

Mise en correspondance stéréoscopique par approches variationnelles convexes ; application à la détection d'obstacles routiers

Wided Souid-Miled

▶ **To cite this version:**

Wided Souid-Miled. Mise en correspondance stéréoscopique par approches variationnelles convexes ; application à la détection d'obstacles routiers. Traitement du signal et de l'image [eess.SP]. Université Paris-Est, 2007. Français. NNT : . tel-00738363

HAL Id: tel-00738363

https://theses.hal.science/tel-00738363

Submitted on 4 Oct 2012

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.

THÈSE

soutenue le 17/12/2007 pour obtenir

le grade de **Docteur en Sciences** de l'Université Paris-Est

Spécialité : Traitement du Signal

par

Wided Souid-Miled

Mise en correspondance stéréoscopique par approches variationnelles convexes ; application à la détection d'obstacles routiers.

Équipe d'accueil : IMARA - INRIA Rocquencourt

Composition de la commission d'examen :

Président :	Frédéric Truchetet
Rapporteurs :	Fawzi Nashashibi
	Luce Morin
Examinateurs :	Amel Benazza-Benyahia
	Michel Parent
	Didier Aubert
Directeur de thèse :	Jean-Christophe Pesquet

À Myriam

Remerciements

Je tiens tout d'abord à exprimer ma profonde reconnaissance pour mon directeur de thèse, Monsieur Jean-Christophe Pesquet, Professeur à l'université Paris-Est Marne-la-Vallée, qui m'a ouvert les portes de la recherche et m'a offert la possibilité de faire une thèse. Je le remercie pour son extrême disponibilité, son soutien, ses conseils précieux et l'aide scientifique qu'il a su m'apporter. Grâce à ses qualités humaines et professionnelles, j'ai pu mener ces travaux de recherche dans les meilleures conditions possibles.

Mes sincères remerciements vont également à Monsieur Michel Parent, Directeur de Recherche à l'INRIA Rocquencourt, qui m'a accueillie au sein de son équipe IMARA et m'a donné toutes les facilités pour mener à bien ce travail de thèse. Sans son aide et le soutien qu'il m'a porté, cette thèse n'aurait jamais vu le jour.

Je remercie par ailleurs tous les membres du jury pour le soin qu'ils ont porté à la lecture de mon manuscrit, pour leurs questions pertinentes et leurs remarques constructives. Je remercie Monsieur Frédéric Truchetet, Professeur à l'université de Bourgogne, pour m'avoir fait l'immense honneur de présider le jury de thèse. Toute ma gratitude va également à Monsieur Fawzi Nashashibi, Ingénieur de recherche à l'École des Mines de Paris et Madame Luce Morin, Maître de conférences à IRISA Rennes, d'avoir acceptés d'être rapporteur de cette thèse. Je remercie également Madame Amel Benazza, Maître de conférences à Sup'Com Tunis et Monsieur Didier Aubert, Directeur de recherche au LIVIC, pour l'intérêt qu'ils ont porté à ce travail en acceptant de faire partie du jury.

Je tiens à remercier chaleureusement tous les membres de l'équipe IMARA que j'ai côtoyés au quotidien pendant trois ans : un grand merci à Laurent, Arnauld, Armand, André, Olivier, Mickael, Yann, Nicolas, Rodrigo et tous ceux qui étaient de passage le temps d'un stage. Je voudrais également remercier Chantal Chazelas, notre assistante, pour m'avoir aidée dans toutes les démarches administratives.

Cette thèse a été menée en collaboration avec l'équipe Signal et Communications de l'Institut Gaspard Monge de l'université de Marne-la-Vallée. Je remercie tous les chercheurs et les doctorants de cette équipe. Je remercie tout particulièrement Caroline Chaux qui m'a toujours apporté de l'aide quand j'en avais besoin depuis mon stage de DEA et jusqu'à ma thèse. Merci encore Caroline pour tes encouragements, ton soutien et ta disponibilité.

Je remercie toute ma famille et mes proches pour l'amour et le soutien qu'ils m'ont toujours manifestés. Je remercie ma belle famille et tous mes amis qui ont toujours pensé à moi. Un grand merci à mes parents pour tout l'amour et l'éducation qu'ils m'ont donné. Merci aussi pour leurs encouragements dans les moments difficiles et leur disponibilité malgré la distance qui me sépare d'eux. Plein de mercis à mon mari pour l'amour, l'attention, le soutien et l'aide qu'il m'a apporté. Je te remercie Wissem pour ton incomparable patience, ta compréhension et ton écoute. Enfin, merci à ma fille, source première de motivation et de bonheur.

Table des matières

Table des matières **5**

Résumé **9**

Abstract **11**

Glossaire **13**

1 Introduction **15**

2 Généralités sur la mise en correspondance stéréoscopique **21**
2.1 Introduction . . . 21
2.2 Le modèle géométrique de la caméra . . . 21
2.2.1 Le modèle sténopé . . . 21
2.2.2 Paramètres intrinsèques et extrinsèques de la caméra . . . 22
2.2.3 Le modèle du capteur stéréoscopique . . . 27
2.3 La mise en correspondance stéréoscopique . . . 36
2.3.1 Les difficultés de la mise en correspondance . . . 36
2.3.2 Les primitives à mettre en correspondance . . . 37
2.3.3 Les contraintes pour la mise en correspondance . . . 38
2.4 Un rapide état de l'art sur la mise en correspondance . . . 43
2.4.1 Les approches locales . . . 43
2.4.2 Les approches progressives . . . 45
2.4.3 Les approches coopératives . . . 46
2.4.4 Les approches globales . . . 46
2.5 Conclusion . . . 52

3 Approche variationnelle convexe pour l'estimation de la disparité **53**
3.1 Introduction . . . 53
3.2 Formulation du problème . . . 54
3.2.1 Hypothèse de conservation de la luminance . . . 54
3.2.2 Formulation : critère convexe . . . 54
3.2.3 Un problème mal posé . . . 55
3.3 Optimisation d'une fonction convexe différentiable . . . 57
3.3.1 Formulation ensembliste . . . 57
3.3.2 Quelques rappels d'analyse convexe . . . 59
3.3.3 Méthode d'optimisation . . . 64

3.4 Application à l'estimation de la disparité . . . 69
3.4.1 Fonction objectif globale . . . 70
3.4.2 Contrainte de plage de valeurs . . . 72
3.4.3 Régularisation par variation totale . . . 72
3.4.4 Régularisation suivant les contours de l'image de référence . . . 76
3.4.5 Régularisation dans le domaine des ondelettes . . . 78
3.5 Conclusion . . . 85

4 Résultats expérimentaux 87
4.1 Quelques détails de mise en œuvre . . . 87
4.1.1 Formulation du critère . . . 87
4.1.2 Choix des paramètres . . . 88
4.1.3 Critère d'arrêt . . . 89
4.2 Résultats sur des images avec vérité terrain disponible . . . 89
4.2.1 Images testées et critères évalués . . . 89
4.2.2 Expérimentations avec des contraintes spatiales . . . 90
4.2.3 Introduction de contraintes définies dans le domaine des ondelettes . . . 101
4.3 Résultats sur des images réelles . . . 107
4.3.1 Séquence *Flower garden* . . . 109
4.3.2 Paire d'images *Parking meter* . . . 110
4.4 Conclusion . . . 110

5 Prise en compte des variations d'illumination 111
5.1 Introduction . . . 111
5.2 Passage en revue des méthodes existantes . . . 112
5.2.1 Méthodes avec prétraitement . . . 112
5.2.2 Méthodes utilisant des mesures de corrélation robustes . . . 113
5.2.3 Méthodes avec modélisation affine . . . 113
5.3 Modélisation des variations d'illumination . . . 114
5.3.1 Équation de formation des images . . . 115
5.3.2 Un modèle d'illumination multiplicatif . . . 118
5.4 Mise en correspondance robuste aux changements de luminance . . . 120
5.4.1 Formulation ensembliste du problème . . . 120
5.4.2 Définition d'une fonction objectif quadratique . . . 122
5.4.3 Introduction de contraintes convexes . . . 123
5.4.4 Méthode d'optimisation . . . 125
5.5 Résultats expérimentaux . . . 128
5.5.1 Exemple synthétique avec des variations d'illumination simulées . . . 129
5.5.2 Images réelles avec des changements d'illumination réels . . . 134
5.6 Conclusion . . . 138

6 Application à la détection d'obstacles routiers 139
6.1 Introduction . . . 139
6.2 Un survol des méthodes de détection d'obstacles par vision . . . 140
6.2.1 Les approches basées sur la vision monoculaire . . . 140
6.2.2 Les approches basées sur la vision stéréoscopique . . . 141
6.3 Notre méthode de détection d'obstacles par stéréovision . . . 142

6.3.1 Matériel d'expérimentation 143
6.3.2 Estimation de la carte de disparité 144
6.3.3 Classification des pixels de la carte de disparité 146
6.3.4 Caractérisation des obstacles détectés 151
6.4 Conclusion 155

7 Conclusion et perspectives **157**

Annexes **160**

A Un bref rappel sur les ondelettes **161**

Table des figures **166**

Liste des tableaux **169**

Bibliographie **184**

Résumé

Cette thèse porte sur la mise en correspondance stéréoscopique dense ainsi que sur son application à la détection des obstacles routiers à partir d'un système de vision stéréoscopique embarqué à bord d'un véhicule. La mise en correspondance est une étape cruciale dans la reconstruction de la structure tridimensionnelle de la scène observée. Elle consiste à retrouver les pixels homologues dans deux images de la même scène, prises de deux points de vue différents, et se ramène à un problème d'estimation d'un champ de disparité.

Le problème d'estimation de la disparité, formulé comme un problème inverse *mal posé*, a été appréhendé, dans le cadre d'une approche ensembliste, en minimisant une fonction objectif convexe sur l'intersection d'ensembles convexes. Ces ensembles sont associés à des contraintes convexes modélisant des informations *a priori* tenant compte des propriétés du champ à estimer et des données observées. Dans ce travail, nous avons mis l'accent sur une estimation d'un champ de disparité lisse et homogène, tout en respectant les discontinuités présentes dans la scène. Pour cela, nous avons considéré différentes contraintes de régularisation satisfaisant cette propriété. Ces contraintes sont définies dans le domaine spatial ou exprimées à partir des coefficients en ondelettes de l'image. Pour résoudre le problème d'optimisation sous contraintes considéré, nous avons mis en place un algorithme efficace itératif par bloc, utilisant des développements récents d'analyse convexe et mettant en œuvre des projections sous-différentielles sur des ensembles convexes. Cet algorithme offre une grande flexibilité quant à l'incorporation de contraintes multiples et peut être efficacement implanté sur une architecture à processeurs parallèles.

La plupart des méthodes de mise en correspondance reposent sur l'hypothèse de conservation de la luminance qui établit que les projections d'un même point de la scène ont la même intensité dans les deux images stéréoscopiques. Cependant, cette hypothèse physique n'est pas toujours vérifiée en pratique, notamment en présence de zones d'ombres ou de surfaces non lambertiennes. Notre approche pour aborder ce problème consiste à modéliser les variations spatiales de luminosité de la scène, en considérant un modèle d'illumination multiplicatif variant spatialement. Elle permet de mesurer les changements d'illumination et obtenir ainsi une estimation de la disparité très robuste vis-à-vis de ces changements. Des résultats sur des images synthétiques et réelles ont permis de valider la méthode proposée.

Dans l'objectif d'aider la navigation autonome dans un environnement routier des systèmes de transport intelligents, nous avons proposé une méthode de détection d'obstacles par stéréovision, en exploitant notre approche d'estimation de la disparité et en s'appuyant sur un critère approprié d'orientation de surface .

Abstract

This thesis deals with the problem of stereo matching and its application to road obstacle detection from a stereovision system mounted on a vehicle. The main goal of stereo matching is to recover the depth information of a scene from a pair of left and right images taken from two different locations. It involves finding corresponding pixels in both images and reduces to a disparity field estimation problem.

The dense disparity estimation problem, formulated as an inverse *ill-posed* problem, was addressed, in a set theoretic framework, by minimizing a quadratic objective function under multiple convex constraints. These constraints arise from prior knowledge and rely on various properties of the field to be estimated. In order to obtain a smooth disparity field, while preserving discontinuities around object edges, we consider appropriate spatial and wavelet based regularization constraints. The resulting optimization problem is solved with a parallel block iterative algorithm using subgradient projections on convex constraint sets and based on recently developed convex analysis tools. This algorithm offers great flexibility in the incorporation of several constraints and can be efficiently implemented on parallel computing architectures.

In most classical dense disparity estimation techniques, the scene is assumed to be Lambertian so that corresponding pixels in the left and right images have identical intensity values. However, in the presence of illumination changes often encountered in practice, this assumption is violated which may largely reduce the efficiency of depth recovery. To address the problem of disparity estimation under varying illumination conditions, we develop a spatially varying multiplicative model that accounts for photometric changes between both images in the stereo pair. Results on synthetic and real stereo pairs demonstrate the efficiency of the proposed method to recover illumination changes and disparity map simultaneously, making disparity estimation very robust w.r.t. such changes.

The ability to detect obstacles on the road is essential for intelligent vehicle navigation systems. In this work, we develop an accurate and efficient vision based obstacle detection method that relies on the reliability of the depth information obtained from the disparity estimation stage. Object segmentation based on the depth map and making use of a surface orientation criterion enables efficient and robust obstacle detection.

Glossaire

$\mathbb{N}$:	Espace des entiers naturels.
$\mathbb{R}^n$:	Espace euclidien de dimension n.
$\mathbb{P}^n$:	Espace projectif de dimension n.
$\mathcal{H}$:	Espace de Hilbert réel.
$\mathcal{D}$:	Support discret de l'image.
Ω :	Domaine continu de l'image.
$\langle . \mid . \rangle$:	produit scalaire de $\mathcal{H}$.
$\| . \|$:	la norme euclidienne de $\mathcal{H}$ induite par le produit scalaire $\langle . \mid . \rangle$.
$d(u, C)$:	la distance d'un point $u \in \mathcal{H}$ à l'ensemble $C \subset \mathcal{H}$.
P_C :	le projecteur sur l'ensemble convexe fermé non vide $C \subset \mathcal{H}$.
S :	L'ensemble des solutions admissibles.
$\text{lev}_{\leq \eta} f$:	l'ensemble de niveau η de f.
$\partial f(u)$:	le sous-différentiel en u d'une fonction $f : \mathcal{H} \to \mathbb{R}$.
card I :	le nombre d'éléments d'un ensemble I.
$[p]_\times$:	matrice de produit vectoriel d'un vecteur p.
$p_1 \wedge p_2$:	produit vectoriel entre les vecteurs p_1 et p_2.
$M^\top$:	le transposé de la matrice M.
T^* :	adjoint d'un opérateur T linéaire borné.

Chapitre 1

Introduction

Contexte de l'étude

Les récentes évolutions technologiques dans de nombreux domaines, comme la médecine, la physique, les mathématiques et plus récemment l'informatique, ont fortement contribué à l'émergence de nouveaux robots, *intelligents* et fiables. Les progrès dans le domaine de la neurophysiologie, permettant de mieux comprendre les mécanismes de fonctionnement du système visuel humain, ont suscité, en particulier, l'ambition de doter ces machines d'une capacité visuelle, proche de celle du système humain, pour accomplir des tâches dans des domaines aussi variés que la navigation autonome, l'imagerie médicale ou la vidéo surveillance. Cette finalité, aux applications nombreuses, constitue ce qu'on appelle la *vision par ordinateur*.

Dans le domaine de la route intelligente, les recherches portent, depuis plusieurs années, sur le développement de nouveaux systèmes de transport fournissant une aide à la conduite et, par là même, une amélioration de la sécurité routière. De tels systèmes, dits *Systèmes de Transport Intelligents* (ITS [1]), doivent être capables d'interagir, en temps réel, avec l'environnement dans lequel ils se déplacent. Les tâches nécessaires à cette interaction comprennent :

- la perception de l'environnement routier,
- le traitement des données,
- l'action à réaliser.

Outre la localisation du véhicule sur la chaussée, la détection d'obstacles constitue un élément essentiel dans la perception de l'environnement routier du véhicule. Différentes technologies existent pour accomplir cette tâche, suivant qu'elles font appel à des capteurs actifs ou passifs. Les capteurs actifs, comme les radars, les télémètres laser ou à ultrasons, présentent l'inconvénient d'avoir un coût élévé et un champ de perception restreint. De plus, lorsque plusieurs véhicules existent dans un même environnement, ces capteurs embarqués peuvent interférer entre eux et compromettre ainsi la fiabilité des informations qu'ils apportent. Avec l'amélioration des techniques d'analyse d'images, la vision, monoculaire ou stéréoscopique, est devenue un moyen attractif de détection et d'évitement d'obstacles, dans la mesure où

[1] Intelligent Transportation Systems

elle permet d'acquérir des informations pertinentes sur l'environnement du véhicule à partir d'informations photométriques issues d'images de scènes routières.
Bien que les approches basées sur la vision monoculaire permettent, dans la majorité des cas, de détecter les obstacles routiers, leurs distances ne peuvent pas être calculées précisément sans ajouter un autre capteur ou une autre connaissance *a priori* sur la scène. L'information de distance est primordiale pour localiser les objets qui représentent un obstacle potentiel pour le véhicule. Les approches basées sur la vision stéréoscopique permettent de retrouver cette information par un procédé de triangulation, qui consiste à calculer les coordonnées 3D d'un point à partir de ses projections dans les images.

La vision stéréoscopique permet de reconstruire la structure tridimensionnelle de la scène observée à partir de deux images prises de deux points de vue différents. Chez l'être humain, la vision binoculaire est un élément essentiel de la perception tridimensionnelle : chaque œil est un capteur qui fournit au cerveau sa propre image de la scène ; la connaissance simultanée de ces deux images lui permet de reconstituer la profondeur de la scène. Ce processus ne pose aucune difficulté pour l'être humain, mais constitue, en revanche, un enjeu majeur en vision par ordinateur. Une étape cruciale dans la résolution de ce problème est la *mise en correspondance stéréoscopique*, qui consiste à retrouver dans les images gauche et droite, les pixels homologues qui sont les projections d'un même point physique de la scène. La différence entre les coordonnées de pixels homologues est appelée *disparité*. Le problème de mise en correspondance se ramène ainsi à un problème d'estimation d'un champ de disparité.

Le cadre de cette thèse, qui s'est déroulée à cheval entre le projet IMARA (Informatique, Mathématiques, Automatique, pour la Route Automatisée) de l'Institut National de Recherche en Informatique et en Automatique et l'équipe Signal et Communication du laboratoire d'Informatique de l'Institut Gaspard Monge de l'Université Paris-Est, est la mise en correspondance stéréoscopique dense pour la détection des obstacles routiers, à partir d'un système de vision stéréoscopique embarqué à bord d'un véhicule en mouvement.

Vue d'ensemble du travail effectué

Mise en correspondance

Le problème de mise en correspondance étant fondamental en vision par ordinateur, il a fait l'objet de nombreux travaux au cours des dernières années. La résolution numérique de ce problème, qui se ramène généralement à celui de l'optimisation d'un critère global, entraîne un coût de calcul qui devient d'autant plus important que la taille des images considérées est grande. Pour limiter cette complexité numérique, le recours aux avancées récentes dans le domaine des méthodes d'optimisation dédiées aux problèmes de grande taille devient ainsi d'un intérêt majeur.

Dans ce contexte, notre contribution consiste à explorer l'apport que peuvent constituer certaines de ces techniques d'optimisation récentes, qui ont été appliquées avec succès dans le domaine de traitement d'images, pour résoudre des problèmes de vision par ordinateur. L'efficacité des algorithmes d'optimisation qui en découlent dépend fortement de la structure du critère à optimiser, qui prend en compte les données observées et les informations connues

a priori sur la solution. Ainsi, de par l'importance de la propriété de convexité en optimisation, nous limitons le cadre de notre étude aux méthodes d'optimisation convexe. Nous nous intéressons, en particulier, aux algorithmes parallèles, itératifs, se basant sur de récents outils d'analyse convexe et mettant en œuvre des projections approchées sur des ensembles convexes. L'utilisation de ces algorithmes, dans le cadre de notre application, nécessite une formulation appropriée du problème de mise en correspondance.

Prise en compte des variations d'illumination

Un problème, relativement peu traité dans la littérature de mise en correspondance, est celui de la prise en compte des variations d'illumination induites entre les images stéréoscopiques. Il s'agit d'un problème d'importance, puisqu'en présence de ces variations, l'hypothèse de conservation de la luminance sur laquelle reposent la plupart des méthodes de mise en correspondance, est mise en défaut. L'étude que nous avons menée dans ce contexte montre que ces changements peuvent être compensés par une modélisation locale des variations spatiales de luminosité. En proposant un modèle d'illumination multiplicatif, nous améliorons ainsi la robustesse de notre méthode de mise en correspondance vis-à-vis des variations d'illumination de la scène.

Détection des obstacles routiers

Du fait des problématiques que sont la congestion, la pollution et la sécurité, liées à l'usage de la voiture individuelle comme moyen de transport principal, l'utilisation de modes de transport alternatifs devient une nécessité, en particulier en milieu urbain. Pour offrir un mode de transport plus approprié, l'équipe de recherche IMARA travaille activement sur le développement de véhicules urbains avec une conduite totalement automatisée. L'objectif, à long terme, est d'équiper les grandes villes de flottes de véhicules automatiques, en libre service, mis à disposition des usagers qui passeront commande en fonction de leurs besoins en déplacement. Le Cycab, actuellement commercialisé par la société Robosoft[2], est l'un de ces véhicules novateurs. Il exploite les nouvelles technologies de l'information et de la communication et bénéficie d'une certaine autonomie. La fonction la plus élémentaire, et certainement la plus cruciale, permettant à ce véhicule, équipé d'un système de stéréovision Bumblebee[3], de naviguer de façon autonome sûre et fiable, est la détection d'obstacles. Dans ce contexte, notre objectif est d'appliquer notre approche de mise en correspondance stéréoscopique dense, en prenant en compte le problème des variations de luminosité, pour la perception des objets qui peuvent entraver la voie de circulation du Cycab.

Organisation du document

Bien que l'objectif final de cette thèse soit la détection d'obstacles dans un milieu routier par vision stéréoscopique, nous avons choisi de ne traiter ce problème qu'en dernière partie du mémoire. En effet, les méthodes de mise en correspondance proposées ici sont assez génériques

[2]http ://www.robosoft.fr/eng/
[3]http ://www.ptgrey.com

et peuvent donc être appliquées sur tous types d'images stéréoscopiques acquises par deux caméras.

Dans le chapitre 2, nous introduisons le modèle géométrique du capteur stéréoscopique et nous abordons quelques aspects généraux de la mise en correspondance, qui seront utiles par la suite. Nous évoquons en particulier les contraintes fréquemment utilisées pour pallier les difficultés de ce problème. Nous dressons finalement un rapide état de l'art des méthodes d'estimation de la disparité, nous conduisant à adopter une formulation variationnelle de ce problème.

Le chapitre 3 présente notre méthode d'estimation de la disparité. Partant de l'hypothèse de conservation de la luminance entre les deux projections d'un même point de l'espace, nous formulons ce problème comme celui de la minimisation d'une fonction objectif convexe, garantissant ainsi la possibilité d'obtenir la solution globale. Pour résoudre ce problème de minimisation vu comme un problème inverse mal posé, nous mettons en œuvre une nouvelle approche ensembliste. Cette approche, qui consiste à chercher la solution qui minimise un certain critère convexe sur l'intersection d'ensembles convexes construits à partir des connaissances *a priori* et des observations, constitue une nouvelle façon d'appréhender le problème de l'estimation de la disparité. Une fois la formulation du problème effectuée, un aspect important de nos contributions consiste, par la suite, à définir un critère global convexe à minimiser ainsi que des contraintes convexes pertinentes sur la solution. Dans ce chapitre, nous considérons un critère différentiable et nous introduisons des contraintes, associées à des fonctionnelles non nécessairement différentiables, à la fois dans le domaine spatial et dans le domaine des ondelettes. Au niveau de l'ajout de l'information *a priori*, nous modélisons l'image de disparité comme étant constituée de zones homogènes, séparées par de franches discontinuités. Dans le domaine spatial, ceci peut se traduire en supposant que le champ de disparité appartient à l'espace des fonctions à variations bornées. La contrainte convexe de variation totale bornée permet ainsi de régulariser la solution tout en préservant les discontinuités autour des contours des objets. Des contraintes définies dans le domaine des ondelettes peuvent également servir à imposer la régularité du champ de disparité. L'idée de cette régularisation est de supposer que la semi-norme de l'image de disparité, définie à partir de ses coefficients en ondelettes dans un espace approprié, est bornée. Deux contraintes principales garantissant les propriétés de régularisation avec préservation des discontinuités sont ainsi identifiées. Nous proposons également, dans ce chapitre, un algorithme itératif très efficace, que nous mettons en place, pour résoudre le problème d'optimisation considéré.

Dans le chapitre 4, nous validons notre méthode d'estimation de la disparité sur des données synthétiques et réelles. L'étude comparative établie avec les approches concurrentes actuellement disponibles confirme l'efficacité et la fiabilité de la méthode proposée. Ce chapitre popose également de mener des comparaisons entre les différentes contraintes de régularisation considérées.

Le chapitre 5 traite du problème des variations d'illumination dans la scène observée. Nous considérons pour cela un modèle d'illumination multiplicatif qui permet de compenser les variations spatiales de luminosité de la scène. Ce modèle local est pris en compte, dans la formulation du problème de mise en correspondance, au travers d'un facteur d'illumination multiplicatif que nous estimons conjointement avec la disparité. Motivés par les résultats précédents, une approche ensembliste est à nouveau adoptée. La résolution du problème de l'estimation conjointe des champs de disparité et d'illumination implique alors de définir

un critère à minimiser et d'introduire des informations *a priori* sur les champs à estimer. Le critère convexe à minimiser découle du modèle d'illumination établi et les contraintes sur le champ de disparité déjà proposées sont mises à profit. Une des contributions de ce chapitre consiste donc à formuler des contraintes convexes traduisant les propriétés connues *a priori* sur le champ d'illumination. Pour apporter une solution numérique à ce problème d'optimisation, nous utilisons un algorithme itératif par bloc que nous adaptons pour calculer simultanément la disparité et les changements d'illumination. Des expérimentations sur divers types de couples stéréoscopiques, soumis à des variations d'illumination simulées ou réelles, démontrent la robustesse de notre approche vis-à-vis de ces variations.

Dans le chapitre 6, nous appliquons notre méthode d'estimation de la disparité robuste aux variations d'illumination pour la détection des obstacles routiers. En utilisant un critère approprié d'orientation de surface, nous classifions les pixels de la carte de disparité en deux catégories : les pixels appartenant à la route et ceux appartenant aux obstacles. Notre classification tire profit de la densité et de la précision de la carte de disparité estimée pour extraire des caractéristiques discriminantes entre les deux éléments constituant la scène routière, à savoir le plan de la route et les obstacles. Dans ce chapitre, des résultats expérimentaux, sur des images stéréoscopiques acquises par le système de vision embarqué à bord du Cycab, permettent de valider la méthode proposée.

Dans la conclusion générale, nous présentons une synthèse de notre étude et nous dressons quelques perspectives se situant dans le prolongement des travaux menés.

Publications

Articles soumis ou en préparation

⋆ Wided Miled, Jean-Christophe Pesquet, Michel Parent. A Convex Optimization Approach for Depth Estimation Under Illumination Variation. *En préparation*, 27 pages.

⋆ Wided Miled, Jean-Christophe Pesquet, Michel Parent. Wavelet-constrained stereo matching under photometric variations. *Soumis à SPIE International Symposium on Optical and Digital Image Processing*, Avril 2008.

Articles de conférences

⋆ Wided Miled, Jean-Christophe Pesquet, Michel Parent. Robust obstacle detection based on dense disparity maps. *11th International Conference on Computer Aided Systems Theory*, (EUROCAST'07), *Springer LNCS 4739*, pp. 1142-1150, Las Palmas de Gran Canaria, 12-16 Février, 2007.

⋆ Wided Miled, Jean-Christophe Pesquet, Michel Parent. Dense disparity estimation from stereo images. *International symposium on Image/Video Communications* (ISIVC'06), Hammamet, Tunisie, 13-15 Septembre 2006, 4 pages.

⋆ Wided Miled, Jean-Christophe Pesquet, Michel Parent. Wavelet constrained regularization for disparity map estimation. *European Signal Processing Conference*, (EUSIPCO'06),

Florence, Italie, 4-8 Septembre 2006, 5 pages.

⋆ Wided Miled, Jean-Christophe Pesquet, Michel Parent. Disparity map estimation using a total variation bound. *Third Canadian Conference on Computer and Robot Vision*, (CRV'06), pages 48-55, Québec, Canada, 7-9 Juin 2006, 8 pages.

Prix

EUSIPCO 2006 : *Best Student Paper Award* pour le papier intitulé : Wavelet constrained regularization for disparity map estimation.

Chapitre 2

Généralités sur la mise en correspondance stéréoscopique

2.1 Introduction

Un des problèmes classiques en vision par ordinateur est la stéréovision ou la reconstruction tridimensionnelle d'une scène à partir de deux images prises de deux points de vue différents. Une étape cruciale dans la reconstruction 3D est la mise en correspondance stéréoscopique qui consiste à retrouver dans les images gauche et droite, les primitives homologues, qui sont les projections d'un même point physique de la scène. La différence entre les coordonnées de primitives homologues est appelée disparité. Ce problème d'appariement se ramène ainsi à un problème d'estimation d'un champ bidimensionnel de disparité.

L'objectif de ce premier chapitre est d'introduire certaines notions nécessaires à la compréhension des problématiques de mise en correspondance stéréoscopique. Les notions abordées ne sont pas exhaustives et ont été choisies afin de présenter clairement certains aspects qui seront utilisés dans la suite. Des descriptions plus détaillées peuvent être trouvées dans [Horaud, Monga, 1993, Faugeras, 1993, Hartley, Zisserman, 2004].

Nous présentons d'abord le modèle projectif de la caméra ainsi que la géométrie épipolaire d'un système stéréoscopique. Nous évoquons ensuite les primitives et les contraintes les plus utilisées pour pallier les difficultés de la mise en correspondance stéréoscopique. Enfin, nous terminons par un état de l'art succinct des méthodes d'estimation de la disparité.

