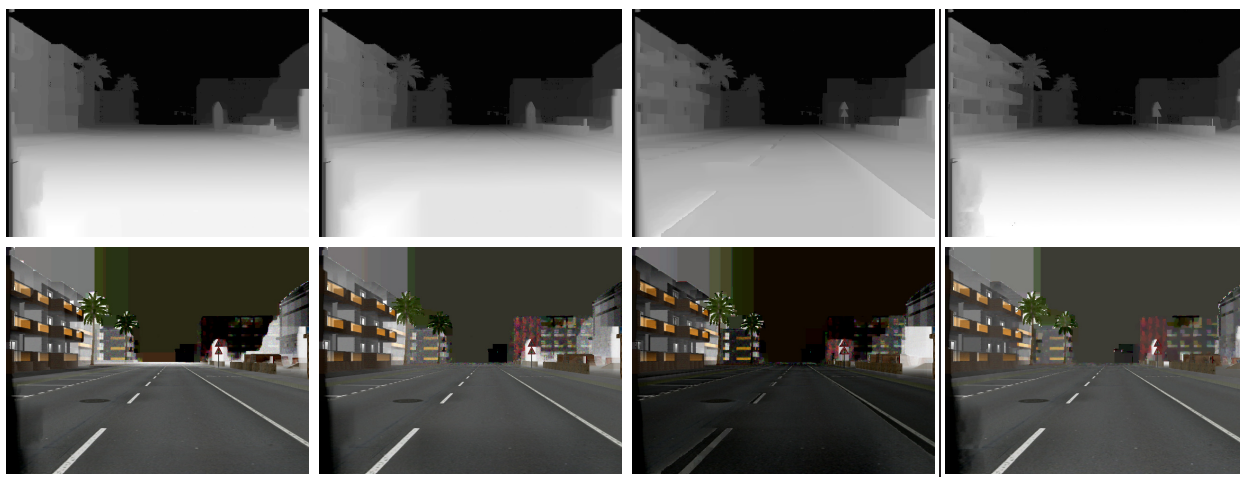


FIGURE 5.3 – Effet des paramètres sur les résultats de la méthode de reconstruction et restauration. La première colonne correspond au résultat avec $\alpha = 0$ et $\sigma_p \sqrt{\lambda_{I_0}} = 2$, la seconde avec $\alpha = 0$ et $\sigma_p \sqrt{\lambda_{I_0}} = 5$, la troisième avec $\alpha = 0$ et $\sigma_p \sqrt{\lambda_{I_0}} = 10$ et la dernière avec $\alpha = 4$ et $\sigma_p \sqrt{\lambda_{I_0}} = 20$.

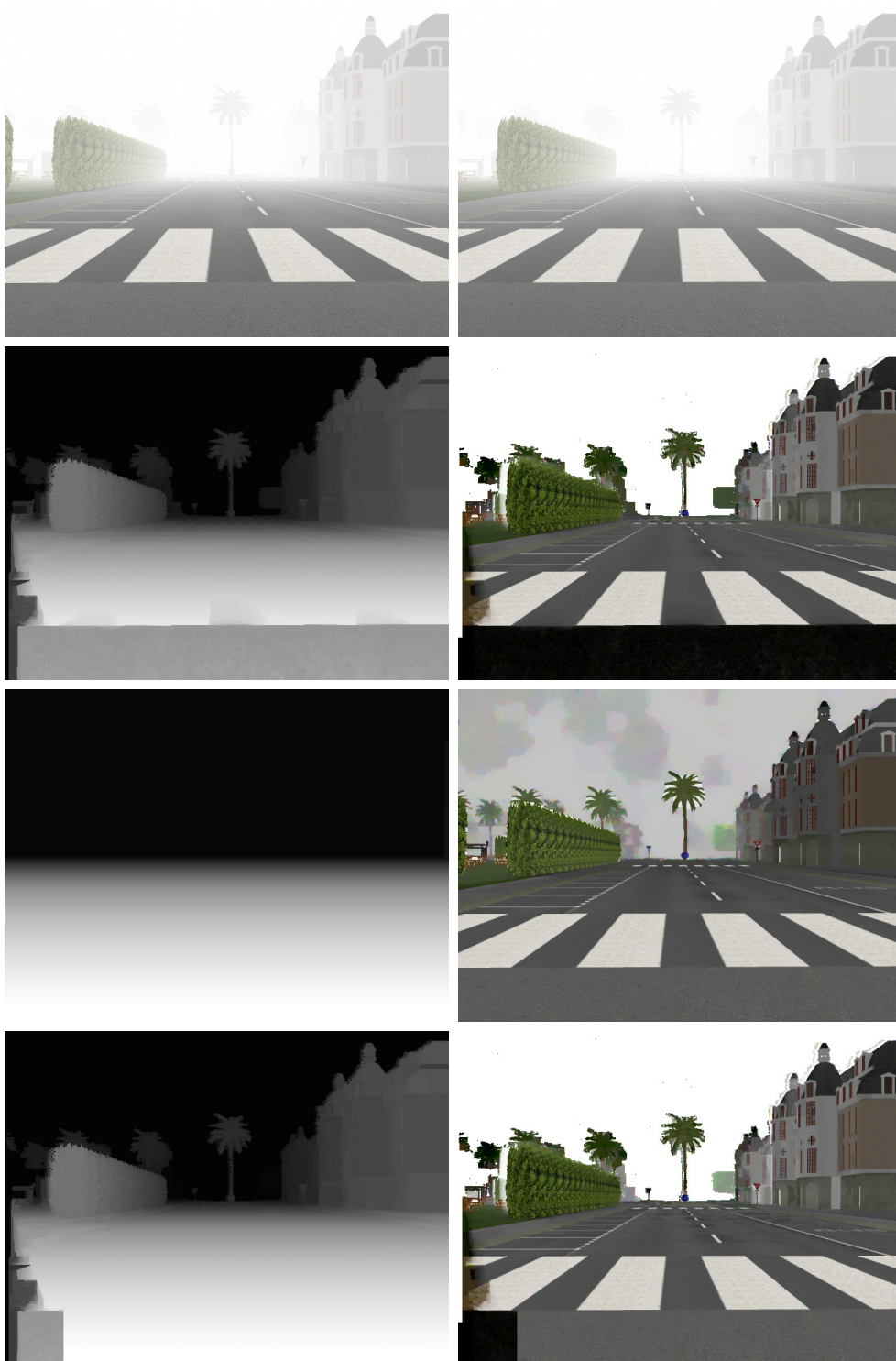


FIGURE 5.4 – Exemple de fusion avec l'*a priori* du second ordre. Première ligne, l'image gauche et droite de la paire stéréoscopique. Seconde ligne, la carte de disparité et l'image restaurée obtenue avec l'ordre premier. Troisième ligne, le plan de la route et l'image restaurée obtenue avec l'algorithme de restauration du chapitre précédent. Dernière ligne, la carte de disparité et la restauration obtenue après la fusion entre les deux résultats des précédentes lignes.

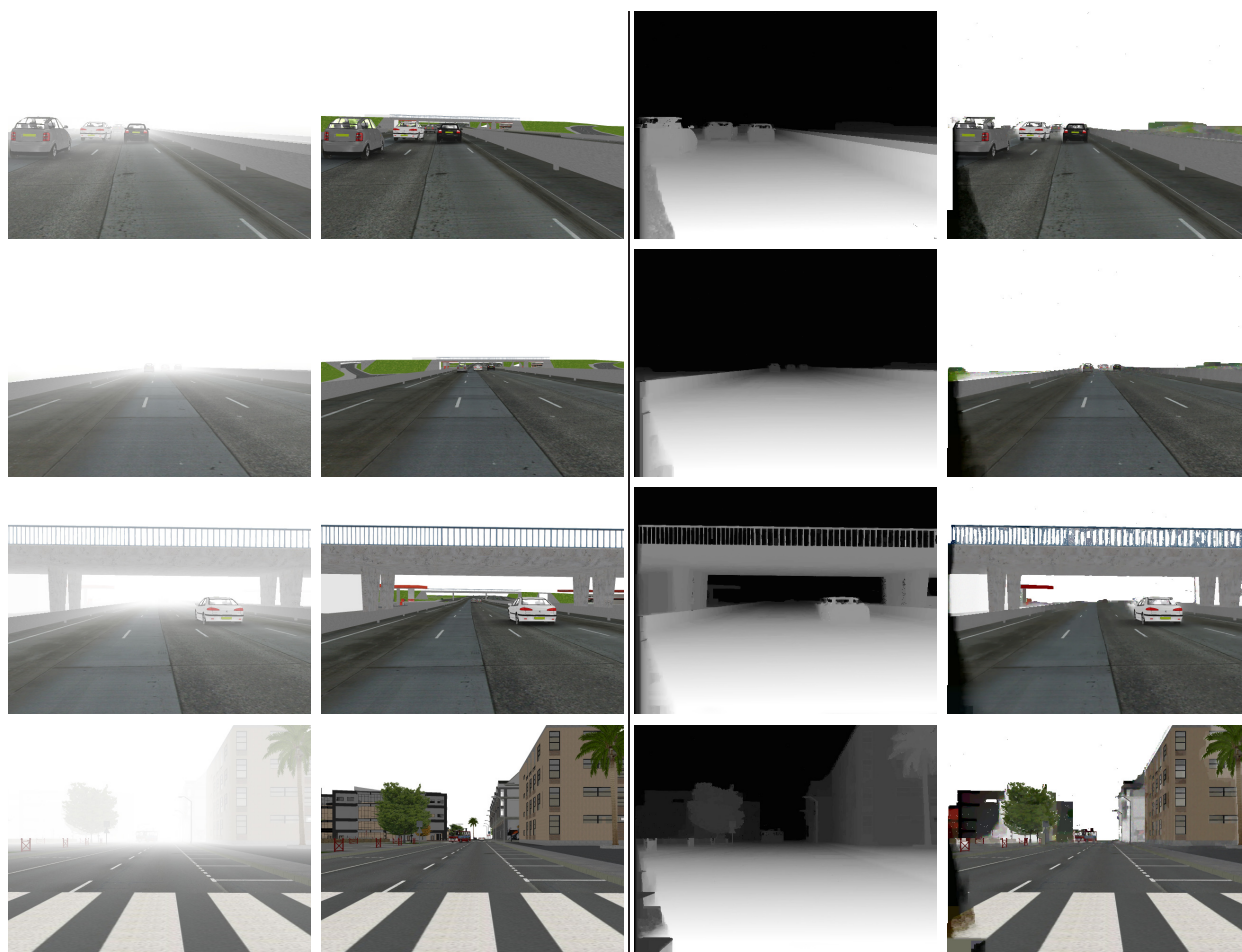


FIGURE 5.5 – Résultat de l'algorithme sur quatre images stéréoscopiques de synthèse de la base FRIDA3. Première colonne : l'image gauche en présence de brouillard. Deuxième colonne : la même scène sans brouillard. Troisième colonne la carte de disparité. Dernière colonne : l'image restaurée obtenue avec la méthode proposée.

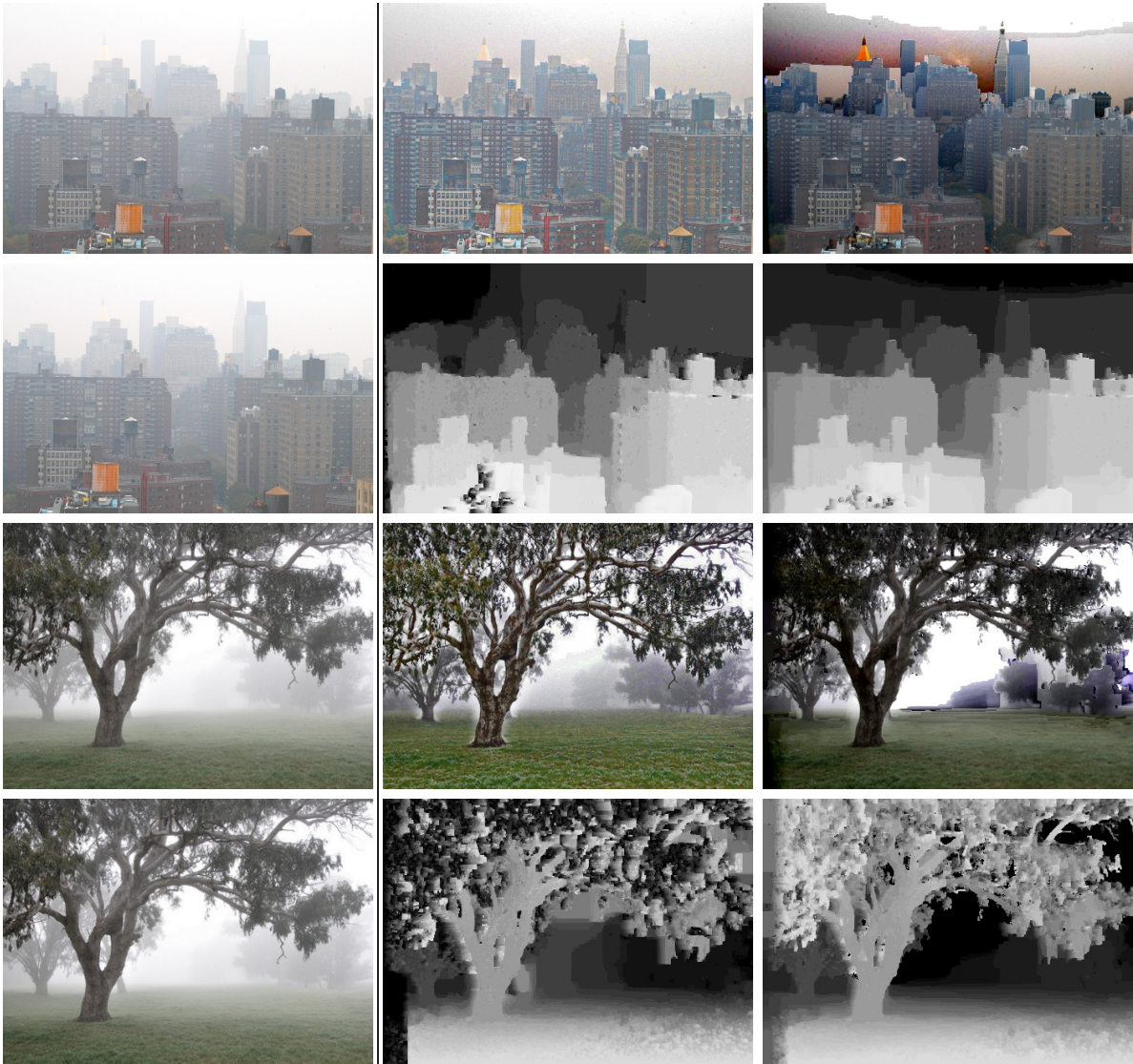


FIGURE 5.6 – Résultats sur des images caméras. Première colonne : la paire stéréoscopique en présence de brouillard. Seconde colonne : résultat de la restauration avec l'algorithme monoculaire [TH09] et la carte de disparité obtenue avec l'algorithme de reconstruction à partir de paires stéréoscopiques sans prendre en compte le brouillard. Troisième colonne : les résultats obtenus avec l'algorithme proposé.

Chapitre 6

Estimation du coefficient d'extinction du brouillard par stéréovision

Une méthode d'estimation du coefficient d'extinction du brouillard à partir d'une paire stéréoscopique est proposée. La méthode est fondée sur la minimisation de l'entropie de l'image restaurée.

L'algorithme de reconstruction décrit dans le chapitre précédent nécessite la connaissance du coefficient d'extinction. D'autres applications nécessitent de pouvoir estimer rapidement et efficacement le coefficient d'extinction. Les méthodes d'estimation présentées dans la section 2.4 ont des inconvénients comme l'obligation pour la scène d'avoir certaines caractéristiques (comme les marquages au sol ou encore la visibilité du point d'inflexion). Une méthode moins contraignante serait donc utile.

Dans la partie 6.1, nous proposons une première méthode à partir du modèle présenté dans le chapitre précédent, mais l'algorithme qui en découle est très coûteux en temps de calcul. Dans la partie 6.2, un autre algorithme, beaucoup plus rapide en utilisant une autre approche est proposé et approfondi.

6.1 Estimation du coefficient d'extinction à partir du modèle global

Il est premièrement possible d'envisager une approche à partir du modèle proposé dans ce chapitre en posant β comme variable. En effet, pour une carte de profondeur donnée et pour une certaine échelle, on peut chercher quel est le β minimisant l'énergie. Pour cela, il faut dans un premier temps calculer pour chaque β candidat, l'image restaurée minimisant l'énergie du modèle. Et par la suite, regarder quel est le β avec image restaurée correspondante qui minimise l'énergie globale du modèle (5.8).