2.2 Le modèle géométrique de la caméra

2.2.1 Le modèle sténopé

Le processus d'acquisition d'une image par une caméra peut être modélisé à l'aide d'un système à projection perspective, aussi appelé sténopé (cf. figure 2.1). Ce modèle de projection

centrale, qui transforme un point de l'espace 3D en un point image 2D, correspond au modèle de caméra le plus couramment utilisé. Il se compose d'un plan rétinien et d'un centre de projection ou centre optique C. Le centre de projection correspond à la position de la caméra et le plan rétinien est le plan de formation des images. On appelle

- distance focale f : la distance du centre de projection C au plan rétinien,
- plan focal : le plan situé en C et parallèle au plan rétinien,
- axe optique : l'axe orthogonal au plan rétinien passant par C.

Remarque 2.1 *Ce modèle de caméra crée pour tout objet observé de la scène, une image inversée sur le plan rétinien situé derrière le centre de projection C. Cependant, il est courant de placer, par commodité, le plan image Π à la même distance f devant le centre optique de manière à obtenir une image non inversée de la scène. L'axe optique intersecte Π en un point c appelé point principal.*

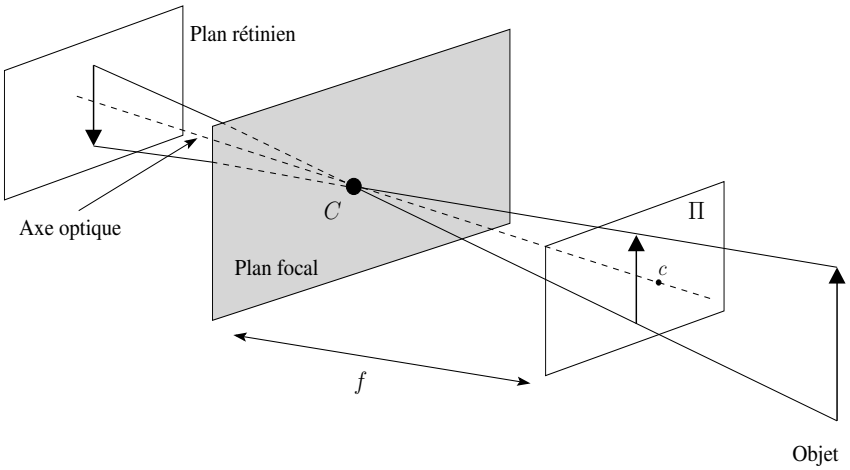


FIG. 2.1 – Le modèle de caméra sténopé.

Tout point M de la scène se projette en un point m sur le plan image par l'intersection de ce dernier avec le rayon optique (CM) (cf. figure 2.2). Seuls les points situés sur le plan focal n'ont pas de projection sur le plan image. L'expression de cette projection dépend des systèmes de coordonnées choisis dans l'espace et dans le plan image.

2.2.2 Paramètres intrinsèques et extrinsèques de la caméra

La projection d'un point de l'espace 3D en un point image 2D nécessite trois transformations :

- une projection perspective de l'espace euclidien $\mathbb{R}^3$ dans l'espace euclidien $\mathbb{R}^2$,
- un changement de repère,
- une transformation des coordonnées métriques exprimées dans un repère lié à la caméra à des coordonnées pixelliques dans un repère lié à l'image.

2.2.2.1 La projection perspective

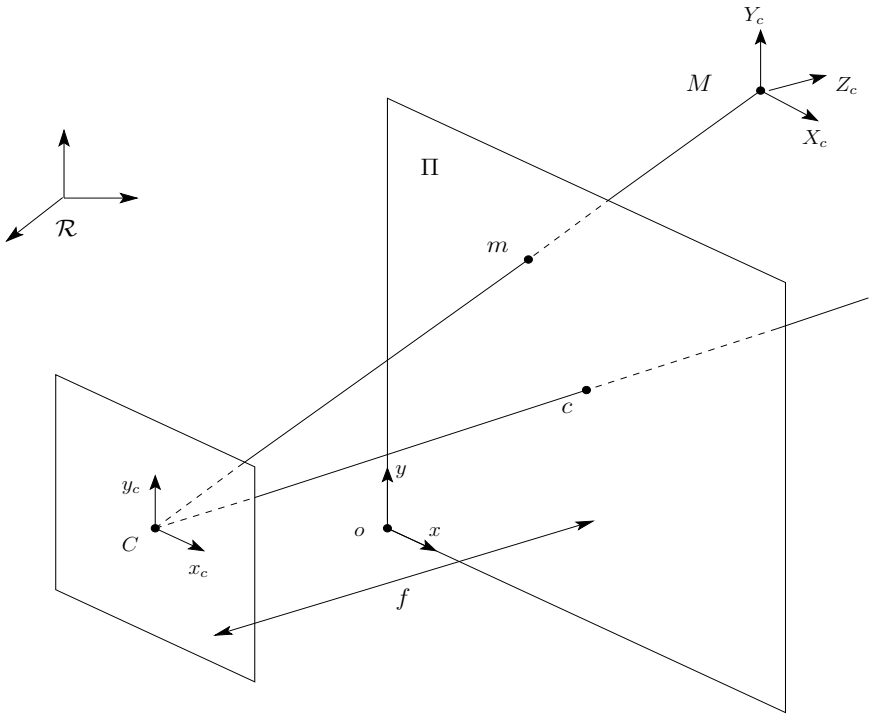


FIG. 2.2 – La projection perspective.

Soit $\mathcal{R}_c$ un repère lié à la caméra, d'origine le centre optique C et tel que l'axe des Z est confondu avec l'axe optique. Le repère $\mathcal{R}_c$ est appelé système de coordonnées standard de la caméra et est formé des axes horizontal, vertical du plan image et de l'axe optique. Soient X_c, Y_c et Z_c les coordonnées d'un point M de l'espace dans le repère qu'on vient de décrire. Dans ce même repère, l'expression des coordonnées de m, projection de M sur le plan image, résulte de l'application du théorème de Thalès :

$$\begin{aligned} x_c &= f\frac{X_c}{Z_c}, \\ y_c &= f\frac{Y_c}{Z_c}. \end{aligned} \tag{2.1}$$

Ces équations ne sont pas linéaires par rapport aux coordonnées euclidiennes X_c, Y_c et Z_c mais elles le deviennent si l'on utilise les coordonnées projectives dans l'espace et dans le plan image.

Si (X, Y, Z) sont les coordonnées euclidiennes d'un point de $\mathbb{R}^3$, ses coordonnées projectives (homogènes) sont définies dans l'espace projectif par $(\lambda.X, \lambda.Y, \lambda.Z, \lambda)$, pour tout réel λ non nul.

Définition 2.1 *On appelle espace projectif $\mathbb{P}^3$, l'espace quotient de $\mathbb{R}^4 \setminus \{0\}$ par la relation d'équivalence :*

$$(X, Y, Z, W) \simeq (X', Y', Z', W') \quad \Leftrightarrow \quad \exists\ \lambda \neq 0 \ \mid\ (X, Y, Z, W) = \lambda.(X', Y', Z', W').$$

Les coordonnées d'un point de $\mathbb{P}^3$ sont définies à un coefficient multiplicatif près. Ce sont des coordonnées homogènes.

On note $\widetilde{\mathbf{M}} = (X, Y, Z, 1) \simeq (\lambda.X, \lambda.Y, \lambda.Z, \lambda) \in \mathbb{P}^3$, où $\lambda \neq 0$ arbitraire, les coordonnées projectives d'un point M de coordonnées euclidiennes (X, Y, Z) de $\mathbb{R}^3$ et, de manière similaire, $\widetilde{\mathbf{m}} = (x, y, 1) \simeq (\lambda.x, \lambda.y, \lambda) \in \mathbb{P}^2$ celles d'un point m de coordonnées euclidiennes (x, y) dans le plan image.

En adoptant les coordonnées homogènes, la relation (2.1) peut s'écrire sous forme matricielle comme suit

$$\begin{pmatrix} s.x_c \\ s.y_c \\ s \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}. \tag{2.2}$$

Remarque 2.2 *Tous les points appartenant au rayon optique (CM) se projettent en un unique point m sur le plan image. Ils sont, de ce fait, équivalents en coordonnées projectives.*

2.2.2.2 La transformation caméra-image

Soit $\mathcal{R}_a$ un repère affine du plan image, d'origine le point o (localisé par exemple au coin de l'image en bas à gauche) et dont les axes sont parallèles à ceux du repère associé à la caméra (cf. figure 2.2). Les coordonnées des points projetés sur le plan image dans le repère caméra $\mathcal{R}_c$ seront désormais exprimées en pixels dans le repère $\mathcal{R}_a$ lié à l'image.

La transformation des coordonnées métriques aux coordonnées pixelliques est réalisée par un changement de repère affine

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} k_x & 0 & x_0 \\ 0 & k_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix}. \tag{2.3}$$

Ce passage du repère caméra au repère image repose sur la connaissance des paramètres intrinsèques de la caméra :

- (x_0, y_0) sont les coordonnées du point principal c exprimées en pixels dans le repère image,
- k_x et k_y sont respectivement les facteurs d'échelles horizontale et verticale exprimés en pixels/mm.

Ces paramètres sont qualifiés d'intrinsèques car ils ne dépendent que des caractéristiques de la caméra.

La relation entre les coordonnées du point $\widetilde{\mathbf{M}}$ dans le repère caméra et les coordonnées de sa projection $\widetilde{\mathbf{m}}$ exprimées dans le repère image s'écrit

$$\begin{pmatrix} s.x \\ s.y \\ s \end{pmatrix} = \begin{pmatrix} k_x & 0 & x_0 \\ 0 & k_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}. \tag{2.4}$$

En introduisant $\alpha_x = k_x f$ et $\alpha_y = k_y f$, cette équation est souvent réécrite sous la forme suivante

$$\begin{pmatrix} s.x \\ s.y \\ s \end{pmatrix} = \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}. \tag{2.5}$$

2.2.2.3 Le changement de repère

Soient $\mathcal{R}$ un repère orthonormé de l'espace et (X, Y, Z) les coordonnées du point M dans ce même repère. Les coordonnées (X_c, Y_c, Z_c) de ce point exprimées dans le repère $\mathcal{R}_c$ sont obtenues en introduisant les paramètres extrinsèques de la caméra qui traduisent le changement de repère de $\mathcal{R}$ à $\mathcal{R}_c$. Ce changement de repère est un déplacement dans $\mathbb{R}^3$, que l'on peut décomposer en une rotation $\mathbf{R}$ suivie d'une translation $\mathbf{t}$, ainsi

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = \mathbf{R} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \mathbf{t}. \tag{2.6}$$

Il s'exprime en coordonnées homogènes de la façon suivante

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \tag{2.7}$$

On appelle donc paramètres extrinsèques, l'expression de la position et de l'orientation de la caméra dans le repère de l'espace, c'est-à-dire le vecteur $\mathbf{t}$ de taille 3×1 et la matrice $\mathbf{R}$ de taille 3×3. Ils dépendent chacun de trois paramètres.

Remarque 2.3 *L'expression des coordonnées du centre optique C (aussi notées C par abus de notation) dans le système de coordonnées de l'espace $\mathcal{R}$ est donnée à partir de la relation (2.6) par*

$$\mathbf{0}_3 = \mathbf{R}C + \mathbf{t}, \tag{2.8}$$

ce qui implique que

$$C = -\mathbf{R}^{-1}\mathbf{t}. \tag{2.9}$$

2.2.2.4 La transformation perspective

La transformation perspective est la composition des trois transformations précédentes. Elle s'écrit sous la forme d'une matrice qui peut se décomposer comme suit

$$\mathbf{P} = \mathbf{A}\mathbf{P_0}\mathbf{K}, \tag{2.10}$$

où

$$\mathbf{A} = \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{P_0} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{et} \quad \mathbf{K} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}. \tag{2.11}$$

Les matrices $\mathbf{A}$ et $\mathbf{K}$ sont respectivement les matrices des paramètres intrinsèques et extrinsèques de la caméra, et $\mathbf{P_0}$ est la matrice dite de projection centrale.

La matrice $\mathbf{P}$ peut être réécrite d'une façon plus concise comme suit :

$$\mathbf{P} = \mathbf{A}(\mathbf{R} \quad \mathbf{t}). \tag{2.12}$$

Définition 2.2 *On appelle ainsi matrice de projection perspective* $\mathbf{P}$ *associée à une caméra la matrice de taille* 3×4 *représentant, en coordonnées projectives, la projection des points de l'espace sur le plan image. On a alors*

$$\widetilde{\mathbf{m}} = \mathbf{P}\widetilde{\mathbf{M}}. \tag{2.13}$$

La matrice $\mathbf{P}$ est définie à un coefficient multiplicatif près. Estimer cette matrice constitue le problème de calibration de la caméra.

Proposition 2.1 *Les coordonnées du centre optique* C *exprimées dans le repère de l'espace peuvent être obtenues à partir de la matrice de projection perspective associée en résolvant l'équation suivante :*

$$\mathbf{P}\begin{pmatrix} C \\ 1 \end{pmatrix} = \mathbf{0}_3. \tag{2.14}$$

Démonstration. En remplaçant $\mathbf{P}$ par son expression dans (2.12), on obtient

$$\begin{aligned} \mathbf{P}\begin{pmatrix} C \\ 1 \end{pmatrix} &= \mathbf{A}(\mathbf{R} \quad \mathbf{t})\begin{pmatrix} C \\ 1 \end{pmatrix} \\ &= \mathbf{A}(\mathbf{R}C + \mathbf{t}) \qquad \text{(d'après (2.8))}, \\ &= \mathbf{0}_3. \end{aligned}$$

■

Si l'on peut décomposer la matrice $\mathbf{P}$ de taille 3×4 en une matrice $\mathbf{Q}$ de taille 3×3 et un vecteur $\mathbf{p}$ de taille 3×1, telle que $\mathbf{P} = (\mathbf{Q} \quad \mathbf{p})$ et si on suppose que la matrice $\mathbf{Q}$ est de rang 3, l'équation (2.8) peut être réécrite comme suit [Faugeras, 1993] :

$$C = -\mathbf{Q}^{-1}\mathbf{p}. \tag{2.15}$$

2.2.3 Le modèle du capteur stéréoscopique

Nous considérons le cas de la stéréovision binoculaire, où deux images, appelées image de gauche et image de droite, sont prises par deux caméras observant la même scène à partir de deux positions différentes (cf. figure 2.3). Chacune des deux caméras est caractérisée par un centre optique, une matrice de projection perspective et un repère ayant pour origine le centre optique de la caméra. Par un changement de repère adapté, nous pouvons ramener le repère de la caméra de gauche au repère de l'espace.

Tout point M de l'espace se projette dans les images de gauche et de droite selon l'équation projective (2.13) par

$$\widetilde{\mathbf{m}} = \mathbf{P}\ \widetilde{\mathbf{M}}, \tag{2.16}$$

et

$$\widetilde{\mathbf{m}}' = \mathbf{P}'\widetilde{\mathbf{M}}, \tag{2.17}$$

où $\mathbf{P}$ et $\mathbf{P}'$ sont les deux matrices de projection perspective associées aux caméras de gauche et de droite. Dans le système de coordonnées de l'espace que nous considérons, ces matrices sont données par

$$\mathbf{P} = \mathbf{A}\,(\mathbf{I}_3 \quad \mathbf{0}) \qquad \text{et} \qquad \mathbf{P}' = \mathbf{A}'\,(\mathbf{R} \quad \mathbf{t})\,. \tag{2.18}$$

où $\mathbf{I}_3$ est la matrice identité de taille 3×3, $\mathbf{A}$ et $\mathbf{A}'$ sont les matrices des paramètres intrinsèques des deux caméras et $\mathbf{R}$ et $\mathbf{t}$ caractérisent la rotation et la translation s'opérant entre le repère de la caméra de gauche et celui de la caméra de droite.

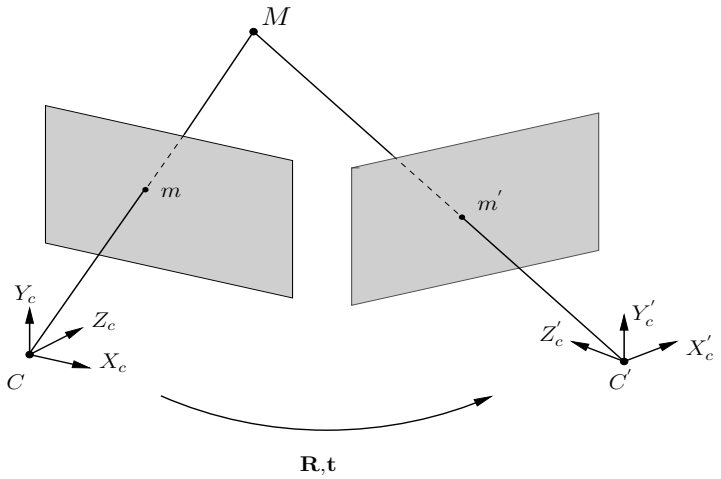


FIG. 2.3 – Le modèle géométrique du capteur stéréoscopique.

2.2.3.1 La géométrie épipolaire

Nous nous intéressons maintenant aux relations géométriques qui existent entre les deux images stéréoscopiques d'une même scène. Plus précisément, nous établissons une contrainte géométrique forte entre un point dans une image et son correspondant dans l'autre image. Cette contrainte est illustrée par la figure 2.4.

Soient C et $C^{'}$ les centres optique des deux caméras de gauche et de droite. Nous pouvons observer sur cette figure qu'un point M de l'espace, ses projections m et $m^{'}$ dans chacune des deux images et les points C et $C^{'}$ sont coplanaires et définissent un plan Π_e appelé *plan épipolaire*.

Supposons à présent qu'on connaît seulement le point m de l'image de gauche et qu'on cherche à déterminer son correspondant dans l'image de droite. Le point m représente la projection dans l'image de gauche de tous les points de l'espace appartenant au rayon (Cm). Ce rayon définit avec la droite $(CC^{'})$, appelée ligne de base, le plan Π_e. Si $m^{'}$ est le correspondant de m dans l'image de droite alors le rayon $(C^{'}m^{'})$ se situe également dans le plan Π_e. Le point $m^{'}$ appartient donc nécessairement à la droite d'intersection de Π_e avec l'image de droite, qui n'est autre que la projection du rayon (Cm) dans l'image de droite. Cette droite s'appelle *ligne épipolaire* associée à m et est notée $l_{m'}$.

La géométrie qui décrit la formation d'une paire d'images stéréoscopiques s'appelle *géométrie épipolaire*. Elle est essentiellement basée sur les définitions suivantes :

- l'*épipole* est le point d'intersection de la ligne de base $(CC^{'})$ avec le plan de l'image. L'épipole e (resp. $e^{'}$) correspond à la projection perspective dans l'image de gauche (resp. droite) du centre optique $C^{'}$ (resp. C) (cf. figure 2.4).
- un *plan épipolaire* est un plan contenant la ligne de base.
- une *ligne épipolaire* est l'intersection d'un plan épipolaire avec le plan de l'image. Toutes les lignes épipolaires d'une image se coupent en l'unique point épipole. L'intersection d'un plan épipolaire avec chacune des deux plans image sont deux lignes *épipolaires conjuguées*.

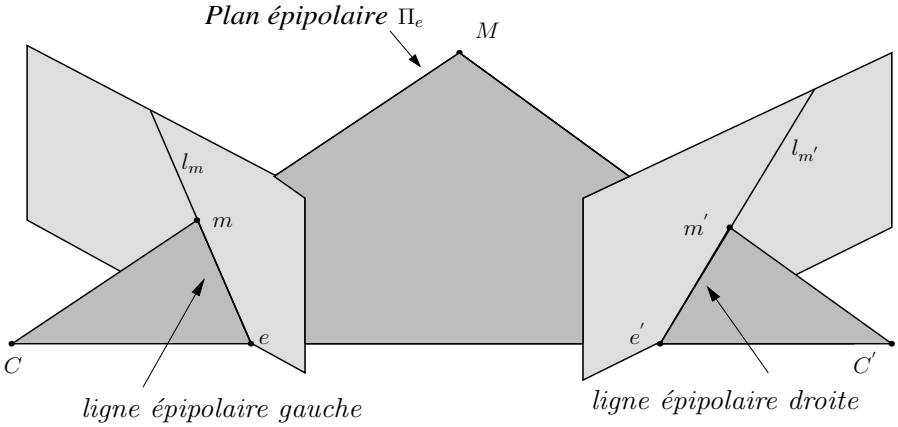


FIG. 2.4 – La géométrie épipolaire.

La géométrie épipolaire fournit la *contrainte épipolaire* qui s'énonce comme suit.

Propriété 2.1 *Un point m, qui se trouve sur la ligne épipolaire gauche l_m doit nécessairement correspondre à un point $m^{'}$ sur sa ligne épipolaire conjuguée $l_{m'}$, et réciproquement.*

La contrainte épipolaire nous permet de restreindre le domaine de recherche des correspondants, de l'image entière à la ligne épipolaire. En effet, lorsqu'on cherche pour un point dans l'image de gauche un correspondant dans l'image de droite, on peut se limiter à effectuer cette recherche le long de la ligne épipolaire associée. La contrainte épipolaire est donc utilisée dans tous les travaux de mise en correspondance.

2.2.3.2 La matrice fondamentale

La contrainte épipolaire permet d'obtenir une relation entre un pixel dans une image et son correspondant dans l'autre image. Ainsi, à chaque point m de l'image de gauche est associée une ligne épipolaire $l_{m'}$ dans l'image de droite. Le correspondant $m^{'}$ de m appartient nécessairement à cette droite. La matrice qui représente cette relation est appelée matrice fondamentale $\mathbf{F}$. Elle décrit l'application linéaire en coordonnées projectives, liant le point m de l'image de gauche à sa ligne épipolaire $l_{m'}$ dans l'image de droite, suivant la relation

$$\widetilde{\mathbf{l}}_{\mathbf{m'}} = \mathbf{F}\widetilde{\mathbf{m}}. \tag{2.19}$$

Géométriquement, $\mathbf{F}$ définit une application de l'image de gauche dans le faisceau de lignes épipolaires passant par l'épipole de l'image de droite, soit une application du plan projectif $\mathbb{P}^2$ vers une droite projective de $\mathbb{P}^1$.

Nous allons maintenant exprimer la matrice fondamentale $\mathbf{F}$, définie à un facteur d'échelle près, en fonction des matrices de projections perspectives $\mathbf{P}$ et $\mathbf{P}^{'}$ associées respectivement aux caméras de gauche et de droite. La formulation (2.26) donnée ci-après revient à Zhang et Xu [Zhang, Xu, 1997].

Proposition 2.2 [Faugeras, 1993] *Une droite $\widetilde{\mathbf{l}}$ du plan projectif $\mathbb{P}^2$ passant pas deux points m_1 et m_2, de coordonnées homogènes $\widetilde{\mathbf{m}}_{\mathbf{1}}$ et $\widetilde{\mathbf{m}}_{\mathbf{2}}$, est représentée par le produit vectoriel*

$$\widetilde{\mathbf{l}} = \widetilde{\mathbf{m}}_{\mathbf{1}} \wedge \widetilde{\mathbf{m}}_{\mathbf{2}}. \tag{2.20}$$

On vérifie, en effet, aisément que les deux premières composantes de $\widetilde{\mathbf{l}}$ fournissent un vecteur normal à la droite $(\widetilde{\mathbf{m}}_{\mathbf{1}}\widetilde{\mathbf{m}}_{\mathbf{2}})$.

Soit à présent m un point de l'image de gauche et $m^{'}$ son correspondant dans l'image de droite, alors la ligne épipolaire $l_{m'}$ associée à m passe nécessairement par le point $m^{'}$ et par l'épipole $e^{'}$ de l'image de droite. D'après la proposition 2.2, elle peut donc être définie, en coordonnées projectives, par le produit vectoriel de ces deux points, soit

$$\widetilde{\mathbf{l}}_{\mathbf{m'}} = \widetilde{\mathbf{e}}^{\,\prime} \wedge \widetilde{\mathbf{m}}^{\prime} = [\widetilde{\mathbf{e}}^{\,\prime}]_{\times}\widetilde{\mathbf{m}}^{\prime}, \tag{2.21}$$

où $[\widetilde{\mathbf{e}}\,']_\times$ est la matrice antisymétrique relative à $\widetilde{\mathbf{e}}\,' = (\widetilde{e}_1', \widetilde{e}_2', \widetilde{e}_3')^\top$ et donnée par

$$[\widetilde{\mathbf{e}}\,']_\times = \begin{pmatrix} 0 & -\widetilde{e}_3' & \widetilde{e}_2' \\ \widetilde{e}_3' & 0 & -\widetilde{e}_1' \\ -\widetilde{e}_2' & \widetilde{e}_1' & 0 \end{pmatrix}. \tag{2.22}$$

En éliminant $\widetilde{\mathbf{M}}$ dans les équations (2.16) et (2.17), on obtient l'équation suivante liant $\widetilde{\mathbf{m}}$ et $\widetilde{\mathbf{m}}'$

$$\widetilde{\mathbf{m}}' = \mathbf{P}'\mathbf{P}^+\widetilde{\mathbf{m}}, \tag{2.23}$$

où $\mathbf{P}^+$ est le pseudo-inverse de $\mathbf{P}$ défini par :

$$\mathbf{P}^+ = \mathbf{P}^\top(\mathbf{P}\mathbf{P}^\top)^{-1}. \tag{2.24}$$

L'équation (2.21) devient ainsi

$$\widetilde{\mathbf{l}}_{\mathbf{m}'} = [\widetilde{\mathbf{e}}\,']_\times \mathbf{P}'\mathbf{P}^+\widetilde{\mathbf{m}}. \tag{2.25}$$

En combinant les équations (2.19) et (2.25), la matrice fondamentale s'écrit en fonction des matrices de projection $\mathbf{P}$, $\mathbf{P}'$ et de l'épipole de l'image de droite e' comme suit

$$\mathbf{F} = [\widetilde{\mathbf{e}}\,']_\times \mathbf{P}'\mathbf{P}^+. \tag{2.26}$$

La matrice fondamentale dépend donc seulement de la configuration des caméras (paramètres intrinsèques et extrinsèques) et est indépendante des coordonnées du point de la scène. Dans le cas où le repère de l'espace est celui de la caméra de gauche, les expressions des matrices de projection perspectives sont données par l'équation (2.18). Ainsi,

$$\mathbf{P}^+ = \begin{pmatrix} \mathbf{A}^{-1} \\ \mathbf{0}^\top \end{pmatrix} \quad \text{et} \quad \widetilde{\mathbf{C}} = \begin{pmatrix} \mathbf{0} \\ \mathbf{1} \end{pmatrix}. \tag{2.27}$$

Par ailleurs, l'épipole e', représentant la projection dans l'image de droite du centre optique C de la caméra de gauche, est défini par

$$\widetilde{\mathbf{e}}\,' = \mathbf{P}'\widetilde{\mathbf{C}}. \tag{2.28}$$

En remplaçant dans (2.26) et en utilisant la propriété suivante [Hartley, Zisserman, 2004] :

$$[\mathbf{A}'\mathbf{t}]_\times \mathbf{A}' = \det(\mathbf{A}')\mathbf{A}'^{-\top}[\mathbf{t}]_\times, \tag{2.29}$$

et l'équation (2.18), la matrice fondamentale $\mathbf{F}$ s'écrit

$$\begin{aligned} \mathbf{F} &= [\mathbf{P}'\widetilde{\mathbf{C}}]_\times \mathbf{P}'\mathbf{P}^+ \\ &= [\mathbf{A}'\mathbf{t}]_\times \mathbf{A}'\mathbf{R}\mathbf{A}^{-1} \\ &= \mathbf{A}'^{-\top}[\mathbf{t}]_\times \mathbf{R}\mathbf{A}^{-1}. \end{aligned} \tag{2.30}$$

Notons que puisque $\mathbf{F}$ est définie à un facteur d'échelle près, le coefficient $\det(\mathbf{A}')$ a été omis.

Proposition 2.3 *Pour chaque couple de points homologues $(m, m^{'})$, la matrice fondamentale vérifie la propriété*

$$\widetilde{\mathbf{m}}^{'\top}\mathbf{F}\widetilde{\mathbf{m}} = 0. \tag{2.31}$$

Démonstration. Le correspondant $m^{'}$ d'un point m de l'image de gauche appartient nécessairement à la ligne épipolaire $l_{m'}$ associée à m, soit en coordonnées projectives

$$\widetilde{\mathbf{m}}^{'\top}\widetilde{\mathbf{l}}_{\mathbf{m}'} = 0.$$

Il s'ensuit alors d'après (2.19) la relation (2.31). ■

La matrice fondamentale permet donc de vérifier la correspondance entre les points images. Elle possède les propriétés suivantes

- $\mathbf{F}$ est une matrice 3×3, de rang 2, définie à un coefficient multiplicatif près et à 7 degrès de liberté.
- Si $\mathbf{F}$ est la matrice fondamentale de couple d'images (gauche, droite) alors $\mathbf{F}^{\top}$ est la matrice fondamentale de couple d'images (droite, gauche) et vérifie

$$\widetilde{\mathbf{m}}^{\top}\mathbf{F}^{\top}\widetilde{\mathbf{m}}^{'} = 0. \tag{2.32}$$

- Pour tout point m de l'image de gauche, $\widetilde{\mathbf{l}}_{m'} = \mathbf{F}\widetilde{\mathbf{m}}$ représente la ligne épipolaire sur laquelle se trouvent les correspondants possibles de m dans l'image de droite. Réciproquement, $\widetilde{\mathbf{l}}_{m} = \mathbf{F}^{\top}\widetilde{\mathbf{m}}^{'}$ traduit la ligne épipolaire associée à $m^{'}$ dans l'image de gauche.
- Les épipoles e et $e^{'}$ de l'image gauche et droite, respectivement, vérifient

$$\mathbf{F}\,\widetilde{\mathbf{e}} = 0 \quad \text{et} \quad \mathbf{F}^{\top}\widetilde{\mathbf{e}}^{\,'} = 0.$$

L'estimation de la matrice fondamentale a fait l'objet de nombreux travaux. Une synthèse des différentes techniques est proposée dans [Zhang, 1996]. D'une manière générale, celle-ci peut être estimée à partir des paramètres de calibration ou, si l'on ne dispose pas de ceux-ci, en utilisant un ensemble d'appariements de points entre les deux images [Zhang *et al.*, 1994, Oisel *et al.*, 2003]. Dans ce dernier cas et lorsque la matrice fondamentale est disponible, on dit que le système binoculaire est *faiblement calibré*.

2.2.3.3 La matrice essentielle

La matrice essentielle a été introduite avant la matrice fondamentale par Longuet-Higgins [Longuet-Higgins, 1981]. Elle ne prend pas en compte les caractéristiques des caméras (paramètres intrinsèques) et s'exprime simplement à l'aide de la rotation $\mathbf{R}$ et la translation $\mathbf{t}$ s'opérant entre les repères des deux caméras :

$$\mathbf{E} = [\mathbf{t}]_{\times}\mathbf{R}. \tag{2.33}$$

Supposons que les paramètres intrinsèques $\mathbf{A}$ et $\mathbf{A}^{'}$ des deux caméras sont connues, alors la matrice essentielle $\mathbf{E}$ vérifie la propriété suivante.

Propriété 2.2 *Pour tout point M de l'espace, associée à ses projections m et m' dans les deux images, on a*

$$\widetilde{\mathbf{m}}_{\mathbf{n}}^{'\top}\mathbf{E}\widetilde{\mathbf{m}}_{\mathbf{n}} = 0, \tag{2.34}$$

où

$$\widetilde{\mathbf{m}}_{\mathbf{n}} = \mathbf{A}^{-1}\widetilde{\mathbf{m}} \quad et \quad \widetilde{\mathbf{m}}'_{\mathbf{n}} = \mathbf{A}^{'-1}\widetilde{\mathbf{m}}', \tag{2.35}$$

sont les coordonnées normalisées, exprimées en mètre, de m et m', respectivement.

La matrice $\mathbf{E}$ est de taille 3×3, de rang 2 et est déterminée à une constante multiplicative près. Elle a des propriétés identiques à celles de $\mathbf{F}$ mais a seulement 5 degrès de liberté. En particulier, d'après (2.30) et (2.31) la matrice essentielle est reliée à la matrice fondamentale par la relation linéaire suivante

$$\mathbf{F} = \mathbf{A}^{'-\top}\mathbf{E}\mathbf{A}^{-1}. \tag{2.36}$$

2.2.3.4 Différences entre les matrices essentielle et fondamentale

Si aucun calibrage des caméras n'est disponible, la géométrie épipolaire est complètement décrite par la matrice fondamentale. L'estimation de cette matrice, qui contient toute l'information du système binoculaire, permet de faciliter la mise en correspondance de points d'images.

La matrice essentielle contient l'information sur le déplacement entre les deux vues gauche et droite. Lorsque les paramètres intrinsèques de chaque caméra sont connues, on estime la matrice essentielle, à partir de laquelle on peut extraire la transformation rigide entre les deux caméras, en la décomposant en un produit d'une matrice antisymétrique et une matrice de rotation. En général, cinq correspondances de points suffisent pour estimer la matrice essentielle à un facteur d'échelle près.

2.2.3.5 La rectification épipolaire

Il existe une configuration géométrique particulière du capteur stéréoscopique, appelée configuration parallèle. Cette configuration exige que les axes optiques des deux caméras soient parallèles entre eux et qu'ils soient tous deux perpendiculaires à la droite joignant les deux centres optiques C et C'.

Cette géométrie de caméras est particulièrement intéressante car elle simplifie le processus de mise en correspondance des images. En effet, dans cette configuration, les lignes épipolaires sont toutes parallèles et correspondent aux lignes horizontales des images. Deux pixels qui se correspondent se trouvent donc sur la même ligne dans les deux images, comme l'illustre la figure 2.5.

Toutefois, il est rare en pratique que les caméras soient parfaitement alignées selon ce modèle projectif parallèle. Une transformation géométrique des images devra être effectuée de façon à rendre les lignes épipolaires parallèles et horizontales dans les deux images. Cette transformation est appelée *rectification épipolaire* et consiste à reprojeter les images formées sur les plans images de chaque caméra sur un nouveau plan de projection commun entre les deux caméras. Pour que les droites épipolaires soient parallèles, il faut que ce nouveau plan

de projection soit parallèle à l'axe reliant les deux centres de projection. On obtient après rectification, une géométrie épipolaire simplifiée où le changement de repère entre les deux caméras se réduit à une simple translation.

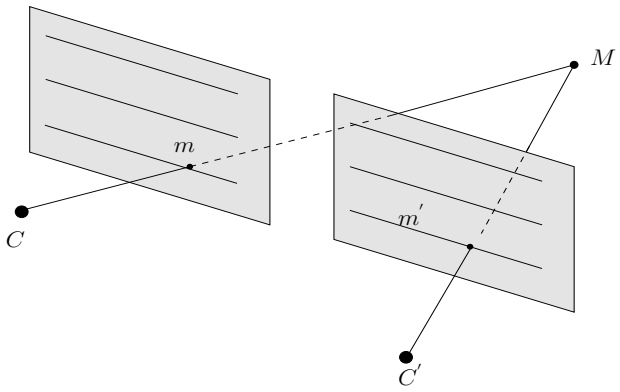


FIG. 2.5 – La rectification épipolaire.

L'avantage de la rectification est de simplifier considérablement la mise en correspondance de points entre les deux images en limitant la zone de recherche des correspondants à des paires de lignes horizontales. La difficulté réside dans l'estimation des paramètres de la transformation à appliquer pour rectifier les images stéréoscopiques. Plusieurs méthodes ont été proposées, selon l'information dont on dispose :

- Si l'on dispose des paramètres intrinsèques et extrinsèques du capteur stéréoscopique, les matrices de projection perspective associées aux deux images peuvent être calculées. Nous pouvons alors en déduire la transformation à appliquer de façon à rendre les deux plans images coplanaires [Fusiello *et al.*, 2000b] et parallèles à la droite passant par les deux centres optiques.

- Lorsque les paramètres de calibration ne sont pas connus, la technique de rectification se base sur la connaissance de la matrice fondamentale calculée à partir d'appariements de points entre les deux images. La connaissance de cette matrice permet, par la suite, de calculer les paramètres de la rectification [Hartley, Gupta, 1993].

2.2.3.6 Définition de la disparité

Nous nous plaçons dans le cas où la mise en correspondance est effectuée de l'image de gauche vers l'image de droite et nous appelons image de référence l'image de gauche. Un point dans l'image de référence peut être associé à un point dans l'image de droite s'ils correspondent au même point physique de la scène : ce sont deux points homologues. La différence entre les coordonnées des pixels homologues est appelée *disparité*.

Nous pouvons donc assimiler le problème de mise en correspondance à la recherche d'une

fonction u qui attribue une disparité à chaque pixel (x, y) de l'image de gauche :

$$\begin{array}{rccl} u: & \mathbb{N}^2 & \to & \mathbb{R}^2 \\ & (x,y) & \mapsto & u(x,y) = (x - x^{'}, y - y^{'}). \end{array} \tag{2.37}$$

où $(x^{'}, y^{'})$ sont les coordonnées dans l'image de droite du pixel correspondant à (x, y).

Propriété 2.3 *Lorsque les images sont rectifiées, les points en correspondance sont sur des droites horizontales de même ordonnée, c'est à dire $y = y^{'}$. Dans ce cas, la disparité correspond à la différence d'abscisses et l'expression (2.37) devient*

$$\begin{array}{rccl} u: & \mathbb{N}^2 & \to & \mathbb{R} \\ & (x,y) & \mapsto & u(x,y) = x - x^{'}. \end{array} \tag{2.38}$$

Tout au long de ce manuscrit, nous allons travailler avec des images rectifiées et nous retenons ainsi la définition (2.38) de la disparité.

Remarque 2.4 *Dans le cas d'images rectifiées et lorsque la mise en correspondance est effectuée de la droite vers la gauche, la disparité en tout point de coordonnées $(x^{'}, y)$ dans l'image de droite et dont l'homologue a pour coordonnées (x, y) dans l'image de gauche peut s'écrire :*

$$\begin{array}{rccl} u_d: & \mathbb{N}^2 & \to & \mathbb{R} \\ & (x^{'},y) & \mapsto & u_d(x^{'},y) = x - x^{'}. \end{array} \tag{2.39}$$

A partir des équations (2.38) et (2.39), on peut établir, pour tout point (x, y) de l'image de gauche qui admet un correspondant dans l'image de droite, l'égalité suivante

$$u(x, y) = u_d(x - u(x, y), y). \tag{2.40}$$

Pour visualiser le résultat de la mise en correspondance, nous utilisons dans ce mémoire, comme dans tous les travaux de mise en correspondance, une image appelée *carte de disparité*. Chaque pixel de la carte représente l'amplitude de la disparité : plus le pixel est clair et plus la disparité est importante. La figure 2.6 montre un couple d'images, dénommé *cones*, et la carte de disparité associée.



(a) (b) (c)

FIG. 2.6 – Couple d'images Cones. (a) Image gauche. (b) Image droite. (c) Carte de disparité.

2.2.3.7 Reconstruction 3D par triangulation

L'objectif de la reconstruction 3D par triangulation est de calculer un modèle tridimensionnel de la scène observée. L'opération de triangulation consiste à estimer, à partir des résultats de la mise en correspondance, les coordonnées (X, Y, Z) d'un point M de la scène. La profondeur Z est la distance entre le point M et le plan des caméras (cf. figure 2.7).
Dans ce qui suit, nous montrons, en particulier, que la profondeur du point M peut être directement calculée en fonction de la disparité de ses projections dans les deux images.

En appliquant le théorème de Thalès dans les triangles rectangles d'hypoténuses (CM) et $(C^{'}M)$ de la figure 2.7, on obtient :

$$\frac{x}{f} = \frac{X}{Z}, \tag{2.41}$$

et

$$\frac{x^{'}}{f} = \frac{X - B}{Z},$$

où B est la distance qui sépare les deux caméras et f est la distance focale. De la définition (2.38) de la disparité u, il résulte que

$$u = x - x^{'} = \frac{B\,f}{Z}. \tag{2.42}$$

On constate d'après cette équation que la disparité u est inversement proportionnelle à la profondeur Z. Sur la carte de disparité de la figure 2.6, la zone claire correspond à des disparités élevées et donc à des profondeurs faibles, et inversement pour la zone foncée qui est plus loin des caméras que la zone claire. Calculer une carte de disparité suffit donc pour estimer la profondeur des objets 3D de la scène.

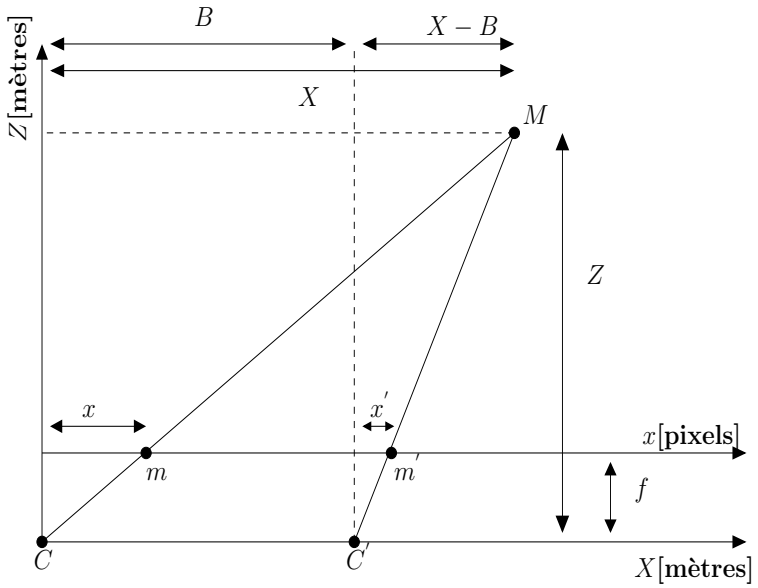


Fig. 2.7 – Relation entre la disparité et la profondeur.

2.3 La mise en correspondance stéréoscopique

Malgré la simplification qu'apporte la rectification épipolaire, le problème de mise en correspondance reste un problème difficile à résoudre, en raison notamment de la présence de changements de luminosité entre les deux images, d'occultations et de surfaces non texturées. Pour pallier ces difficultés, les méthodes de mise en correspondance sont amenées à exploiter toutes les informations disponibles afin de faciliter la recherche et la détermination des correspondants.

Dans ce paragraphe, nous détaillons tout d'abord les difficultés de la mise en correspondance, puis nous présentons les primitives d'images et les contraintes les plus utilisées pour surmonter ces difficultés.

2.3.1 Les difficultés de la mise en correspondance

La mise en correspondance consiste à retrouver, à partir de deux images, les couples de points qui se correspondent. Si l'on dispose d'images de niveaux de gris, il paraît raisonnable de faire l'hypothèse que les projections d'un même point de la scène ont la même intensité dans les deux images [Marr, Poggio, 1976]. Cette contrainte physique connue sous le nom de conservation de la luminance, est aujourd'hui utilisée dans la plupart des approches de mise en correspondance. Cependant, elle est uniquement vérifiée pour les surfaces lambertiennes qui reflètent la lumière identiquement suivant toutes les directions. Outre l'existence de surfaces spéculaires, la violation de cette hypothèse peut être causée par la présence de bruit de mesure ou simplement par des différences de sensibilités entre les caméras utilisées. Par conséquent, l'un des problèmes importants des méthodes de mise en correspondance, utilisant l'hypothèse de conservation de la luminance, est leur sensibilité aux changements d'illumination.

Un autre problème très important provient de la présence d'objets dont la surface est uniforme ou la présence de textures répétitives. En effet, les pixels dans ces régions ne représentent pas des primitives discriminantes puisque plusieurs correspondants potentiels ayant la même valeur de luminance existent dans l'autre image. Les régions homogènes sont donc difficiles à apparier. Elle sont prises en compte et détectées en utilisant le plus souvent le gradient de l'image [Scharstein, Szeliski, 2002].

Par ailleurs, une autre difficulté rencontrée par les méthodes de mise en correspondance concerne les occultations. En effet, certains points de la scène peuvent être visibles dans une image de la paire stéréoscopique mais pas dans l'autre. C'est à dire qu'une portion de la scène observée n'est visible que depuis un seul point de vue : il s'agit d'un problème d'occultation. Ce problème est illustré par la figure 2.8 : le point m est visible uniquement depuis la caméra de gauche. Les occultations représentent une difficulté majeure pour les techniques de mise en correspondance car ces zones d'images n'ont pas de correspondants dans l'autre image et engendrent souvent des ambiguïtés d'appariement. Dans la littérature, de nombreux articles ont été proposés pour prendre en compte cette difficulté [Weng *et al.*, 1992, Kolmogorov, Zabih, 2005, Kang *et al.*, 2001, Zitnick, Kanade, 2000].

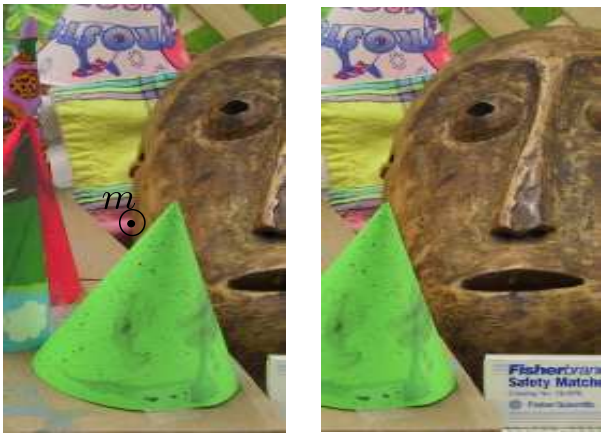


Image de gauche Image de droite

FIG. 2.8 – Le problème d'occultation : le point m est visible dans l'image de gauche mais pas dans l'image de droite.

Brown et al. [Kang *et al.*, 2001] classent ces méthodes en trois catégories consistant à :

- Détecter les régions occultées et les remplacer par une interpolation basée sur la disparité du voisinage de la zone occultée [Weng *et al.*, 1992]. Une étude détaillée et comparative de cinq méthodes de détection d'occlusions a été menée dans [Egnal, Wildes, 2002].
- Réduire la sensibilité de la mise en correspondance aux occultations en utilisant des mesures de corrélation robustes comme la corrélation normalisée ou les M-estimateurs [Oisel *et al.*, 2003]. Zabih et Woodfill [Zabih, Woodfill, 1994] appliquent des transformations non paramétriques aux intensités des pixels avant de procéder à la mise en correspondance par corrélation. Une autre approche pour prendre en compte le problème des occultations est d'adapter la taille et la forme de la zone de recherche pour optimiser la mesure de ressemblance au voisinage d'une occultation [Zitnick, Kanade, 2000].
- Calculer un modèle géométrique d'occultation afin de l'intégrer au processus d'appariement. Ce modèle peut ensuite être utilisé dans le cadre d'une mise en correspondance par programmation dynamique ou coupures de graphes pour définir un coût d'occultation [Birchfield, Tomasi, 1998].

2.3.2 Les primitives à mettre en correspondance

Le choix des primitives, c'est à dire les caractéristiques extraites des images, est déterminant dans le processus de mise en correspondance. Ces primitives peuvent être des pixels, des points d'intérêt, des points de contours, des segments, etc.

Les pixels de l'image

Le pixel est la primitive la plus simple, dite de bas niveau. Sa mise en correspondance permet d'obtenir une carte de profondeur dense. Elle offre la plus grande densité mais c'est aussi la primitive la plus pauvre en information, dont les seuls attributs sont la position et le niveau de gris.

Les points d'intérêt

Ce sont les pixels qui correspondent à des points particuliers dans l'image et qui possèdent des caractéristiques particulières les rendant plus discriminants pour la mise en correspondance. Comme type de pixel particulier, on peut citer les points de contours qui proviennent des changements brusques de niveau de gris, les coins, les jonctions, etc. De nombreuses méthodes de mise en correspondance s'appuient sur une détection de points d'intérêt [Konrad, Lan, 2000, Kim *et al.*, 2005a, Wei, Quan, 2004]. Les détecteurs les plus populaires sont celui de Canny [Canny, 1986] et de Harris et Stephens [Harris, Stephens, 1988].

Les primitives structurées

Ce sont par exemple les segments, les courbes, les régions, etc. Leurs attributs sont nombreux et peuvent être de types géométriques ou photométriques : la position, l'orientation, la longueur, le niveau de gris, la surface, etc. Les travaux de mise en correspondance utilisant des segments sont nombreux, on peut citer à titre d'exemple [Boukir *et al.*, 1992]. Cependant, les régions sont des primitives difficiles à utiliser pour la mise en correspondance car, d'une image à l'autre, elles peuvent avoir des formes et des tailles différentes, dûes au phénomène de distorsion projective.

Les approches de mise en correspondance sont souvent classées en deux grandes catégories suivant les primitives à apparier. Les méthodes basées sur la mise en correspondance de pixels, appelées *pixel-based matching*, sont très sensibles aux occultations et aux changements d'illumination entre les deux vues. Elles sont aussi, souvent, coûteuses en temps de calcul. Cependant, l'avantage de ces méthodes est qu'elles permettent d'obtenir directement un appariement très dense. Elles concernent de ce fait la plupart des publications sur la mise en correspondance. Toutefois, un certain nombre de publications utilise des primitives éparses, comme les points d'intérêt ou les régions, pour guider la mise en correspondance de pixels. Ces méthodes, appelées *feature-based matching*, nécessitent de détecter au préalable ces primitives qui doivent être caractérisées par des attributs suffisamment invariants et discriminants.

Dans le cadre de notre application où l'obtention d'une reconstruction dense de la scène apparaît comme une nécessité, les primitives que nous considérons sont les pixels de l'image dans leur totalité.

2.3.3 Les contraintes pour la mise en correspondance

Il est nécessaire d'utiliser des contraintes pour réduire le nombre de correspondants potentiels et pour éliminer les faux appariements. On peut distinguer trois types de contraintes :

- géométriques liées à la nature du capteur stéréoscopique : la contrainte la plus couramment utilisée est la contrainte épipolaire.
- géométriques liées à la nature de la scène : elles concernent plus particulièrement les objets formant la scène. Il s'agit des contraintes d'unicité, d'ordre et du gradient de disparité. Elles sont aussi appelées contraintes de compatibilité.
- physiques dues aux propriétés d'illumination : ce type de contrainte découle des hy-

pothèses faites sur la réflectance de la surface des objets et la position de la source lumineuse. Si l'on suppose que les sources lumineuses sont fixes et que les objets sont parfaitement mats, ou lambertiens, un point sur un objet de la scène se projette toujours avec la même intensité dans les deux images.

Dans ce paragraphe, nous étudierons seulement les contraintes géométriques. La contrainte physique fera l'objet d'une discussion plus approfondie au chapitre 5.

La contrainte épipolaire

Cette contrainte géométrique, décrite au paragraphe 2.2.3, nous permet de réduire l'espace de recherche des correspondants à une seule dimension en considérant que les points homologues se situent toujours le long des lignes épipolaires conjuguées.
Lorsque les images sont rectifiées, les lignes épipolaires conjuguées correspondent à la même ligne dans les deux images. Dans ce cas, la zone de recherche peut être davantage réduite si l'on fixe les limites inférieures et supérieures des disparités qui correspondent aux objets de la scène respectivement les plus éloignés et les plus proches du capteur stéréoscopique.

La contrainte de disparité positive

Cette contrainte provient directement de la géométrie du capteur stéréoscopique. Si les caméras sont parfaitement alignées ou les images sont rectifiées, la disparité est toujours positive. En effet, d'après l'équation (2.42), la disparité u est inversement proportionnelle à la profondeur Z. Or la profondeur des objets situés devant la caméra étant positive, la disparité doit donc aussi être positive.

La contrainte d'unicité

La contrainte d'unicité [Marr, Poggio, 1976, Marr, Poggio, 1979] impose qu'une primitive dans une image ne peut être appariée qu'avec une seule primitive dans l'autre image. Cela suppose donc qu'une primitive est la projection d'un unique élément 3D. Dans le cas d'images rectifiées, cette contrainte s'exprime analytiquement comme suit

$$\text{Si } x_1 - u(x_1, y) = x_1^{'}, \quad \text{alors } \forall x_2 \neq x_1 \;\; x_2 - u(x_2, y) \neq x_1^{'}, \tag{2.43}$$

où (x_1, y) et (x_2, y) sont deux points de l'image de gauche. La contrainte d'unicité n'est mise en défaut que dans le cas où un plan de la scène est très incliné. Plusieurs éléments 3D se projettent ainsi en une même primitive dans une image. Un exemple est fourni sur la figure 2.9. Les points M_1 et M_2 se projettent en un même point dans l'image de droite et en deux points différents dans l'image de gauche.

La contrainte d'ordre

La contrainte d'ordre [Yuille, Poggio, 1984], illustrée par la figure 2.10, stipule que l'ordre des projections le long des lignes épipolaires doit être le même que celui de leurs correspondants.

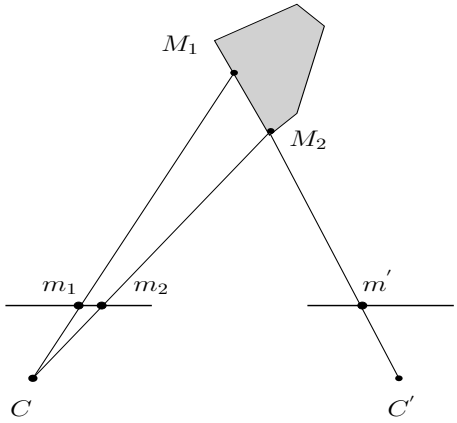


FIG. 2.9 – Cas de non respect de la contrainte d'unicité : tous les pixels qui se trouvent sur le segment de droite $[M_1, M_2]$ se projettent sur le pixel m' dans l'image de droite.

Dans le cas d'images rectifiées, cette contrainte est définie par :

$$\text{Si } x_1 - u(x_1, y) = x_1' \text{ et } x_2 - u(x_2, y) = x_2' \text{ alors } (x_1 - x_2)(x_1' - x_2') > 0, \quad (2.44)$$

où (x_1, y) et (x_2, y) sont deux points de l'image de gauche et (x_1', y) et (x_2', y) leurs correspondants dans l'image de droite.

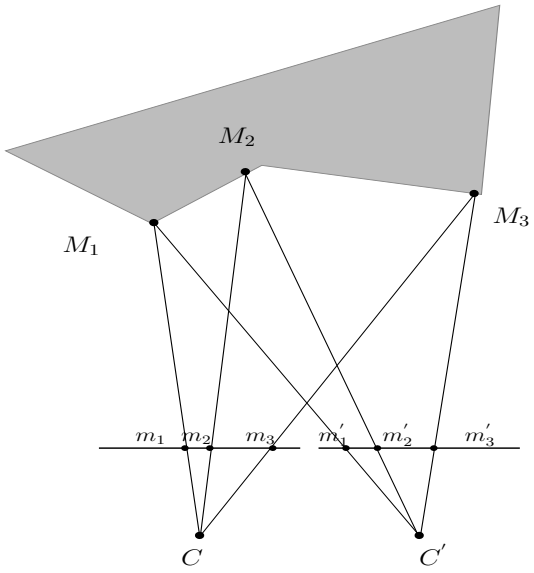


FIG. 2.10 – La contrainte d'ordre.

A chaque point M de la scène 3D est associée une *zone interdite* défini par les deux rayons optiques (CM) et $(C'M)$ [Faugeras, 1993, Yuille, Poggio, 1984]. Tout point N appartenant à cette zone aura l'ordre de ses projections inversé dans les deux plans

images par rapport aux projections du point M (cf. figure 2.11(a)).

Remarque 2.5 *Si M et N se trouvent sur la surface d'un même objet transparent, comme c'est le cas dans la figure 2.11(b), le fait que N appartienne à la zone interdite de M correspond au fait que l'objet est très incliné par rapport au plan des images.*

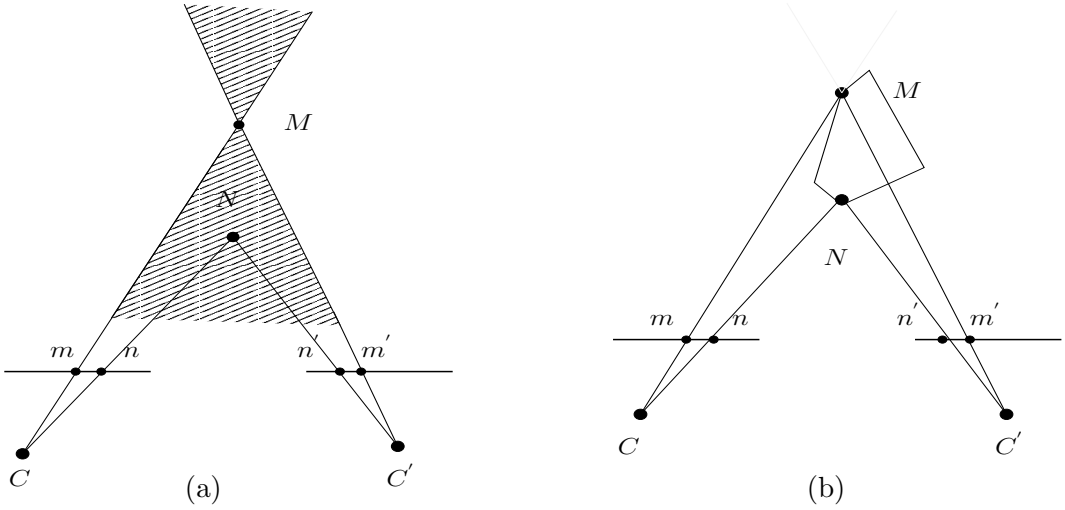


FIG. 2.11 – (a) Zone interdite associée au point M. (b) Cas d'un objet transparent fortement incliné.

La limite du gradient de disparité

Le gradient de disparité est défini comme la différence des disparités de deux couples de points divisée par leur *séparation cyclopéenne*. Celle-ci est égale à la moyenne arithmétique de la distance entre deux pixels dans l'image de gauche et de la distance entre leurs correspondants respectifs dans l'image de droite.
Soient (x_1, y), (x_2, y) deux pixels dans l'image de gauche et $(x_1^{'}, y)$, $(x_2^{'}, y)$ leurs correspondants respectifs dans l'image de droite, le gradient de disparité est défini par l'expression suivante [Burt, Julesz, 1980] :

$$G_u\left((x_1, y), (x_2, y)\right) = \frac{|u(x_1, y) - u(x_2, y)|}{|(\frac{1}{2}\left((x_1 + x_1^{'}) - (x_2 + x_2^{'})\right)|}. \tag{2.45}$$

La contrainte de la limite du gradient de disparité consiste à imposer une limite l_u sur la valeur du gradient de disparité associée à deux couples de points. Elle a été introduite par Burt et Julesz comme contrainte sur le système visuel humain où la limite l_u est égale à 1 [Burt, Julesz, 1980], puis Pollard et al. ont proposé une application de cette contrainte en stérévision [Pollard *et al.*, 1985] et ont suggéré une limite qui peut varier entre 0 et 2.

Remarque 2.6 *Une valeur de la limite l_u strictement inférieure à deux garantit que la*

contrainte d'ordre et celle d'unicité sont satisfaites.

Démonstration. Soient (x_1, y) et (x_2, y) deux pixels de l'image de gauche, tel que

$$x_1 - u(x_1, y) = x_1^{'} \quad \text{et} \quad x_2 - u(x_2, y) = x_2^{'}.$$

Supposons que G_u défini par l'équation (2.45) est inférieur à 2 et montrons que la contrainte d'ordre est vérifiée. Pour cela, on réécrit le gradient de la disparité de la façon suivante

$$G_u = \frac{|(x_1 - x_2) - (x_1^{'} - x_2^{'})|}{|\frac{1}{2}\left((x_1 - x_2) + (x_1^{'} - x_2^{'})\right)|}. \tag{2.46}$$

Comme $G_u < 2$, on a

$$|(x_1 - x_2) - (x_1^{'} - x_2^{'})| < |(x_1 - x_2) + (x_1^{'} - x_2^{'})|,$$

il s'ensuit l'inégalité suivante

$$\left((x_1 - x_2) - (x_1^{'} - x_2^{'})\right)^2 < \left((x_1 - x_2) + (x_1^{'} - x_2^{'})\right)^2,$$

et donc

$$(x_1 - x_2)(x_1^{'} - x_2^{'}) > 0.$$

D'après la définition (2.44) de la contrainte d'ordre, on conclut que cette dernière est valide.