La figure 6.1 montre l'énergie totale du modèle pour différents β après avoir, pour chaque β , calculé l'image restaurée minimisant l'énergie globale. La figure 6.1a montre l'énergie en fonction de β pour différentes échelles σ_P . Nous constatons que, quelle que soit l'échelle du

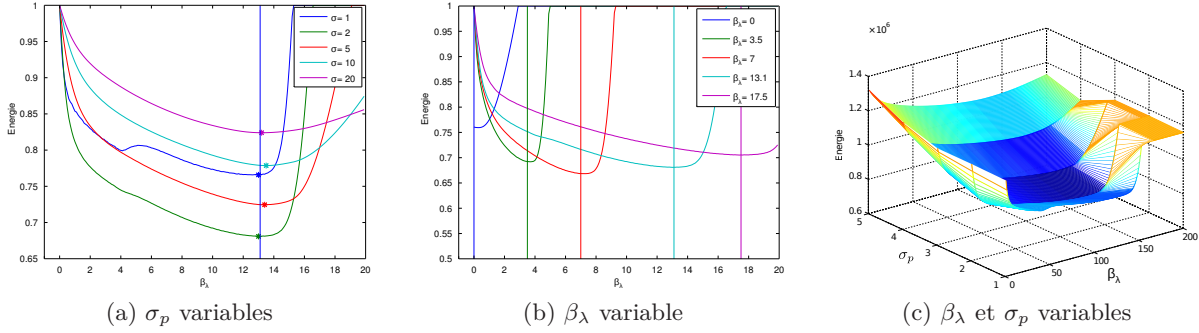


FIGURE 6.1 – Variation de l'énergie en fonction du coefficient d'extinction et de σ_p . La figure 6.1a montre l'énergie globale normalisée du modèle pour différentes valeurs de σ_p et β_λ avec la vérité terrain comme carte de profondeur ainsi que l'image restaurée minimisant l'énergie pour chaque couple de paramètres. La figure 6.1b et 6.1c montre le même résultat, pour un σ_p constants et différents β_λ . Les lignes verticales montrent la vérité terrain pour le coefficient d'extinction.

terme de vraisemblance, le coefficient d'extinction β minimisant l'énergie est proche de celui correspondant à la vérité terrain. Quand σ_p est grand, nous pouvons voir que l'énergie décroît progressivement jusqu'à atteindre le minimum proche du β correspondant à la vérité terrain pour ensuite, remonter progressivement. Lorsque le coefficient est trop petit ($\beta = 1$), c'est à dire proche du bruit, de nombreux minima locaux apparaissent. Nous pouvons remarquer que les paramètres β et σ_p qui minimisent l'énergie globale sont ceux recherchés, c'est à dire un σ_p permettant d'obtenir une image sans bruit ainsi que le β correspondant à la vérité terrain.

L'intérêt de cette approche est d'obtenir un modèle complet pour la reconstruction 3D, la restauration et l'estimation du coefficient d'extinction. Néanmoins, cette méthode est extrêmement longue du fait du nombre important d'inconnues et du temps de calcul des algorithmes d'optimisation. Ces temps de traitement sont incompatibles avec certaines applications en temps réel. Pour cela, une nouvelle approche, permettant d'estimer rapidement le coefficient d'extinction est proposée.

6.2 Estimation du coefficient d'extinction par minimisation d'entropie

Dans ce chapitre, nous proposons une méthode pouvant s'appliquer à une grande diversité de scènes. Pour cela, nous considérons le cas où une carte de profondeur est estimée jusqu'à une certaine distance. On suppose donc qu'un algorithme rapide de reconstruction est appliqué préalablement et que le nombre de bons appariements est suffisamment conséquent jusqu'à une certaine distance, comme un pourcentage de la distance de visibilité. Quand la profondeur est estimée, l'idée est de calculer β en observant l'atténuation du gradient des zones homogènes de la scène. Comme pour la suppression de l'ombre [FDL09, KHD12], l'entropie de Shannon est utilisée comme un critère à minimiser. Cette méthode présente un double avantage par rapport aux méthodes précédentes : premièrement, elle s'applique de façon générale pour un grand nombre de scènes, et, deuxièmement, toutes les profondeurs dans la scène ne sont pas nécessaires pour estimer le coefficient d'extinction. Cette dernière caractéristique est un point clé. En effet, il est très difficile de connaître l'ensemble des profondeurs dans la scène, surtout au loin, quand la caméra est mobile. L'algorithme proposé, combiné avec un algorithme de

reconstruction à partir de paires stéréoscopiques en temps réel, apporte donc une solution rapide pour estimer le coefficient d'extinction quand la caméra est en mouvement. Comme le montrent les expérimentations, cette méthode donne des résultats très satisfaisants, également dans le cas d'une caméra statique.

Dans la première partie nous décrivons le principe de l'algorithme de détection et de caractérisation du brouillard avec l'estimation du coefficient d'extinction. L'algorithme est détaillé dans la seconde partie. Deux évaluations de l'algorithme sont réalisées : une première évaluation sur 200 images d'une scène statique avec des conditions changeantes, une deuxième évaluation sur la base de données FRIDA3 pour simuler une situation en embarqué.

6.3 Problème

Quand une région d'intensité uniforme est observée en présence de brouillard, cette région ne possède plus une intensité constante si elle n'est pas fronto-parallèle. En effet, avec l'augmentation de la distance, l'image devient de plus en plus blanche. Cela est dû comme nous l'avons vu au voile atmosphérique dans l'équation de Koschmieder (2.1) qui devient de plus en plus important avec la distance. L'entropie de Shannon est une mesure de la redondance dans une distribution de probabilités. En conséquence, l'entropie de la distribution d'une région d'intensité constante (c'est à dire sans brouillard), est très faible. Inversement, l'entropie de la même région en présence de brouillard est plus importante si la région s'étend sur une grande plage de distances. L'idée principale avec ce constat, est d'estimer le coefficient d'extinction β qui produit l'image restaurée ayant la plus petite entropie dans les régions d'intensités constantes.

La figure 6.2 illustre le principe de l'algorithme. Une image en présence de brouillard est montrée dans la figure 6.2a. Pour trouver une région d'intensité constante sans brouillard, nous proposons de segmenter l'image avec brouillard en régions où l'intensité varie lentement (figure 6.2b). Les régions, comme celles fronto-parallèles, avec une faible variation de profondeurs ne sont pas utilisées pour l'estimation de β . Dans la figure 6.2c, ces critères font qu'il ne reste plus que trois régions sélectionnées. Dans la figure 6.2d, 6.2e et 6.2f, l'image en présence de brouillard est restaurée avec trois β différents, respectivement $\beta = 0.0024$, $\beta = 0.0086$ et $\beta = 0.0167$. Ces restaurations sont obtenues en inversant l'équation de Koschmieder (2.1). La bonne valeur de β est la seconde valeur (obtenue avec un visibilimètre). Nous constatons que lorsque β est trop faible, l'image restaurée est encore brumeuse et que le gradient dû au brouillard est encore perceptible. Inversement, quand β est trop important. Les zones au loin sont beaucoup trop sombres. Pour la vérité terrain $\beta = 0.0086$, l'image restaurée est homogène et paraît mieux restaurée que les deux autres. Dans les figures 6.2g, 6.2h et 6.2i, les distributions des niveaux de gris sont affichées pour chaque restauration. Nous pouvons voir que pour $\beta = 0.0086$, l'histogramme est beaucoup plus serré autour d'un pic que pour les deux autres valeurs de β , pour les trois régions sélectionnées. La figure 6.2j montre l'entropie de chaque région sélectionnée en fonction de β . Le minimum de chaque entropie est localisé à une portion assez proche de la vraie valeur de β pour les trois zones sélectionnées. Finalement, la figure 6.2k montre la somme des trois entropies en fonction de β . L'entropie totale de l'image est aussi montrée. Les valeurs de β utilisées précédemment sont représentées par les lignes verticales. Il est clair d'après cette figure, que la somme des entropies permet d'estimer avec plus de précision le coefficient d'extinction. En effet, la valeur du coefficient β pour lequel le minimum est atteint est très proche de la vérité terrain.

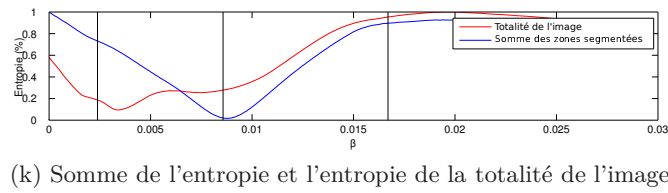
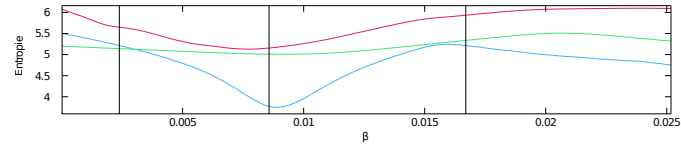
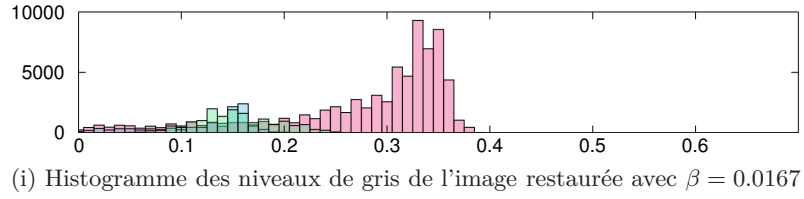
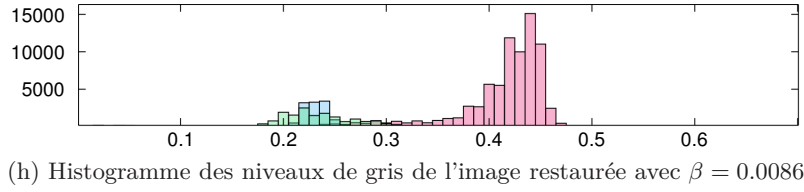
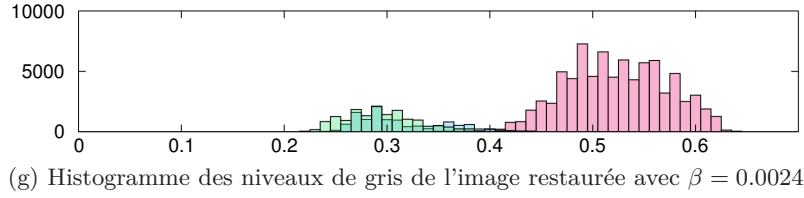
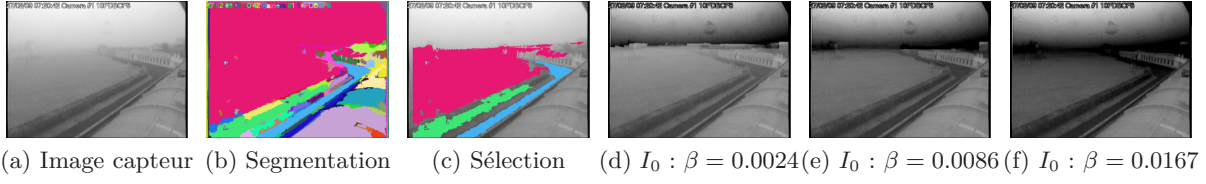


FIGURE 6.2 – Illustration du principe de la méthode d'estimation du coefficient d'extinction β . Première ligne, 6.2a (a) est l'image en présence de brouillard avec la profondeur connue, le β mesuré par le visibilimètre est $\beta = 0.0086$, 6.2b (b) est la segmentation en régions et 6.2c (c) sont les trois régions sélectionnées. (d) (e) (f) sont les images après restauration pour trois valeurs de β . (g) (h) (i) montrent l'historgramme de la distribution de l'intensité en niveaux de gris pour les trois régions sélectionnées et pour les trois valeurs de β . La figure 6.2j (j) montre l'entropie des trois régions sélectionnées en fonction de β et 6.2k (k) montre la somme de l'entropie des trois régions sélectionnées comparée à l'entropie globale de l'image. Les lignes verticales représentent les trois valeurs de β utilisées pour les restaurations des images en haut de la figure.

6.4 Méthode

Pour chaque paire stéréoscopique, la carte de disparité peut être calculée avec n'importe quelle méthode. Le choix de l'algorithme est discuté dans la section expérimentation où il est montré que la carte de disparité doit être correcte uniquement jusqu'à une certaine distance.

Les données d'entrée de l'algorithme sont : une des images de la paire stéréoscopique I ainsi que la carte de disparité D relative à cette image obtenue grâce à la reconstruction à partir de paires stéréoscopiques.

Si I et D sont connues et β est donné, il reste comme inconnues dans l'équation (2.1) : le coefficient d'extinction, l'image restaurée I_0 et l'intensité du ciel I_S . La première étape est donc d'estimer I_S . Pour cela, nous proposons une nouvelle solution dans la section 6.4.1. Ensuite, pour un β donné, l'image restaurée I_0 peut être calculée à partir de I , D et I_S en inversant (2.1). La façon de sélectionner les régions est décrite dans la partie 6.4.2, et l'estimation de I_0 et β dans la partie 6.4.3. La structure de l'algorithme est décrite dans 6.4.4.

6.4.1 Estimation de I_s

Les méthodes de restauration calculent souvent I_S de manière simple : I_S est choisie comme l'intensité maximale de l'image, ou comme la valeur correspondant à un certain quantile de l'histogramme des intensités [Tan08, TH09]. Ces approches donnent des estimations approximatives de I_S , qui, dans le cas de la restauration d'image, sont tolérables. En revanche, dans le cas de l'estimation du coefficient d'extinction, nous avons observé que la précision de I_S est importante pour estimer correctement la valeur de β . Le brouillard à longue distance n'étant pas forcément homogène, il est préférable d'estimer I_S à une distance raisonnable.

Nous avons donc introduit une nouvelle méthode, assez simple à mettre en oeuvre, fondée sur la détection de contraste en utilisant l'opérateur de Sobel. L'opérateur de Sobel est d'abord calculé sur l'image d'entrée. Ensuite, les pixels dont l'amplitude du gradient est supérieure à un seuil assez faible sont gardés. Finalement, nous choisissons I_S comme l'intensité maximale des pixels sélectionnés par l'étape précédente.

Cette méthode s'applique aussi bien à une caméra statique qu'en mouvement. Lorsque la carte de profondeur est connue avec précision sur toute la scène (cas d'une scène statique), une méthode plus précise peut être utilisée : la valeur de I_S peut être estimée en prenant la valeur maximale de l'intensité des pixels, les plus éloignés, après sélection.

6.4.2 Sélection des régions

La minimisation de l'entropie n'a de sens que pour un ensemble de régions d'intensité constante sans brouillard. De plus, la région doit être étendue sur une plage de profondeur conséquente. L'image d'entrée est d'abord segmentée en régions dont l'intensité varie faiblement. Les régions fronto-parallèles ne doivent pas être sélectionnées ainsi que les régions trop petites. Nous avons choisi comme critères de ne garder que celles dont l'étendue de disparité est supérieure à 10% de l'étendue globale de la profondeur de la scène et ayant une surface supérieure à 2% de celle de l'image.

La segmentation peut être réalisée par temps dégagé lorsqu'un capteur statique est utilisé.

6.4.3 Estimation de β et de I_0

Après avoir estimé I_s , I_0 peut être obtenue pour un β donné en inversant (2.1). L'image restaurée, I_0 étant obtenue, il reste à calculer l'entropie sur les régions sélectionnées. L'entropie

de Shannon pour une variable aléatoire $y \in Y$ est :

$$H(y) = -\mathbf{E}[\log p(y)] = - \sum_{y \in Y} p(y) \log p(y) \quad (6.1)$$

où $\mathbf{E}$ est l'espérance et $p(y)$ est la densité de probabilité de la variable y . Pour plusieurs variables aléatoires décorréées, l'entropie de l'ensemble des variables est la somme des entropies individuelles.

La scène est composée de plusieurs objets, et donc, de plusieurs régions. Chaque région correspond à une variable aléatoire. Soit S_c l'ensemble des régions sélectionnées, l'entropie totale sur cet ensemble est :

$$TH = \sum_{y_i \in S_c} H(y_i) = - \sum_{y_i \in S_c} \sum_{y_i \in Y} p(y_i) \log p(y_i) \quad (6.2)$$

L'entropie totale est calculée pour différentes valeurs de β , le β qui minimise (6.2) est choisi comme estimation du coefficient d'extinction.

Lorsque le β est mal choisi, l'image restaurée peut avoir de larges zones noires, d'entropie faible. Un exemple peut être observé dans la figure 6.2e, où certains objets au loin à droite de l'image sont totalement noirs du fait que le brouillard est plus léger dans ces zones. Pour ne pas être gêné par ces cas, les solutions qui présentent trop de pixels noirs ne sont pas considérées.