Prouvons à présent que la contrainte d'unicité est également vérifiée. Pour cela, on suppose que $x_1 \neq x_2$ et on se propose de montrer que $x_1^{'} \neq x_2^{'}$. Si $x_1^{'} = x_2^{'}$, alors d'après (2.46), on a

$$G_u = \frac{|(x_1 - x_2)|}{|\frac{1}{2}(x_1 - x_2)|} < 2,$$

or, puisque $x_1 - x_2 \neq 0$, ceci est absurde. On en déduit donc que si $x_1 \neq x_2$ alors $x_1^{'} \neq x_2^{'}$, ce qui montre d'après (2.43) que la contrainte d'unicité est respectée. ■

La contrainte de continuité de disparité

Cette contrainte repose sur l'observation que les surfaces des objets constituant la scène observée sont continues et que la variation locale de profondeur est faible. Elle impose que la disparité varie continûment quasiment partout sur une image [Marr, Poggio, 1979]. Cette contrainte de cohérence spatiale est vérifiée pour les variations lentes de profondeur, mais elle ne l'est pas en présence de discontinuités de profondeur qui correspondent aux frontières des objets. Il existe une autre contrainte moins forte appelée contrainte de continuité figurale qui impose la continuité uniquement le long des contours de l'image [Horaud, Monga, 1993].

La contrainte de symétrie

La contrainte de symétrie découle de la contrainte d'unicité. Elle impose qu'un point $m^{'}$ de l'image de droite est le correspondant d'un point m de l'image de gauche, si et seulement si m est le correspondant de $m^{'}$ [Fua, 1993]. Elle nécessite à la fois une mise en correspondance de la gauche vers la droite et de la droite vers la gauche qui est

souvent appelée *appariement bidirectionnel*.
Soit (x, y) un pixel de l'image de gauche, cette contrainte s'écrit :

$$\text{Si } \; x - u(x,y) = x^{'} \; \text{ alors } \; x^{'} + u_d(x^{'}, y) = x, \tag{2.47}$$

où u_d est obtenue par mise en correspondance de la droite vers la gauche. La contrainte de symétrie permet ainsi de valider l'appariement entre les éléments des deux images [Fusiello *et al.*, 2000a], [Kim *et al.*, 2005a], [Cochran, Médioni, 1992]. Elle est aussi souvent utilisée pour détecter les pixels occultés ou les pixels ambigus [Konrad, Lan, 2000], [Fua, 1993].

2.4 Un rapide état de l'art sur la mise en correspondance

De nombreux algorithmes de mise en correspondance existent dans la littérature. En effet, le nombre de publications dans ce domaine ne cesse de croître depuis les années 1980. Un état de l'art assez exhaustif des différentes méthodes existantes est réalisé par Scharstein et Szeliski dans [Scharstein, Szeliski, 2002]. Les auteurs distinguent quatre éléments essentiels pour caractériser les algorithmes de stéréovision :

(i) le coût de correspondance,

(ii) la zone d'agrégation,

(iii) la méthode d'optimisation,

(iv) l'affinement des résultats.

Dans notre synthèse effectuée ci-après, on classe les approches de mise en correspondance suivant leur méthode d'optimisation, et on les sépare principalement en quatre catégories : les approches locales, les approches progressives, les approches coopératives et les approches globales.

Dans les approches locales, la mise en correspondance est basée sur l'appariement des éléments les plus semblables grâce à une mesure de similarité. Le principe de ces méthodes repose généralement sur la définition d'une fenêtre appropriée contenant l'élément à mettre en correspondance et sur le choix d'un critère de ressemblance. L'avantage des méthodes dans cette catégorie est que la mise en œuvre est simple et efficace pour un temps de calcul relativement faible. Cependant, les méthodes locales sont très sensibles aux ambiguïtés provoquées par les occultations, une faible texture ou une variation d'illumination entre les différentes vues. Les approches globales tentent de pallier ces inconvénients en minimisant une fonction de coût globale qui fait intervenir l'ensemble des pixels des images, mais au prix d'une complexité de calcul plus élevée. Les méthodes progressives et les méthodes coopératives correspondent à une combinaison d'une méthode locale et d'une méthode globale, dans le but de profiter de leurs avantages respectifs.

2.4.1 Les approches locales

Les approches locales reposent sur la ressemblance entre deux ensembles de pixels et sont aussi appelées approches par corrélation. Ces méthodes sont très populaires et ont été

largement utilisées en vision par ordinateur pour des applications tels que le recalage d'images, l'estimation du mouvement, l'indexation vidéo, etc.

Dans le cadre de la mise en correspondance stéréoscopique, le principe de ces méthodes est de comparer le voisinage du point à apparier avec le voisinage des correspondants potentiels, situés sur la ligne épipolaire associée dans l'autre image. Si l'on cherche le correspondant d'un point (x, y) dans l'image de gauche, la première étape consiste à définir une zone de recherche et une fenêtre glissante qui va parcourir cette zone dans l'image de droite (cf. figure 2.12). Un score de corrélation est ensuite calculé, au point (x, y) et pour une disparité u, en comparant les intensités des pixels d'une fenêtre centrée sur le point (x, y) dans l'image de gauche et une fenêtre centrée sur le point $(x - u, y)$ dans l'image de droite. L'ensemble des scores de mise en correspondance calculés sur la zone de recherche définit l'image de l'espace des disparités, notée DSI (Disparity Space Image). Le point choisi comme correspondant sera celui qui aura le meilleur score de corrélation dans cet espace.

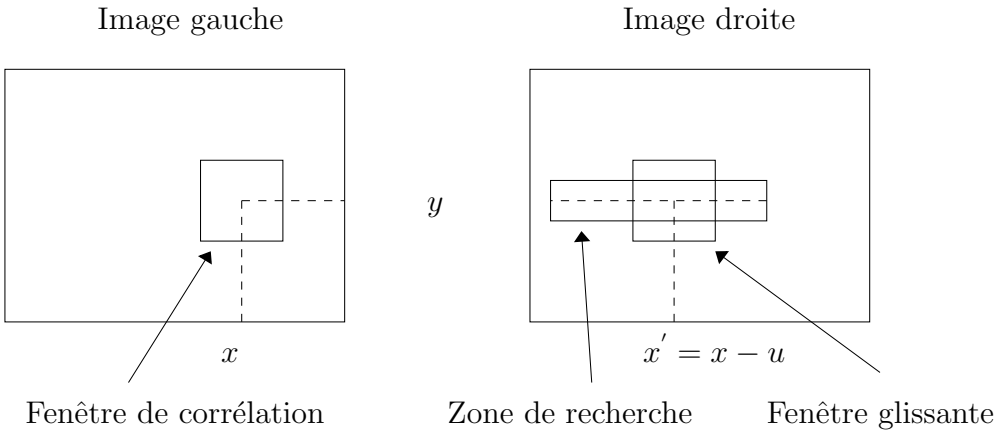


Fig. 2.12 – Mise en correspondance par corrélation.

Afin de pouvoir identifier les pixels des deux images correspondant au même élément physique de la scène, il faut utiliser une mesure de corrélation qui évalue la ressemblance entre les voisinages de deux pixels. Les mesures les plus couramment utilisées sont la somme des carrés des différences, notée SSD (Sum of Squared Differences) et la somme des valeurs absolues des différences, notée SAD (Sum of Absolute Differences). Une autre mesure très populaire est la corrélation normalisée, notée NCC (Normalized Cross-Correlation). Cette mesure est insensible aux transformations affines d'intensité des images, ce qui permet de pallier certains changements d'illumination qui peuvent se produire entre les prises de vues.

D'autres mesures non-paramétriques ont été aussi largement utilisées. Elles sont fondées sur des transformations qui se basent sur l'ordre des niveaux de gris des pixels dans la fenêtre de corrélation et non sur leurs valeurs. Zabih et Woodfill [Zabih, Woodfill, 1994] proposent deux mesures qui s'appuient sur le rang et sur le recensement. La transformation de rang correspond au nombre de pixels dans la fenêtre de corrélation dont le niveau de gris est inférieur au niveau de gris du pixel central et la transformation de recensement remplace les pixels de la fenêtre de corrélation par une chaîne de bits selon que leurs niveaux de

gris est inférieur ou supérieur au niveau de gris du pixel central. Certains travaux ont effectué une évaluation des mesures de corrélation en comparant plusieurs mesures différentes [Aschwanden, Guggenbül, 1992, Roma *et al.*, 2002]. Roma et al. [Roma *et al.*, 2002] ont également étudié l'influence de la taille de la fenêtre de corrélation sur la performance des mesures de corrélation ainsi que sur leurs coût de calcul.

Le choix approprié de la taille et de la forme de la fenêtre de corrélation est crucial pour garantir l'efficacité des méthodes locales. Dans les régions peu texturées, il est avantageux d'utiliser une fenêtre de grande taille pouvant contenir suffisamment d'éléments afin de rendre l'appariement fiable. Cependant, la taille doit être suffisamment petite au voisinage des discontinuités pour éviter les problèmes liés aux occultations et aux fortes variations de disparité. L'approche classique de mise en correspondance par corrélation consiste à définir un voisinage rectangulaire de taille fixe. Ce choix de taille fixe n'est pas très judicieux car il ne convient pas à toutes les zones de l'image. Le principe de méthodes plus performantes est de faire varier la taille et la forme des fenêtres de corrélation [Kanade, Okutomi, 1994, Veksler, 2001, Kang *et al.*, 2001]. Kanade et Okutomi [Kanade, Okutomi, 1994] adaptent itérativement la taille d'une fenêtre rectangulaire en fonction de la variation locale des niveaux de gris ainsi que des valeurs des disparités initialement estimées. Dans [Veksler, 2001], Veksler estime la taille et la forme optimales de la fenêtre de corrélation en effectuant un choix parmi une classe de fenêtres rectangulaires et non rectangulaires. Certains autres travaux utilisent plusieurs fenêtres de même taille placées à différents endroits [Fusiello *et al.*, 2000a, Kang *et al.*, 2001]. Par exemple, Fusiello et al. [Fusiello *et al.*, 2000a] utilisent neuf fenêtres pour un pixel. Un seul coût est alors calculé à partir de ces dernières en choisissant le minimum de tous les scores. Ces méthodes sont souvent coûteuses en temps de calcul mais Fusiello et al. ont montré que leur méthode est beaucoup plus rapide que la méthode de Kanade et Okutomi [Kanade, Okutomi, 1994]. Dans la méthode de Yoon et Kweon [Yoon, Kweon, 2005], au lieu de trouver la fenêtre de corrélation optimale, les auteurs ajustent le poids de chaque pixel dans la fenêtre en se basant sur une mesure de similarité (différence de couleur) et une mesure de proximité (distance spatiale) entre ce pixel et le pixel central. Cette méthode est robuste au choix de la taille de la fenêtre. Cependant, elle est coûteuse en temps de calcul et elle est très sensible au bruit.

2.4.2 Les approches progressives

Le principe des approches progressives, aussi appelées approches par croissance de germes, est de mettre en correspondance une partie des pixels dans les régions non ambiguës et de propager le résultat de ces pixels appariés sur les pixels voisins, de manière itérative, jusqu'à obtenir une mise en correspondance dense [Cochran, Médioni, 1992, Konrad, Lan, 2000, Zhang, Shan, 2000, Wei, Quan, 2004, Kim *et al.*, 2005a, Bleyer, Gelautz, 2004].

Dans [Konrad, Lan, 2000], les auteurs effectuent une mise en correspondance de points d'intérêt et sélectionnent les correspondances les plus fiables en utilisant la contrainte de symétrie. Ils identifient ensuite les blocs qui contiennent ces points appariés et leurs affectent la moyenne des disparités contenues dans chaque bloc. A partir de ces blocs, appelés blocs de contrôle, ils définissent un terme de régularisation qui mesure la similarité entre les disparités des blocs voisins. Pour estimer la disparité en chaque pixel, ils minimisent enfin une fonction de coût qui est la somme de ce terme de régularisation et d'un terme d'attache aux données.

Wei et Quan [Wei, Quan, 2004] déterminent les points de contrôle (Ground Control Points) en sélectionnant les correspondances les plus fiables. Ils effectuent ensuite une segmentation couleur de l'image en régions et à chaque région, ils attribuent une mesure de confiance qui dépend du nombre de points de contrôle dans la région. Pour chaque région ayant une confiance non nulle, une mesure d'ambiguïté qui dépend du nombre de disparités différentes dans la région est aussi calculée. Si cette ambiguïté est inférieure à un certain seuil, la région est considérée comme appariée, sinon, elle est divisée en deux régions et les deux régions sont considérées comme non appariées. La propagation s'effectue des régions texturées aux régions occultées et régions homogènes et s'arrête lorsque toutes les régions sont appariées.

L'avantage des approches progressives réside dans leur coût de calcul, puisqu'elles évitent généralement une optimisation globale coûteuse en temps de calcul. Cependant, elles propagent les erreurs d'appariement si un point de contrôle initial n'a pas été bien apparié. La méthode de Wei et Quan [Wei, Quan, 2004] permet de limiter cet inconvénient en mettant en correspondance des régions issues d'une segmentation couleur de l'image. Puisque l'information contenue dans les régions est plus riche que celle contenue dans des pixels individuels, la probabilité qu'un appariement initial soit erroné est réduite.

2.4.3 Les approches coopératives

Les méthodes coopératives utilisent à la fois les méthodes locales et les méthodes globales. Zitnick et Kanade [Zitnick, Kanade, 2000] ont proposé un algorithme coopératif qui suit l'approche de Marr et Poggio [Marr, Poggio, 1976]. Ils définissent un espace de disparité tridimensionnel (x, y, u) dont chaque élément correspond au pixel (x, y) de l'image de référence et toutes les disparités possibles u. L'algorithme est initialisé en calculant, pour tous les appariements possibles, un score de mise en correspondance initial grâce à une méthode de corrélation qui utilise une mesure de similarité (SAD ou NCC). Une fonction itérative qui s'appuie sur la contrainte d'unicité et la contrainte de continuité permet d'améliorer, de manière itérative, le score de mise en correspondance. Ce processus est répété jusqu'à ce que la contrainte d'unicité ne soit plus violée. Pour chaque pixel (x, y), l'élément (x, y, u) qui a le score le plus élevé est enfin sélectionné. L'algorithme de Zhang et Kambhamettu [Zhang, Kambhamettu, 2002] étend l'approche de Zitnick et Kanade en utilisant une segmentation en régions de l'image pour garantir que les appariements initiaux soient corrects.

2.4.4 Les approches globales

Une méthode de mise en correspondance est dite globale lorsque la fonction de coût est évaluée sur l'ensemble de l'image. Ce coût global a pour forme générale :

$$E(u) = E_1(u) + \alpha\ E_2(u). \tag{2.48}$$

où $E_1(u)$ est le terme d'attaches aux données, aussi appelé terme d'observation, $E_2(u)$ est un terme de régularisation et α est un paramètre réel positif pondérant l'importance relative accordée au deuxième terme vis-à-vis du premier.

Résoudre le problème de mise en correspondance, consiste donc à trouver la fonction de disparité u qui minimise ce coût global. Plusieurs méthodes d'optimisation ont été proposées

dans la littérature. Nous pouvons distinguer les méthodes à variables discrètes, qui effectuent un appariement de couples de pixels, comme la programmation dynamique ou les coupures de graphes, des méthodes à variables continues, où la disparité estimée est réelle, comme dans les approches variationnelles.

2.4.4.1 Programmation dynamique

La première méthode d'optimisation utilisée dans le cadre de la mise en correspondance et, également, celle qui est la plus populaire est la programmation dynamique [Sun, 2002, Zhang, Kambhamettu, 2002, Ishikawa, Geiger, 1998, Birchfield, Tomasi, 1998, Veksler, 2005]. Elle est appliquée entre deux séquences de pixels sur les lignes épipolaires conjuguées tout en respectant l'ordre des pixels à l'intérieur de chaque séquence. Les éléments des séquences à apparier définissent les deux dimensions d'une matrice qui contient les coûts des correspondances possibles entre les pixels. Le but est de trouver le chemin de coût optimal dans cette matrice à partir d'une mesure de similarité (SAD, SSD, etc.) entre deux éléments. Le coût d'un chemin est la somme des coûts de ses éléments. La figure 2.13 représente la matrice de recherche, où les axes sont définis par les lignes de recherche des images gauche et droite.

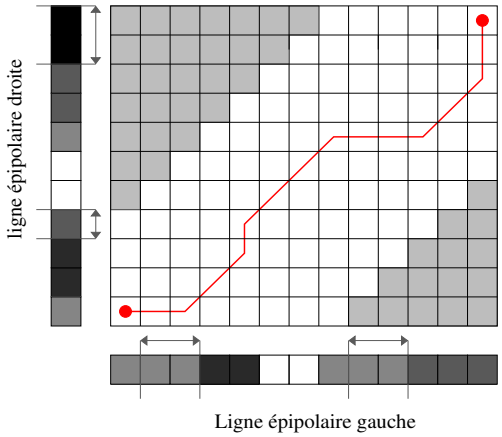


FIG. 2.13 – Programmation dynamique.

Cette méthode présente l'intérêt d'introduire un lissage le long d'une ligne épipolaire donnée. Par contre, son inconvénient majeur résulte de l'absence de lissage entre les lignes épipolaires voisines. De nombreuses déclinaisons de cet algorithme sont apparues pour prendre en compte les droites épipolaires adjacentes qui se situent respectivement au dessus et au dessous de la droite épipolaire considérée [Kim *et al.*, 2005a, Cox *et al.*, 1996, Ohta, Kanade, 1985, Rziza *et al.*, 2001].

2.4.4.2 Coupures de graphes

L'utilisation des coupures de graphes (*Graph cuts*) en vision par ordinateur a été introduite par Roy et Cox [Roy, Cox, 1998] pour s'attaquer au problème d'incohérence inter-lignes

épipolaires posé par la programmation dynamique. Elle a été ensuite formalisée par Veksler [Veksler, 1999] et depuis, plusieurs travaux qui s'appuient sur les techniques de coupures de graphes ont été proposées [Agrawal, Davis, 2004, Kang *et al.*, 2001, Hong, Chen, 2004, Bleyer, Gelautz, 2004, Bleyer, Gelautz, 2005, Kolmogorov, Zabih, 2005]. Ces approches formulent le problème de mise en correspondance comme un problème de recherche d'un flot maximum dans un graphe en minimisant une fonction de coût attachée à ce graphe. Ce flot maximum est atteint pour la coupure minimale du graphe.

Le problème de flot dans un graphe peut être vu comme un problème d'écoulement d'eau dans un réseau de tuyaux, c'est à dire qu'étant donné une source d'eau s et un puits p, on cherche le flot maximum qui peut passer dans le réseau de tuyaux reliant la source au puits. Pour symboliser ce réseau, les méthodes de coupures de graphes utilisent un graphe, noté $\mathcal{G}$, composé d'un ensemble de noeuds $\mathcal{N}$, qui correspondent aux connexions, et un ensemble d'arcs orientés $\mathcal{A}$ qui représentent les tuyaux.
Lorsqu'il s'agit d'utiliser ce type de méthode pour réaliser une mise en correspondance de pixels, la coupure de graphe est effectuée à partir d'un coût de mise en correspondance global. Les noeuds du graphe correspondent aux pixels et le coût associé à chaque arc correspond aux termes de la fonction de coût (2.48). On distingue donc deux types d'arcs : les arcs reliant un pixel (x, y) de l'image de gauche et un pixel $(x - u, y)$ de l'image de droite pour une disparité u avec pour attribut le premier terme de la fonction de coût et les arcs reliant les pixels voisins avec pour attribut le deuxième terme de la fonction d'énergie. La source et le puits correspondent respectivement aux pixels qui sont les projections des points les plus proches et les plus éloignés du plan image. La capacité d'un arc correspond au coût de mise en correspondance associé aux deux noeuds qu'il relie et le coût d'une coupure est la somme des capacités de ses arcs.

2.4.4.3 Approches variationnelles

Les approches variationnelles modélisant le problème de mise en correspondance par des équations aux dérivées partielles (EDP) ont été aussi largement utilisées pour minimiser la fonctionnelle (2.48). Introduites en traitement d'images pour résoudre quelques problèmes de base tels que la restauration ou la segmentation d'images [Aubert, Vese, 1997], les approches à base d'EDP ont été ensuite, largement employées par la communauté de vision par ordinateur pour l'estimation du mouvement dans une séquence d'images [Deriche *et al.*, 1996, Aubert *et al.*, 1999, Alvarez *et al.*, 2000b, Alvarez *et al.*, 2002]. Elles avaient déjà connu un véritable essor dans ce domaine avec la formulation variationnelle introduite par Horn et Shunck [Horn, Schunck, 1981] et qui s'appuie sur deux hypothèses : la conservation de la luminance et la cohérence spatiale du champ de mouvement. Cependant, la régularisation de *Tikhonov* [Tikhonov, Arsenin, 1977] utilisée par Horn et Shunck, faisant intervenir la norme quadratique, ne permet pas de bien préserver les discontinuités de la solution. Ceci a posé le problème du choix du terme de régularisation à utiliser afin d'assurer au mieux leur préservation. Ainsi, de nombreux auteurs se sont intéressés au développement de termes de régularisation permettant de lisser la solution recherchée dans les zones homogènes de l'image, tout en préservant les contours (cf. tableau 2.1). Les approches proposées ont été ensuite appliquées au domaine de la stéréovision pour l'estimation de la disparité [Alvarez *et al.*, 2000a, Robert, Deriche, 1996, Pons *et al.*, 2003, Slesareva *et al.*, 2005]. Un état de l'art sur l'utilisa-

Auteurs	$\Phi(t)$
Tikhonov [Tikhonov, Arsenin, 1977]	t^2
Rudin *et al.*[Rudin *et al.*, 1992]	t
Perona et Malik [Perona, Malik, 1990]	$-\frac{k^2}{2}e^{-(t/k)^2}$
Geman et Reynolds [Geman, Reynolds, 1992]	$\frac{(t/k)^2}{1+(t/k)^2}$
Charbonnier *et al.* [Charbonnier *et al.*, 1997]	$\sqrt{1+(t/k)^2}-1$

TAB. 2.1 – Quelques termes de régularisation utilisés en traitement d'images.

tion des EDP dans les domaines du traitement d'images et celui de la vision par ordinateur est présenté dans [Deriche, Faugeras, 1995].

Les approches variationnelles développées en vision par ordinateur pour l'estimation de la disparité considèrent la régularisation avec préservation des discontinuités comme un phénomène de diffusion, isotrope dans les zones de faible gradient de l'image et anisotrope au bord des discontinuités. Le terme de régularisation $E_2(u)$ dépend généralement de la norme du gradient de u, notée $|\nabla u|$, et s'exprime par :

$$E_2(u) = \Phi(|\nabla u|), \tag{2.49}$$

où $\Phi : \mathbb{R} \mapsto \mathbb{R}$ est une fonction de régularisation à définir. Dans le cas où cette fonction est quadratique ($\Phi(|\nabla u|) = |\nabla u|^2$), on retrouve la régularisation de *Tikhonov* qui pénalise les discontinuités et qui ne constitue donc pas un choix très pertinent. Plusieurs stratégies de recherche ont alors été développées afin de définir des conditions permettant de mieux choisir une telle fonction. Ces conditions sont données par l'équation d'Euler-Lagrange associée au terme de régularisation (2.49) :

$$\mathrm{div}\left(\frac{\Phi^{'}(|\nabla u|)}{|\nabla u|}\nabla u\right) = 0 \tag{2.50}$$

où div est l'opérateur classique de divergence et $\Phi^{'}$ représente la dérivée première de Φ supposée deux fois différentiable. Cette expression peut être développée et réécrite sous la forme suivante :

$$\Phi^{''}(|\nabla u|)u_{\xi\xi} + \frac{\Phi^{'}(|\nabla u|)}{|\nabla u|}u_{\eta\eta} = 0 \tag{2.51}$$

où $u_{\xi\xi}$ (resp. $u_{\eta\eta}$) représente la dérivée seconde directionnelle de u dans la direction du (resp. orthogonale au) gradient $\xi = \frac{\nabla u}{|\nabla u|}$.

Cette réécriture permet d'interpréter la fonction de régularisation comme étant liée à un processus de diffusion anisotrope opérant dans deux directions orthogonales : les directions de la normale et de la tangente aux courbes de niveau de u. L'anisotropie provient des pondérations différentes $\Phi^{''}(|\nabla u|)$ et $\frac{\Phi^{'}(|\nabla u|)}{|\nabla u|}$ accordées au processus de diffusion dans ces deux directions. Afin de régulariser la solution et préserver les discontinuités, il est souhaitable d'opérer de manière isotrope dans les régions homogènes et préserver les contours dans les régions à forts gradients où des discontinuités peuvent se trouver. Le lissage isotropique dans les zones homogènes conduit à imposer les conditions suivantes sur $\Phi(|\nabla u|)$ lorsque $|\nabla u|$ tend vers 0 :

$$\lim_{t\to 0^+} \frac{\Phi^{'}(t)}{t} = \lim_{t\to 0^+} \Phi^{''}(t) = \Phi^{''}(0) > 0, \tag{2.52}$$

ce qui implique $\Phi^{'}(0) = 0$. Par ailleurs, afin de préserver les discontinuités dans les zones à fort gradient, la condition de lissage anisotropique le long des directions perpendiculaires au gradient s'énonce comme :

$$\lim_{t\to\infty} \Phi^{''}(t) = 0 \tag{2.53}$$

$$\lim_{t\to\infty} \frac{\Phi^{'}(t)}{t} = \text{cste} > 0 \tag{2.54}$$

Mathématiquement, aucune fonction Φ ne peut satisfaire simultanément les conditions (2.53) et (2.54). En revanche, il est possible de privilégier une décroissance plus rapide du poids de lissage dans la direction du gradient que celui associé au lissage dans la direction perpendiculaire. Ainsi pour préserver les discontinuités, la fonction Φ doit respecter les conditions suivantes :

$$\Phi^{''}(0) > 0 \tag{2.55}$$

$$\lim_{t\to\infty} \Phi^{''}(t) = \lim_{t\to\infty} \frac{\Phi^{'}(t)}{t} = 0 \tag{2.56}$$

$$\lim_{t\to\infty} \frac{\Phi^{''}(t)}{\frac{\Phi^{'}(t)}{t}} = 0 \tag{2.57}$$

Plusieurs fonctions de régularisation satisfaisant les conditions (2.55)-(2.57) ont déjà été proposées dans la littérature, comme celle de Charbonnier *et al.* [Charbonnier *et al.*, 1997] (cf. tableau 2.1).

Un autre modèle de régularisation avec prise en compte des discontinuités a été proposé par Nagel et Enkelmann [Nagel, Enkelmann, 1986] pour l'estimation du flot optique. L'utilisation de ce modèle en stéréovision a été introduite par Alvarez et al. [Alvarez *et al.*, 2000a]. L'idée fondamentale consiste à régulariser l'image de disparité en accord avec les discontinuités de l'image de référence I_g au travers d'un terme de régularisation basé sur l'opérateur de Nagel et Enkelmann. Ce terme ne dépend plus uniquement de ∇u mais aussi des gradients spatiaux d'intensité de l'image de référence I_g. Il est donné par

$$\Phi(\nabla u, \nabla I_g) = (\nabla u)^\top D(\nabla I_g)(\nabla u), \tag{2.58}$$

où $D(\nabla I_g)$ est une matrice symétrique da taille 2×2 définie par :

$$D(\nabla I_g) = \frac{1}{|\nabla I_g|^2 + 2\nu^2} \left\{ \begin{bmatrix} \frac{\partial I_g}{\partial y} \\ -\frac{\partial I_g}{\partial x} \end{bmatrix} \begin{bmatrix} \frac{\partial I_g}{\partial y} \\ -\frac{\partial I_g}{\partial x} \end{bmatrix}^\top + \nu^2 I_{2,2} \right\}. \tag{2.59}$$

$I_{2,2}$ étant la matrice identité de taille 2×2 et ν une constante qui contrôle le poids de la diffusion anisotrope. Les vecteurs propres de cet opérateur sont ∇I_g et $(\nabla I_g)^\perp$ et les valeurs propres ν_1 et ν_2 associées sont :

$$\begin{cases} \nu_1 = \frac{\nu^2}{|\nabla I_g|^2 + 2\nu^2}, \\ \\ \nu_2 = \frac{|\nabla I_g|^2 + \nu^2}{|\nabla I_g|^2 + 2\nu^2}. \end{cases}$$

Lorsque le gradient $|\nabla I_g|$ tend vers zéro, nous obtenons $\nu_1 = \nu_2 = 1/2$. Dans ce cas correspondant aux zones homogènes, le comportement est isotrope car les directions ∇I_g et $(\nabla I_g)^\perp$ sont uniformément traitées. Cependant, lorsque le gradient $|\nabla I_g|$ tend vers l'infini, on a $\nu_1 = 0$ et $\nu_2 = 1$. Dans ce cas relatif aux zones comportant des frontières, la direction de diffusion privilégiée est alors celle de $(\nabla I_g)^\perp$, c'est à dire le long de la normale au gradient. La diffusion est donc anisotrope et tend à rendre ∇u colinéaire à ∇I_g.

L'opérateur de Nagel et Enkelmann D contrôle le lissage de u tout en autorisant ses discontinuités au niveau des variations importantes d'intensités de l'image, c'est à dire lorsque $|\nabla I_g| \gg \nu$. La diffusion de cet opérateur dépend donc du choix de paramètre ν. Si sa valeur est élevée, alors la diffusion est isotrope partout et se ramène à celle de Tikhonov et si ν est choisi faible, l'opérateur D opère de manière uniformément anisotrope. Alvarez et al. [Alvarez *et al.*, 2000a] proposent une adaptation de ce paramètre en fonction de la dynamique d'intensité présente dans l'image de référence. Ils ajustent ainsi la valeur de ν en spécifiant la valeur d'un autre paramètre z tel que

$$z = \int_0^\nu H_{|\nabla I_g|}(s) ds, \tag{2.60}$$

où $H_{|\nabla I_g|}(s)$ représente l'histogramme normalisé de $|\nabla I_g|$ et z est le rapport d'isotropie compris entre 0 et 1. Lorsque z est proche de zéro, l'opérateur de Nagel et Enkelmann conduit à une diffusion anisotrope dans toutes les zones de l'image et lorsque z approche la valeur 1, il se comporte de manière complètement isotrope.

Une fois fixée la fonction de régularisation qui remplit les conditions de préservation des discontinuités, la minimisation de la fonction d'énergie (2.48) se fait en déduisant l'équation d'Euler-Lagrange associée. Cette équation est ensuite résolue à l'aide d'un schéma itératif issu de la discrétisation de l'EDP. Dans [Alvarez *et al.*, 2000a], une approche multi-échelle a été mise en œuvre afin d'augmenter les performances de convergence du schéma proposé.

Les approches à base d'EDP permettent d'obtenir dans de nombreux cas des résultats d'existence et d'unicité pour la solution recherchée mais nécessitent la mise en place de schémas numériques qui peuvent se révéler lourds. En outre, la discrétisation des EDP est une opération délicate et le choix du paramètre α dans l'équation (2.48) pose souvent un problème majeur dans la méthode.

2.5 Conclusion

Nous avons présenté dans ce chapitre quelques notions très générales, fréquemment utilisées pour la mise en correspondance stéréoscopique. Après quelques considérations géométriques, nous avons classé les approches de mise en correspondance selon la méthode d'optimisation associée. Ainsi nous avons distingué quatre catégories : les approches locales, les approches progressives, les approches coopératives et les approches globales.

Les approches locales utilisent le voisinage d'un point dans une image pour trouver son correspondant dans l'autre image grâce à une mesure de similarité. Elles nécessitent la définition d'une fenêtre de taille et de forme appropriées et le choix d'un critère de ressemblance. Elles sont relativement efficaces, mais elles sont sensibles aux occultations, zones homogènes, etc.

Le principe des approches progressives est d'apparier une partie de pixels, appelés points de contrôle, dans les régions non ambiguës puis d'augmenter progressivement le nombre de pixels mis en correspondance en propageant le résultat des pixels appariés aux pixels voisins.

Il existe aussi des approches qui correspondent à une combinaison d'une méthode locale et d'une méthode globale, qui sont appelées approches coopératives.

Enfin, nous avons répertorié un très grand nombre d'articles utilisant les approches globales. La programmation dynamique est la technique la plus populaire. Elle est appliquée entre deux suites de pixels sur les droites épipolaires. Une autre approche pour aborder le problème de mise en correspondance de manière globale utilise la théorie des graphes. Contrairement à la programmation dynamique, cette approche permet d'inclure une cohérence bidimensionnelle par rapport à la scène observée. Parmi les méthodes permettant d'estimer une disparité à valeurs continues, nous avons mentionné les approches variationnelles qui modélisent le problème de mise en correspondance par des équations aux dérivées partielles pour minimiser une fonction d'énergie globale. Les efforts menés pour cette classe de méthodes ont principalement porté sur la définition de termes de régularisation avec préservation de discontinuités.

Dans ce mémoire, nous nous intéressons aux méthodes à variables continues car nous cherchons à estimer un champ de disparité le plus précis possible. Les approches variationnelles nous paraissent ainsi les plus adaptées à notre problème. Nous avons donc orienté notre choix vers ce type de méthodes pour développer une nouvelle approche de mise en correspondance dense, mais en adoptant plus une approche de type optimisation qu'EDP. Ceci fera l'objet du chapitre suivant.

Chapitre 3

Approche variationnelle convexe pour l'estimation de la disparité

3.1 Introduction

Dans ce chapitre, nous allons décrire l'approche que nous proposons pour l'estimation de la disparité à partir d'une paire d'images stéréoscopiques. Le but de notre travail étant d'obtenir un champ de disparité dense et précis, la méthode d'estimation considérée est basée sur une formulation variationnelle et fournit des valeurs de disparité variant continûment.

En s'appuyant sur des travaux récemment menés dans le domaine de l'optimisation, nous proposons une nouvelle formulation ensembliste du problème de l'estimation de la disparité. Dans un cadre variationnel, nous posons ainsi le problème comme un problème de programmation convexe consistant à minimiser une fonction objectif convexe sur l'intersection d'ensembles convexes. Ces ensembles sont associés à des contraintes convexes modélisant des informations *a priori* tenant compte des propriétés du champ à estimer et des données observées. La fonction objectif repose sur l'hypothèse de conservation de la luminance entre les deux projections d'un même point 3D de l'espace. Nous considérons dans ce chapitre une fonction objectif différentiable et convexe nous conduisant à un problème d'optimisation convexe qui serait relativement facile à résoudre s'il n'était de grande taille. Pour apporter une solution numérique à ce problème d'optimisation, nous utilisons un algorithme efficace et itératif par bloc, adapté à la minimisation de critères différentiables sous contraintes convexes. Cet algorithme permet de converger vers une solution optimale du problème, tout en permettant d'introduire des contraintes sur la solution avec une grande flexibilité.

Nous présentons, dans un premier paragraphe, notre formulation du problème d'estimation de la disparité. Nous rappelons ensuite quelques notions utiles d'optimisation convexe et nous donnons la forme générale de l'algorithme d'optimisation que nous considérons. Enfin, dans la dernière partie de ce chapitre, nous décrivons l'application de cet algorithme au cas de l'estimation de la disparité en détaillant la fonction objectif à minimiser et les contraintes convexes que nous introduisons sur la solution.

3.2 Formulation du problème

Dans le précédent chapitre, nous avons passé en revue les méthodes de mise en correspondance existantes et nous avons vu que ce problème ne peut être correctement résolu que par l'introduction de contraintes ou d'hypothèses supplémentaires sur la solution recherchée. La principale contrainte utilisée consiste à considérer que les projections d'un même point 3D de la scène ont la même valeur de luminance dans les deux images de la paire stéréoscopique.

3.2.1 Hypothèse de conservation de la luminance

Si nous représentons chaque point de l'image de gauche et de l'image de droite par l'intensité de la lumière (luminance) en ce point, la formulation mathématique de l'hypothèse de conservation de la luminance entre les deux projections d'un même point de l'espace s'écrit

$$I_g(s) = I_d(x - u(s), y), \tag{3.1}$$

où $s = (x, y)$ est la position spatiale d'un point dans l'image, I_g et I_d représentent respectivement l'intensité lumineuse de l'image de gauche et de droite, vues comme des fonctions différentiables, et $u(s)$ est la disparité associée au point s.

Il est à noter qu'en réalité, d'autres éléments entrent dans le modèle (3.1), tels que les distorsions géométriques et photométriques des images, la réflectance des surfaces, la radiosité, le bruit de mesure, etc. En effet, l'hypothèse de conservation de la luminance peut être remise en cause dans certains cas, notamment en présence de :

- zones de reflets dans la scène,
- zones d'occultations,
- différence de sensibilité entre les caméras utilisées,
- bruit provenant du capteur stéréoscopique.

Cependant, bien que cette hypothèse ait des limites de validité, la plupart des méthodes de mise en correspondance la considèrent suffisante et raisonnable.

3.2.2 Formulation : critère convexe

La ressemblance entre les deux projections d'un même point 3D de la scène, traduite par l'équation de conservation de la luminance (3.1), peut être mesurée grâce à une fonction de corrélation. La fonction la plus couramment employée utilisant la somme des différences d'intensités dans les deux images, s'exprime par

$$\tilde{J}(u) = \sum_{s \in \mathcal{D}} \varphi \left(I_g(s) - I_d(x - u(s), y) \right) \ , \tag{3.2}$$

où $\varphi : \mathbb{R} \mapsto \mathbb{R}$ est une mesure d'écart qui est convexe et différentiable et $\mathcal{D} \subset \mathbb{N}^2$ est le support de l'image. Notons ici que la méthode d'estimation de la disparité que nous proposons permet de calculer la valeur de la disparité $u(s)$ en tout point s de l'image de gauche, c'est pourquoi

la fonction d'erreur est évaluée sur l'ensemble de l'image. La valeur optimale de la disparité $\widehat{u}$ est celle pour laquelle cette fonction est minimum :

$$\widehat{u} = \arg\min_{u} \quad \tilde{J}\ (u)\ . \tag{3.3}$$

Cependant, la minimisation par des méthodes numériques de la corrélation (3.2) pose des problèmes de garantie de convergence vers l'optimum global du fait que cette fonctionnelle n'est pas convexe et peut donc admettre plusieurs minima locaux. Cette difficulté peut être contournée si nous supposons qu'une estimée initiale $\bar{u}$ de u est accessible. Cette dernière permet en effet d'approximer le critère (3.2) par un critère convexe en utilisant le développement en série de Taylor de $I_d(x-u,y)$ autour de $\bar{u}$ au premier ordre :

$$I_d(x-u,y) \simeq I_d(x-\bar{u},y) - (u-\bar{u})\ \ \nabla I_d^x(x-\bar{u},y) \tag{3.4}$$

où $\nabla I_d^x(x-\bar{u},y)$ est le gradient horizontal de $I_d(x-\bar{u},y)$. Afin de simplifier les notations, nous n'écrivons pas explicitement la dépendance en s de u et $\bar{u}$. L'estimée initiale $\bar{u}$ peut être obtenue à l'aide d'une méthode locale par corrélation ou générée par une procédure itérative en prenant pour $\bar{u}$ le résultat de l'itération précédente (cf. paragraphe 3.4.1.1).

Le critère à minimiser (3.2) peut maintenant être remplacé par :

$$J(u) = \sum_{s\in\mathcal{D}} \varphi\left(L(s)\ u(s) - r(s)\right)\ , \tag{3.5}$$

avec

$$L(s) = \nabla I_d^x(x-\bar{u}(s),y), \tag{3.6}$$

$$r(s) = I_d(x-\bar{u}(s),y) + \bar{u}(s)\ L(s) - I_g(s). \tag{3.7}$$

Estimer le meilleur champ de disparité u revient donc à résoudre un problème de minimisation globale de la fonction d'énergie convexe (3.5). Ce problème se formule comme un *problème inverse* consistant à estimer la disparité inconnue u à partir des champs d'observation L et r. La difficulté de la résolution de ce problème inverse réside dans le fait que L peut être nul pour certains points de l'image et que l'opérateur linéaire associé n'est donc pas inversible. Ceci signifie que le critère J n'est pas strictement convexe et qu'il n'existe pas de solution unique au problème d'optimisation.

Nous pouvons observer que la linéarisation (3.4) en vue d'obtenir un critère convexe de la forme (3.5) nous sera très utile par la suite pour pouvoir profiter de la fiabilité et de la robustesse des algorithmes d'optimisation convexe. En effet, il est bien connu que la minimisation efficace d'un critère non convexe nécessite la mise en œuvre d'algorithmes d'optimisation plus complexes et donc d'un coût de calcul bien plus élevé que la minimisation d'un critère convexe, sans offrir souvent de garantie de convergence globale.

3.2.3 Un problème mal posé

Le problème d'estimation de la disparité formulé comme un problème de minimisation de la fonctionnelle convexe (3.5) est un problème inverse. De nombreux autres problèmes en traitement d'images se formulent comme des problèmes inverses, où l'on doit reconstituer la forme

originelle d'une image à partir d'informations et d'observations *a priori* [Bertero *et al.*, 1988]. Il s'agit par exemple des problèmes de débruitage, de déconvolution, de restauration et de reconstruction d'images [Feraud, 2000]. La modélisation mathématique de ces problèmes permet de lier l'image inconnue u aux données disponibles L et r par une équation du type

$$r = Lu, \tag{3.8}$$

où u et r représentent respectivement les données originales et observées et où L est un opérateur linéaire représentant le phénomène de dégradation. L'équation (3.8) liant les observations à l'image recherchée est appelée *modèle d'observation*.

La résolution des problèmes inverses est rendue difficile par leur caractère *mal posé* en général. Au début du XX^e siècle, Hadamard a défini les trois conditions suivantes pour qu'un problème soit *bien posé* [Hadamard, 1902] :

(i) existence d'une solution,

(ii) unicité de la solution,

(iii) stabilité de la solution par rapport aux données.

Un problème qui ne satisfait pas une de ces trois conditions est dit *mal posé*.

Lorsque l'opérateur d'observation L est inversible, une méthode d'inversion directe qui consiste à estimer la solution par

$$\widehat{u} = L^{-1}r,$$

peut être envisagée. Cependant, l'inverse de l'opérateur L peut ne pas exister (ou il peut être mal conditionné), comme c'est le cas dans notre problème, et l'inversion directe se trouve donc inapplicable.

Une autre manière de résoudre le problème inverse est la régularisation au sens où Tikhonov [Tikhonov, Arsénine, 1976, Tikhonov, Arsenin, 1977] l'a introduite. Elle consiste à reformuler le problème d'inversion comme un problème de minimisation d'une fonctionnelle qui comprend deux termes : un premier terme de fidélité aux données qui approche la solution vers celle des moindres carrées et un second terme qui régularise cette solution :

$$\widehat{u} = \arg\min_{u} \{ \|Lu - r\|^2 + \alpha\, \Phi(u) \}, \tag{3.9}$$

où Φ est une fonction de pénalisation choisie sur la base de connaissances *a priori* sur la solution recherchée et α est le paramètre de régularisation. L'idée de base de la régularisation est donc de définir un critère qui permet de sélectionner une solution approchée parmi l'ensemble des solutions admissibles. Elle permet de transformer ainsi un problème mal posé en un problème bien posé, au sens de Hadamard.

En conclusion, la régularisation d'un problème mal posé joue un rôle primordial dans l'obtention d'une solution unique et ceci en exploitant les informations *a priori* sur les propriétés de la solution à estimer. Si les contraintes modélisant ces *a priori* peuvent être décrites comme des ensembles convexes fermés, une façon différente d'appréhender un problème mal posé consiste à rechercher l'image qui minimise un certain critère sous la contrainte

qu'elle appartienne à l'intersection de ces ensembles de contraintes [Trussell, Civanlar, 1984, Combettes, 1996].

Ce type d'approche est dénommé, dans la littérature sur l'optimisation sous contraintes, *approche ensembliste* et consiste à résoudre un problème *d'admissibilité* (ou encore de *faisabilité*) qui se formule ainsi :

minimiser une fonction objectif convexe sur l'intersection d'ensembles convexes construits à partir des connaissances a priori et des observations.

Notons que par formulation lagrangienne, il existe des équivalences entre cette approche et la résolution de problèmes d'optimisation non contraints tels que (3.9) [Ekeland, Temam, 1974].

Pour résoudre le problème d'estimation de la disparité, vu comme un problème inverse mal posé, nous avons mis en œuvre de nouvelles approches ensemblistes. Nous détaillerons dans ce qui suit le principe de cette approche et nous proposerons un algorithme qui permet de résoudre efficacement le problème d'admissibilité considéré, dans le cadre d'une optimisation convexe différentiable, en s'appuyant sur de puissants outils d'analyse convexe.

3.3 Optimisation d'une fonction convexe différentiable

3.3.1 Formulation ensembliste

Une approche ensembliste utilise un critère d'admissibilité plutôt qu'un critère d'optimalité en ce sens qu'une solution est dite admissible si elle satisfait à toutes les contraintes disponibles $(\Psi_i)_{i\in I}$ [Combettes, Pesquet, 1993, Combettes, 1996]. Ces contraintes proviennent des connaissances *a priori* sur les propriétés de la solution à estimer et des données observées.
Si l'on se place dans un espace Hilbertien réel $\mathcal{H}$ qui représente l'espace des champs 2D considérés, doté du produit scalaire $\langle . \mid . \rangle$, de la norme $\| \, . \, \|$ et de la distance d, chaque propriété connue de la solution restreint celle-ci à appartenir à un sous-ensemble de cet espace. En désignant par S_i le sous-ensemble de $\mathcal{H}$ vérifiant la contrainte Ψ_i, le problème d'admissibilité se met sous la forme

$$\text{trouver} \quad \widehat{u} \in S = \bigcap_{i\in I} S_i = \{u \in \mathcal{H} \mid (\forall\, i \in I)\ u \text{ satisfait à } \Psi_i\}. \tag{3.10}$$

Le principal intérêt de l'approche ensembliste est la simplicité avec laquelle il est possible d'introduire des informations *a priori* de natures très variées sur la solution désirée. Il suffit pour cela de définir des ensembles qui les incorporent, l'intersection de tous ces ensembles constituant l'ensemble des solutions admissibles. Considérons la figure 3.1 pour illustrer le principe de ces approches. L'espace des solutions admissibles est dénoté par S et les S_i sont des sous-ensembles de $\mathcal{H}$ qui traduisent des propriétés particulières sur la solution.

Dans l'approche ensembliste de type (3.10), toute image admissible qui appartient à l'ensemble S, et qui satisfait de ce fait toutes les contraintes qui sont imposées, est une solution acceptable. Dans certains problèmes, il peut être souhaitable de sélectionner une image admissible qui minimise une certaine fonction objectif J. Ce problème s'écrit :

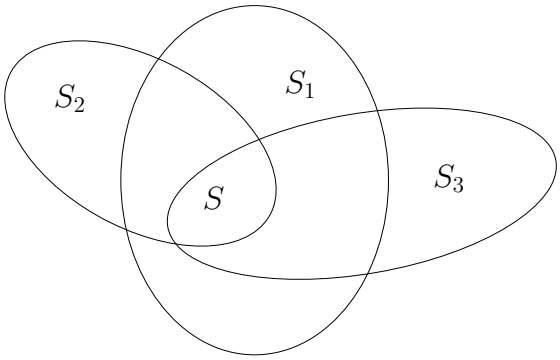


FIG. 3.1 – Le principe d'une approche ensembliste.

$$\text{trouver } \widehat{u} \in S = \bigcap_{i \in I} S_i, \quad \text{tel que } J(\widehat{u}) = \inf J(S). \tag{3.11}$$

Remarque 3.1 *Le problème* (3.10) *se ramène à celui de l'équation* (3.11) *en adoptant une fonction de coût J constante sur S.*

On a souvent intérêt à choisir l'objectif J et les ensembles de contraintes $(S_i)_{i \in I}$ convexes pour pouvoir disposer d'algorithmes d'optimisation efficaces. En effet, un problème d'optimisation avec ces caractéristiques possède des propriétés théoriques très agréables et peut être résolu efficacement, ce qui n'est, généralement, pas le cas pour des problèmes non convexes.

En prenant comme cadre l'optimisation convexe, le problème de programmation (3.11) implique trois étapes

1. Définir une fonction objectif convexe.

2. Définir des contraintes et des *a priori* qui peuvent être décrits comme des ensembles convexes fermés.

3. Résoudre le problème le plus souvent de manière numérique.

La définition du critère à minimiser et l'introduction de l'information *a priori* sur la solution sont deux étapes essentielles pour résoudre un tel problème d'optimisation. Ces deux éléments définissent la modélisation mathématique du problème physique et ont pour vocation de dépeindre la réalité de la manière la plus précise possible. Il est donc naturel, avant de résoudre le problème, de commencer par fixer le modèle mathématique. Les propriétés du critère et des contraintes choisis conditionnent par la suite le type de méthode employé pour résoudre le problème tout en répondant au mieux aux besoins du modèle. On en vient ainsi à la troisième étape du schéma de résolution décrit ci-dessus. Cette dernière étape qui consiste à résoudre le problème de programmation convexe implique naturellement d'effectuer le choix d'un algorithme d'optimisation.

Dans ce qui suit, après avoir rappelé quelques notions utiles d'analyse convexe, nous pré-

sentons l'algorithme que nous utilisons. Cet algorithme a été initialement proposé par Combettes [Combettes, 2000] pour résoudre des problèmes d'optimisation convexe de la forme (3.11) où l'objectif est de minimiser une fonction convexe sur l'intersection d'un grand nombre d'ensembles convexes. Nous présentons tout d'abord l'algorithme dans sa forme la plus générale. Nous décrivons ensuite une variante de cet algorithme qui a été adaptée pour la résolution des problèmes quadratiques convexes sous contraintes convexes.

3.3.2 Quelques rappels d'analyse convexe

L'analyse convexe fournit un cadre puissant permettant d'aborder de nombreux problèmes d'optimisation, notamment en traitement du signal et d'image. Dans cette partie, on se contentera de donner les définitions de base de l'analyse convexe (convexité, sous gradient, projection, etc). On se réfèrera à l'ouvrage de Rockafellar [Rockafellar, 1970] et à celui de Ekeland et Temam [Ekeland, Temam, 1974] pour plus d'informations.

Ensemble convexe et fonction convexe

Définition 3.1 *Un ensemble $C \subset \mathcal{H}$ est dit convexe si et seulement si*

$$\forall\ (u,v) \in C^2,\quad \forall\ \lambda \in [0,1],\quad \lambda u + (1-\lambda)v\ \in\ C. \tag{3.12}$$

Définition 3.2 *On dit qu'une fonction $f : \mathcal{H} \mapsto]-\infty,+\infty]$ est convexe si elle vérifie :*

$$\forall\ (u,v) \in \mathcal{H}^2,\ \forall\ \lambda \in [0,1],\quad f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda) f(v). \tag{3.13}$$

Le domaine effectif de la fonction f est l'ensemble

$$\text{dom}\ f = \{u \in \mathcal{H}\ \mid\ f(u) < +\infty\}. \tag{3.14}$$

La fonction f est dite strictement convexe si l'inégalité (3.13) est stricte pour tout $(u,v) \in (\text{dom}\ f)^2$ tel que $u \neq v$ et pour tout $\lambda \in]0,1[$.

Définition 3.3 *Soit une fonction $f : \mathcal{H} \mapsto]-\infty,+\infty]$, l'ensemble*

$$\text{epi}\ f = \{(u,\mu) \in \mathcal{H} \times \mathbb{R} \mid f(u) \leq \mu\},$$

est appelé l'épigraphe de f (noté epi f*).*

Théorème 3.1 *L'ensemble de niveau $\eta \in \mathbb{R}$ d'une fonction convexe $f : \mathcal{H} \mapsto]-\infty,+\infty]$ (noté* $\text{lev}_{\leq\eta} f$*), défini de la façon suivante*

$$\text{lev}_{\leq\eta} f = \{u \in \mathcal{H} \mid f(u) \leq \eta\}, \tag{3.15}$$

est fermé et convexe.

Définition 3.4 *Soit $f : \mathcal{H} \mapsto]-\infty, +\infty]$, on dit que*

(i) f est propre si dom $f \neq \emptyset$

(ii) f est coercive si $\lim_{\|u\|\to\infty} f(u) = +\infty$.

Propriétés :

- Une fonction $f : \mathcal{H} \mapsto]-\infty, +\infty]$ est convexe si et seulement si son épigraphe est convexe.
- Une fonction $f : \mathcal{H} \mapsto]-\infty, +\infty]$ est semi-continue inférieurement (s.c.i) si et seulement si son épigraphe est fermé.
- Toute fonction convexe propre sur un espace de dimension finie est continue sur l'intérieur de son domaine effectif.
- Si f est finie et s.c.i. alors elle est continue.

Sous-gradient et sous-différentiel

Définition 3.5 *Un vecteur $t \in \mathcal{H}$ est un sous-gradient de la fonction convexe f en un point $u \in \mathcal{H}$ si*

$$\forall\ v \in \mathcal{H}, \quad f(u) + \langle t \mid v - u \rangle \leq f(v), \tag{3.16}$$

Du point de vue géométrique, cette condition, que l'on nomme *inégalité du sous-gradient*, exprime le fait que le graphe de la fonction affine $f_{u,t}(v) = f(u) + \langle t \mid v - u \rangle$ est un *hyperplan d'appui* de l'ensemble convexe epi f au point $(u, f(u))$.

Il existe au moins un sous-gradient de f en tout point $u \in \mathcal{H}$ où f est continue. L'ensemble des sous-gradients de f en u est appelé *sous-différentiel* de f au point u et est noté $\partial f(u)$.

Théorème 3.2 *Pour tout $u \in \mathcal{H}$, le sous-différentiel $\partial f(u)$ d'une fonction convexe propre f est convexe et fermé. Il est borné si f est continue en u. De plus,*

$$\forall\ u \in \mathcal{H}, \quad f(u) = \inf_{v\in\mathcal{H}} f(v) \Leftrightarrow 0 \in \partial f(u). \tag{3.17}$$

Proposition 3.1 *Soient $f_1, \ldots, f_p$ des fonctions convexes de $\mathcal{H}$ dans $]-\infty, +\infty]$, on a*

$$\forall u \in \bigcap_{i=1}^{p} \text{dom } f_i, \quad \sum_{i=1}^{p} \partial f_i(u) \subset \partial \left(\sum_{i=1}^{p} f_i \right)(u). \tag{3.18}$$

Définition 3.6 *On dit qu'une fonction $f : \mathcal{H} \mapsto]-\infty, +\infty]$ est Gâteaux-différentiable en $u \in \text{dom } f$ si et seulement si la dérivée directionnelle*

$$f^{'}(u; v) = \lim_{t \to 0^+} \frac{f(u + tv) - f(u)}{t}, \tag{3.19}$$

existe dans toute direction v de $\mathcal{H}$ et si l'application

$$v \mapsto f^{'}(u; v) \tag{3.20}$$

est linéaire continue.

Théorème 3.3 *Si f est convexe et Gâteaux-différentiable en $u \in \text{dom } f$, alors il existe un unique sous-gradient appelé gradient de f en u et noté $\nabla f(u)$. Dans ce cas, $\partial f(u) = \{\nabla f(u)\}$.*

Minimum local et minimum global

Définition 3.7 *Un vecteur u est un minimum local d'une fonction $f : \mathcal{H} \mapsto]-\infty, +\infty]$ s'il existe $\epsilon > 0$ tel que*

$$f(u) \leq f(v), \quad \forall\ v \in \mathcal{H} \quad \text{tel que} \quad \|u - v\| < \epsilon. \tag{3.21}$$

Le vecteur u est un minimum global de f si

$$f(u) \leq f(v), \quad \forall\ v \in \mathcal{H}. \tag{3.22}$$

Lorsque les deux inégalités ci-dessus sont strictes pour $v \neq u$ alors le minimum local ou global est dit strict.

Pour qu'un point u soit un minimum local d'une fonction différentiable f, il est nécessaire que la condition suivante soit vérifiée

$$\nabla f(u) = 0. \tag{3.23}$$

Remarque 3.2 *Si f est convexe et différentiable en u alors la condition (3.23) est une condition nécessaire et suffisante assurant que le point u est un minimum local (cf. (3.17)). Par ailleurs, tous les minima locaux de f sont des minima globaux. Si de plus la fonction f est strictement convexe alors u est l'unique minimum global.*

Projection et projection sous-différentielle

Soit $C \in \mathcal{H}$ un ensemble convexe fermé et non vide.

Définition 3.8 *La distance d'un point $u \in \mathcal{H}$ à C est définie par*

$$d(u, C) = \inf\{\|u - v\| \ \mid \ v \in C\}. \tag{3.24}$$

Théorème 3.4 *Pour tout $u \in \mathcal{H}$, il existe un unique point $P_C(u) \in C$ tel que*

$$\|u - P_C(u)\| = d(u, P_C(u)) = d(u, C). \tag{3.25}$$

Ce point $P_C(u)$ est appelé projection de u sur C. Il est caractérisé par l'inégalité

$$\forall\ v \in C, \quad \langle P_C(u) - u \mid v - P_C(u) \rangle \geq 0. \tag{3.26}$$

Définition 3.9 *L'opérateur $P_C : \mathcal{H} \to C$ est appelé opérateur de projection sur C.*

Calculer la projection sur C revient à résoudre un problème de minimisation sous contrainte et nécessite généralement l'emploi d'une méthode itérative. Ce problème peut être dans certains cas difficile à résoudre. Toutefois, dans quelques cas particuliers, l'expression de la projection exacte peut être donnée explicitement. Considérons par exemple le demi-espace suivant :

$$Q = \{v \in \mathcal{H}, \mid \langle v \mid w \rangle \leq \eta\}, \quad \text{où} \quad w \in \mathcal{H} \setminus \{0\} \quad \text{et} \quad \eta \in \mathbb{R}. \tag{3.27}$$

La projection de u sur Q est donnée par [Combettes, 1997] :

$$P_Q u = \begin{cases} u + \dfrac{\eta - \langle u \mid w \rangle}{\|w\|^2} w, & \text{si} \quad \langle u \mid w \rangle > \eta, \\ u, & \text{si} \quad \langle u \mid w \rangle \leq \eta. \end{cases} \tag{3.28}$$

Remarque 3.3 *Si une approximation de C par un demi-espace Q peut être effectuée, la projection sur C devient facile à approximer. En effet, le calcul de la projection $P_Q u$ de u sur le demi-espace Q est direct et donc, en général, beaucoup plus économique que celui de la projection exacte $P_C u$.*

Nous décrivons à présent un contexte général dans lequel cette « projection » par approximations affines est possible. Supposons que $C = \text{lev}_{\leq 0} f \neq \emptyset$ où $f : \mathcal{H} \mapsto \mathbb{R}$ est continue et convexe. Soit $u \in \mathcal{H}$, d'après la remarque 3.3, une manière plus facile pour calculer la projection de u sur C est de l'approximer par un demi-espace Q_u fermé tel que

$$C \subset Q_u, \tag{3.29}$$

et

$$\text{si } u \notin C \text{ alors } u \notin Q_u. \tag{3.30}$$

La projection de $u \notin C$ sur C est donc tout simplement approximée par la projection de u sur l'hyperplan H_u délimitant Q_u et séparant u de C (cf. figure 3.2). Montrons à présent comment on peut définir une telle approximation pour $C = \text{lev}_{\leq 0} f$. Pour cela, considérons l'ensemble convexe fermé suivant :

$$Q_u = \{v \in \mathcal{H} \mid \langle t \mid u - v \rangle \geq f(u)\}, \tag{3.31}$$

où $t \in \partial f(u)$.

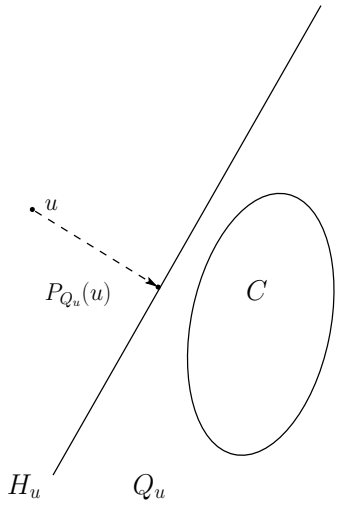


FIG. 3.2 – Projection sur l'hyperplan de séparation H_u délimitant Q_u.

Notons que

- $u \notin C \Leftrightarrow f(u) > 0$ et par conséquent $u \notin Q_u$;
- si $v \in C$ alors $f(v) \leq 0$ et il découle de (3.16) que, puisque $t \in \partial f(u)$, $\langle t \mid u - v \rangle \geq f(u)$, ce qui implique que v appartient à Q_u et donc $C \subset Q_u$;
- si $u \notin C$ alors $t \neq 0$ (car $f(u) > 0$), Q_u est donc un demi-espace.

Dans ce dernier cas, le demi-espace Q_u vérifie les conditions (3.29) et (3.30) et constitue donc une approximation valide de C.

La projection de u sur Q_u est appelée *projection sous-différentielle* de u sur C et est définie à partir de (3.28) par

$$P_{Q_u} u = \begin{cases} u - \dfrac{f(u)}{\|t\|^2} t, & \text{si } f(u) > 0; \\ u, & \text{si } f(u) \leq 0. \end{cases} \tag{3.32}$$

Le calcul de la projection sous-différentielle est plus facile que celui de la projection exacte car il nécessite simplement le calcul d'un sous-gradient t de f en u. Si f est différentiable, on prend $t = \nabla f(u)$.