6.4.4 Algorithme

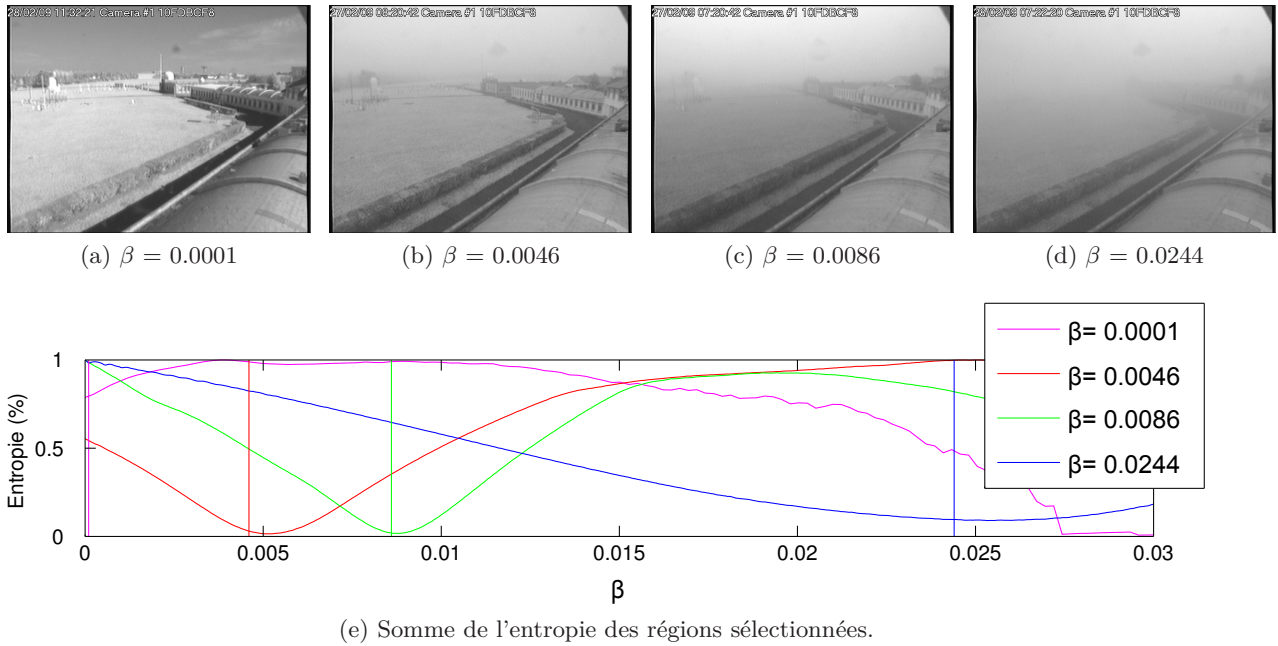


FIGURE 6.3 – L'entropie totale normalisée (6.2) en fonction de β sur la même scène avec différentes conditions météorologiques. Les vérités terrain sur la valeur de β , mesurées avec un visibilimètre, sont représentées par des lignes verticales (0.0001, 0.0046, 0.0086 and 0.0244). On remarque l'adéquation entre ces valeurs et le minimum de l'entropie.

En résumé, la structure de l'algorithme d'estimation du coefficient d'extinction β est la suivante :

- Détection des pixels avec une amplitude du gradient supérieur à un certain seuil sur l'image d'entrée I . I_s est estimée comme l'intensité maximale sur les pixels détectés.
- Segmentation de l'image originale I en régions et sélection des régions S_c ayant une grande plage de disparités et une grande surface.
- Pour chaque β dans une liste de valeurs possibles et dans un ordre croissant :
 - Calculer l'image restaurée pour les pixels des régions sélectionnées S_c en chaque site s de profondeur $p(s)$ connue, grâce à :

$$I_0(s) = I(s)e^{\beta p(s)} - I_s(e^{\beta p(s)} - 1)$$

- Calculer l'entropie totale HT sur toutes les régions sélectionnées après restauration.
- Calculer le nombre de pixels noirs de l'image restaurée I_0 . Si ce nombre est supérieur à un certain seuil limite, la recherche du β est stoppée.
- La solution est donnée par le β qui minimise l'entropie totale (6.2).

Dans la figure 6.3, l'entropie totale sur les trois régions sélectionnées dans la figure 6.2 est montrée en fonction de β . La vérité terrain du β est indiquée par des lignes verticales. Cet exemple illustre le fait que le β qui minimise l'entropie est assez proche de la vérité terrain mesurée par le visibilimètre.

6.5 Évaluation

L'évaluation de l'algorithme précédent nécessite d'avoir des images avec une vérité terrain sur la visibilité. Un visibilimètre peut être utilisé pour avoir cette vérité terrain dans le cas d'une scène statique. Jusqu'à présent les mesures obtenues par visibilimètre embarqué sont perturbées par les turbulences de l'air autour du capteur. En conséquence, dans le cas embarqué l'évaluation est faite à partir d'images synthétiques. Dans le cas statique, une base de donnée est utilisée.

6.5.1 Capteur statique

La base de données Matilda réalisée dans [HBD⁺11] est composée d'une séquence de plus de 300 images acquises par une caméra à bas coût pendant 3 jours (jour et nuit). La visibilité est obtenue par visibilimètre. Une approximation de la carte de profondeur est aussi donnée. Dans cette base de données, la visibilité varie de 100m à 15km. La précision du visibilimètre utilisé est d'environ 10%. La distance maximale de la carte de profondeur est de 1300m.

Afin de contrôler si l'estimation est correcte pour des distances proches, l'évaluation est réalisée en tronquant la carte de distance à une distance limite maximale. L'évaluation est réalisée selon deux protocoles : dans le premier cas, toute la carte de profondeur est utilisée (notée $D_{max}(I_s)$), et dans le second cas, seuls les pixels avec une profondeur inférieure à 50m sont utilisés (noté $D_{max} = 50m$). La caméra étant statique, la segmentation des régions n'est effectuée qu'une seule fois lors d'un ciel dégagé en utilisant l'algorithme proposé dans [FH06]. La même segmentation est alors utilisée pour toute la séquence. Cela nous permet d'évaluer le calcul de la distance de visibilité en évitant d'avoir une source supplémentaire d'erreur due à la segmentation. Ensuite la segmentation est faite à chaque image.

Dans le tableau 6.1, les résultats de l'évaluation sont présentés en terme d'erreur moyenne relative entre la distance de visibilité estimée et vérité terrain. Nous observons que l'algorithme produit des résultats avec une précision de l'ordre de celle du visibilimètre jusqu'à 1000m. Les résultats s'améliorent quand la courbe de distance est limitée à 50m. Cela peut s'expliquer par le fait que le brouillard, à longue distance, n'est plus homogène dans certaines zones. Les distances

Étendue (mètres)	0-250	0-500	0-1000	Complète
Nombre d'images	9	19	24	198
Avec une seule segmentation				
$D_{max}(I_s)$	10.0%	9.1%	11.7%	97.3%
$D_{max} = 50m$	7.1%	7.1%	10.0%	91.5%
Avec différentes segmentations				
$D_{max}(I_s)$	25.4%	17.2%	20.7%	101.7%

TABLE 6.1 – L'erreur moyenne relative entre la distance de visibilité estimée et la vérité terrain sur la base de données Matilda avec une seule segmentation obtenue sur une image sans brouillard. L'évaluation est effectuée en tronquant la carte de distance à deux valeurs différentes. La première distance est obtenue en calculant la distance maximale où I_s est observée. La seconde distance, plus courte est fixée à 50m. Dans la dernière ligne, la segmentation est appliquée sur chaque image, pour s'approcher de la situation où le capteur est situé sur un véhicule en mouvement.

	Erreur relative moyenne
Avec la vraie distance	
Segmentation	4.67%
Avec la distance estimée par stéréo reconstruction	
Segmentation	16.03%

TABLE 6.2 – L'erreur moyenne relative entre la visibilité estimée et la vérité terrain sur 66 images stéréoscopiques de synthèse de la base de données FRIDA3. Dans le premier cas, la vérité terrain est utilisée pour la carte de disparité. Dans le second cas, la carte de disparité est estimée à partir d'un algorithme de reconstruction 3D à partir de paires stéréoscopiques (ici [GRU10]).

proches reflètent mieux les données acquises par le visibilimètre qui est placé à proximité de la caméra. Dans cette situation, c'est la seconde méthode d'estimation de I_s qui est utilisée pour obtenir de meilleurs résultats. La figure 6.4 montre les résultats pendant deux jours, avec un intervalle de 10 minutes entre les images, sans prendre en compte les périodes de nuit. Ces résultats montrent que l'algorithme proposé est clairement capable d'évaluer le β dans les cas de visibilité faible ou supérieure à 500m. Les visibilités au-dessus de 500m ne sont pas considérées dans les situations de conduite.

6.5.2 Capteur en mouvement

Dans le cas d'un capteur en mouvement, nous avons évalué la méthode proposée sur 66 images stéréoscopiques de synthèse de la base FRIDA3 utilisée dans le chapitre précédent.

Pour étudier l'effet sur l'estimation de β de mauvais appariements lors de la reconstruction stéréoscopique, nous avons comparé le cas où la vérité terrain est utilisée pour la carte de profondeur avec le cas où la carte de profondeur provient d'un algorithme de reconstruction 3D. Nous avons utilisé la méthode de reconstruction Elas [GRU10]. Le tableau 6.2 montre l'erreur relative moyenne entre la visibilité estimée et la référence par les différentes restaurations. Les résultats obtenus avec la reconstruction stéréo sont moins bons que ceux de la référence mais

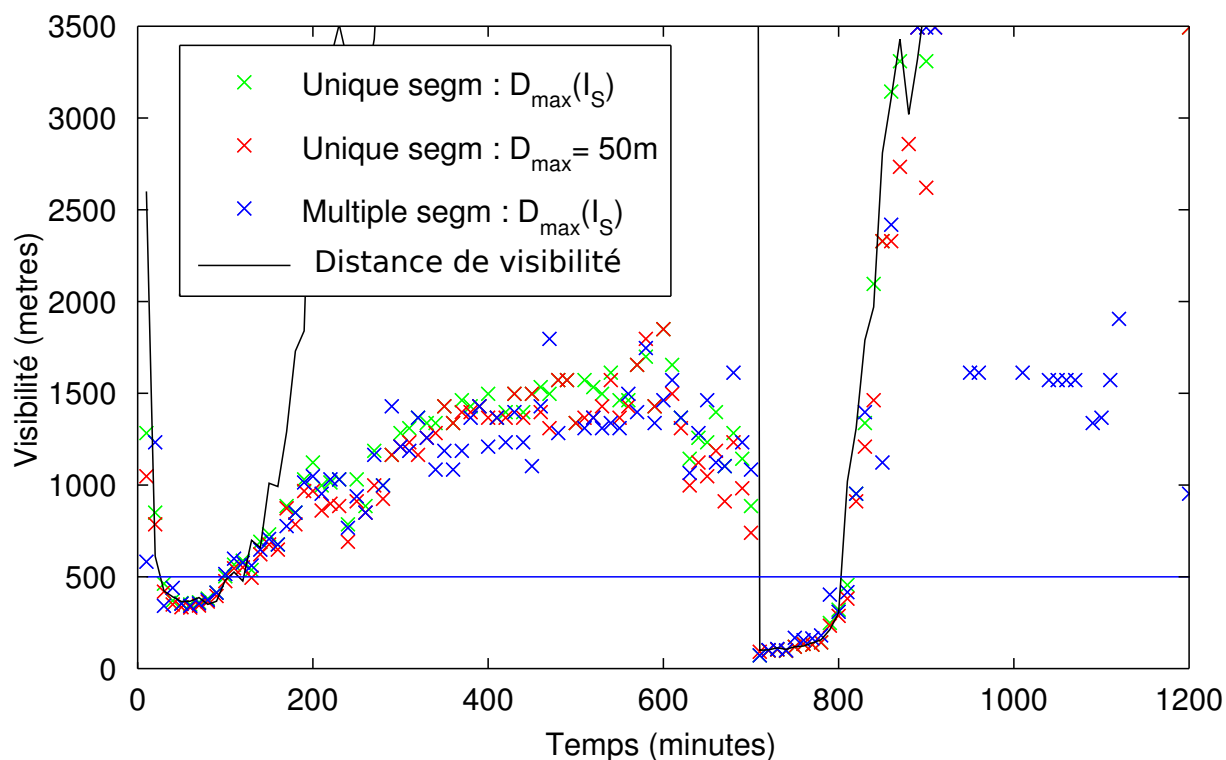


FIGURE 6.4 – Comparaison entre la visibilité estimée et la vérité terrain pendant 1200 minutes (une image toutes les 10 minutes, les images de nuit ne sont pas prises en compte) de la base de données Matilda. La méthode proposée est testée avec deux distances maximales $D_{max}(I_s)$ et $D_{max} = 50m$, toutes deux en utilisant une seule segmentation. Le cas avec une segmentation sur chaque image est aussi testé et affiché (*Multiple segm : $D_{max} = (I_s)$*). La vérité terrain est représentée par la courbe noire. La ligne horizontale indique la visibilité à 500m qui est considérée comme la distance à laquelle le brouillard n'a plus de conséquence dans les situations de conduite.

restent néanmoins de bonne qualité.

6.6 Conclusion

Nous avons proposé une méthode originale pour l'estimation du coefficient d'extinction du brouillard (ce qui équivaut à estimer la distance de visibilité). La méthode proposée est fondée sur la minimisation de l'entropie de l'image restaurée sur un ensemble de régions bien choisies. La carte de profondeur peut provenir d'un modèle 3D de la scène, ou d'un algorithme de reconstruction à partir de paires stéréoscopiques. Les résultats obtenus sur une scène statique sont précis, d'une précision proche de ceux obtenus avec un visibilimètre quand la carte de profondeur est de bonne qualité. L'algorithme peut aussi être utilisé avec un capteur en mouvement. Cette condition a été testée avec succès sur des images de synthèse. L'algorithme présente plusieurs avantages. Une distance faible peut être utilisée pour calculer une distance de visibilité dix fois plus importante que le maximum de la distance donnée. Ceci est un point clé pour les applications en embarqué où d'autres véhicules peuvent cacher une grande partie de la scène. Enfin, l'algorithme proposé est très rapide, son temps d'exécution est proche des conditions temps réel.

Il est donc parfaitement envisageable de l'utiliser pour une application en temps réel. Il peut aussi être facilement implanté en parallèle.

Chapitre 7

Conclusion et perspectives

Dans ce travail, nous nous sommes intéressés au problème de la reconstruction à partir de paires stéréoscopiques en présence de brouillard. De nombreux algorithmes existent pour faire la reconstruction stéréo, mais en général, ils ne sont pas adaptés aux conditions dégradées et en particulier, au brouillard. De ce fait, la qualité de la carte de reconstruction produite par ces algorithmes se dégrade fortement à partir d’une certaine distance. L’une des principales causes est la diminution du contraste avec la distance.

Depuis quelques années, la restauration du contraste en présence de brouillard à partir d’une seule image connaît un intérêt croissant pour diverses applications en traitement d’image. Un nombre conséquent d’algorithmes ont été proposés. Ils apportent des solutions approchées. La plupart sont fondés sur la loi de Koschmieder qui permet de lier les variations de profondeur dans la scène avec l’image, avec et sans brouillard.

L’un des problèmes de la restauration du contraste monoculaire est l’ambiguïté entre la profondeur et l’intensité originale de la scène, cela rend le problème mal posé. Une grande partie des algorithmes utilisent des *a priori* forts pour pouvoir faire l’estimation de la profondeur lors de la restauration. Or, cette estimation est souvent éloignée de la vraie profondeur, en particulier pour les zones proches et les zones grises.

C’est à partir de ce constat que l’idée d’un algorithme de reconstruction et de restauration stéréo simultanées a été proposée pour la première fois.

7.1 Résultats et contributions

L’état de l’art des méthodes pour la reconstruction 3D à partir de paires stéréoscopiques a été traité dans le chapitre 1. L’état de l’art des différentes méthodes de restauration du contraste à partir d’une seule image est donné dans le chapitre 2. Il en ressort d’après les deux bases d’évaluation *Middlebury Stereo Evaluation* et *KITTI Vision Benchmark Suite*, que les méthodes probabilistes et les méthodes *globales* qui en découlent apportent les meilleurs résultats pour la reconstruction stéréoscopique, aussi bien dans le cadre général que pour des scènes spécifiques comme l’environnement urbain. En effet, elles répondent à toutes les contraintes spécifiques de ces scènes telles que l’hypothèse d’une scène en partie plane ou encore la prise en compte de grandes zones homogènes. Pour la problématique de la restauration d’image en présence de brouillard, les techniques fondées sur le modèle de Koschmieder prévalent sur les autres, en particulier en environnement routier. Ceci est évalué qualitativement dans [THC⁺12].

Pour faire la reconstruction 3D en présence de brouillard, nous nous sommes fondés sur une approche probabiliste, et plus particulièrement sur les champs de Markov, qui, depuis plus

d'une vingtaine d'années, ont connu un essor important en traitement d'images. Cette approche a l'avantage de proposer un cadre complet pour modéliser et optimiser un grand nombre de problèmes. Ainsi, elle permet d'effectuer la reconstruction stéréo, mais aussi la restauration. Nous montrons qu'elle permet aussi de combiner reconstruction et restauration. A partir du modèle présenté, une énergie est dérivée puis optimisée. Parmi les nombreuses approches existantes, une des plus efficaces est celle qui utilise l'optimisation de fonctions pseudo-booléennes. Dans le chapitre 3, cette théorie est rappelée. Cela nous a permis de caractériser les modèles, et ainsi, de savoir quel modèle peut être facilement optimisé, et, les différentes alternatives possibles selon les propriétés du modèle.

7.1.1 Modèle markovien pour la restauration d'image monoculaire en présence de brouillard

Dans le chapitre 4, l'algorithme fondé sur la contrainte *no-black-pixel* (NBPC) proposé dans [THC⁺12] est reformulé avec un modèle probabiliste. Cet algorithme est décomposé en deux étapes : le voile atmosphérique est d'abord calculé, et dans un second temps, l'image restaurée est calculée à partir du voile. Nous proposons de modéliser le voile atmosphérique et l'image restaurée par un champ de Markov. Afin de lever l'ambiguïté dans les zones grises qui risqueront d'être restaurées en noir, un biais de la solution par l'initialisation a été introduit pour obtenir un effet homogène avec la distance. Le modèle proposé permet aussi d'améliorer la prise en compte du bruit lors de la restauration.