Proposition 3.2 *Si la projection $P_C u$ de u sur C est facile à calculer, on peut prendre $f = d(., C)$ et obtenir ainsi $P_{Q_u} u = P_C u$.*

Démonstration.

(i) Si $u \in C$ alors $f(u) = 0$ et donc $P_{Q_u} u = P_C u = u$.

(ii) si $u \notin C$, on a

$$\partial f(u) = \nabla d(u, C) = \frac{u - P_C(u)}{\|u - P_C(u)\|}. \tag{3.33}$$

Or $d(u,C) > 0$ donc d'après (3.25) et (3.32), on obtient

$$\begin{aligned} P_{Q_u} u &= u - \frac{d(u,C)}{\|\nabla d(u,C)\|^2} \nabla d(u,C), \\ &= u - \frac{d(u,C)}{\|u - P_C(u)\|} (u - P_C(u)), \\ &= P_C(u). \end{aligned}$$

■

3.3.3 Méthode d'optimisation

On note $\Gamma_0(\mathcal{H})$ la classe des fonctions convexes propres et semi-continues inférieurement de $\mathcal{H}$ vers $]-\infty, +\infty]$. Le problème qu'on se propose de résoudre est le suivant :

$$\text{trouver } \ \widehat{u} \in S = \bigcap_{i=1}^{m} S_i, \quad \text{tel que } \ J(\widehat{u}) = \inf J(S), \tag{3.34}$$

en se plaçant dans le cadre d'un problème d'optimisation convexe où l'on suppose que le critère J à minimiser appartient à $\Gamma_0(\mathcal{H})$ et que les ensembles de contraintes $(S_i)_{1 \leq i \leq m}$ sont des convexes fermés de $\mathcal{H}$. On fait aussi l'hypothèse que le problème est consistant, ce qui signifie que l'ensemble S des solutions admissibles est non vide.

Pour résoudre le problème (3.34), un *algorithme itératif* doit généralement être mis en œuvre. A partir d'un point initial u_0, cet algorithme génère une suite d'itérées $(u_n)_{n \in \mathbb{N}}$ convergeant vers la (ou une) solution $\widehat{u}$ lorsque les conditions de convergence associées sont vérifiées. Plusieurs algorithmes itératifs peuvent être envisagés et il s'agit alors de choisir un algorithme itératif qui soit le mieux adapté possible au problème d'optimisation considéré.

Ici, nous nous intéressons aux algorithmes itératifs mettant en œuvre des projections sur des ensembles convexes. Ces algorithmes peuvent être regroupés en trois familles. La première famille se compose des algorithmes *séquentiels*, qui consistent à calculer une seule projection à chaque itération en alternant les projections périodiquement. La deuxième famille se compose des algorithmes *parallèles* qui utilisent des projections simultanées en invoquant tous les ensembles à chaque itération. Enfin la troisième famille se compose des algorithmes *par blocs* qui consistent à calculer des « blocs » de projections dont la structure peut varier au cours des itérations.

L'algorithme séquentiel le plus simple et le plus connu pour résoudre le problème d'admissibilité (3.34) dans le cas où J est constante sur S est l'algorithme de projection sur des ensembles convexes (POCS : *Projection Onto Convex Sets*) [Youla, Webb, 1982]. Il consiste à projeter itérativement sur chacun des ensembles S_i une image estimée initiale, puis de répéter le procédé jusqu'à l'obtention d'un point appartenant à l'intersection de tous les ensembles $(S_i)_{1 \leq i \leq m}$.

Algorithme 3.1 *Soit $u_0 \in \mathcal{H}$, l'algorithme POCS est décrit par les itérations*

$$\forall n \in \mathbb{N}, \quad u_{n+1} = u_n + \lambda_n (P_{i(n)}(u_n) - u_n), \tag{3.35}$$

où $i(n) = n(\text{modulo } m) + 1$, $P_{i(n)}$ représente l'opérateur de projection sur $S_{i(n)}$ et λ_n est un paramètre de relaxation à valeur dans $[\varepsilon, 2 - \varepsilon]$ avec $0 < \varepsilon < 1$.

La convergence de cet algorithme est assurée sous les hypothèses que les ensembles $(S_i)_{1 \leq i \leq m}$ sont fermés et convexes et que leur intersection S est non vide.

L'algorithme POCS a connu un grand succès dans le domaine de traitement d'images pour résoudre de nombreux problèmes tels que la restauration ou la reconstruction d'images. Il a été également utilisé par la communauté de vision par ordinateur pour l'estimation du flot optique [G. E. Mailloux, Bertrand, 1989, Simard, Mailloux, 1988, Simard, Mailloux, 1990]. Les raisons de l'intérêt porté à l'algorithme POCS sont sa simplicité de mise en œuvre et sa facilité à incorporer de l'information *a priori*. En effet, il est en général simple de définir des ensembles de contraintes modélisant les caractéristiques de la solution recherchée. Cet algorithme présente cependant plusieurs inconvénients :

- il est lent à converger.
- il est séquentiel et ne peut traiter qu'une seule contrainte par itération.
- il nécessite le calcul de projections exactes, ce qui est souvent coûteux numériquement.

Pour s'affranchir des limitations de POCS, des algorithmes améliorés ont été proposés. Ces améliorations ont surtout porté sur l'aspect séquentiel de l'algorithme. Ainsi la technique de reconstruction itérative simultanée SART (*Simultaneous Iterative Reconstruction Technique*) a été développée pour la reconstruction d'images en tomographie [Andersen, Kak, 1984]. Elle permet le calcul parallèle des projections. Le nouvel itéré est alors calculé en moyennant toutes les projections

$$\forall n \in \mathbb{N}, \quad u_{n+1} = \frac{1}{m} \sum_{i=1}^{m} P_i(u_n), \tag{3.36}$$

Bien que cet algorithme puisse donner de meilleurs résultats, sa convergence est lente également. D'autre part, cet algorithme parallèle nécessite que tous les ensembles de contraintes soient activés à chaque itération, ce qui, d'un point de vue numérique, peut être délicat si le nombre d'ensembles dépasse le nombre de processeurs dont on dispose. D'où l'intérêt des algorithmes par bloc où seulement des blocs de contraintes sont considérés à chaque itération.

L'algorithme que nous utilisons dans le cadre de ce travail est parallèle et itératif par bloc. Il peut être vu comme une extension des méthodes de projection sur des ensembles convexes et permet de résoudre le problème d'optimisation général (3.34). Il présente l'avantage de pouvoir gérer simultanément un grand nombre de contraintes convexes et grâce à sa structure itérative par bloc, il peut être efficacement mis en œuvre sur une architecture à processeurs parallèles. Cet algorithme offre une méthode de résolution puissante et efficace pour l'optimisation d'une fonction convexe et différentiable.

Nous allons à présent décrire l'algorithme. Il est tiré de [Combettes, 2000] et repose sur trois hypothèses fondamentales.

3.3.3.1 Hypothèses

1. $J \in \Gamma_0(\mathcal{H})$.

2. Il existe $z \in S$ tel que $J(z) < +\infty$ et $C = \mathrm{lev}_{\leq J(z)} J$ est borné. De plus, J est différentiable et strictement convexe sur C.

3. Les contraintes convexes sont modélisées sous la forme

$$(\forall i \in \{1, \ldots, m\}) \quad S_i = \{u \in \mathcal{H} \mid f_i(u) \leq \delta_i\}, \tag{3.37}$$

où $(f_i)_{1 \leq i \leq m}$ est une famille de fonctions convexes et continues de $\mathcal{H}$ dans $\mathbb{R}$ et $(\delta_i)_{1 \leq i \leq m} \in \mathbb{R}^m$ des paramètres réels.

Il est à noter que les hypothèses **1** et **2** sur le critère à minimiser J sont des conditions classiques et permettent d'assurer que le problème (3.34) admet une solution unique. L'hypothèse **3** fournit une description explicite des ensembles de contraintes $(S_i)_{1 \leq i \leq m}$ comme étant des ensembles de niveau des fonctions $(f_i)_{1 \leq i \leq m}$.

Proposition 3.3 *Supposons que $S \neq \emptyset$ et que $J : \mathcal{H} \to]-\infty, +\infty]$ est strictement convexe, différentiable et coercive alors les hypothèses* **1** *et* **2** *sont vérifiées.*

3.3.3.2 Algorithme

Avant de décrire l'algorithme, nous rappelons ici que si $t_{i,n}$ est un sous-gradient de f_i en $u_n \in \mathcal{H}$, alors la projection sous-différentielle de u_n sur $S_i = \mathrm{lev}_{\leq \delta_i} f_i$ est donnée par

$$p_{i,n} = \begin{cases} u_n + \dfrac{\delta_i - f_i(u_n)}{\|t_{i,n}\|^2} t_{i,n}, & \text{si } \; f_i(u_n) > \delta_i; \\ u_n, & \text{si } \; f_i(u_n) \leq \delta_i. \end{cases} \tag{3.38}$$

Par ailleurs, étant donné $I_n \subset \{1, \ldots, m\}$ et des poids convexes $(\omega_{i,n})_{i \in I_n}$ tels que $(\omega_{i,n})_{i \in I_n} \in]\varepsilon, 1]$ et $\sum_{i \in I_n} \omega_{i,n} = 1$, on définit

$$\lambda_n = \begin{cases} \dfrac{\sum_{i \in I_n} \omega_{i,n} \|p_{i,n} - u_n\|^2}{\left\|\sum_{i \in I_n} \omega_{i,n} p_{i,n} - u_n\right\|^2}, & \text{si } u_n \notin \bigcap_{i \in I_n} S_i; \\ 1 & \text{sinon.} \end{cases} \tag{3.39}$$

L'algorithme peut maintenant être écrit de la manière suivante :

Algorithme 3.2

Etape 0 Fixer $\varepsilon \in]0, 1/m[$. Choisir u_0 comme le minimiseur de J sur $\mathcal{H}$ et poser $n = 0$.

Etape 1 Prendre un ensemble d'indices non vide $I_n \subset \{1, \ldots, m\}$.

Etape 2 Poser $z_n = u_n + \lambda_n(\sum_{i \in I_n} \omega_{i,n} p_{i,n} - u_n)$, où :

a. $(\forall i \in I_n)$, $p_{i,n}$ est donné par (3.38).

b. $(\omega_{i,n})_{i \in I_n}$ varient dans $]\varepsilon, 1]$ et sont tels que $\sum_{i \in I_n} \omega_{i,n} = 1$.

c. λ_n est donné par (3.39).

Etape 3 Poser $\begin{cases} D_n = \{u \in \mathcal{H} \mid \langle u_n - u \mid \nabla J(u_n) \rangle \leq 0\} \\ H_n = \{u \in \mathcal{H} \mid \langle u - z_n \mid u_n - z_n \rangle \leq 0\}. \end{cases}$

Etape 4 Prendre pour u_{n+1} le minimiseur de J sur $D_n \bigcap H_n$.

Etape 5 Incrémenter n et aller à l'étape 1.

Théorème 3.5 [Combettes, 2000] *Supposons que*

$$(\exists K \in \mathbb{N}^*)(\forall n \in \mathbb{N}) \quad \bigcup_{l=n}^{n+K-1} I_l = \{1, \ldots, m\}, \tag{3.40}$$

alors la suite $(u_n)_{n \in \mathbb{N}}$ générée par l'algorithme 3.2 converge vers l'unique solution de (3.34) sous les hypothèses 3.3.3.1.

Remarques 3.1

- *La suite $(I_n)_{n \in \mathbb{N}}$ définit les blocs d'ensembles balayés à chaque itération. La règle (3.40) impose que chaque ensemble S_i soit utilisé au moins une fois au cours de K itérations consécutives. L'algorithme 3.2 peut être donc efficacement mis en œuvre sur une architecture à processeurs parallèles en adaptant la charge de calcul à chaque itération au nombre de processeurs parallèles disponibles.*

- *Si m processeurs parallèles sont disponibles, on peut activer toutes les contraintes à chaque itération en prenant $I_n = \{1, \ldots, m\}$. Cependant, si l'on dispose d'un seul processeur, on peut alterner les projections périodiquement. On obtient ainsi un algorithme séquentiel de type POCS où $(\forall n \in \mathbb{N})$ card $I_n = 1$.*

- *L'étape clé de l'algorithme 3.2 est l'étape 4 où l'itérée courante est obtenue comme le minimiseur de J sur l'intersection de deux demi-espaces D_n et H_n, qui inclut S. Notons que si l'on ignore l'étape 3 et on remplace l'étape 4 par $u_{n+1} = z_n$, on se ramènera à l'algorithme proposé dans* [Combettes, 1997] *qui produit une image admissible quelconque dans S.*

3.3.3.3 Minimisation d'un critère quadratique

L'algorithme d'optimisation que nous venons de décrire a été développé pour la minimisation de critères différentiables sous des contraintes convexes. Cet algorithme a ensuite été adapté pour la minimisation de critères quadratiques donnant naissance à une nouvelle variante de l'algorithme plus simple à implémenter [Combettes, 2003]. Le problème quadratique considéré se formule comme suit

$$\begin{aligned} &\text{Trouver} \quad \widehat{u} \in S = \bigcap_{i=1}^{m} S_i, \quad \text{tel que} \quad J(\widehat{u}) = \inf J(S), \\ &\text{où} \qquad J : u \mapsto \langle u - u_0 \;\mid\; R(u - u_0) \rangle, \quad u_0 \in \mathcal{H}, \end{aligned} \tag{3.41}$$

et R est un opérateur borné, auto-adjoint, défini positif.

Il s'agit ici d'un problème sensiblement plus simple que le problème généralisé (3.34) où J est un critère différentiable quelconque. En effet, lorsqu'une fonction objectif quadratique est adoptée, l'étape 4 de l'algorithme 3.2, qui consiste à minimiser J sur l'intersection de deux demi-espaces peut s'exprimer d'une manière analytique simple, comme expliqué ci-après.

Dans ce cas spécifique, le demi-espace D_n s'écrit :

$$\begin{aligned} D_n &= \{u \in \mathcal{H} \mid \langle u_n - u \mid \nabla J(u_n) \rangle \leq 0\}, \\ &= \{u \in \mathcal{H} \mid \langle R(u - u_n) \mid u_0 - u_n \rangle \leq 0\}, \end{aligned} \tag{3.42}$$

l'algorithme est initialisé avec u_0 et l'expression du nouvel itéré u_{n+1}, minimiseur de J sur $D_n \bigcap H_n$, est ainsi donnée explicitement grâce au lemme suivant.

Lemme 3.1 [Haugazeau, 1968] *Soit $(u, v, w) \in \mathcal{H}^3$, on note $A = \{h \in \mathcal{H} \mid \langle h - v \mid u - v \rangle \leq 0\} \bigcap \{h \in \mathcal{H} \mid \langle h - w \mid v - w \rangle \leq 0\}$ et $Q(u, v, w)$ la projection de u sur A.*
On définit $\pi = \langle u - v | v - w \rangle$, $\mu = \|u - v\|^2$, $\nu = \|v - w\|^2$, et $\varrho = \mu\nu - \pi^2$, alors

$$Q(u, v, w) = \begin{cases} w, & \text{si } \varrho = 0 \text{ et } \pi \geq 0; \\ u + (1 + \frac{\pi}{\nu})(w - v), & \text{si } \varrho > 0 \text{ et } \pi\nu \geq \varrho; \\ v + \frac{\nu}{\varrho}(\pi(u - v) + \mu(w - v)), & \text{si } \varrho > 0 \text{ et } \pi\nu < \varrho. \end{cases}$$

Nous donnerons à présent une description de l'algorithme proposé par Combettes pour résoudre le problème quadratique (3.41) [Combettes, 2003]. Cet algorithme, qui est une variante de l'algorithme 3.2, repose sur les mêmes hypothèses 3.3.3.1.

Algorithme 3.3

Etape 0 Fixer $\varepsilon \in]0, 1/m[$ et poser $n = 0$.

Etape 1 Prendre un ensemble d'indices non vide $I_n \subset \{1, \ldots, m\}$.

Etape 2 $\forall i \in I_n$, poser $a_{i,n} = -f_i(u_n)t_{i,n}/\|t_{i,n}\|^2$, où $t_{i,n} \in \partial f_i(u_n)$ si $f_i(u_n) > 0$; $a_{i,n} = 0$ sinon.

Etape 3 Choisir $(\omega_{i,n})_{i\in I_n} \subset]\varepsilon, 1]$ tels que $\sum_{i\in I_n} \omega_{i,n} = 1$. Prendre $v_n = \sum_{i\in I_n} w_{i,n}a_{i,n}$ et $L_n = \sum_{i\in I_n} w_{i,n}\|a_{i,n}\|^2$.

Etape 4 Si $L_n = 0$, poser $u_{n+1} = u_n$ et sortir. Sinon, poser $b_n = u_0 - u_n$, $c_n = Rb_n$, $d_n = R^{-1}v_n$, et $\widetilde{L}_n = L_n/\langle d_n \mid v_n\rangle$.

Etape 5 Choisir $\lambda_n \in [\varepsilon\widetilde{L}_n, \widetilde{L}_n]$ et poser $d_n = \lambda_n d_n$.

Etape 6 Poser $\pi_n = -\langle c_n \mid d_n\rangle$, $\mu_n = \langle b_n \mid c_n\rangle$, $\nu_n = \lambda\langle d_n \mid v_n\rangle$, et $\varrho_n = \mu_n\nu_n - \pi_n^2$.

Etape 7 Poser

$$u_{n+1} = \begin{cases} u_n + d_n, & \text{si } \varrho_n = 0 \text{ et } \pi_n \geq 0; \\ u_0 + (1 + \frac{\pi_n}{\nu_n})d_n, & \text{si } \varrho_n > 0 \text{ et } \pi_n\nu_n \geq \varrho_n; \\ u_n + \frac{\nu_n}{\varrho_n}(\pi_n b_n + \mu_n d_n), & \text{si } \varrho_n > 0 \text{ et } \pi_n\nu_n < \varrho_n. \end{cases}$$

Etape 8 Incrémenter n et aller à l'étape 1.

Théorème 3.6 *Supposons que* (3.40) *soit vérifiée, alors la suite engendrée par l'algorithme 3.3 converge vers l'unique solution du problème* (3.41) *sous les hypothèses 3.3.3.1.*

Dans la littérature de programmation quadratique, de nombreuses méthodes ont été proposées pour résoudre le problème (3.41). Parmi les méthodes par projection sur des convexes fermés, on peut citer à titre d'exemple la méthode de projections alternées due à *Boyle et Dykstra* [Boyle, Dykstra, 1986] et la méthode du *point d'ancrage* [Yamada *et al.*, 1998]. Une étude comparative des performances numériques de ces méthodes et de celles de l'algorithme 3.3 est établie dans [Combettes, 2003]. Cette étude montre que dans le cas où il y a un grand nombre de contraintes ou les projections sont difficiles à calculer, l'algorithme 3.3 est plus avantageux car il n'utilise que des projections approchées (sous-différentielles). D'autre part, cet algorithme permet de traiter les contraintes par blocs en parallèle et offre donc plus de flexibilité en terme d'implémentation parallèle. Enfin, il a été observé que sa convergence est beaucoup plus rapide que celle des algorithmes de *Dykstra* et du *point d'ancrage*.

3.4 Application à l'estimation de la disparité

L'approche ensembliste consistant à résoudre un problème de programmation convexe de la forme (3.34) s'est déjà révélée efficace dans diverses applications en traitement du signal et d'image [Trussell, Civanlar, 1984, Combettes, 1996]. Des résultats récents en reconstruction tomographique [Capricelli, Combettes, 2005], en restauration [Combettes, Pesquet, 2004a] et en débruitage d'images [Combettes, Pesquet, 2003] illustrent les potentialités de cette approche, et en particulier celles de l'algorithme itératif 3.2 qui atteint de très bonnes performances à la fois en terme de qualité du résultat fourni mais aussi en terme de convergence.

Ces exemples variés prouvent aussi qu'un grand nombre de contraintes couramment utilisées en traitement d'images (positivité, amplitude, énergie finie, lissage, etc) s'expriment à partir de sous-ensembles convexes et fermés de $\mathcal{H}$.

Dans cette partie, nous allons décrire une nouvelle application de l'approche ensembliste pour résoudre le problème de l'estimation de la disparité à partir d'une paire d'images stéréoscopiques. Ce problème déjà formulé comme un problème inverse au paragraphe 3.2.2, sera donc posé comme un problème de programmation convexe de la forme (3.34) et résolu par l'algorithme 3.2 bien adapté à ce type de problèmes.

Pour cela, nous définissons dans ce qui suit la fonction objectif globale à minimiser et nous introduisons les ensembles de contraintes convexes que nous proposons dans le cadre spécifique de notre application.

3.4.1 Fonction objectif globale

La formulation mathématique du problème de mise en correspondance nous a conduit au paragraphe 3.2 à chercher l'image de disparité u comme la solution du problème de minimisation de la fonction objectif suivante :

$$J_{\mathcal{D}}(u) = \sum_{s \in \mathcal{D}} \varphi(L(s)\ u(s) - r(s)), \tag{3.43}$$

où $\mathcal{D}$ est le support de l'image et L et r sont donnés par les équations (3.6) et (3.7). La fonction φ étant supposée convexe et différentiable, le problème qu'on se propose de résoudre est donc un problème d'optimisation convexe différentiable.

3.4.1.1 Calcul d'une estimée initiale

On rappelle que la linéarisation (3.4) en vue d'obtenir un critère convexe suppose qu'une estimée initiale $\bar{u}$ de la disparité u est accessible. Une étape préliminaire de calcul d'un champ de disparité initial est donc nécessaire. Cette étape permet d'attribuer une valeur initiale cohérente de la disparité et de s'assurer ainsi que le développement limité effectué autour de $\bar{u}$ est valide. Une manière de procéder courante pour l'initialisation est d'utiliser une pyramide d'images à différents niveaux de résolution. Au niveau de résolution le plus grossier, l'initialisation est fixée par un champ nul et aux niveaux de résolution suivants, le processus d'estimation est initialisé par le champ estimé au niveau précédent. Pour fournir un champ de disparité initial proche de la solution, Oisel *et al.* [Oisel *et al.*, 2003] proposent, en alternative, une initialisation par interpolation de points d'intérêt préalablement appariés.

Dans ce travail, nous avons utilisé comme estimée initiale celle donnée par la méthode de mise en correspondance par corrélation décrite au paragraphe 2.4.1. Nous avons choisi cette approche pour sa simplicité de mise en œuvre et son temps de calcul relativement faible. Par ailleurs, l'estimée initiale sera affinée au fur et à mesure des estimations en prenant le résultat de l'estimation précédente comme initialisation à l'étape suivante. Cette stratégie améliore considérablement la qualité de la solution recherchée et diminue sa dépendance vis-à-vis de l'initialisation.

3.4.1.2 Gestion des occultations

La fonction objectif (3.43) repose sur l'hypothèse de conservation de la luminance entre les deux projections d'un même point de l'espace. Dans les zones d'occultations où un point n'est visible que dans l'une des deux images (cf. paragraphe 2.3.1), ce critère de ressemblance est mis en défaut. Pour limiter l'influence de ces pixels occultés dont le coût de correspondance peut être important, nous les avons exclus lors du processus de minimisation.

En notant $\mathcal{O}$ l'ensemble de points de l'image occultés, on limite ainsi la sommation de l'équation (3.43) aux pixels visibles :

$$J_{\mathcal{D}\setminus\mathcal{O}}(u) = \sum_{s\in\mathcal{D}\setminus\mathcal{O}} \varphi(L(s)\ u(s) - r(s)), \tag{3.44}$$

Pour la détection des zones des occultations $\mathcal{O}$, nous avons exploité la contrainte d'unicité et la contrainte d'ordre : ainsi tout pixel ne respectant pas ces deux contraintes est considéré comme occulté.

Remarque 3.4 *Il est à noter que bien que les zones des occultations aient été éliminées du critère (3.44) à minimiser, elles seront prises en compte dans la formulation des contraintes de régularisation avec préservation des discontinuités, qui seront détaillées ci-après.*

3.4.1.3 Fonction objectif strictement convexe

Étant donné que les composantes de l'opérateur L peuvent s'annuler en certains points de l'image et que le champ d'occultation $\mathcal{O}$ est toujours non vide, le critère (3.44) est convexe mais pas strictement. Pour se ramener à la stricte convexité en accord avec les hypothèses 3.3.3.1 de l'algorithme que nous utilisons, on introduit un terme additif strictement convexe dans la fonction objectif globale à minimiser :

$$J(u) = \sum_{s\in\mathcal{D}\setminus\mathcal{O}} \varphi(L(s)\ u(s) - r(s)) + \alpha \sum_{s\in\mathcal{D}} \varphi(u(s) - \bar{u}(s)), \tag{3.45}$$

où φ est supposée strictement convexe et le paramètre $\alpha > 0$, choisi d'une façon empirique, pondère l'influence du deuxième terme par rapport au premier.

A présent, nous allons décrire et modéliser les contraintes traduisant les propriétés connues *a priori* sur le champ de disparité à estimer comme des sous-ensembles convexes de la forme (3.37), tout en respectant ainsi l'hypothèse **3** de l'algorithme que nous utilisons. Pour chaque sous-ensemble convexe ainsi défini, nous établissons une expression de la projection sous-différentielle sur cet ensemble.

En considérant des images discrètes, comme c'est le cas en pratique, on peut représenter une image par une matrice réelle à deux dimensions, de taille $N_\mathrm{l} \times N_\mathrm{c}$, et dont les éléments donnent les niveaux de gris de l'image. L'espace des images sera donc désormais l'espace euclidien $\mathcal{H} = \mathbb{R}^{N_\mathrm{l}\times N_\mathrm{c}}$.

3.4.2 Contrainte de plage de valeurs

Une interprétation de la contrainte épipolaire et de la contrainte de disparité positive décrites au paragraphe 2.3.3, permet de conclure que la fonction de disparité u est bornée et comprise entre une valeur minimale $u_{\min} \geq 0$ et une valeur maximale $u_{\max}$.

Cette information sur l'espace de valeurs permet donc de placer la solution u sur l'ensemble convexe suivant :

$$S_1 = \{u \in \mathcal{H} \mid u_{\min} \leq u \leq u_{\max}\}. \tag{3.46}$$

Notons que la projection exacte sur l'ensemble S_1 ci-dessus peut être calculée aisément, d'une manière explicite.

3.4.3 Régularisation par variation totale

Dans la plupart des applications de stéréovision, le champ de disparité u qu'on cherche à retrouver a vocation à être lisse dans les zones homogènes et discontinu aux frontières des objets. D'un point de vue théorique, ceci peut se modéliser en supposant que la disparité u appartient à l'espace BV des fonctions à variation bornée. L'intérêt de cet espace, souvent utilisé dans les domaines du traitement d'images et celui de la vision par ordinateur, est d'autoriser les discontinuités le long des contours de l'image.

Nous donnerons dans ce qui suit une formulation de cette information *a priori* qui considère l'image de disparité u à estimer comme étant constituée de différentes zones homogènes séparées par de franches discontinuités.

On rappelle ici la définition de l'espace BV et on renvoie le lecteur à [Evans, Gariepy, 1992, Giusti, 1984, Ambrosio *et al.*, 2000] pour plus de détails. On suppose que Ω, le domaine de l'image, est un ouvert borné de $\mathbb{R}^2$ et on appelle $C_0^1(\Omega)$ l'ensemble des fonctions continûment différentiables à support compact dans Ω.

Définition 3.10 *L'espace des fonctions à variations bornées (noté $BV(\Omega)$) est défini par :*

$$BV(\Omega) = \{u \in L^1(\Omega) \mid \sup \int_\Omega u(s)\ \mathrm{div}\psi(s)ds < \infty,\ \psi \in C_0^1(\Omega) \text{ et } \sup_{s\in\Omega}|\psi(s)| \leq 1\}, \tag{3.47}$$

où div *est l'opérateur de divergence.*

Définition 3.11 *On appelle variation totale de u sur Ω et on note* $\mathrm{tv}(u)$ *la mesure suivante*

$$\mathrm{tv}(u) = \sup\{\int_\Omega u(s)\ \mathrm{div}\ \psi(s)ds \mid \psi \in C_0^1(\Omega) \text{ et } \sup_{s\in\Omega}|\psi(s)| \leq 1\}, \tag{3.48}$$

Propriétés : On a :

1.
$$BV(\Omega) \subset L^2(\Omega). \tag{3.49}$$

2.
$$BV(\Omega) \subset L^1(\Omega). \tag{3.50}$$

3. si u appartient à l'espace de Sobolev $W^{1,1}(\Omega) = \{u \in L^1(\Omega) \mid \nabla u \in L^1(\Omega)^2\}$, alors

$$\mathrm{tv}(u) = \int_\Omega |\nabla u(s)|_2 \, ds, \tag{3.51}$$

où $|\ .\ |_2$ désigne la norme euclidienne de $\mathbb{R}^2$.

Le succès de l'espace BV en traitement d'images est étroitement lié à celui du modèle de régularisation proposé par Rudin, Osher et Fatemi [Rudin *et al.*, 1992] pour la restauration d'images. La particularité de ce modèle est de permettre un lissage de l'image tout en préservant ses discontinuités. En effet, pour pallier aux inconvénients de la régularisation quadratique introduite par Tikhonov, les auteurs ont proposé d'utiliser une régularisation fondée sur la variation totale en formulant le problème inverse initial comme un problème de minimisation de la variation totale sous contraintes. L'avantage de cette régularisation est qu'une fonction à variation bornée peut présenter des discontinuités au travers de courbes, et donc, représenter les objets avec des bords nets.

Ces dernières années, la minimisation de la variation totale pour des fins de régularisation a été appliquée pour résoudre divers problèmes inverses mal posés et s'est particulièrement imposée comme une bonne approche en traitement d'images [Dobson, Vogel, 1997, Malgouyres, 2002, Osher *et al.*, 2003, Durand, Froment, 2003, Dey *et al.*, 2004], et plus récemment en vision par ordinateur [Aubert *et al.*, 1999, Slesareva *et al.*, 2005].