7.1.2 Restauration et reconstruction simultanées en présence de brouillard à partir de paires stéréoscopiques

Le chapitre 5, présente un modèle complet pour la reconstruction et la restauration par paires stéréoscopiques en présence de brouillard, fondé en partie sur le modèle précédent. Ce modèle peut être optimisé de façon efficace par un algorithme alternant une optimisation sur la carte de profondeur et sur l'image restaurée. Les résultats obtenus avec des images de synthèse montrent une nette amélioration par rapport aux algorithmes qui ne prennent pas en compte le brouillard.

7.1.3 Estimation du coefficient d'extinction du brouillard à partir de paires stéréoscopiques

La méthode précédente nécessite la connaissance du coefficient d'extinction du brouillard. Dans le chapitre 6, une nouvelle méthode est proposée pour caractériser le brouillard, en se fondant sur la minimisation de l'entropie de l'image restaurée sachant la distance. Cette méthode permet, à partir d'une carte de profondeur, d'estimer le coefficient d'extinction du brouillard. Cette méthode présente le double avantage de ne pas faire d'hypothèse sur la scène observée et ne nécessite pas d'avoir les distances éloignées. L'évaluation sur la base de données Maltida composée de 200 images avec un grand nombre de configurations météorologiques montre que l'on obtient des résultats très satisfaisants sur l'ensemble des distances de visibilité.

7.1.4 Implantation générique pour l'optimisation de champ de Markov

Un programme complet permettant d'optimiser des modèles assez génériques de champ de Markov a été implanté. Ce programme possède une structure générique de champ de Markov à plusieurs niveaux appelé *factorial Markov random field*. Cette structure a la particularité de lier

plusieurs champs de Markov par un même terme, ce qui est le cas de l'image restaurée avec la carte de profondeur.

L'avantage de ce code est que la formulation du modèle est assez dissociée de la façon de l'optimiser. Ainsi, il est possible de substituer très facilement un algorithme d'optimisation par un autre. Le programme développé permet de créer assez facilement de nouveaux modèles héritant directement d'un autre modèle où d'un ensemble de termes déjà créés tels que les termes de régularisation. Il suffira, pour créer un nouveau modèle, de créer une classe fille en redéfinissant le terme d'attache aux données.

7.2 Perspectives

7.2.1 Base de tests d'images réelles

Durant ce travail, les paires stéréoscopiques réelles en présence de brouillard utilisées ont été obtenues sur internet⁸. Nous disposons aussi d'une séquence stéréo en présence de brouillard, mais de qualité insuffisante pour permettre d'obtenir un résultat acceptable (image floues, prise de vue intérieur au véhicule et donc sujettes aux reflets du pare brise). Les algorithmes classiques de reconstruction produisent de très mauvais résultats sur cette séquence même sur les zones proches sans brouillard.

Une base de tests avec des images réelles de bonne qualité en conditions dégradées est nécessaire pour permettre d'évaluer de façon plus approfondie l'algorithme proposé. Les bases d'évaluation telle que *KITTI Vision Benchmark Suite* utilisent la technologie LIDAR afin d'obtenir la vérité terrain. Cette approche fonctionne très bien en conditions dégagées, mais il est difficile de garantir son bon fonctionnement avec du brouillard. Le caractère intermittent du brouillard permet aussi difficilement de planifier ces acquisitions.

Actuellement, même sans avoir la connaissance des caractéristiques des images recueillies sur Internet (telles que la courbe de transfert, le coefficient d'extinction du brouillard, la compression effectuée, etc.), les résultats obtenus avec l'algorithme proposé sont encourageants.

7.2.2 Accélération graphique

Un des points critiques des méthodes proposées de reconstruction et restauration est le temps nécessaire à la recherche de la solution. Les algorithmes, fondés sur les champs de Markov, proposés dans cette étude conduisent à un traitement temps réel. Le calcul d'un champ de Markov, par programmation séquentielle, avoisine la minute sur une image de taille courante. Quand il est nécessaire d'alterner la minimisation entre plusieurs champs, jusqu'à la convergence, cela peut nécessiter des temps de traitement importants incompatibles avec du temps réel. Outre l'incidence sur le fonctionnement de l'application, des temps de calcul trop longs peuvent empêcher le déploiement d'un algorithme. La rapidité d'exécution est donc importante.

Pour améliorer la rapidité, il est possible d'utiliser une heuristique moins performante lors de l'optimisation d'un champ de Markov. Cependant, la qualité des résultats est impactée si le nombre d'itérations n'est pas suffisant. La programmation parallèle est une des solutions à ce problème de temps de calcul, plus particulièrement la programmation sur processeur graphique (GPU). De nombreux algorithmes d'optimisation tels que les *Belief propagation* ont un facteur de parallélisation important. Dès lors, des algorithmes ont été proposés sur GPU [LCL⁺09, kCWP12] où le temps de calcul est de l'ordre de la centaine de millisecondes. Les algorithmes de coupe de graphe et de flot maximal, mentionnés dans le chapitre 3, peuvent aussi être implantés

8. www.flickr.com

en parallèle. Il existe, actuellement, des implantations de ces algorithmes sur GPU. Cependant, le code source n'est pas toujours disponible ou les interfaçages sont compliqués. Nous n'avons donc pas pu en profiter lors de nos travaux. Compte tenu de l'importance des champs de Markov et des développements en traitement parallèle, de nouvelles bibliothèques seront probablement disponibles dans un avenir proche, laissant espérer une accélération importante de l'algorithme proposé.

7.2.3 Restauration d'image monoculaire

La méthode de restauration, présentée dans la partie 4, avait comme objectif de réinterpréter l'algorithme introduit dans [THC⁺12], et de tester un nouvel *a priori* sur l'image restaurée afin d'apporter un terme de régularisation adéquat pour le modèle couplé avec la stéréo-vision.

Néanmoins, nous aurions pu définir un modèle bayésien complet sur le voile et l'image restaurée pouvant être minimisé de façon alternée, comme pour l'algorithme proposé dans [NKL12] et dans le chapitre 5, en ajoutant le biais avec un *a priori* supplémentaire sur l'intensité à estimer. Cette voie n'a pas pu être explorée car la convergence est extrêmement longue et les résultats étaient déjà satisfaisants. Cependant, il aurait été intéressant de les comparer avec ceux proposés.

Si les algorithmes d'optimisations sont plus rapides, il serait aussi possible de créer des modèles plus sophistiqués. En effet, le brouillard n'est pas forcément homogène dans toute l'image.

Grâce à l'implantation générique proposée, il est possible d'ajouter assez facilement un champ de variables aléatoires. Il serait donc possible d'essayer d'estimer le coefficient d'extinction à un facteur près en chaque pixel, avec pour chaque pixel, une variable aléatoire pour prendre en compte les variations d'un brouillard hétérogène. En supposant que le brouillard ait une densité continue dans le voisinage, un terme quadratique sur la régularisation serait alors judicieux.

7.2.4 Restauration du contraste et reconstruction simultanées à partir de paires stéréoscopiques en présence de brouillard

Remplacer la profondeur par le voile

La comparaison des résultats de l'algorithme de reconstruction et de restauration simultanées avec ceux de la restauration monoculaire estimant le voile atmosphérique, nous a montré que le voile, à longue distance, doit être échantillonné avec une plus grande précision. Si visuellement ces restaurations sont de meilleure qualité que celles obtenues en monoculaire dans les zones proches, elles contiennent encore des artéfacts. La création d'un modèle fondé sur le voile en prenant en compte la disparité serait envisageable pour améliorer la précision de la reconstruction et de la restauration au loin.

Étudier plus en détail le modèle avec la régularisation du second ordre et les occultations

Nous avons montré qu'il est possible, grâce à la méthode de fusion, d'optimiser le modèle avec un terme de régularisation du second ordre plus adapté aux scènes urbaines. Il serait donc intéressant par la suite, d'expérimenter différentes heuristiques pour l'optimisation d'un modèle avec un terme de régularisation du second ordre en prenant en compte, en plus, les occultations.

Approfondir l'étude de l'estimation du coefficient d'extinction

Les premiers résultats ont montré de bonnes perspectives pour estimer le coefficient d'extinction à partir du modèle. Il serait donc intéressant de poursuivre dans cet axe, dans l'optique de proposer un algorithme complet, pour estimer en même temps le coefficient d'extinction, la carte de profondeur et l'image restaurée.

7.2.5 Prendre en compte d'autres conditions dégradées

Si le développement d'un modèle spécifique pour le brouillard améliore de façon significative la qualité des algorithmes de reconstruction, une démarche similaire peut être envisagée pour les autres conditions dégradées. Le point commun entre le brouillard et les autres perturbations météorologiques est la dégradation de l'intensité originale de la scène par une perturbation autre que celles rencontrées habituellement tel que le bruit du capteur.

La pluie entraînera des spécularités sur les surfaces, et l'apparition de réflexions sur les surfaces mouillées. Pour traiter cela, un modèle possible consisterait, comme pour le brouillard, à modéliser le phénomène observé et à estimer l'intensité de la scène sans les spécularités ainsi que les caractéristiques de la perturbation. Dans ce cas, la BRDF (*Bidirectional reflectance distribution function* en anglais) est une fonction permettant de modéliser le comportement de l'intensité lumineuse quand elle atteint une surface en un point. Comme l'équation de Koschmieder, cette fonction permettrait de faire le lien entre l'intensité observée par la caméra, les caractéristiques du matériau observé, et celle de l'objet réfléchi. On pourrait donc imaginer un modèle global liant la carte de profondeur d'image et l'image restaurée et ainsi, chercher la solution qui minimise ce modèle global.

En poursuivant ce raisonnement, il faudrait modéliser chaque perturbation et combiner l'ensemble dans un seul modèle, afin d'obtenir simultanément l'intensité originale de la scène observée et la carte de profondeur.

Annexe A

Publications

Conférences :

- Markov Random Field Model for Single Image Defogging
 - **L. Caraffa** et J.-P. Tarel,
 - IEEE Intelligent Vehicles Symposium (IV'13),
 - Juin 2013. **Best Student Paper Award**. FRIDA2 comparison dataset.
- Stereo Reconstruction and Contrast Restoration in Daytime Fog
 - **L. Caraffa** et J.-P. Tarel,
 - IEEE Asian Conference on Computer Vision (ACCV'12),
 - Présentation **Orale**.

Revue :

- Vision Enhancement in Homogeneous and Heterogeneous Fog
 - J.-P. Tarel, N. Hautière, **L. Caraffa**, A. Cord, H. Halmaoui et D. Gruyer,
 - IEEE Intelligent Transportation Systems Magazine,
 - 2012.
- Combining Stereo and Atmospheric Veil Depth Cues for 3D Reconstruction
 - **L. Caraffa** et J.-P. Tarel,
 - IPSJ Transactions on Computer Vision and Applications - Special Issue on ACCV2012
 - Statut : à paraître en février 2014.

Chapitre de livre :

- Optimisation de fonctions pseudo-bouliennes quadratiques.
 - **L. Caraffa** et J.-P. Tarel,
 - Chapitre du livre "Optimisation de formes", études et recherches des laboratoires des ponts et chaussées"
 - (ERLPC), série "sciences pour le génie civil".
 - Statut : à paraître en 2014.

Annexe B

Filtre bilatéral guidé

B.1 Introduction

Le filtre bilatéral est un filtre très utilisé en imagerie numérique, il permet de lisser une image tout en préservant les contours. L'idée principale du filtre bilatéral, est l'ajout d'une pondération photométrique au filtre Gaussien. Grâce à ces propriétés, à sa facilité d'implantation et à sa flexibilité, le filtre bilatéral est actuellement utilisé dans de nombreux domaines, comme le rehaussement de contour [ZA08], le sur-échantillonnage [KCLU07], l'affinement de la carte de disparité en reconstruction 3D [YYDN07], l'édition d'image [KRFB06], la restauration d'image en condition de brouillard, etc.

Un résumé très complet sur l'interprétation, les améliorations et les extensions du filtre bilatéral est fait dans [PKTD07]. A notre avis, cinq points importants sont à relever. Premièrement, le filtre bilatéral est connecté à l'estimation robuste de l'intensité moyenne dans le voisinage, à l'équation de Perona-Malik [PM90], et du « local mode filtering » [vdWvdB01]. Deuxièmement, l'effet escalier qui peut être observé dans certains cas peut être évité en utilisant un modèle linéaire et non un modèle constant [BCM06]. Troisièmement, le filtre bilatéral adaptatif est une extension du filtre bilatéral où la pondération photométrique est sélectionnée en fonction de l'intensité du voisinage [ZA08]. Quatrièmement, le temps d'exécution du filtre bilatéral peut être amélioré de plusieurs façons, comme par exemple l'utilisation d'une structure de grille [DP06] ou une distribution d'histogrammes [Wei06]. Cinquièmement, le filtre bilatéral croisé [ED04, PSA⁺04] peut être utilisé pour introduire une information supplémentaire apportée par une image *guide*. L'idée est d'utiliser la pondération de l'image guide à la place de celle de l'image. Plus récemment, le filtre bilatéral dual [BMM07] a été proposé pour gérer deux images guide simultanément.

Les améliorations apportées par les points 2, 3 et 4 restent cohérentes avec l'interprétation robuste du filtre bilatéral, contrairement à l'introduction *ad hoc* du guide dans le filtre bilatéral croisé. Après un rappel dans la partie B.2 de l'interprétation du filtre bilatéral en terme d'estimation robuste et le rappel du filtre bilatéral robuste, nous proposons dans la section B.2 de relier le filtre bilatéral croisé à l'estimation robuste. Ce lien permet d'étendre le filtre bilatéral croisé en un nouveau filtre que nous nommons le *filtre bilatéral guidé*. Par la suite, une étude expérimentale est effectuée pour étudier la robustesse en terme statistique du filtre proposé en présence de différents types de bruits sur l'image originale et sur le guide. Enfin, une étude comparative montre l'avantage de la réintroduction du terme photométrique comparé aux méthodes actuelles.

B.2 Filtre bilatéral

Le filtre bilatéral, introduit dans [TM98] est une extension du filtrage linéaire où une pondération photométrique w_p est introduite en supplément du poids w_s . A partir de l'image originale E , le résultat du filtre bilatéral est l'image F donnée par :

$$F(x) = \frac{\sum_{t \in S_m} w_s(\|t\|) w_p(E(x) - E(x+t)) E(x+t)}{\sum_{t \in S_m} w_s(\|t\|) w_p(E(x) - E(x+t))} \quad (\text{B.1})$$

où S_m est une fenêtre de taille m , dans notre cas, une fenêtre carrée de taille $[-m, m] \times [-m, m]$. le poids w_s est une fonction symétrique, décroissante sur la distance $\|t\|$ à partir du centre de S_m . Généralement, le poids w_p est aussi une fonction symétrique et décroissante.

B.2.1 Lien avec l'estimation robuste

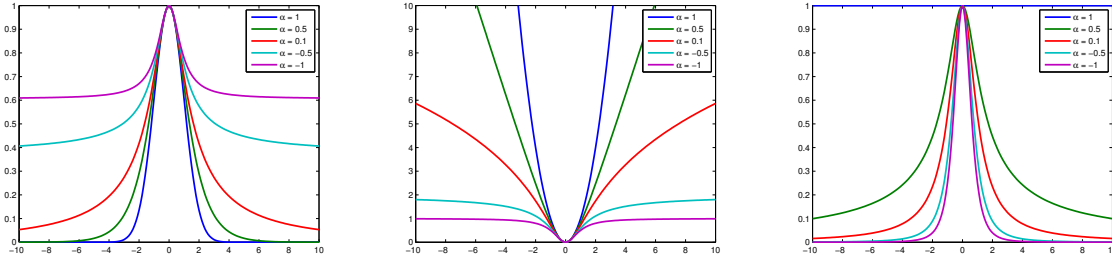


FIGURE B.1 – De gauche à droite, pour différentes valeurs de α dans la famille SEF : la densité de probabilité $\exp(-\frac{1}{2}\phi_\alpha(b^2))$, la fonction d'erreur $\phi_\alpha(b^2)$, et la pondération correspondante $\phi'_\alpha(b^2)$

Selon [Ela02], l'image résultat F du filtre bilatéral peut être interprétée comme la première étape de la minimisation du coût local :

$$\underset{F(x)}{\operatorname{argmin}} \sum_{t \in S_m} w_s(\|t\|) \phi((F(x) - E(x+t))^2) \quad (\text{B.2})$$

Quand $w_s(\|t\|) = 1$, ce coût implique que $F(x)$ est la moyenne robuste des valeurs $E(x+t)$, $t \in S_m$. Dans ce cas, ϕ caractérise le modèle du bruit additif sur les intensités dans le voisinage S_m . En annulant la dérivée de la somme dans (B.2) en fonction de $F(x)$, nous avons :

$$\sum_{t \in S_m} w_s(\|t\|) \phi'((F(x) - E(x+t))^2) (F(x) - E(x+t)) = 0$$

Après réécriture, et en utilisant la substitution « one-step-late » pour le terme non linéaire, l'algorithme des moindres carrés itératifs pondérés (Iterative Reweighted Least Squares : IRLS) est obtenu :

$$F_{k+1}(x) = \frac{\sum_{t \in S_m} w_s(\|t\|) \phi'((F_k(x) - E(x+t))^2) E(x+t)}{\sum_{t \in S_m} w_s(\|t\|) \phi'((F_k(x) - E(x+t))^2)} \quad (\text{B.3})$$

En comparant (B.1) avec (B.3), on constate que le filtre bilatéral est la première étape de l'algorithme IRLS quand $w_p(u) = \phi'(u^2)$ avec comme initialisation l'image originale (soit $F_0(x) = E(x)$).