Ainsi, deux modèles variationnels de régularisation par variation totale sont généralement utilisés :

- le premier modèle repose sur la minimisation de la variation totale sous contraintes, et se formule comme suit :

$$\text{trouver } u \text{ minimisant } \ \mathrm{tv}(u) \quad \text{tel que} \quad \|Lu - r\|^2 \leq \sigma, \quad \sigma > 0. \tag{3.52}$$

- le second modèle se traduit par l'ajout d'un terme de pénalisation basé sur la variation totale au terme d'attache aux données du critère à minimiser. Le problème devient :

$$\text{trouver } u \text{ minimisant } \ \mathrm{tv}(u) + \alpha\|Lu - r\|^2, \quad \alpha > 0. \tag{3.53}$$

Récemment, Combettes et Pesquet [Combettes, Pesquet, 2003, Combettes, Pesquet, 2004a] ont proposé une nouvelle formulation pour prendre en compte la variation totale en restauration d'images. L'idée est d'utiliser la variation totale non plus comme un critère à minimiser mais comme une contrainte dans le cadre d'un problème de faisabilité convexe. Cette idée est motivée par l'observation que, dans de nombreux types de problèmes, la variation totale de l'image originale, qui est une mesure de la somme des amplitudes des discontinuités, ne dépasse pas une certaine borne connue τ. Cette information restreint la solution à appartenir à l'ensemble convexe suivant :

$$S_2 = \{u \in \mathcal{H} \ \mid \ \mathrm{tv}(u) \leq \tau\}. \tag{3.54}$$

En pratique, la borne τ peut être estimée à partir d'expérimentation ou en exploitant des bases de données d'images du même type. Dans [Combettes, Pesquet, 2004a], les auteurs

soulignent la robustesse de la contrainte (3.54) vis-à-vis du choix de la borne τ sur la variation totale pour une application en restauration d'images. Dans le chapitre suivant, nous présentons, au travers d'expérimentations, l'influence du choix de τ sur le calcul de la solution dans le cadre de notre application.

3.4.3.1 Discrétisation de la variation totale

Pour définir une variation totale discrète, nous introduisons d'abord une version discrète de l'opérateur gradient. Si $u \in \mathcal{H}$, le gradient de $\widehat{\nabla} u$ est un vecteur de $\mathcal{H} \times \mathcal{H}$ donné par :

$$\widehat{\nabla} u = \big((\widehat{\nabla} u)^1, (\widehat{\nabla} u)^2\big)^\top, \tag{3.55}$$

avec

$$\begin{aligned} (\widehat{\nabla} u)^1(x,y) &= \begin{cases} u(x+1,y) - u(x,y) & \text{si } \ 1 \leq x < N_\mathrm{l}, \\ 0 & \text{si } \ x = N_\mathrm{l}, \end{cases} \\ &\text{et} \\ (\widehat{\nabla} u)^2(x,y) &= \begin{cases} u(x,y+1) - u(x,y) & \text{si } \ 1 \leq y < N_\mathrm{c}, \\ 0 & \text{si } \ y = N_\mathrm{c}. \end{cases} \end{aligned} \tag{3.56}$$

La variation totale discrète de u est alors donnée par :

$$\mathrm{tv}(u) = \sum_{x=1}^{N_\mathrm{l}} \sum_{y=1}^{N_\mathrm{c}} |\widehat{\nabla} u(x,y)|_2, \tag{3.57}$$

En prenant en compte les effets de bords, la définition de la variation totale discrète de u devient :

$$\begin{aligned} \mathrm{tv}(u) = &\sum_{x=1}^{N_\mathrm{l}-1} \sum_{y=1}^{N_\mathrm{c}-1} \sqrt{|(\widehat{\nabla} u)^1(x,y)|^2 + |(\widehat{\nabla} u)^2(x,y)|^2} \\ &+ \sum_{x=1}^{N_\mathrm{l}-1} |(\widehat{\nabla} u)^1(x,N_\mathrm{c})| + \sum_{y=1}^{N_\mathrm{c}-1} |(\widehat{\nabla} u)^2(N_\mathrm{l},y)|. \end{aligned} \tag{3.58}$$

Si l'on note :

$$\begin{cases} \Gamma_{x,y} : u \mapsto \Big(\ u(x+1,y) - u(x,y), u(x,y+1) - u(x,y) \ \Big), \\ \Gamma_{x,N_\mathrm{c}} : u \mapsto u(x+1,N_\mathrm{c}) - u(x,N_\mathrm{c}), \\ \Gamma_{N_\mathrm{l},y} : u \mapsto u(N_\mathrm{l},y+1) - u(N_\mathrm{l},y), \end{cases}$$

l'expression de la variation totale peut être réécrite comme suit :

$$\mathrm{tv}(u) = \sum_{x=1}^{N_\mathrm{l}-1}\sum_{y=1}^{N_\mathrm{c}-1} |\Gamma_{x,y}(u)|_2 + \sum_{x=1}^{N_\mathrm{l}-1} |\Gamma_{x,N_\mathrm{c}}(u)| + \sum_{y=1}^{N_\mathrm{c}-1} |\Gamma_{N_\mathrm{l},y}(u)|. \tag{3.59}$$

Proposition 3.4 *La fonction* $\mathrm{tv} : \mathcal{H} \to \mathbb{R}$ *est convexe et non différentiable.*

Démonstration.

Comme les normes dans (3.59) sont convexes, leur somme est convexe [Rockafellar, 1970]. Ceci montre que la variation totale est une fonction convexe. Par ailleurs, ces normes sont non différentiables à l'origine. Il s'ensuit alors que, si l'une des conditions suivantes est satisfaite :

(i) $1 \leq x \leq N_\mathrm{l} - 1$, $1 \leq y \leq N_\mathrm{c} - 1$ et $\Gamma_{x,y}(u) = 0$, i.e., $u(x+1, y) = u(x, y) = u(x, y+1)$.

(ii) $1 \leq x \leq N_\mathrm{l} - 1$ et $\Gamma_{x,N_\mathrm{c}}(u) = 0$, i.e., $u(x+1, N_\mathrm{c}) = u(x, N_\mathrm{c})$.

(iii) $1 \leq y \leq N_\mathrm{c} - 1$ et $\Gamma_{N_\mathrm{l},y}(u) = 0$, i.e., $u(N_\mathrm{l}, y+1) = u(N_\mathrm{l}, y)$.

la variation totale est non différentiable en $u \in \mathcal{H}$. ■

3.4.3.2 Projection sous-différentielle sur S_2

Établir une expression de la projection sous-différentielle de $u \in \mathcal{H}$ sur l'ensemble S_2 nécessite le calcul d'un sous-gradient de $\mathrm{tv}(u)$. D'après la définition (3.59) de la variation totale et la proposition 3.1, il vient

$$\sum_{x=1}^{N_\mathrm{l}-1}\sum_{y=1}^{N_\mathrm{c}-1} \partial|\Gamma_{x,y}(u)|_2 + \sum_{x=1}^{N_\mathrm{l}-1} \partial|\Gamma_{x,N_\mathrm{c}}(u)| + \sum_{y=1}^{N_\mathrm{c}-1} \partial|\Gamma_{N_\mathrm{l},y}(u)| \subset \partial\mathrm{tv}(u), \tag{3.60}$$

où $\partial\mathrm{tv}(u)$ est le sous-différentiel de la variation totale en u.

Le calcul sous-différentiel des termes de gauche de l'inclusion (3.60) diffère selon que les conditions (i)-(iii) evoquées plus haut sont satisfaites ou non [Combettes, Luo, 2002b]. Notons que ces conditions sont satisfaites si l'image u est constante au voisinage d'un pixel (x, y). Examinons par exemple le cas du premier terme :

- Si la condition (i) est satisfaite, la fonction $u \mapsto \partial|\Gamma_{x,y}(u)|_2$ n'est pas différentiable et son sous-différentiel en u est donné par $\partial|\Gamma_{x,y}(u)|_2 = \Gamma_{x,y}^\top(B_2(0;1))$, où $B_2(0;1)$ est le disque fermé de norme unité dans $\mathbb{R}^2$ [Rockafellar, 1970]. Étant donné que $0 \in \partial|\Gamma_{x,y}(u)|_2$, nous pouvons choisir d'ignorer la contribution de ce terme dans la formulation finale du sous-différentiel. On pose alors $\partial|\Gamma_{x,y}(u)|_2 = 0$.

- Si la condition (i) n'est pas satisfaite, la fonction $u \mapsto \partial|\Gamma_{x,y}(u)|_2$ est différentiable en u et $\nabla|\Gamma_{x,y}(u)|_2 = \Gamma_{x,y}^\top\Gamma_{x,y}(u)/|\Gamma_{x,y}(u)|_2$ est son unique sous-gradient.

Nous pouvons définir, d'une manière similaire, le sous-différentiel de la fonction $u \mapsto |\Gamma_{x,N_\mathrm{c}}(u)|$ (resp. $u \mapsto \partial|\Gamma_{N_\mathrm{l},y}(u)|$). Ainsi, si la condition (ii) (resp. (iii)) est satisfaite, nous pouvons

ignorer la contribution du terme correspondant. Sinon, la contribution est tout simplement $\nabla|\Gamma_{x,N_c}(u)|$ (resp. $\nabla|\Gamma_{N_l,y}(u)|$).

Analytiquement, ceci revient à écrire :

$$\left\{\begin{array}{lll} \partial|\Gamma_{x,y}(u)|_2 & = \left\{\begin{array}{ll} 0, & \text{si } \Gamma_{x,y}(u)=0; \\ \Gamma_{x,y}^\top \Gamma_{x,y}(u)/|\Gamma_{x,y}(u)|_2, & \text{sinon ;} \end{array}\right. \\ \partial|\Gamma_{x,N_c}(u)| & = \left\{\begin{array}{ll} 0, & \text{si } \Gamma_{x,N_c}(u)=0; \\ \Gamma_{x,N_c}^\top \Gamma_{x,N_c}(u)/|\Gamma_{x,N_c}(u)|, & \text{sinon ;} \end{array}\right. \\ \partial|\Gamma_{N_l,y}(u)| & = \left\{\begin{array}{ll} 0, & \text{si } \Gamma_{N_l,y}(u)=0; \\ \Gamma_{N_l,y}^\top \Gamma_{N_l,y}(u)/|\Gamma_{N_l,y}(u)|, & \text{sinon .} \end{array}\right. \end{array}\right. \tag{3.61}$$

Nous pouvons désormais définir un sous-gradient $t_2 \in \partial \mathrm{tv}(u)$ comme suit :

$$t_2 = \sum_{x=1}^{N_l-1}\sum_{y=1}^{N_c-1} \partial|\Gamma_{x,y}(u)|_2 + \sum_{x=1}^{N_l-1} \partial|\Gamma_{x,N_c}(u)| + \sum_{y=1}^{N_c-1} \partial|\Gamma_{N_l,y}(u)|. \tag{3.62}$$

Enfin, en utilisant l'équation (3.38), la projection sous-différentielle d'une image u sur l'ensemble S_2 s'écrit :

$$p_2 = \left\{\begin{array}{ll} u + \dfrac{\tau - \mathrm{tv}(u)}{\|t_2\|^2} t_2, & \text{si } \mathrm{tv}(u) > \tau; \\ u, & \text{si } \mathrm{tv}(u) \leq \tau. \end{array}\right. \tag{3.63}$$

3.4.4 Régularisation suivant les contours de l'image de référence

Le principe de la régularisation suivant les contours de l'image de référence est de supposer que les discontinuités de l'image de disparité constituent un sous-ensemble des discontinuités en niveaux de gris présentes dans l'image de référence de la paire stéréoscopique. Une formulation de cette régularisation, décrite en détails dans le paragraphe 2.4.4.3, pour une utilisation dans le contexte de notre application nous conduit à définir l'ensemble convexe suivant :

$$S_3 = \{u \in \mathcal{H} \mid \int_\Omega \big(\nabla u(s)\big)^\top D(\nabla I_g)(s)\, \nabla u(s)\, ds \leq \delta\}\,, \tag{3.64}$$

où δ est une constante positive qui peut être estimée à partir d'expérimentations. L'idée de cette contrainte est donc d'atténuer le lissage le long de la direction perpendiculaire au gradient de l'intensité, lorsque $|\nabla I_g|$ tend vers l'infini, et d'encourager ce lissage pour les faibles valeurs du gradient de I_g correspondant aux zones homogènes de l'image.

3.4.4.1 Discrétisation

Pour définir une version discrète du terme de régularisation suivant

$$\int_\Omega \big(\nabla u(s)\big)^\top D(\nabla I_g)(s)\,\nabla u(s)\,ds, \tag{3.65}$$

on introduit d'abord une version discrète $\widehat{\nabla} I_g = \big((\widehat{\nabla} I_g)^1, (\widehat{\nabla} I_g)^2\big)^\top$ du gradient de l'image de gauche I_g, de manière similaire à (3.56). On introduit ensuite une version discrète $D(\widehat{\nabla} I_g)$ de l'opérateur de Nagel-Enkelmann. Ainsi, pour tout pixel (x, y), cet opérateur discret est une matrice de taille 2×2, que l'on note :

$$\widehat{D}(x,y) = \begin{pmatrix} D_1(x,y) & D_3(x,y) \\ D_3(x,y) & D_2(x,y) \end{pmatrix}, \tag{3.66}$$

Soit $u \in \mathcal{H}$, la fonction discrète associée à (3.65), que l'on notera f_3, s'écrit :

$$\begin{aligned} f_3(u) &= \sum_{x=1}^{N_\mathrm{l}} \sum_{y=1}^{N_\mathrm{c}} \big(\widehat{\nabla} u(x,y)\big)^\top \widehat{D}(x,y) \widehat{\nabla} u(x,y). \\ &= \sum_{x=1}^{N_\mathrm{l}} \sum_{y=1}^{N_\mathrm{c}} \big((\widehat{\nabla} u)^1(x,y)\big)^2 D_1(x,y) + \big((\widehat{\nabla} u)^2(x,y)\big)^2 D_2(x,y) \\ &\qquad + 2\,\big((\widehat{\nabla} u)^1(x,y)\big)\big((\widehat{\nabla} u)^2(x,y)\big) D_3(x,y). \end{aligned} \tag{3.67}$$

3.4.4.2 Projection sous-différentielle sur S_3

On vérifie aisément que la fonction f_3 est différentiable sur $\mathcal{H}$. D'après le théorème 3.3, le sous-différentiel ∂f_3 de cette fonction est non vide et se réduit au gradient de f_3. On note t_3 ce gradient défini, pour tout $(x, y) \in \{1, \ldots, N_\mathrm{l}\} \times \{1, \ldots, N_\mathrm{c}\}$, par :

$$\begin{aligned} t_3(x,y) = 2\Big(& - D_1(x,y)(\widehat{\nabla} u)^1(x,y)\,\mathbf{1}_{\{x<N_\mathrm{l}\}} + D_1(x-1,y)(\widehat{\nabla} u)^1(x-1,y)\,\mathbf{1}_{\{x\geq 2\}} \\ & - D_2(x,y)(\widehat{\nabla} u)^2(x,y)\,\mathbf{1}_{\{y<N_\mathrm{c}\}} + D_2(x,y-1)(\widehat{\nabla} u)^2(x,y-1)\,\mathbf{1}_{\{y\geq 2\}} \\ & - D_3(x,y)\big((\widehat{\nabla} u)^1(x,y) + (\widehat{\nabla} u)^2(x,y)\big)\mathbf{1}_{\{x<N_\mathrm{l}, y<N_\mathrm{c}\}} \\ & + D_3(x-1,y)(\widehat{\nabla} u)^2(x-1,y)\mathbf{1}_{\{x\geq 2, y<N_\mathrm{c}\}} \\ & + D_3(x,y-1)(\widehat{\nabla} u)^1(x,y-1)\mathbf{1}_{\{x<N_\mathrm{l}, y\geq 2\}}\Big), \end{aligned} \tag{3.68}$$

où $\mathbf{1}_A$ est égale à 1 si la condition A est vérifiée et 0 sinon.

On tire de (3.38) l'expression suivante de la projection sous-différentielle de $u \in \mathcal{H}$ sur l'ensemble S_3 :

$$p_3 = \begin{cases} u + \dfrac{\delta - f_3(u)}{\|t_3\|^2} t_3, & \text{si } f_3(u) > \delta; \\ u, & \text{si } f_3(u) \leq \delta. \end{cases} \tag{3.69}$$

3.4.5 Régularisation dans le domaine des ondelettes

En s'inspirant de travaux récents effectués dans le cadre de la restauration d'images [Combettes, Pesquet, 2004b], nous avons proposé d'autres formes de contraintes de régularisation servant à conférer une certaine régularité à l'image de disparité.

L'idée de base de cette régularisation est de supposer qu'une semi-norme de l'image de disparité, définie à partir de coefficients d'ondelettes de u dans un espace $\mathcal{X}$ approprié, est bornée. La forme générale de la contrainte associée à cette information *a priori* s'écrit :

$$S_4 = \{u \in \mathcal{H} \mid \|u\|_{\mathcal{X}} \leq \kappa\}, \tag{3.70}$$

où $\kappa > 0$ et $\|\cdot\|_{\mathcal{X}}$ est une semi-norme dans l'espace $\mathcal{X}$.

Pour régulariser la solution et préserver les discontinuités, l'espace $\mathcal{X}$ qui sert à définir la régularité *a priori* de u doit être choisi de façon à garantir les propriétés de régularisation désirées, à savoir lisser de manière isotrope dans les zones homogènes et préserver les discontinuités aux endroits importants comme les bords des objets. Lorsque $\mathcal{X}$ est l'espace BV des fonctions à variation bornée, $\|u\|_{\mathcal{X}}$ est simplement la variation totale de u que nous avons introduite précédemment et qui assure un lissage isotrope dans les zones homogènes de la solution tout en préservant les discontinuités. Cependant, la variation totale d'une image ne peut pas s'exprimer en fonction de ses coefficients en ondelettes. Une famille d'espaces bien adaptée aux ondelettes est celle des espaces de Besov. La régularisation dans ces espaces a déjà montré son efficacité dans de nombreux problèmes en traitement d'images [Choi, Baraniuk, 1999b, Chambolle *et al.*, 1998, Choi, Baraniuk, 1999a, Bect *et al.*, 2004].

Une des contributions de ce chapitre consiste donc à choisir pour $\mathcal{X}$ dans l'expression de la contrainte (3.70) des espaces de Besov $B^{\sigma}_{p,q}$ (pour $0 < \sigma < \infty$, $1 \leq p \leq \infty$ et $1 \leq q \leq \infty$). Un des arguments pour ce choix d'espaces est qu'ils contiennent des fonctions pouvant être discontinues le long de courbes. Une autre raison est que les normes dans ces espaces sont équivalentes à des expressions relativement simples en terme des coefficients de l'image dans une décomposition en ondelettes [Meyer, 1992b]. Enfin, notre choix est motivé par l'aptitude de l'algorithme que nous utilisons à pouvoir facilement combiner des contraintes définies dans le domaine spatial et d'autres définies dans un domaine transformé, notamment le domaine des ondelettes.

Nous donnons à présent une description des contraintes de régularisation que nous formulons dans le domaine des ondelettes. Quelques rappels sur les ondelettes sont effectués dans l'Annexe A.

On note $W^{\mathcal{B}}$ la décomposition en ondelettes 2D dans une base séparable $\mathcal{B} = \{\psi_{j,k,o} = 2^{-j}\psi^o(\cdot/2^j - k)\}_{j\in\mathbb{Z},k\in\mathbb{Z}^2,o\in\{1,2,3\}}$ et $(c^{\mathcal{B}}_{j,k,o}(u))_{j\in\mathbb{Z},k\in\mathbb{Z}^2,o\in\{1,2,3\}}$ les coefficients en ondelettes de

$u \in \mathcal{H}$ issus de cette décomposition, où $j \in \mathbb{Z}$ est le niveau de résolution et $o \in \{1,2,3\}$ désigne l'orientation suivant les directions horizontale, verticale et diagonale. Le niveau de résolution le plus grossier de cette décomposition est noté l.

Rappelons tout d'abord la définition suivante.

Définition 3.12 *Deux normes A et B sur $\mathcal{H}$ sont dites équivalentes si :*

$$\exists (C_1, C_2) \in (\mathbb{R}_+^*)^2 \text{ tel que } \forall u \in \mathcal{H}, \quad C_1 A(u) \leq B(u) \leq C_2 A(u). \tag{3.71}$$

On note alors $A \simeq B$ sur $\mathcal{H}$.

Une fonction $u \in \mathcal{H}$ admet la décomposition suivante :

$$u = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}^2, o \in \{1,2,3\}} c^{\mathcal{B}}_{j,k,o} \psi_{j,k,o}. \tag{3.72}$$

Si l'on utilise des ondelettes possédant $N+1$ moments nuls et de régularité au moins C^{N+1}, avec $N \geq \sigma$, alors si u appartient à l'espace de Besov $B^{\sigma}_{p,q}$, la norme de u sur cet espace est équivalente à la norme suivante de ses coefficients en ondelettes :

$$\|u\|_{B^{\sigma}_{p,q}} \simeq \Big(\sum_{j \in \mathbb{Z}} \Big(\sum_{k \in \mathbb{Z}^2, o \in \{1,2,3\}} 2^{-j(\sigma p + p - 2)} |c^{\mathcal{B}}_{j,k,o}(u)|^p \Big)^{q/p} \Big)^{1/q}. \tag{3.73}$$

Les fonctions dans les espaces de Besov $B^{\sigma}_{p,q}$, définis principalement par l'indice de régularité σ et l'indice de tolérance p, ont σ dérivées dans $L^p(\mathbb{R}^2)$. Notons que l'indice q permettant de jouer plus finement sur la régularité des fonctions, est d'importance secondaire par comparaison aux indices σ et p. Un espace de Besov qui joue un rôle central en traitement d'images est l'espace $B^{\sigma}_{p,p}$ avec $1/p = \sigma/2 + 1/2$. Il s'agit des espaces de Besov de régularité minimale qui s'injectent dans $L^2(\mathbb{R}^2)$ [DeVore, Lucier, 1992]. La norme sur ces espaces est définie par :

$$\forall u \in B^{\sigma}_{p,p}, \quad \|u\|_{B^{\sigma}_{p,p}} \simeq \Big(\sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}^2, o \in \{1,2,3\}} |c^{\mathcal{B}}_{j,k,o}(u)|^p \Big)^{1/p}. \tag{3.74}$$

Dans ce travail, nous nous sommes particulièrement intéressés à l'espace de Besov $B^1_{1,1}$, en prenant $p = \sigma = 1$. Cet espace est très proche de l'espace des fonctions à variations bornées BV du fait des inclusions suivantes (pour des fonctions définies sur $\Omega \subset \mathbb{R}^2$ borné) :

$$B^1_{1,1} \subset BV \subset L^2(\Omega). \tag{3.75}$$

L'inclusion de gauche implique que l'espace $B^1_{1,1}$ est presque aussi bien adapté que l'espace BV pour modéliser les images géométriques, comme étant constituées de différentes zones homogènes séparées par de franches discontinuités. L'idée dans ce travail est alors d'utiliser pour $\mathcal{X}$ dans (3.70) l'espace de Besov $B^1_{1,1}$. En annulant les coefficients de détails diagonaux,

où la contribution du bruit est souvent plus forte que celle du signal utile, on obtient l'ensemble convexe suivant associé à une semi-norme sur cet espace :

$$S_4^{(0,0)} = \{u \in \mathcal{H} \mid \sum_{j\in\mathbb{Z},k\in\mathbb{Z}^2,o\in\{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u)| \leq \kappa_4\}, \tag{3.76}$$

où $\kappa_4 > 0$.

À l'instar de ces semi-normes sur les espaces de Besov caractérisés par les coefficients en ondelettes, nous avons proposé une nouvelle semi-norme équivalente, et qui par conséquent satisfait les propriétés de régularisation avec préservation de discontinuités. Imposer une borne sur cette semi-norme, définie également à partir des coefficients en ondelettes de l'image, restreint la solution à appartenir à l'ensemble convexe suivant :

$$S_5^{(0,0)} = \{u \in \mathcal{H} \mid \sum_{j\in\mathbb{Z},k\in\mathbb{Z}^2} (\sum_{o\in\{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u)|^2)^{1/2} \leq \kappa_5\}, \tag{3.77}$$

où $\kappa_5 > 0$.

3.4.5.1 Invariance par translation

Un point faible de la transformation en ondelettes discrète est qu'elle fait appel à un processus de décimation d'un facteur 2 afin de réduire le nombre de coefficients au fur et à mesure que la résolution devient plus grossière. De ce fait, elle n'est pas invariante par translation. La propriété d'invariance par translation garantit qu'en cas de translation spatiale de l'image, les coefficients d'ondelettes sont également simplement translatés. Dans le cadre de la mise en correspondance, cette propriété est essentielle [Kim *et al.*, 1997, Shi *et al.*, 2001]. En effet, lorsque les images d'un couple stéréoscopique sont rectifiées, le déplacement entre un pixel de l'image de gauche et son correspondant dans l'image de droite se réduit à une simple translation. Cette translation correspond à la disparité associée à ce pixel, qui dépend fortement de sa position dans l'image. La non-invariance par translation se traduit donc par une prise en compte inhomogène d'un point de vue spatial des valeurs de disparité dans l'image, ce qui nuit aux performances de la mise en correspondance.

Remarque 3.5 *L'invariance par rotation de la décomposition en ondelettes est, au même titre que l'invariance par translation, une propriété nécessaire dans bon nombre d'applications en traitement d'images pour garantir une bonne représentation de l'image. Notons, qu'en revanche, ce problème ne se pose pas dans le cadre de la mise en correspondance, dès lors que les images de la paire stéréoscopique sont réctifiées.*

Une manière simple de pallier le problème d'invariance par translation de la transformée en ondelettes discrète est d'utiliser des transformées continues ou non décimées. Ceci introduit malheureusement une redondance accrue qui peut se révéler pénalisante en termes de coût de calcul et de nombre de coefficients à manipuler. La décomposition en ondelettes invariante par translation que nous avons adoptée ici consiste à effectuer, à chaque niveau de résolution, une transformation décimée classique, pour tous les décalages circulaires possibles de l'image originale [Coifman, D.Donoho, 1995, Pesquet *et al.*, 1996] ou un sous-ensemble d'entre eux.

Définition 3.13 *Pour une image u donnée, on appelle décalage circulaire de u d'une position $\mathrm{d} \in \mathbb{Z}^2$ et on note $u^{(\mathrm{d})}$ la version décalée de u par d. Pour une décomposition discrète en ondelettes 2D sur l niveaux de résolutions, il y a 2^{2l} décalages circulaires possibles de u.*

Par exemple, si $l = 1$, le décalage d peut prendre quatre valeurs :

$$\{(0,0),(0,1),(1,0),(1,1)\}. \tag{3.78}$$

Il est clair que si $\mathrm{d} = (0,0)$, l'image n'est pas translatée et $u^{(\mathrm{d})}$ est simplement l'image originale u.

L'invariance par translation étant réalisée par la combinaison des différentes translations de l'image, la contrainte définie par l'ensemble (3.76) devient :

$$S_4^{'} = \{u \in \mathcal{H} \mid \frac{1}{2^{2l}} \Big(\sum_{\mathrm{d} \in \{0,\ldots,2^l-1\}^2} \sum_{j=1}^{l} \sum_{k \in \mathbb{Z}^2, o \in \{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u^{\mathrm{d}})| \Big) \leq \kappa_4\}. \tag{3.79}$$

Cette contrainte de régularisation globale permet de conserver la propriété d'invariance par translation. Cependant, pour limiter la complexité de calcul, nous avons choisi de la diviser en plusieurs contraintes faisant intervenir chacune la semi-norme de l'image décalée. Plus précisément, ceci revient à considérer la contrainte suivante :

$$S_4 = \bigcap_{\mathrm{d} \in \{0,\ldots,2^l-1\}^2} S_4^{(\mathrm{d})}, \tag{3.80}$$

avec

$$S_4^{(\mathrm{d})} = \{u \in \mathcal{H} \mid \sum_{j=1}^{l} \sum_{k \in \mathbb{Z}^2, o \in \{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u^{(\mathrm{d})})| \leq \kappa_4\}. \tag{3.81}$$

D'une manière similaire, on peut définir, une contrainte convexe S_5 associée à la contrainte (3.77), soit :

$$S_5 = \bigcap_{\mathrm{d} \in \{0,\ldots,2^l-1\}^2} S_5^{(\mathrm{d})}, \tag{3.82}$$

avec

$$S_5^{(\mathrm{d})} = \{u \in \mathcal{H} \mid \sum_{j=1}^{l} \sum_{k \in \mathbb{Z}^2} \Big(\sum_{o \in \{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u^{(\mathrm{d})})|^2 \Big)^{1/2} \leq \kappa_5\}. \tag{3.83}$$

Il est intéressant de noter que le choix que nous avons effectué de diviser les contraintes de régularisation globales est motivé par la capacité de l'algorithme que nous utilisons à incorporer, avec une grande flexibilité, un grand nombre de contraintes convexes. Un autre gros avantage qu'offre cet algorithme est la possibilité de le mettre en œuvre sur une architecture à processeurs parallèles. Dans le contexte de ces contraintes, l'intérêt d'une implémentation parallèle est renforcé. En effet, si 2^{2l} processeurs parallèles sont disponibles, les contraintes de régularisation (3.81) et (3.83) associées à tous les décalages possibles de l'image originale peuvent être traitées simultanément, ce qui permet de réduire les temps de calcul.

3.4.5.2 Projection sous-différentielle sur S_4

Avant de formuler une expression de la projection différentielle sur l'ensemble S_4, il peut être intéressant d'énoncer la proposition suivante sur le sous-différentiel des fonctions définies à partir des coefficients en ondelettes :

Proposition 3.5 *Soient f et g deux fonctions convexes de $\mathcal{H}$ dans $]-\infty,+\infty]$, tel que*

$$f = g \circ W^{\mathcal{B}}, \tag{3.84}$$

où $W^{\mathcal{B}}$ est la décomposition dans une base orthonormale d'ondelettes $\mathcal{B}$. Supposons qu'il existe $u \in \mathcal{H}$ tel que g est finie et continue en $W^{\mathcal{B}}(u)$, alors

$$\partial f(u) = (W^{\mathcal{B}})^* \partial g(W^{\mathcal{B}}(u)) . \tag{3.85}$$

La base $\mathcal{B}$ étant orthonormale, on a $(W^{\mathcal{B}})^* = (W^{\mathcal{B}})^{-1}$.

Ce résultat se déduit des propriétés élémentaires des sous-différentielles [Ekeland, Temam, 1974].