Pour modéliser différents types de bruits, la densité de probabilité suivante est utilisée :

$$p_{\alpha,s}(b) \propto \frac{1}{s} e^{-\phi_{\alpha}((\frac{b}{s})^2)} \quad (\text{B.4})$$

où $\phi_{\alpha}(t) = \frac{1}{2\alpha}((1+t)^{\alpha} - 1)$. C'est la *famille des exponentielles lissées* utilisée dans [TIC02, ITC07]. La forme de la densité de probabilité est réglée par deux paramètres : s , qui est l'échelle du bruit et α , qui permet une transition continue entre les différents types de densités connues (SEF), (voir Fig. B.1). En particulier, la densité de probabilité Gaussienne est obtenue pour $\alpha = 1$, la densité de probabilité Laplace est approximée avec $\alpha = 0.5$, et la densité de probabilité Geman & McClure est obtenue pour $\alpha = -1$. La densité de probabilité correspondant à un bruit de Cauchy est obtenue quand $\alpha \rightarrow 0$, par l'intégration de ϕ' [ITC07].

Le coût (B.2) peut être dérivé du maximum de vraisemblance sur les résidus $b = F(x) + E(x+t)$ donné par $\|t\|$:

$$p(b|\|t\|) \propto \frac{1}{s(\|t\|)} e^{-\phi_{\alpha}((\frac{b}{s(\|t\|)})^2)} \quad (\text{B.5})$$

Avec ce modèle, en suivant la dérivation de (B.3) à partir de (B.2), un autre filtre est dérivé à partir de la vraisemblance. En comparant l'expression obtenue avec (B.5), la pondération spatiale $w_s(\|t\|)$ est associée avec $\frac{1}{s^2(\|t\|)}$ et donc la pondération photométrique $w_p(b)$ est associée à $\phi'_{\alpha}((\frac{b}{s(\|t\|)})^2)$. Remarquons qu'avec cette dérivation, l'échelle photométrique doit être l'inverse de la racine carré de la pondération spatiale. Cette dernière contrainte n'est pas prise en compte par le filtre bilatéral.

B.2.2 Filtre bilatéral robuste

En dépit de sa relation avec l'estimation robuste, le filtre bilatéral n'est pas robuste au bruit non Gaussien dans l'image originale. L'application itérative du filtre bilatéral n'est pas plus robuste, comme illustré dans la seconde colonne de la Fig. B.2. Au contraire, l'algorithme IRLS décrit dans la section précédente est robuste. En conséquence, nous avons appelé ce dernier le filtre bilatéral robuste. La dérivation de l'IRLS précédemment présentée est rapide et succincte, mais ne permet pas d'expliquer pourquoi l'algorithme IRLS converge vers un minimum local. Des preuves de convergence peuvent être trouvées par exemple dans [TIC08, CBFAB97].

Il est important de faire la différence entre le filtre bilatéral robuste et le filtre bilatéral itéré. En effet, dans le filtre bilatéral robuste (B.3), F_k est comparé à la valeur fixe E , tandis que le filtre bilatéral itéré compare F_k à la valeur F_{k-1} précédemment calculée. Cependant, comme cela est expliqué dans [ITC07], la convergence vers un bon minimum local n'est pas garantie, en effet, la fonction d'erreur n'étant pas convexe pour un $\alpha < 0.5$, l'initialisation nécessite quelques précautions. Une façon correcte de procéder consiste à initialiser $F_0(x)$ comme le résultat d'un filtre médian sur $E(x)$. Plus généralement, l'approche par non-convexité graduelle (Graduate non convexity (GNC)) [BZ87, ITC07] peut être utilisée. L'idée est d'utiliser la famille SEF ϕ_{α_p, s_p} et de forcer la convexité en fixant $\alpha_p = 1$ ou $\alpha_p = 0.5$, et ensuite de décrémenter α_p pas à pas en utilisant à chaque fois l'image résultante de l'itération précédente. Ce processus est itéré jusqu'à atteindre la valeur α_p cible. Une alternative consiste à minimiser le coût B.1 (en énumérant tous les niveaux de gris) afin de trouver le minimum global. Cette solution n'est cependant possible uniquement dans le cas où l'on estime un terme constant. Dans le cas où un degré plus important est estimé, il devient alors impossible d'énumérer toutes les solutions. En utilisant l'approche GNC, nous avons observé en pratique que l'énergie finale est très proche de la valeur de l'énergie globale, le *PSNR* étant souvent meilleur. L'approche GNC permet en plus

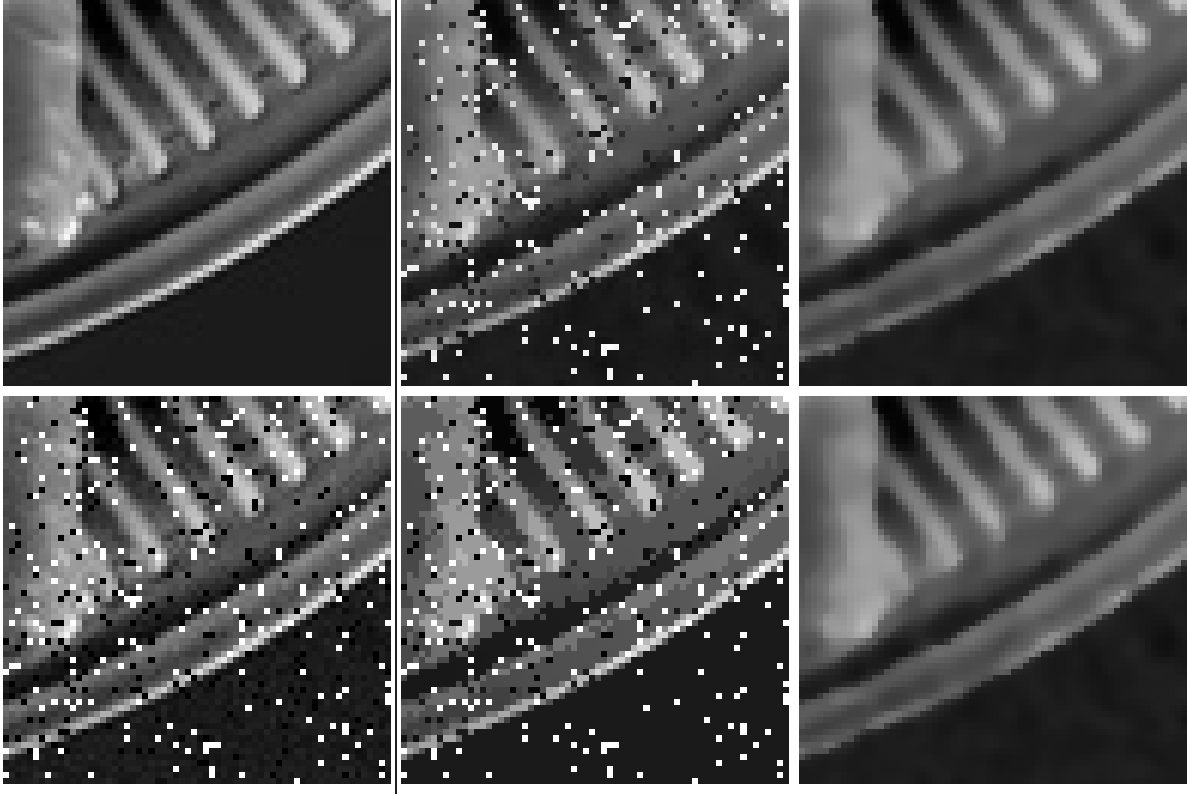


FIGURE B.2 – Première colonne : l'image originale et l'image bruitée. Seconde colonne : résultats avec le filtre bilatéral après 1 et 100 itérations. Troisième colonne : résultats du filtre bilatéral robuste avec 10 et 100 itérations.

de faire une régression linéaire et pas seulement constante comme nous l'expliquons par la suite, le coût en calcul est réduit d'un facteur 30 environ par rapport à l'approche exhaustive.

B.2.3 Estimation linéaire

Du fait de la modélisation par une constante dans (B.2), le filtre bilatéral itéré et le filtre bilatéral robuste sont sujets à un effet d'escalier dans les régions de faible gradient. A partir de [BCM06], l'effet escalier peut être supprimé par une estimation linéaire du voisinage de chaque pixel. Chaque itération consiste alors à effectuer l'algorithme IRLS sur les paramètres d'un plan 2D local :

$$\begin{pmatrix} H_{k+1}(x) \\ F_{k+1}(x) \end{pmatrix} = \left(R + \sum_{t \in S_m} w_s(\|t\|) w_{p,k} \begin{pmatrix} t \\ 1 \end{pmatrix} \begin{pmatrix} t \\ 1 \end{pmatrix}^t \right)^{-1} \sum_{t \in S_m} w_s(\|t\|) w_{p,k} E(x+t) \begin{pmatrix} t \\ 1 \end{pmatrix} \quad (\text{B.6})$$

Avec $w_{p,k} = \phi'((H_k^t(x)t + F_k(x) - E(x+t))^2)$, où R est une matrice diagonale assurant que la matrice soit bien inversible. $H_k(x)$ sont les coefficients directeurs du plan 2D à l'itération k .

Le résultat du filtre est la constante $F_k(x)$ du plan 2D. Comme dans la section précédente, la convergence vers un minimum local intéressant requiert l'utilisation du GNC.

B.3 Le filtre bilatéral guidé



FIGURE B.3 – Première colonne : l'image originale avec et sans flash. Seconde colonne : en haut, le résultat obtenu dans [PSA⁺04]. En bas, en utilisant le filtre bilatéral robuste sur l'image sans flash et avec comme guide l'image avec flash

Dans plusieurs applications, une autre image de la scène, prise dans une condition différente, peut être une source d'information supplémentaire d'autant plus lorsqu'elle est moins bruitée. Cette image est donc appelée le guide, mais elle est photométriquement distordue par rapport à l'image de la scène (exemple Fig. B.3). Le filtre bilatéral croisé a été introduit d'une façon *ad hoc* dans [PSA⁺04, ED04], à partir du filtre bilatéral, en substituant dans (B.1) la pondération photométrique w_p liée à l'image E avec celle de l'image guide G . L'image guide indique quels sont les pixels similaires dans chaque voisinage. Cette substitution a pour conséquence la perte du lien avec l'estimation robuste. Pour permettre de garder ce lien, nous proposons d'introduire l'image guide G en plus de w_p comme une troisième pondération dans (B.2). Nous estimons, contrairement à [PSA⁺04, ED04], que le poids w_g est une sorte de pondération spatiale, et non une pondération photométrique. L'image guide peut être considérée comme un étiquetage décrivant la structure de l'image dans le voisinage. Cela nous amène à introduire un nouveau

filtre : le filtre bilatéral guidé, qui minimise l'erreur suivante :

$$\sum_{t \in S_m} w_s(\|t\|) w_g(G(x) - G(x+t)) \phi((F(x) - E(x+t))^2) \quad (\text{B.7})$$

où w_s est la pondération spatiale, w_g est la pondération du guide et ϕ caractérise le modèle du bruit. Comme pour le filtre robuste, le filtre bilatéral guidé est itéré, et dérivé de (B.7) ce qui nous donne :

$$F_{k+1}(x) = \frac{\sum_{t \in S_m} \mu_t E(x+t)}{\sum_{t \in S_m} \mu_t} \quad (\text{B.8})$$

avec

$$\mu_t = w_s(\|t\|) w_g(G(x) - G(x+t)) \underbrace{\phi'((F_k(x) - E(x+t))^2)}_{w_p(F_k(x) - E(x+t))}$$

Le filtre bilatéral robuste est un cas particulier du filtre bilatéral guidé où la pondération du guide est constante, de même pour le filtre bilatéral croisé qui est la première itération du filtre bilatéral guidé quand $w_p = 1$. Contrairement au filtre bilatéral croisé, le filtre bilatéral guidé est robuste au bruit non Gaussien sur l'image originale si la pondération w_p , et par conséquent si ϕ' sont choisis de façon adéquate.

Du fait de ces 3 pondérations, le filtre bilatéral guidé est très flexible. En pratique, la pondération spatiale w_s peut être fixée à 1 quand le guide est de bonne qualité. En effet, la pondération du guide w_g donne une information plus précise sur la structure que la pondération spatiale. Néanmoins, si le guide est de mauvaise qualité, une fonction gaussienne peut être utilisée en pondération spatiale pour combler les manques du guide. La fenêtre peut être de différentes tailles et varier pour adapter la force du filtrage en fonction de l'a priori. A partir de maintenant, la fonction $w_g(b)$ a pour forme la fonction $p_{\alpha_g, s_g}(b)$, dans (B.4).

La pondération photométrique w_p est directement liée à la nature du bruit sur l'image E comme pour le filtre bilatéral robuste. Quand le bruit est Gaussien, w_p est égale à un. La fonction w_p est choisie dans la famille des SEF $\phi'_{\alpha_p}((\frac{b}{s_p})^2)$. Comme précédemment, une estimation linéaire du voisinage peut être choisie à la place du modèle constant pour éviter l'effet escalier.

Lorsque la fonction ϕ est non convexe, le GNC est utilisé pour converger vers un minimum local intéressant.

Par exemple, sur la Fig. B.3, nous avons utilisé les images avec et sans flash de [PSA⁺04] avec la méthode citée pour supprimer le bruit dans l'image sans flash avec l'aide de l'image avec flash. Pour obtenir la première image de la seconde colonne, l'image avec flash est filtrée en utilisant le filtre bilatéral, ensuite, l'image sans flash est filtrée en utilisant directement le filtre bilatéral guidé avec l'image sans flash comme guide. Enfin, les détails sur l'image avec flash sont introduits en multipliant le résultat précédant avec le ratio de l'image avec flash originale sur l'image avec flash lissée. La seconde image de la deuxième colonne de la figure B.3 est obtenue en utilisant le filtre robuste avec les paramètres suivants : $w_s = 1$, $\alpha_g = 1$, $s_g = 1$, $\alpha_p = -1$, $s_p = 3$, et finalement en utilisant le filtre bilatéral guidé avec les paramètres $w_s = 1$, $\alpha_g = 0$, $s_g = 10$, $\alpha_p = -1$, $s_p = 40$. Grâce à la robustesse du filtre bilatéral guidé, l'image est plus lisse avec un bruit complètement supprimé, l'intensité originale de l'image est mieux approximée. De plus, les spécularités de l'image avec flash ne sont pas transférées dans l'image sans flash.

B.4 Expérimentations

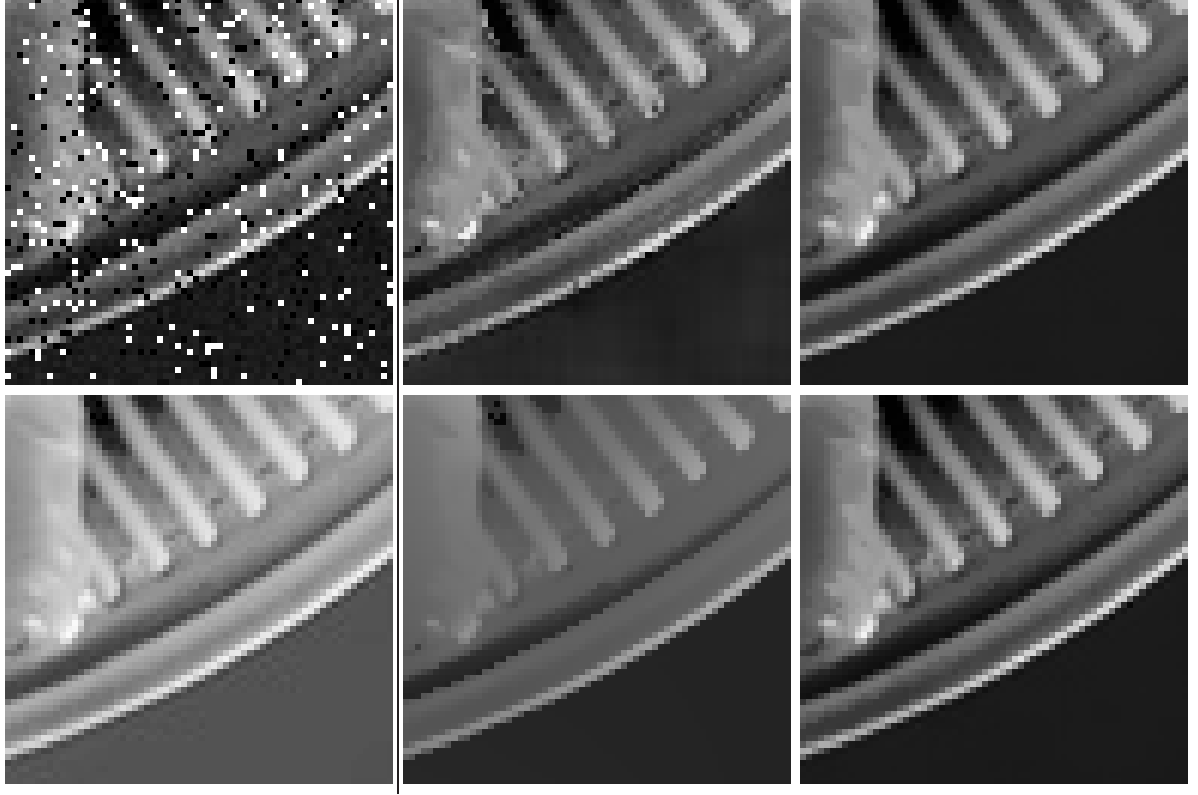


FIGURE B.4 – Première colonne : l'image bruitée E et le guide G sans bruit. Seconde colonne : résultat du filtre bilatéral croisé après 1 et 100 itérations. Troisième colonne : le résultat du filtre bilatéral guidé avec estimation constante et linéaire.

Nous avons étudié les performances du filtre bilatéral guidé comparativement aux autres algorithmes, pour cela, deux heuristiques permettant de converger vers un minimum local satisfaisant seront proposées suivant la présence ou l'absence de bruit sur le guide. Dans la première section, nous étudions le cas où un bruit est ajouté sur l'image originale avec un guide non bruité. Dans la deuxième partie, nous proposons une façon d'estimer les paramètres du filtre à partir des statistiques de l'image d'entrée. Dans la troisième partie, nous étudierons le cas où, du bruit est aussi ajouté sur le guide, d'abord quand les deux bruits sont générés de façon indépendante, puis, le cas où les deux bruits sont corrélés.

B.4.1 Sans bruit sur le guide

Les paramètres du filtre bilatéral guidé sont : la puissance α_g et l'échelle s_g de la pondération du guide w_g , la puissance α_p et l'échelle s_p de la pondération photométrique w_p , s_s est la pondération spatiale qui est supposée Gaussienne. Nous proposons l'heuristique suivante fondée sur le GNC pour permettre d'atteindre un minimum local intéressant :

- Mettre la puissance α_g et l'échelle s_g en fonction de l'intensité des détails et sélectionner la puissance α_p et l'échelle s_p de la pondération photométrique en fonction du bruit sur l'image d'entrée ;

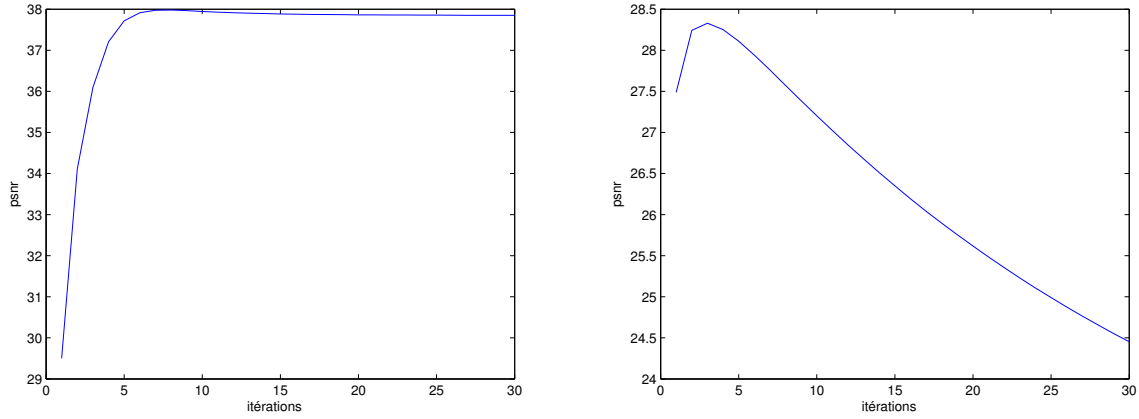


FIGURE B.5 – Comparaison de l'évolution du $PSNR$ au cours des itérations pour le filtre bilatéral guidé et le filtre bilatéral croisé itéré, le $PSNR$ filtre bilatéral croisé (droite) atteint son maximum après quelques itérations avant de diverger du fait des convolutions spatiales successives, le filtre bilatéral guidé, quant à lui, converge vers une valeur fixe et ne diverge que très peu grâce à l'attache aux données

- Effectuer une étape du filtre bilatéral guidé avec $w_p = 1$;
- Sur le résultat, faire une itération avec le filtre bilatéral guidé en changeant seulement α_p à la valeur 0.5
- Si la valeur d' α_p choisie dans la première étape est inférieure à zéro, faire une itération supplémentaire avec $\alpha_p = 0$;
- Finalement, itérer le filtre bilatéral guidé en utilisant α_p jusqu'à ce que le nombre d'itérations maximum 8 soit atteint.

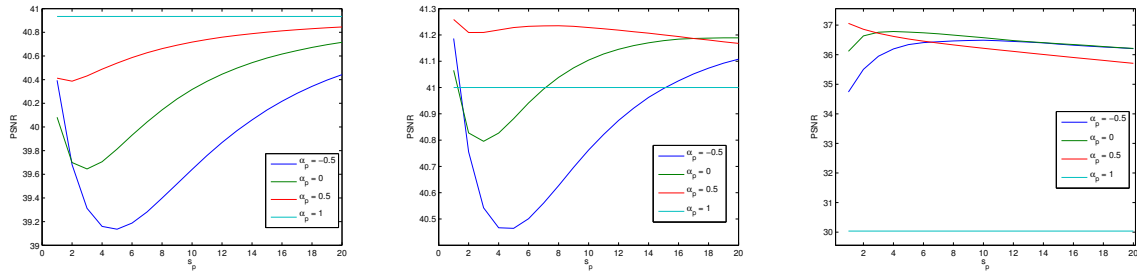


FIGURE B.6 – Variation du $PSNR$ pour différentes valeurs des paramètres de la pondération photométrique. De gauche à droite, trois différents types de bruits sont ajoutés sur l'image : Gaussien ($\alpha = 1$), Laplace ($\alpha = 0.5$) et Cauchy ($\alpha = 0$) avec une même échelle $s = 10$. Notons que les variations du $PSNR$ sont très faibles et que les maxima sont atteints pour des valeurs de la pondération photométrique proches de celles du bruit ajouté sur l'image.

Avec le modèle constant et les paramètres $\alpha_g = 0.5$, $\alpha_p = -1$, et $s_g = s_p = 5$, l'énergie obtenue en utilisant l'heuristique GNC précédente est 1.0022 fois l'énergie obtenue grâce à la méthode exhaustive (Moyenne sur 10 images bruitées avec un bruit Gaussien plus poivre et sel).

Les énergies obtenues grâce au GNC et à celles de la méthode exhaustive, sont par conséquent, très proches. Sur ces images, le $PSNR$ obtenu par la méthode exhaustive est de $37.2 \pm 0.5dB$. Étonnamment, avec l'approche GNC, le $PSNR$ obtenu est supérieur : $39.2 \pm 0.2dB$. Avec une variance diminuée. En effet, avec l'augmentation du nombre d'itérations, le $PSNR$ croît rapidement jusqu'à atteindre un maximum aux alentours d'une dizaine d'itérations, pour diminuer faiblement jusqu'à converger vers une asymptote du maximum atteint au fil des itérations (Fig. B.5). Un comportement similaire est observé quel que soit le type de bruit.

Le filtre bilatéral guidé est appliqué sur une image identique à celle de la Fig. B.2 avec un bruit Gaussien ($s = 5$) plus poivre et sel (10%) ajouté ($PSNR = 14.5dB$). Cette image est choisie pour sa difficulté, en effet, les fines parties formant des angles obtus sont sujettes à disparaître si aucune information de structure est donnée lors d'une restauration de l'image bruitée par une perturbation non Gaussienne. Le guide utilisé est, dans cet exemple, une transformation gamma de l'image originale sans ajout de bruit ($PSNR = 13.1dB$). La figure B.4 montre les résultats du filtre bilatéral croisé (avec $s_s = 1$, $\alpha_g = 1$, $s_g = 5$, $PSNR = 27.48 dB$) après 1 et 100 itérations (avec $s_s = 1$, $\alpha_g = 1$, $s_g = 5$, $PSNR = 20.71dB$). Dans le premier cas, le bruit poivre et sel n'est pas totalement supprimé. Avec 100 itérations, le bruit poivre et sel a été totalement supprimé, mais les convolutions spatiales itérées entraînent aussi une suppression des détails. La troisième colonne correspond au filtre bilatéral guidé avec un modèle constant ($PSNR = 39.6dB$) et linéaire ($PSNR = 38.5dB$). Le $PSNR$ obtenu par le filtre bilatéral guidé est largement supérieur à celui obtenu avec le filtre bilatéral croisé, itéré ou non. Ceci peut être expliqué par le fait que, contrairement au filtre bilatéral croisé, il n'y a pas de convolution spatiale augmentant avec les itérations. Le guide étant sans bruit, une grande valeur pour s_s , i.e. $w_s = 1$, est préférable pour utiliser toute l'information dans le voisinage procuré par le guide. La différence de $PSNR$ entre le modèle constant et linéaire est assez faible. L'avantage du modèle constant est que les détails fins circulaires sont mieux préservés, mais un α_p petit entraîne l'apparition d'un effet escalier dans les régions de faibles gradients. L'avantage du modèle linéaire est la suppression de l'effet escalier, en contre partie, certains détails fins sont supprimés.

Toujours sur la même image, nous avons étudié l'influence de la variation des paramètres $\alpha_p \in [-0.5, \dots, 1.0]$ et $s_p \in [1, \dots, 20]$ en fonction du type de bruit ajouté sur l'image. Trois différents types de bruits ont été ajoutés : Gaussien ($\alpha = 1$), Laplace ($\alpha = 0.5$) et Cauchy ($\alpha = 0$) avec un même paramètre d'échelle $s = 10$. La figure B.6 montre que les variations du $PSNR$ en fonction de α_p et s_p sont légères. Le meilleur $PSNR$ est atteint quand $\alpha_p = 0.5$ quand du bruit Gaussien est ajouté comme attendu. Avec un bruit Laplacien, le meilleur $PSNR$ est atteint pour deux valeurs d'échelle $s_p = 1$ et $s_p \approx 8$. Quand un bruit de Cauchy est ajouté, le meilleur $PSNR$ est atteint quand $s_p = 6$ et $\alpha_p = 0$ mais aussi pour $s_p = 1$ et $\alpha_p = 0.5$. Ces deux maxima peuvent être expliqués par le fait que l'intensité de l'image est bornée entre 0 et 255. Ces résultats montrent que le filtre bilatéral guidé ne nécessite pas une estimation très précise des paramètres pour fournir des résultats proches de l'optimal.

B.4.2 Estimation des paramètres à partir des statistiques de l'image

L'interprétation statistique des poids du filtre bilatéral robuste décrite à la fin de la partie B.2.1 suggère une façon d'estimer le produit entre la forme du produit des pondérations spatiales et photométriques. En effet, si on exclut le centre de la fenêtre, la densité de probabilité de $p(b|||t||)$ dans (B.5) peut être observée approximativement à partir du résidu $E(x) - E(x + t)$, pour chaque $||t||$, à partir de l'image d'entrée. La figure B.7 montre en rouge l'histogramme du résidu pour trois bruits (Gaussien, Laplacien et Cauchy avec comme échelle $s = 10$) ajoutés sur

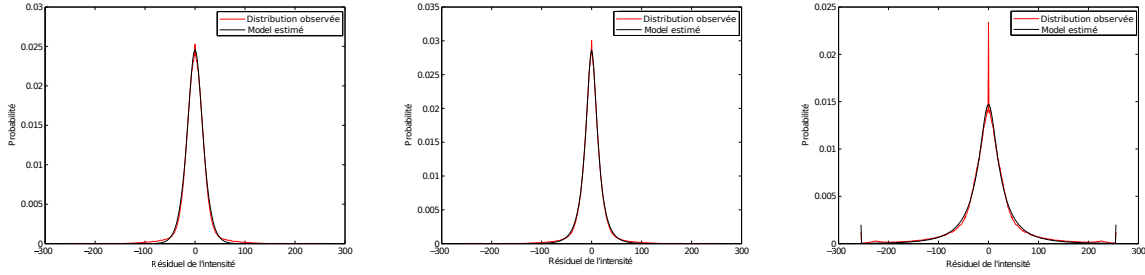


FIGURE B.7 – Estimation des paramètres à partir des statistiques de l'image. En rouge, l'histogramme des résidus des intensités de l'image sur laquelle différents types de bruits (de gauche à droite : Gaussien, Laplace, Cauchy avec une échelle de $s = 10$) ont été ajoutés. En noir, la régression par une loi des SEF.

la même image. La régression d'une loi de type SEF est faite sur chaque histogramme en minimisant la vraisemblance. Ces estimations apportent des valeurs intéressantes des paramètres afin d'initialiser le filtre bilatéral robuste. L'estimation obtenue sur chaque histogramme est tracée en noir sur la Fig B.7. Les valeurs estimées sont $(0.63, 13.6)$, $(0.38, 9.6)$, $(0.23, 15.4)$ respectivement pour les bruits Gaussien, Laplace et Cauchy. En utilisant ces valeurs en entrée du filtre bilatéral guidé, les PSNR moyens sur 10 images sont respectivement : 40.8, 41.2 et 36.3dB. Ces valeurs sont très proches du $PSNR$ optimal atteint sur la figure B.6 (respectivement 40.9, 41.3 et 37.0dB). En moyenne, les $PSNR$ obtenus sont seulement moins bons de 0.8% par rapport aux optimaux.

B.4.3 Bruit sur le guide

Les performances intéressantes obtenues avec le filtre bilatéral guidé en l'absence de bruit sur le guide ont été confirmées lorsqu'un bruit Gaussien ($s = 5$) est ajouté sur le guide. La figure B.8 montre sur quatre images le résultat du filtre bilatéral croisé, du filtre guidé introduit dans [HST10a] et du filtre bilatéral guidé sur une partie de chaque image. Les meilleures performances en terme de $PSNR$ en faisant varier les différents paramètres d'entrée sont affichées dans le Tableau B.10. Le filtre bilatéral guidé surpasse à chaque fois les deux autres filtres de plusieurs dB.

Avec l'image dont un extrait est visible dans la figure B.2, nous avons étudié, le cas où un bruit Gaussien ($s = 5$, $PSNR = 13.1dB$) et poivre et sel (10%, $PSNR = 11.1dB$) ont été ajoutés sur l'image et sur le guide. Des difficultés ont été observées quand le bruit sur l'image E et le guide G ont des valeurs proches. Cela nous a amené à tester le cas plus difficile comme la situation où le bruit ajouté sur le guide et sur l'image originale sont entièrement corrélés (Première colonne de la figure B.9). Dans la dernière colonne de la figure B.9, nous pouvons constater que le bruit poivre et sel n'est pas éliminé par le filtre bilatéral guidé (avec $\alpha_g = 0$, $\alpha_p = -1$, $s_g = s_p = 5$, $PSNR = 14.5dB$) à cause de la corrélation. Le filtre bilatéral guidé n'est pas robuste au bruit quand il est corrélé entre l'image d'entrée et le guide. La solution est donc d'effectuer un pré-traitement en utilisant le filtre bilatéral robuste sur le guide (avec $s_s = 1$, $\alpha_p = -1$, $s_p = 5$). Après ce traitement sur le guide, la procédure utilisant le GNC est appliquée sur l'image d'entrée comme précédemment, mais, avec une pondération spatiale s_s plus petite afin de prendre en compte le fait que le guide est moins précis que dans le cas sans bruit. Un bon $PSNR$ est obtenu (29.0dB) en utilisant le pré-traitement précédemment

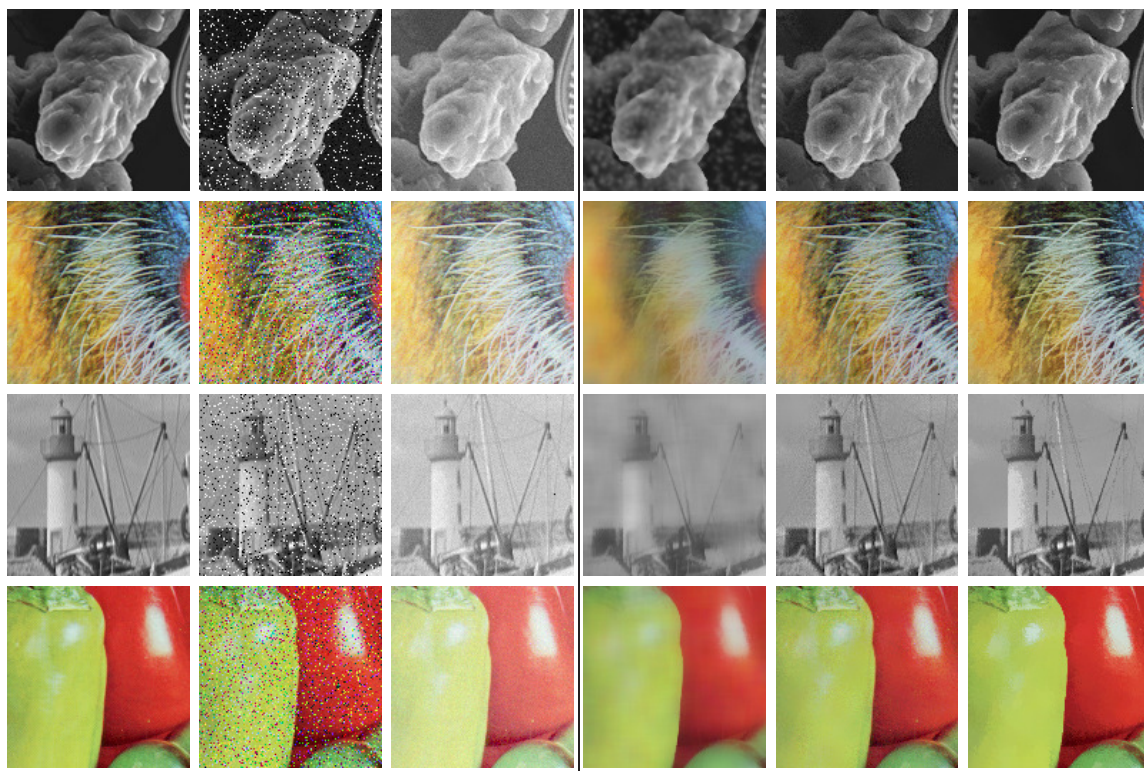


FIGURE B.8 – De gauche à droite : L'image originale, l'image dégradée avec un bruit Gaussien ($s = 5$) plus poivre et sel (10%), l'image guide ($\gamma = 1.4$), le résultat du filtre bilatéral croisé, du filtre guidé [HST10a] et du filtre bilatéral guidé, le résultat du filtre bilatéral croisé itéré, le résultat du filtre guidé, le résultat du filtre bilatéral guidé, notons que l'intensité de l'image originale est mieux restituée avec le filtre bilatéral guidé du fait de son caractère robuste

décrit. En comparaison, sans le guide avec les mêmes paramètres, le $PSNR$ est égal à $27.4dB$, Les résultats obtenus en utilisant le filtre bilatéral croisé sont de moins bonne qualité dans les deux cas, avec ($w_s = 1$, $\alpha_g = 0$, $s_g = 70$, $PSNR = 23.3dB$) et sans pré-traitement sur le guide ($s_s = 1.5$, $\alpha_g = 0$, $s_g = 70$, $PSNR = 21.5dB$).

B.5 Applications

Pour illustrer l'intérêt du filtre bilatéral guidé, nous allons montrer quelques exemples d'application comme l'affinement de carte de profondeur et la restauration d'image en présence de brouillard en plus du cas du débruitage d'image prise avec flash et sans flash de la figure B.3.

B.5.1 Affinement de la carte de profondeur

Le filtre bilatéral croisé a été utilisé dans [YYDN07] pour affiner la carte de profondeur et interpoler en utilisant une des images de la paire stéréo comme guide. La figure B.11 montre une carte de disparité obtenue en utilisant l'algorithme de propagation de croyance (« belief propagation »). L'algorithme a été stoppé volontairement avant la convergence totale. Même avec plusieurs valeurs aberrantes sur la carte de disparité, l'affinement réalisé avec le filtre

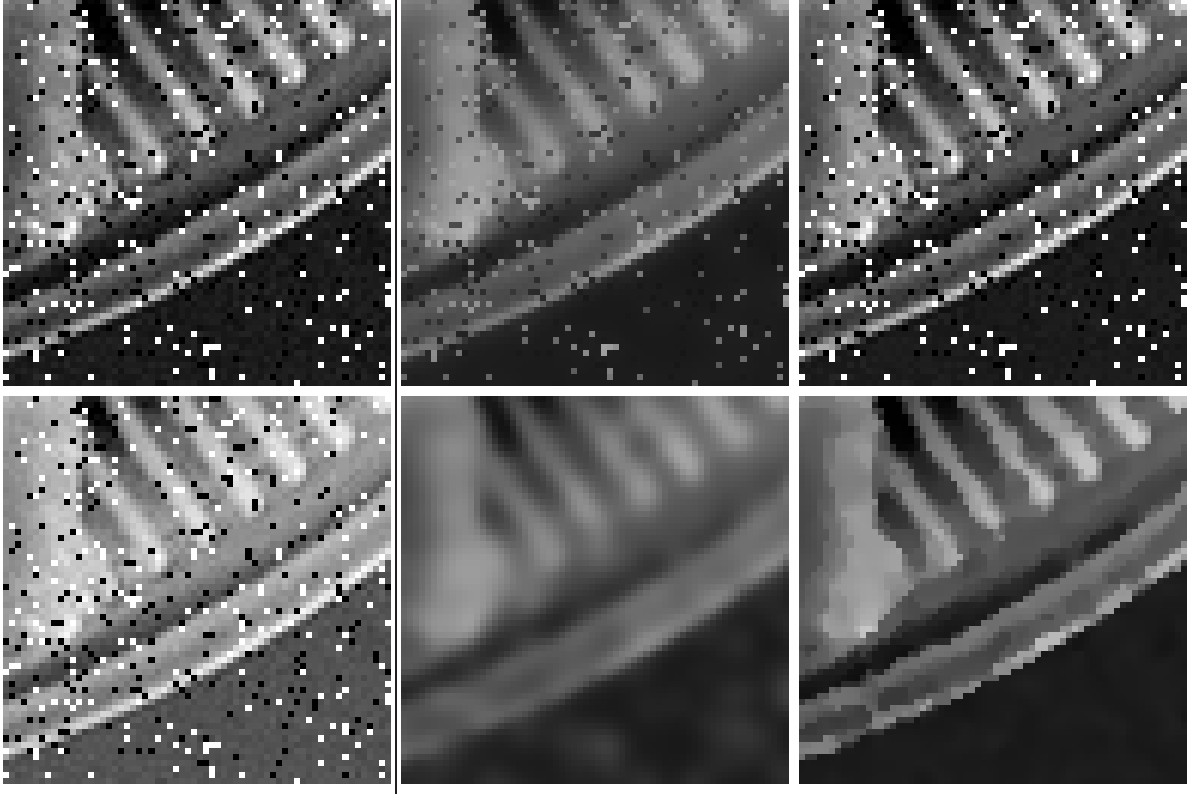


FIGURE B.9 – Première colonne : l'image E à traiter avec un bruit Gaussien plus poivre et sel et l'image guide G avec du bruit Gaussien et poivre et sel corrélé avec celui de l'image à traiter. Seconde colonne : résultat du filtre bilatéral croisé sans et avec le pré-traitement sur le guide. Troisième colonne : le résultat du filtre bilatéral guidé sans et avec le pré-traitement sur le guide.

bilatéral guidé produit de bons résultats. Pour obtenir ce résultat, une approche par GNC a été effectuée avec les paramètres : $w_s = 1$, $\alpha_g = 0$, $s_g = 10$, $\alpha_p = -1$, $s_p = 40$. Le résultat obtenu avec le filtre bilatéral croisé est moins convaincant (cf figure B.11).

B.5.2 Restauration d'image en présence de brouillard

Comme dans [TH09], la restauration d'image en présence de brouillard est effectuée en deux étapes : d'abord, le voile atmosphérique est estimé à partir des couleurs saturées (le maximum des composantes RVB pour chaque pixel), ensuite un filtre médian est effectué. Le voile estimé est utilisé pour supprimer le brouillard en inversant la loi de Koshmieder qui modélise l'effet du brouillard. Entre ces deux étapes, le voile atmosphérique peut être affiné en appliquant le filtre bilatéral guidé sur le voile, en utilisant l'image couleur originale comme guide. Cette étape supplémentaire améliore la suppression du brouillard dans les zones fines et complexes comme les feuilles (figure . B.12.). Les paramètres du filtre bilatéral guidé utilisés sont $w_s = 1$, $\alpha_g = 0.5$, $s_g = 10$, $\alpha_p = 0$, $s_p = 10$ dans cet exemple.

Filtre	filtre bilatéral croisé	filtre guidé	filtre bilatéral guidé
Hudson diathom	24.9	28.3	35.3
Baboon	13.4	28.7	31.6
Fishing boat	14.0	29.9	32.9
Peppers	14.2	28.6	34.03

FIGURE B.10 – Le meilleur $PSNR$ en fonction des valeurs des paramètres pour les 4 images de la figure. B.8 avec le filtre bilatéral croisé, le filtre guidé [HST10a] et le filtre bilatéral guidé.

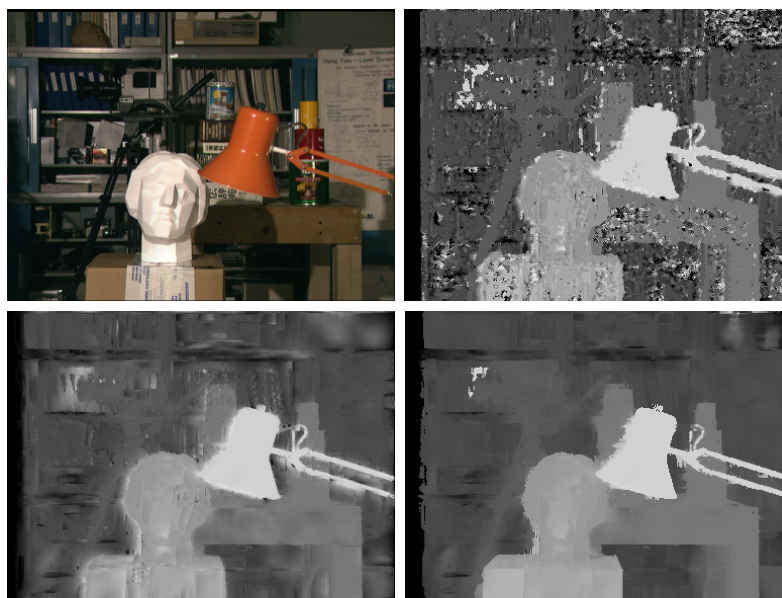


FIGURE B.11 – Première ligne : à gauche, l'image originale de la paire stéréo et la carte de disparité obtenue en utilisant un algorithme de reconstruction dense. Seconde ligne : la carte de disparité affinée par le filtre bilatéral croisé et le filtre bilatéral guidé

B.6 Conclusion

Nous avons décrit les points importants du filtre bilatéral guidé et de ses variantes qui permettent de renforcer le lien entre le filtre bilatéral, le filtre bilatéral robuste et l'estimation robuste. Pour garder le lien entre l'estimation robuste et le filtre bilatéral croisé, nous avons proposé une nouvelle façon d'introduire ce dernier, ce qui nous a amené naturellement à la définition d'un filtre plus générique, que nous appelons le filtre bilatéral guidé. Nous avons décrit le filtre bilatéral guidé et ses variantes. Le filtre a pour cas particuliers le filtre bilatéral, le filtre bilatéral robuste, le filtre bilatéral croisé. Les entrées de ce filtre sont deux images : l'image à traiter et le guide. Le filtre bilatéral guidé nécessite cinq paramètres. Nous proposons une façon de fixer ces paramètres en utilisant les images d'entrée. Ce filtre est itératif, pour cela, nous utilisons une heuristique, basée sur l'approche de non-convexité graduelle (GNC) pour converger vers minimum local intéressant de la fonction de coût en un nombre réduit d'itérations. Expérimentalement, en terme de $PSNR$, le filtre bilatéral guidé surpasse les autres filtres qui sont capables de prendre en compte un guide. Le filtre proposé est robuste aux valeurs



FIGURE B.12 – Première colonne : l'image originale en présence de brouillard. Seconde colonne : suppression du brouillard en utilisant le filtre médian [TH09]. Troisième colonne : suppression du brouillard en utilisant le filtre bilatéral guidé. Les halos produits par le filtre médian sont en grande partie supprimés avec le filtre bilatéral guidé

aberrantes dans l'image traitée mais il ne l'est pas lorsqu'elles sont dans le guide. Dans ce cas, la solution proposée consiste simplement à pré-traiter le guide en utilisant le filtre bilatéral robuste et en prenant en compte la réduction de qualité dans la pondération spatiale du filtre bilatéral guidé. Le filtre proposé peut être utilisé dans plusieurs applications comme le rehaussement de contrastes, le sur-échantillonnage, le traitement d'images avec flash/sans flash, l'affinement de carte de profondeur ou la restauration d'images en présence de brouillard.

Bibliographie

- [AFS⁺11] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, and Richard Szeliski. Building rome in a day. *Commun. ACM*, 54(10) :105–112, 2011.
- [Bar98] Simon A. Barker. *Image Segmentation Using Markov Random Field Models*. PhD thesis, University of Cambridge, 1998.
- [BBBDR05] M. Bertozzi, E. Binelli, A. Broggi, and M. Del Rose. Stereo vision-based approaches for pedestrian detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops - Volume 03*, CVPR '05, pages 16–, Washington, DC, USA, 2005. IEEE Computer Society.
- [BCM06] A. Buades, B. Coll, and J. M. Morel. The staircasing effect in neighborhood filters and its solution. *IEEE Trans. Image Processing*, 15(6) :1499–1505, June 2006.
- [BD98] C. Bush and E. Debes. Wavelet transform for analyzing fog visibility. *IEEE Intelligent Systems*, 13(6) :66–71, 1998.
- [Bes74] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974.
- [BH02] Endre Boros and Peter L Hammer. Pseudo-boolean optimization. *Discrete applied mathematics*, 123(1) :155–225, 2002.
- [BHT06] Endre Boros, Peter L. Hammer, and Gabriel Tavares. Preprocessing of unconstrained quadratic binary optimization. Technical report, 2006.
- [BHT11] J. Bossu, N. Hautière, and J.-P. Tarel. Rain or snow detection in image sequences through use of a histogram of orientation of streaks. *International Journal of Computer Vision*, 93 :348–367, July 2011.
- [BMM07] E. P. Bennett, J. L. Mason, and L. McMillan. Multispectral bilateral video fusion. *IEEE Trans. Image Processing*, 16(5) :1185–1194, May 2007.
- [BRK⁺11] Michael Bleyer, Carsten Rother, Pushmeet Kohli, Daniel Scharstein, and Sudepta N. Sinha. Object stereo - joint stereo matching and object segmentation. In *CVPR*, pages 3081–3088, 2011.
- [BVZ01a] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11) :1222–1239, November 2001.
- [BVZ01b] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23 :1222–1239, November 2001.

- [BZ87] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.
- [CBFAB97] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Processing*, 6(2) :298–311, 1997.
- [CT13] L. Caraffa and J.-P. Tarel. Markov random field model for single image defogging. In *Proceedings of IEEE Intelligent Vehicle Symposium (IV'2013)*, Gold Coast, Australia, 2013.
- [DP06] F. Durand and S. Paris. A fast approximation of the bilateral filter using a signal processing approach. In *ECCV*, pages IV : 568–580, 2006.
- [ED04] Elmar Eisemann and Frédo Durand. Flash photography enhancement via intrinsic relighting. *ACM Trans. on Graphics*, 23(3) :673–678, August 2004.
- [Ela02] M. Elad. On the origin of the bilateral filter and ways to improve it. *IEEE Trans. Image Processing*, 11(10) :1141–1151, October 2002.
- [Fau93] Olivier Faugeras. *Three-dimensional computer vision : a geometric viewpoint*. MIT Press, Cambridge, MA, USA, 1993.
- [FCK09] Philippe Foucher, Pierre Charbonnier, and Houssein Kebbous. Evaluation of a road sign pre-detection system by image analysis. In Alpesh Ranchordas and Helder Araújo, editors, *VISAPP (2)*, pages 362–367. INSTICC Press, 2009.
- [FDL09] Graham D. Finlayson, Mark S. Drew, and Cheng Lu. Entropy minimization for shadow removal. *Int. J. Comput. Vision*, 85(1) :35–57, October 2009.
- [FGBZ11] Alexander Fix, Aritanan Gruber, Endre Boros, and Ramin Zabih. A graph cut algorithm for higher-order markov random fields. In *Proceedings of the 2011 International Conference on Computer Vision, ICCV '11*, pages 1020–1027, Washington, DC, USA, 2011. IEEE Computer Society.
- [FH06] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1) :41–54, October 2006.
- [GBO12] Tom Goldstein, Xavier Bresson, and Stanley Osher. Global minimization of markov random fields with applications to optical flow. *Inverse Problems and Imaging*, 6(4) :623–644, 2012.
- [GBP11] A. C. Gallagher, D. Batra, and D. Parikh. Inference for order reduction in markov random fields. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '11*, pages 1857–1864, Washington, DC, USA, 2011. IEEE Computer Society.
- [GG84] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6) :721–741, November 1984.
- [GLU12] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [GRdL⁺06] D. Gruyer, C. Royere, N. du Lac, G. Michel, and J-M Blosseville. Sivic and rt-maps, interconnected platforms for the conception and the evaluation of driving assistance systems. In *Proc. Intelligent Transport Systems World Congress*, pages 1–8, London, England, 2006.

-
- [GRU10] Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient large-scale stereo matching. In *Asian Conference on Computer Vision*, Queenstown, New Zealand, November 2010.
- [GS05] Pau Gargallo and Peter Sturm. Bayesian 3D Modeling from Images using Multiple Depth Maps. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *IEEE Workshop on Motion and Video Computing*, volume 2, pages 885–891, San Diego, États-Unis, 2005. IEEE Computer Society.
- [HA05] N. Hautière and D. Aubert. Contrast restoration of foggy images through use of an onboard camera. In *Proc. IEEE Conference on Intelligent Transportation Systems (ITSC’05)*, pages 1090–1095, Vienna, Austria, 2005.
- [HB90] F. Heitz and P. Bouthemy. Motion estimation and segmentation using a global bayesian approach. In *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, pages 2305–2308 vol.4, 1990.
- [HB93] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(12) :1217–1232, 1993.
- [HBD⁺11] Nicolas Hautière, Raouf Babari, Éric Dumont, Roland Brémond, and Nicolas Paparoditis. Estimating meteorological visibility using cameras : a probabilistic model-driven approach. In *Proceedings of the 10th Asian conference on Computer vision - Volume Part IV, ACCV’10*, pages 243–254, Berlin, Heidelberg, 2011. Springer-Verlag.
- [HBGR09] Asmaa Hosni, Michael Bleier, Margrit Gelautz, and Christoph Rhemann. Local stereo matching using geodesic support weights. In *Proceedings of the 16th IEEE international conference on Image processing, ICIP’09*, pages 2069–2072, Piscataway, NJ, USA, 2009. IEEE Press.
- [HCH11] Houssam Halmaoui, Aurélien Cord, and Nicolas Hautière. Contrast restoration of road images taken in foggy weather. In *Computational Methods for the Innovative Design of Electrical Devices*, pages 2057–2063, 2011.
- [Hir05] Heiko Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) - Volume 2 - Volume 02, CVPR ’05*, pages 807–814, Washington, DC, USA, 2005. IEEE Computer Society.
- [Hir08] Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information. *PAMI*, 30 :328–41, 2008.
- [HLA06] Nicolas Hautière, Raphaël Labayrade, and Didier Aubert. Estimation of the visibility distance by stereovision : A generic approach. *IEICE Transactions*, 89-D(7) :2084–2091, 2006.
- [How08] Andrew Howard. Real-time stereo visual odometry for autonomous ground vehicles. In *IROS*, pages 3946–3952, 2008.
- [HS09] Heiko Hirschmuller and Daniel Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9) :1582–1599, 2009.
- [HST10a] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In *European Conference on Computer Vision (ECCV’10)*, pages 1–14, Hersonissos, Crete, Greece, 2010.

- [HST10b] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12) :2341–2353, December 2010.
- [HTA07] N. Hautière, J.-P. Tarel, and D. Aubert. Towards fog-free in-vehicle vision systems through contrast restoration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, pages 1–8, Minneapolis, Minnesota, USA, 2007. <http://perso.lcpc.fr/tarel.jean-philippe/publis/cvpr07.html>.
- [HTA09] Nicolas Hautière, Jean-Philippe Tarel, and Didier Aubert. Free space detection for autonomous navigation in daytime foggy weather. In *Proceedings of IAPR Conference on Machine Vision Applications (MVA'09)*, pages 501–504, Yokohama, Japan, 2009. <http://perso.lcpc.fr/tarel.jean-philippe/publis/mva09.html>.
- [HTA10] N. Hautière, J.-P. Tarel, and D. Aubert. Mitigation of visibility loss for advanced camera based driver assistances. *IEEE Transactions on Intelligent Transportation Systems*, 11(2) :474–484, June 2010. <http://perso.lcpc.fr/tarel.jean-philippe/publis/its10.html>.
- [HTLA06] Nicolas Hautière, Jean-Philippe Tarel, Jean Lavenant, and Didier Aubert. Automatic fog detection and estimation of visibility distance through use of an onboard camera. *Machine Vision and Applications*, 17(1) :8–20, 2006. <http://perso.lcpc.fr/tarel.jean-philippe/publis/mva06.html>.
- [Ish09a] Hiroshi Ishikawa. Higher-order clique reduction in binary graph cut. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*, pages 2993–3000. IEEE, 2009.
- [Ish09b] Hiroshi Ishikawa. Higher-order gradient descent by fusion-move graph cut. In *ICCV*, pages 568–574. IEEE, 2009.
- [Ish11] Hiroshi Ishikawa. Transformation of general binary mrf minimization to the first-order case. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(6) :1234–1249, 2011.
- [ITC07] S.-S. Ieng, J.-P. Tarel, and P. Charbonnier. Modeling non-gaussian noise for robust image analysis. In *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP'07)*, pages 183–190, Barcelona, Spain, 2007. <http://perso.lcpc.fr/tarel.jean-philippe/publis/visapp07a.html>.
- [JRW97] D.J. Jobson, Z. Rahman, and G.A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, 6(7) :965–976, 1997.
- [KCLU07] Johannes Kopf, Michael F. Cohen, Dani Lischinski, and Matt Uyttendaele. Joint bilateral upsampling. *ACM Trans. on Graphics*, 26, July 2007.
- [kCWP12] Young kyu Choi, W. Williem, and In Kyu Park. Memory-efficient belief propagation in stereo matching on gpu. In *Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pages 1–4, 2012.
- [KHD12] Vivek Kwatra, Mei Han, and Shengyang Dai. Shadow removal for aerial imagery by information theoretic intrinsic image analysis. In *International Conference on Computational Photography*, 2012.
- [KNC⁺08] Johannes Kopf, Boris Neubert, Billy Chen, Michael F. Cohen, Daniel Cohen-Or, Oliver Deussen, Matt Uyttendaele, and Dani Lischinski. Deep photo : Model-based

-
- photograph enhancement and viewing. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2008)*, 27(5) :116 :1–116 :10, 2008.
- [KO94] Takeo Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window : theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9) :920 – 932, September 1994.
- [Kos24] H Koschmieder. Theorie der horizontalen sichtweite ii : Kontrast und sichtweite [theory of horizontal visibility]. *Meteorologische Zeitschrift*, pages 171–181, 1924.
- [KRFB06] Erum Arif Khan, Erik Reinhard, Roland Fleming, and Heinrich Buelthoff. Image-based material editing. *ACM Trans. on Graphics*, 25(3), July 2006.
- [KS11] Fredrik Kahl and Petter Strandmark. Generalized roof duality for pseudo-boolean optimization. In *Proceedings of the 2011 International Conference on Computer Vision, ICCV '11*, pages 255–262, Washington, DC, USA, 2011. IEEE Computer Society.
- [KZ01] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions using graph cuts. In *International Conference on Computer Vision*, pages 508–515, 2001.
- [KZ02] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? In *Proceedings of the 7th European Conference on Computer Vision-Part III, ECCV '02*, pages 65–81, London, UK, UK, 2002. Springer-Verlag.
- [LAT02] R. Labayrade, D. Aubert, and J.-P. Tarel. Real time obstacle detection on non flat road geometry through ‘v-disparity’ representation. In *Proceedings of IEEE Intelligent Vehicle Symposium*, volume 2, pages 646–651, Versailles, France, 2002.
- [LCL⁺09] Chia-Kai Liang, Chao-Chung Cheng, Yen-Chieh Lai, Liang-Gee Chen, and H.H. Chen. Hardware-efficient belief propagation. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 0 :80–87, 2009.
- [LRRB10] Victor Lempitsky, Carsten Rother, Stefan Roth, and Andrew Blake. Fusion moves for markov random field optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(8) :1392–1405, August 2010.
- [LSR⁺10] Lubor Ladický, Paul Sturges, Christopher Russell, Sunando Sengupta, Yalin Bastanlar, William Clocksin, and Philip Torr. Joint optimisation for object class segmentation and dense stereo reconstruction. In *Proceedings of the British Machine Vision Conference*, pages 104.1–104.11. BMVA Press, 2010. doi :10.5244/C.24.104.
- [LTA02] J. Lavenant, J.-P. Tarel, and D. Aubert. Procédé de détermination de la distance de visibilité et procédé de détermination de la présence d’un brouillard. French patent number 0201822, INRETS/LCPC, February 2002.
- [Nag91] Shojiro Nagata. Pictorial communication in virtual and real environments. chapter How to reinforce perception of depth in single two-dimensional pictures, pages 527–545. Taylor & Francis, Inc., Bristol, PA, USA, 1991.
- [NKL12] Ko Nishino, Louis Kratz, and Stephen Lombardi. Bayesian defogging. *International Journal of Computer Vision*, 98 :263–278, 2012.
- [NN03] S. G. Narashiman and S. K. Nayar. Interactive deweathering of an image using physical model. In *IEEE Workshop on Color and Photometric Methods in Computer Vision*, Nice, France, 2003.

- [PKG99] Marc Pollefeys, Reinhard Koch, and Luc Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *INTERNATIONAL JOURNAL OF COMPUTER VISION*, pages 7–25, 1999.
- [PKTD07] Sylvain Paris, Pierre Kornprobst, Jack Tumblin, and Frédo Durand. A gentle introduction to bilateral filtering and its applications. In *ACM SIGGRAPH 2007 courses*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.
- [PM90] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(7) :629–639, July 1990.
- [Pom97] D. Pomerleau. Visibility estimation from a moving vehicle using the ralph vision system. *IEEE Conference on Intelligent Transportation Systems*, pages 906–911, November 1997.
- [Pri12] Simon J. D. Prince. *Computer Vision : Models, Learning, and Inference*. Cambridge University Press, New York, NY, USA, 1st edition, 2012.
- [PSA⁺04] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. on Graphics*, 23(3) :664–672, August 2004.
- [Ros75] I. Rosenberg. Reduction of bivalent maximization to the quadratic case. *Cahiers du Centre d'Etudes de Recherche Operationnelle*, page 17 :7174, 1975.
- [SS02] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47(1-3) :7–42, April 2002.
- [SSZ02] J. Sun, H. Y. Shum, and N. N. Zheng. Stereo matching using belief propagation. page II : 510 ff., 2002.
- [Str12] Petter Strandmark. *Discrete Optimization in Early Vision*. PhD thesis, Centre for Mathematical Sciences LTH, Lund University, Sweden, 2012.
- [Tan08] R.T. Tan. Visibility in bad weather from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*, pages 1–8, Anchorage, Alaska, 2008.
- [Ter86] Demetri Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(4) :413–424, 1986.
- [TH09] Jean-Philippe Tarel and Nicolas Hautière. Fast visibility restoration from a single color or gray level image. In *ICCV*, pages 2201–2208, Kyoto, Japan, 2009. <http://perso.lcpc.fr/tarel.jean-philippe/publis/iccv09.html>.
- [THC⁺10] J.-P. Tarel, N. Hautière, A. Cord, D. Gruyer, and H. Halmaoui. Improved visibility of road scene images under heterogeneous fog. In *Proceedings of IEEE Intelligent Vehicle Symposium (IV'2010)*, pages 478–485, San Diego, California, USA, 2010. <http://perso.lcpc.fr/tarel.jean-philippe/publis/iv10.html>.
- [THC⁺12] J.-P. Tarel, N. Hautière, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer. Vision enhancement in homogeneous and heterogeneous fog. *IEEE Intelligent Transportation Systems Magazine*, 4(2) :6–20, Summer 2012.
- [TIC02] J.-P. Tarel, S.-S. Ieng, and P. Charbonnier. Using robust estimation algorithms for tracking explicit curves. In *ECCV*, volume I, pages 492–507, Copenhagen, Denmark, 2002. <http://perso.lcpc.fr/tarel.jean-philippe/publis/eccv02.html>.

-
- [TIC07] J.-P. Tarel, S.-S. Ieng, and P. Charbonnier. *Robust Lane Marking Detection by the Half Quadratic Approach*. Collection Etudes et Recherches des Laboratoires des Ponts et Chaussées, CR 49. LCPC, November 2007.
- [TIC08] J.-P. Tarel, S.-S. Ieng, and P. Charbonnier. A constrained-optimization based half-quadratic algorithm for robustly fitting sets of linearly parametrized curves. *Advances in Data Analysis and Classification*, 2(3) :227–239, december 2008. <http://perso.lcpc.fr/tarel.jean-philippe/publis/adac08.html>.
- [TM98] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998.
- [TPP07] R.T. Tan, N. Pettersson, and L. Petersson. Visibility enhancement for roads with foggy or hazy scenes. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV’07)*, pages 19–24, Istanbul, Turkey, 2007.
- [UWSI12] Christian Unger, Eric Wahl, Peter Sturm, and Slobodan Ilic. Stereo Fusion from Multiple Viewpoints. In Axel Pinz, Thomas Pock, Horst Bischof, and Franz Leberl, editors, *DAGM-OAGM Joint Pattern Recognition Symposium*, volume 7476 of *Lecture Notes in Computer Science*, pages 468–477, Graz, Autriche, 2012. Springer.
- [vdWvdB01] Joost van de Weijer and Rein van den Boomgaard. Local mode filtering. In *CVPR*, pages 428–433, 2001.
- [Wei06] Ben Weiss. Fast median and bilateral filtering. *ACM Trans. on Graphics*, 25 :519–526, July 2006.
- [WTRF09] Oliver Woodford, Philip Torr, Ian Reid, and Andrew Fitzgibbon. Global stereo reconstruction under second-order smoothness priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(12) :2115–2128, December 2009.
- [WZ08] Z. F. Wang and Z. G. Zheng. A region based stereo matching algorithm using cooperative optimization. pages 1–8, 2008.
- [YAC⁺10] Wei Yu, Ahmed Bilal Ashraf, Yao-Jen Chang, Congcong Li, and Tsuhan Chen. 3d augmented markov random field for object recognition. In *ICIP*, pages 3889–3892, 2010.
- [YHMu12] Koichiro Yamaguchi, Tamir Hazan, David A. McAllester, and Raquel Urtasun. Continuous markov random fields for robust stereo estimation. *CoRR*, abs/1204.1393, 2012.
- [YMu13a] Koichiro Yamaguchi, David McAllester, and Raquel Urtasun. Robust monocular epipolar flow estimation. In *CVPR*, 2013.
- [YMu13b] Koichiro Yamaguchi, David McAllester, and Raquel Urtasun. Robust monocular epipolar flow estimation. June 2013.
- [YXL10] J. Yu, C. Xiao, and D. Li. Physics-based fast single image fog removal. In *IEEE International Conference on Signal Processing (ICSP’10)*, pages 1048–1052, Amsterdam, The Netherlands, 2010. IEEE.
- [YYDN07] Q. X. Yang, R. G. Yang, J. Davis, and D. Nister. Spatial-depth super resolution for range images. In *CVPR*, pages 1–8, 2007.
- [ZA08] B. Y. Zhang and J. P. Allebach. Adaptive bilateral filter for sharpness enhancement and noise removal. *IEEE Trans. Image Processing*, 17(5) :664–678, May 2008.
- [Zui94] K. Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485, San Diego, USA, 1994. Academic Press Professional, Inc.

- [ZWY10] Chenxi Zhang, Liang Wang, and Ruigang Yang. Semantic segmentation of urban scenes using dense depth maps. In *Proceedings of the 11th European conference on Computer vision : Part IV*, ECCV'10, pages 708–721, Berlin, Heidelberg, 2010. Springer-Verlag.

Résumé

Nous nous sommes intéressés au problème de la reconstruction 3D à partir de paires stéréoscopiques en présence de brouillard. De nombreux algorithmes existent pour effectuer la reconstruction stéréoscopique, mais peu sont adaptés aux conditions dégradées et en particulier, au brouillard. De ce fait, ils produisent des résultats incorrects à partir d'une certaine distance. L'une des principales causes de cette limitation est la diminution du contraste avec la distance due au brouillard.

La restauration du contraste dans une image en présence de brouillard connaît un intérêt croissant en traitement d'image ces dernières années, en particulier pour des applications destinées aux aides à la conduite. De nombreux algorithmes ont été proposés et permettent d'apporter des solutions approchées à ce problème. Un grand nombre d'entre eux sont fondés sur la loi de Koschmieder qui permet de faire le lien entre l'intensité de l'image, l'intensité originale de la scène et la profondeur. L'un des problèmes de la restauration du contraste monoculaire est l'ambiguïté entre l'épaisseur du voile atmosphérique liée à la profondeur et la couleur plus ou moins claire de la scène, qui fait que ce problème est mal posé. Des contraintes arbitraires doivent donc être ajoutées et cela fait que la solution est approchée. Souvent, la profondeur estimée est très différente de la profondeur exacte de la scène, en particulier à courte distance où le voile est faible..

A partir de ce constat, la complémentarité de la reconstruction stéréo et de la restauration apparaît et un algorithme de reconstruction et de restauration simultanées a été proposé pour la première fois. Pour cela, nous proposons une approche probabiliste fondée sur les champs de Markov. Le modèle proposé permet, grâce à l'indice de profondeur de la stéréovision, de restaurer le contraste avec précision à courte distance. De plus, les images restaurées facilitent la reconstruction à longue distance. Une évaluation qualitative de l'algorithme proposé montre une amélioration significative de la qualité de la carte de profondeur générée par rapport aux algorithmes classiques ne prenant pas en compte le brouillard. L'évaluation montre aussi que les restaurations obtenues sont de qualité proche de celles de l'état de l'art.

Mots-clés: Reconstruction 3D, Restauration d'images, Brouillard, Optimisation de fonctions pseudo-booléennes