Soit maintenant f_4 la semi-norme donnée par :

$$f_4 : u \mapsto \sum_{j=1}^{l} \sum_{k \in \mathbb{Z}^2, o \in \{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u)|. \tag{3.86}$$

Grâce à la proposition 3.5, un sous-gradient t_4 de la fonction f_4 en $u \in \mathcal{H}$ peut être calculé dans le domaine des ondelettes comme suit :

$$c^{\mathcal{B}}_{j,k,o}(t_4) = \begin{cases} c^{\mathcal{B}}_{j,k,o}(u), & \text{si } o = 3, \\ \\ \text{sign}(c^{\mathcal{B}}_{j,k,o}(u)), & \text{si } o \in \{1,2\}. \end{cases} \tag{3.87}$$

où sign est la fonction signe définie par :

$$\text{sign}(\xi) = \begin{cases} 1, & \text{si } \xi > 0, \\ 0, & \text{si } \xi = 0, \\ -1, & \text{si } \xi < 0. \end{cases} \tag{3.88}$$

En notant $t_4^{(\mathrm{d})}$ le sous-gradient de f_4 en $u^{(\mathrm{d})}$ pour un décalage d donné, la translatée de d de la projection sous-différentielle sur l'ensemble convexe $S_4^{(\mathrm{d})}$ défini par l'équation (3.81), s'écrit :

$$p_4^{(\mathrm{d})} = \begin{cases} u^{(\mathrm{d})} + \dfrac{\kappa_4 - f_4(u^{(\mathrm{d})})}{\|t_4^{(\mathrm{d})}\|^2} t_4^{(\mathrm{d})}, & \text{si } \ f_4(u^{(\mathrm{d})}) > \kappa_4, \\ \\ u^{(\mathrm{d})}, & \text{si } \ f_4(u^{(\mathrm{d})}) \leq \kappa_4. \end{cases} \tag{3.89}$$

3.4.5.3 Projection sous-différentielle sur S_5

Soit à présent f_5 la semi-norme associée à l'ensemble de contrainte convexe S_5. Elle s'écrit

$$f_5 : u \mapsto \sum_{j=1}^{l} \sum_{k \in \mathbb{Z}^2} \Big(\sum_{o \in \{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u)|^2 \Big)^{1/2}. \tag{3.90}$$

D'après la proposition 3.5, un sous-gradient t_5 de f_5 en $u \in \mathcal{H}$ peut être exprimé dans le domaine des ondelettes de la manière suivante :

$$c^{\mathcal{B}}_{j,k,o}(t_5) = \begin{cases} c^{\mathcal{B}}_{j,k,o}(u) & \text{si } \; o = 3, \\ 0 & \text{si } \; o \in \{1,2\} \text{ et } w_{j,k}(u) = 0, \\ c^{\mathcal{B}}_{j,k,o}(u)/w_{j,k}(u), & \text{si } \; o \in \{1,2\} \text{ et } w_{j,k}(u) > 0, \end{cases} \tag{3.91}$$

où $w_{j,k}(u)$ est donné par :

$$w_{j,k}(u) = \Big(\sum_{o \in \{1,2\}} |c^{\mathcal{B}}_{j,o,k}(u)|^2 \Big)^{1/2}. \tag{3.92}$$

La projection sous-différentielle sur les ensembles convexes $S_5^{(\mathrm{d})}$ de l'équation (3.83) s'obtient de manière similaire à (3.89).

3.4.5.4 Projection exacte sur S_5

Il peut être instructif de comparer la projection sous-différentielle précédente à une projection exacte. Calculer la projection exacte $P_e^{(\mathrm{d})}$ de $u^{(\mathrm{d})} \in \mathcal{H}$ sur chacun des sous-ensembles $S_5^{(\mathrm{d})}$ de l'équation (3.83) revient à résoudre le problème de programmation quadratique sous contrainte suivant :

$$\text{Minimiser} \quad \|u^{(\mathrm{d})} - p^{(\mathrm{d})}\|^2 \quad \text{tel que} \quad p^{(\mathrm{d})} \in S_5^{(\mathrm{d})} \; . \tag{3.93}$$

Nous montrons dans ce qui suit qu'il existe une solution quasi-explicite à ce problème de minimisation et, afin de simplifier les notations, nous omettons l'indice de décalage d.

L'ensemble S_5 étant défini dans le domaine des ondelettes et la décomposition étant orthogonale, le problème (3.93) s'écrit, en termes des coefficients en ondelettes, comme suit :

$$\text{Minimiser} \quad \sum_{j,k,o} |c^{\mathcal{B}}_{j,k,o}(u) - c^{\mathcal{B}}_{j,k,o}(p)|^2 \quad \text{sous la contrainte} \quad \sum_{j,k} w_{j,k}(p) \leq \kappa_5 \; , \tag{3.94}$$

où $w_{j,k}$ est donnée par l'équation (3.92). Supposons que $u \notin S_5$, il existe un multiplicateur de Lagrange $\lambda_l > 0$ tel que ce problème est équivalent à

$$\text{Minimiser} \quad \sum_{j,k,o} |c^{\mathcal{B}}_{j,k,o}(u) - c^{\mathcal{B}}_{j,k,o}(p)|^2 + \lambda_l \, \big(\sum_{j,k} w_{j,k}(p) - \kappa_5 \big). \tag{3.95}$$

La solution de ce problème est donnée explicitement, dans le domaine des ondelettes, par

$$\begin{cases} c^{\mathcal{B}}_{j,k,o}(P_e) = \max\Big(0, \frac{w_{j,k}(u)-\mu_l}{w_{j,k}(u)}\Big) c^{\mathcal{B}}_{j,k,o}(u), \\ \\ \sum_{j,k} w_{j,k}(P_e) = \kappa_5. \end{cases} \tag{3.96}$$

où $\mu_l = \lambda_l/2$. Ce paramètre est ajusté de manière à ce que les relations de Kuhn et Tucker (3.96) soient vérifiées. Ceci revient donc à écrire

$$\begin{aligned} & \sum_{j,k} w_{j,k}(P_e) = \kappa_5 \\ \Leftrightarrow \quad & \sum_{j,k} \max\Big(0, \frac{w_{j,k}(u) - \mu_l}{w_{j,k}(u)}\Big) w_{j,k}(u) = \kappa_5 \\ \Leftrightarrow \quad & \sum_{j,k} \max\big(0, w_{j,k}(u) - \mu_l\big) = \kappa_5. \end{aligned} \tag{3.97}$$

La fonction $\mu_l \mapsto \sum_{j,k} \max\big(0, w_{j,k}(u) - \mu_l\big)$ est décroissante et affine par morceaux. Par ailleurs, pour que l'équation (3.97) soit valide, il est clair que le paramètre μ_l ne doit pas dépasser la valeur $\max(w_{j,k}(u))_{j,k}$. Or μ_l étant strictement positif, une manière de réduire l'intervalle de recherche $]0, \max(w_{j,k}(u))_{j,k}]$ afin de trouver une valeur approchée de μ_l est de procéder par dichotomie. Cette procédure itérative est décrite par les étapes suivantes.

Algorithme 3.4

- Poser $a = 0$, $b = \max(w_{j,k}(u))_{j,k}$.
- Tant que $(b - a) > \epsilon$ faire

  $\mu_l = (a + b)/2$
  - si $\sum_{j,k} \max\big(0, w_{j,k}(u) - \mu_l\big) > \kappa_5$ alors

    $a = \mu_l$
  - sinon

    $b = \mu_l$
  - fin si
- fin tant que.

Une fois déterminé le multiplicateur de Lagrange μ_l, on reporte sa valeur dans l'équation (3.96) pour obtenir l'expression, en termes des coefficients en ondelettes, de la projection de $u \in \mathcal{H}$ sur l'ensemble S_5, soit

$$c^{\mathcal{B}}_{j,k,o}(P_e) = \begin{cases} c^{\mathcal{B}}_{j,k,o}(u), & \text{si} \quad o = 3, \\ \max\Big(0, \frac{w_{j,k}(u)-\mu_l}{w_{j,k}(u)}\Big) c^{\mathcal{B}}_{j,k,o}(u), & \text{si} \quad o \in \{1, 2\}. \end{cases} \tag{3.98}$$

Notons que la projection exacte sur chacun des ensembles $S_4^{(\mathrm{d})}$ s'obtient, en considérant la semi-norme associée, d'une manière similaire à celle calculée précédemment sur les ensembles $S_5^{(\mathrm{d})}$.

3.5 Conclusion

Nous avons, dans ce chapitre, présenté une nouvelle méthode d'estimation de la disparité à partir d'une paire d'images stéréoscopiques. Cette méthode variationnelle, qui s'appuie sur une formulation ensembliste du problème de mise en correspondance, consiste à minimiser une fonction objectif différentiable et strictement convexe sur l'intersection d'ensembles convexes.

Les ensembles convexes sont associés à des contraintes convexes modélisant toute propriété connue *a priori* sur le champ de disparité à estimer. Dans ce travail, nous avons mis l'accent, en particulier, sur une estimation d'un champ lisse tout en respectant les discontinuités présentes dans la scène. Pour cela, différentes contraintes de régularisation satisfaisant cette propriété ont été considérées. Ces contraintes sont définies dans le domaine spatial ou exprimées à partir des coefficients en ondelettes de l'image.

Pour apporter une solution numérique au problème d'optimisation sous contraintes résultant, nous avons adopté un algorithme efficace itératif par bloc. Cet algorithme emploie les sous-gradients associés aux diverses contraintes pour approximer les projections sur les ensembles correspondants, tout en évitant ainsi les projections exactes souvent difficles à calculer. Il offre une grande flexibilité quant à l'incorporation de contraintes multiples, qui doivent cependant toujours être convexes, et peut être mis en œuvre sur une architecture à processeurs parallèles.

Dans le chapitre suivant, nous validons cette méthode d'estimation de la disparité sur différents exemples d'images stéréoscopiques synthétiques et réels.

Chapitre 4

Résultats expérimentaux

L'objectif de ce chapitre est de valider notre méthode d'estimation de la disparité sur différents types d'images stéréoscopiques. Dans un premier temps, nous validons la méthode proposée d'un point de vue quantitatif. Pour cela nous exploitons des données synthétiques et réelles pour lesquelles la « vraie »image de disparité est connue. L'utilisation de ces données de référence, qui constituent par abus de langage *la vérité terrain*, nous permet par des critères quantitatifs d'évaluer la pertinence des champs de disparité calculés et de comparer l'efficacité des différentes contraintes de régularisation considérées. Nous illustrons, dans un deuxième temps, les résultats de la méthode sur des images réelles pour lesquelles aucune vérité terrain n'est disponible.

4.1 Quelques détails de mise en œuvre

4.1.1 Formulation du critère

Nous rappelons que le problème qu'on se propose de résoudre se formule comme suit :

$$\text{trouver} \;\; \widehat{u} \in S = \bigcap_{i=1}^{5} S_i, \;\; \text{tel que} \;\; J(\widehat{u}) = \inf J(S), \tag{4.1}$$

où la fonction objectif J, donnée par l'équation (3.45), traduit l'hypothèse de conservation de la luminance entre les intensités des pixels homologues et les ensembles de contraintes convexes $(S_i)_{1\leq i\leq 5}$ modélisent l'information *a priori* sur le champ de disparité u et sont décrits par les équations (3.46), (3.54), (3.64), (3.80) et (3.82).
Notons que le critère J s'exprime au travers d'une mesure d'écart $\varphi : \mathbb{R} \mapsto \mathbb{R}$ différentiable et strictement convexe. En guise d'illustration et pour des raisons de simplicité de mise en œuvre, nous considérons dans ce chapitre une mesure de similarité quadratique, auquel cas le critère J s'écrit :

$$J(u) = \sum_{\mathrm{s}\in\mathcal{D}\setminus\mathcal{O}} [L(\mathrm{s})\, u(\mathrm{s}) - r(\mathrm{s})]^2 + \alpha \sum_{\mathrm{s}\in\mathcal{D}} [u(\mathrm{s}) - \bar{u}(\mathrm{s})]^2 \,, \tag{4.2}$$

où $\bar{u}$ est l'estimée initiale donnée par la méthode de corrélation et α est un paramètre réel strictement positif.

En posant :

$$R : u(s) \mapsto R(s)\ u(s),$$

où

$$R(s) = \begin{cases} L^2(s) + \alpha & \text{si } s \in \mathcal{D} \setminus \mathcal{O}, \\ \alpha & \text{sinon,} \end{cases} \tag{4.3}$$

et

$$u_0(s) = \begin{cases} (L(s)\ r(s) + \alpha\ \bar{u}(s))\ R^{-1}(s) & \text{si } s \in \mathcal{D} \setminus \mathcal{O}, \\ \bar{u}(s) & \text{sinon,} \end{cases} \tag{4.4}$$

le critère J se formule, d'une manière équivalente, comme suit :

$$J(u) = \langle u - u_0 \ \mid \ R(u - u_0) \rangle. \tag{4.5}$$

Il est clair que $J \in \Gamma_0(\mathcal{H})$, qu'il est quadratique et strictement convexe. Les hypothèses 3.3.3.1 étant ainsi satisfaites, nous avons mis en œuvre l'algorithme 3.3, adapté à la minimisation de critères quadratiques, pour résoudre numériquement (4.1) où, à chaque itération n de l'algorithme :

a) tous les ensembles associés aux contraintes considérées sont exploités, c'est à dire $I_n = \{1, \ldots, m\}$;

b) les poids $(\omega_{i,n})_{i \in I_n}$ sont pris égaux à $1/m$ sur les m contraintes retenues ;

c) le paramètre de relaxation λ_n prend la valeur maximale autorisée L_n, ce qui permet d'accélérer la convergence de l'algorithme.

4.1.2 Choix des paramètres

Les paramètres qui entrent en jeu dans cet algorithme sont :

1. Le paramètre de pondération α de l'équation (4.2) – Il doit être fixé de manière à moduler l'influence du terme strictement convexe introduit dans la fonction objectif globale que nous devons minimiser. En choisissant une forte valeur pour ce paramètre, le critère (4.2) ne sert plus qu'à forcer u d'être proche de $\bar{u}$. À l'inverse, en prenant une valeur de α faible, on privilégie le terme d'attache aux données par rapport au deuxième terme. D'après nos expérimentations, ceci ralentit considérablement la convergence de l'algorithme. Un choix de $\alpha = 50$ apparaît expérimentalement comme un bon compromis entre la rapidité de convergence et la qualité de la solution obtenue ; nous le conserverons donc pour tous les tests effectués dans ce mémoire. Toutefois, notons que des valeurs entre 10 et 100 pour ce paramètre permettent également d'obtenir de bons résultats.

2. Le paramètre de diffusion ν dans l'opérateur de l'équation (2.59) – Le choix de ce paramètre permet de fixer la valeur de $|\nabla I_g|$ au-dessous de laquelle la régularisation induite par la contrainte de Nagel-Enkelmann diffuse de manière isotrope. Dans nos

expériences, nous avons ajusté la valeur de ν de façon à ce que le lissage soit isotrope sur 30 pourcent de l'image. Notons que, pour des valeurs de $\nu \gg |\nabla I_g|$, la contrainte de régularisation (3.64) se comporte de manière équivalente à une régularisation de type Tikhonov.

3. Les bornes sur les différents ensembles de contraintes, à savoir $u_{\min}$, $u_{\max}$, τ, δ, κ_4 et κ_5 − Notons que le choix des amplitudes minimale $u_{\min}$ et maximale $u_{\max}$ peut se faire en mesurant la disparité entre certains points d'intérêt des images, détectés automatiquement ou choisis manuellement. Ceci permet de connaître très précisément l'intervalle de disparité à utiliser pour une paire d'images donnée. Cependant, le choix que l'on peut faire de la valeur des autres paramètres est moins évident et se détermine généralement à partir d'expérimentations ou en exploitant des bases de données d'images du même type que celles traitées. Toutefois, comme il sera illustré plus loin au paragraphe 4.2.2.2, notre méthode est assez robuste vis-à-vis du choix de ces paramètres.

4.1.3 Critère d'arrêt

Étant donné qu'un champ de disparité u_n, généré par l'algorithme 3.3, est solution de (4.1) s'il appartient à l'ensemble des solutions admissibles S, le critère d'arrêt naturel qui peut être utilisé pour notre algorithme itératif est la satisfaction de toutes les contraintes considérées. Ce critère est traduit par le test effectué à l'étape 4 de l'algorithme. En effet, si toutes les contraintes sont simultanément satisfaites, L_n est égal à 0 et donc l'algorithme s'arrête. Cependant, dans nos expérimentations, nous avons remarqué qu'au fil des itérations l'estimation u_n se rapproche de la solution sans vérifier strictement les contraintes prises en compte. C'est pourquoi, afin de stopper l'algorithme, nous avons utilisé le critère d'arrêt suivant :

$$\|u_n - u_{n-1}\| < \varepsilon \|u_n\|, \tag{4.6}$$

où $\varepsilon > 0$ est un seuil à fixer. En général, ce seuil, choisi de façon empirique, doit être suffisamment petit. Pour l'ajuster, nous avons comparé les résultats obtenus par notre algorithme pour diverses valeurs du nombre d'itérations et nous avons relevé la valeur de ϵ qui correspond au meilleur compromis entre la qualité du résultat et le temps de calcul. Pour tous les tests effectués dans ce mémoire, cette valeur est de l'ordre de 10^{-4}.

4.2 Résultats sur des images avec vérité terrain disponible

4.2.1 Images testées et critères évalués

Afin d'évaluer quantitativement la méthode d'estimation de la disparité proposée, nous avons considéré deux catégories d'images avec vérité terrain : des images réelles et des images de synthèse. Dans le cas des images réelles, nous avons utilisé les images exploitées dans le protocole d'évaluation et de comparaison des méthodes de mise en correspondance de Scharstein et Szeliski [Scharstein, Szeliski, 2002]. Les auteurs de ce protocole ont mis à disposition sur le Web[1] dix paires d'images pour lesquelles la vérité terrain est fournie. Parmi ces images,

[1] http ://cat.middlebury.edu/stereo/data.html



FIG. 4.1 – Paire d'images *Corridor* et carte de disparité associée.

seules *Tsukuba*, *Venus*, *Sawtooth*, *Map* et *Teddy* sont utilisées pour évaluer notre méthode. Ces images sont présentées en couleur à la figure 4.2. La paire d'images de synthèse que nous utilisons est la paire *Corridor* générée à l'université de Bonn (cf. figure 4.1).

Lorsqu'on dispose de la vérité terrain, un grand nombre de critères quantitatifs peuvent être évalués. Nous avons, dans ce travail, considéré les deux critères suivants :

1. *Erreur absolue moyenne* – C'est la moyenne des valeurs absolues des erreurs entre l'image de disparité calculée u_c et celle théorique u_t. Elle est donnée par :

$$\text{EAM} = \frac{1}{N_\text{l} \times N_\text{c}} \sum_{s \in \mathcal{D}} |u_t(s) - u_c(s)|, \tag{4.7}$$

   où $N_\text{l} \times N_\text{c}$ est le nombre total de pixels.

2. *Pourcentage d'appariements erronés* – Un appariement est considéré erroné si l'erreur absolue moyenne dépasse un certain seuil δ_s, que nous fixons à 1, soit :

$$\text{Err} = \frac{1}{N_\text{l} \times N_\text{c}} \sum_{s \in \mathcal{D}} \mathbf{1}(|u_t(s) - u_c(s)| > \delta_s), \text{ avec } \mathbf{1}(c) = 1 \text{ si } c \text{ est vrai ou 0 sinon.} \tag{4.8}$$

Notons qu'un critère similaire à (4.8) a été proposé par Scharstein et Szeliski où un appariement est décrété erroné si l'erreur quadratique moyenne dépasse la valeur 1. Pour étudier le comportement des méthodes évaluées face à certaines difficultés, les auteurs proposent, dans leur protocole d'évaluation [Scharstein, Szeliski, 2002], d'appliquer ce critère dans trois zones particulières de l'image, à savoir toute l'image, les zones des discontinuités et les zones non texturées, sans prendre en compte les pixels occultés dans les trois zones (cf. figure 4.3). Pour chaque paire d'images testée, ils fournissent ainsi sur le site du protocole[2], en plus de la carte de disparité théorique, la carte des occultations, la carte des discontinuités et la carte des zones non texturées.

4.2.2 Expérimentations avec des contraintes spatiales

Dans ce paragraphe, nous présentons une série d'expérimentations réalisées pour valider notre méthode d'estimation de la disparité sur les images de la figure 4.2 en ne considérant,

[2]http ://cat.middlebury.edu/stereo/data.html

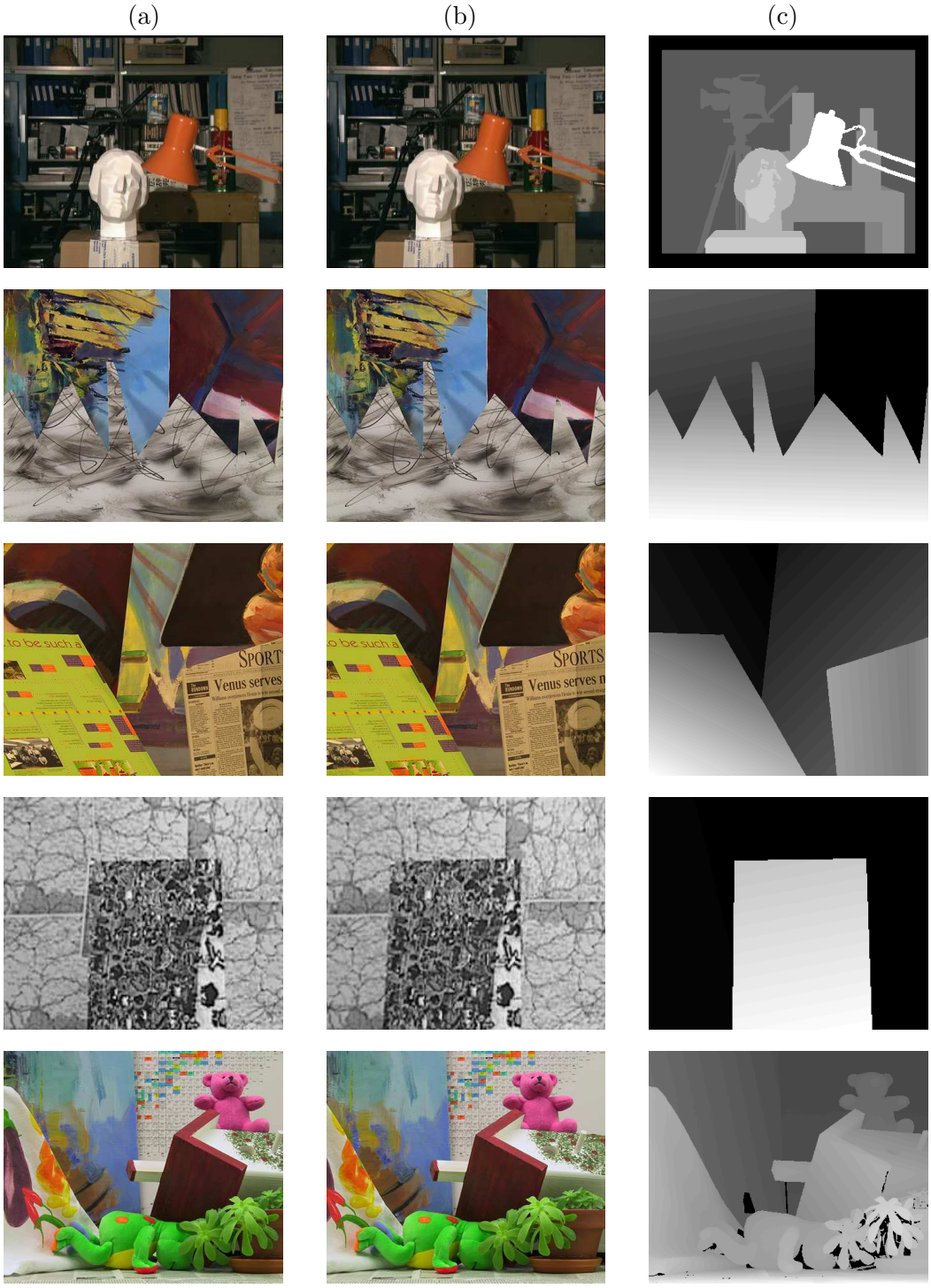


FIG. 4.2 – Exemples de paires stéréoscopiques proposées par Scharstein et Szeliski. De haut en bas : *Tsukuba*, *Sawtooth*, *Venus*, *Map*, *Teddy*. (a) Image de gauche (b) Image de droite (c) Carte de disparité théorique.

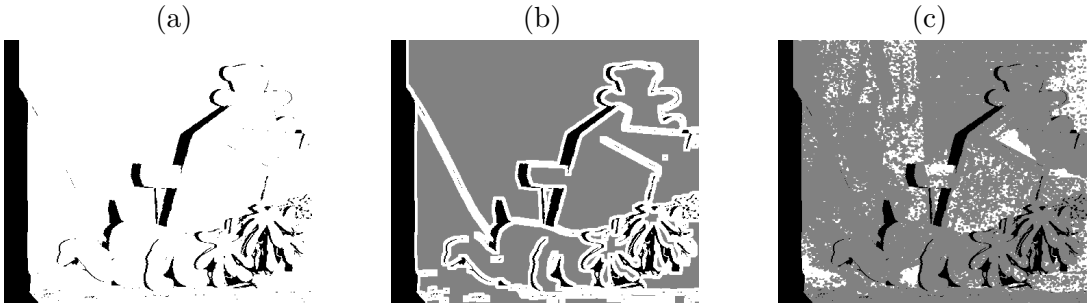


FIG. 4.3 – Zones d'évaluation considérées dans le protocole de Scharstein et Szeliski. Les pixels noirs correspondent aux pixels occultés et les pixels blancs correspondent aux (a) pixels de l'image (b) pixels proches d'une discontinuité (c) pixels dans une zone non texturée.

dans un premier temps, que les contraintes définies dans le domaine spatial, à savoir les contraintes S_1, S_2 et S_3. Notons que nous possédons la vérité terrain de toutes les images testées, nous connaissons aussi les vraies valeurs des bornes $\widehat{u}_{\min}$, $\widehat{u}_{\max}$, $\widehat{\tau}$, $\widehat{\delta}$ sur ces ensembles de contraintes. Ces valeurs sont données dans le tableau 4.1. Pour évaluer la robustesse de notre méthode vis-à-vis du choix des paramètres τ et δ, deux situations ont été considérées : 1) lorsque ces paramètres sont connus et 2) lorsqu'ils ne sont pas connus. Dans ce dernier cas, il s'agit plus précisément d'introduire des perturbations sur les vraies valeurs de ces paramètres.

Paire stéréoscopique	$\widehat{u}_{\min}$	$\widehat{u}_{\max}$	$\widehat{\tau}$	$\widehat{\delta}$
Tsukuba	5	14	1.2×10^4	7.5×10^5
Sawtooth	4	18	1.4×10^4	5.5×10^4
Venus	3	20	9.3×10^3	1.7×10^4
Map	4	29	9.9×10^3	5.7×10^4
Teddy	12.5	54	4.6×10^4	2.6×10^5

TAB. 4.1 – Vraies valeurs des bornes sur les ensembles de contraintes S_1, S_2 et S_3 pour les paires d'images de la figure 4.2.

4.2.2.1 Impact des contraintes sur la qualité de l'estimation

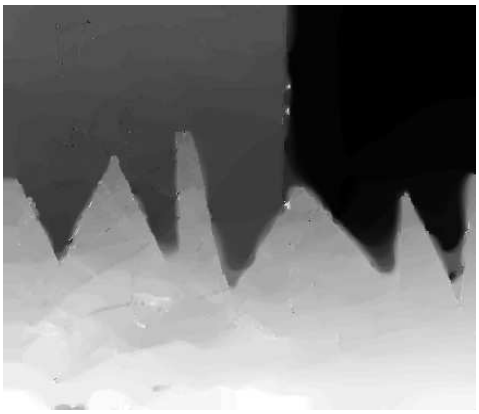
La figure 4.4 présente les cartes de disparités calculées par minimisation de la fonction (4.2) sur l'ensemble des solutions admissibles $S_1 \cap S_2 \cap S_3$. Ces cartes sont obtenues en ajustant les valeurs des bornes sur les ensembles considérés, à partir des valeurs du tableau 4.1 fournies par les cartes de disparité théoriques. Comme nous pouvons l'observer sur cette figure, les résultats obtenus pour les paires d'images *Sawtooth*, *Map* et *Venus* (cf figure 4.5) sont meilleurs que ceux obtenus pour *Tsukuba* et *Teddy*. En effet, ces dernières paires présentent plus de difficultés d'appariements que les autres. Les principales difficultés pour la mise en correspondance de la paire *Tsukuba* sont des structures fines (comme la caméra ou le bras de la lampe) et un grand nombre d'occultations. La paire stéréoscopique *Teddy* est également difficile à apparier du fait de la présence de textures complexes et surtout de la grande plage de variation des disparités. En effet, la disparité maximale pour l'objet le plus proche de cette scène est de 54 pixels (cf. tableau 4.1).

	$S_1 \cap S_2 \cap S_3$		$S_1 \cap S_2$		$S_1 \cap S_3$		$S_2 \cap S_3$	
Paire stéréoscopique	EAM	Err	EAM	Err	EAM	Err	EAM	Err
Tsukuba	**0.43**	**7.89**	0.46	9.23	0.50	11.57	0.53	10.25
Sawtooth	**0.39**	**4.83**	0.42	6.21	0.48	8.42	0.44	6.52
Venus	**0.31**	**6.56**	0.35	8.21	0.40	9.66	0.35	8.23
Map	**0.56**	**3.87**	0.63	4.49	0.79	10.7	0.62	4.51
Teddy	**1.04**	**15**	1.08	16	1.17	18	1.08	17

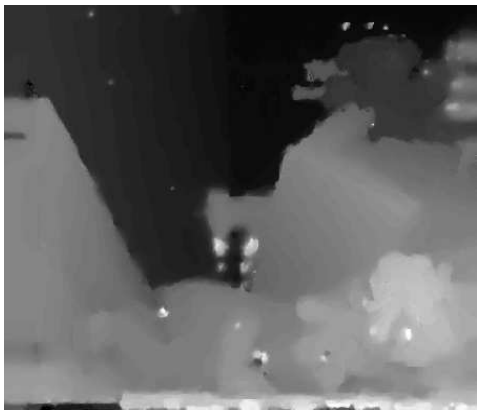
TAB. 4.2 – Impact des contraintes sur les résultats obtenus par notre méthode. Le meilleur résultat est obtenu lorsque toutes les contraintes sont prises en compte.



(a) *Tsukuda*



(b) *Sawtooth*



(c) *Teddy*



(d) *Map*

FIG. 4.4 – Cartes de disparité calculées par projection sur l'ensemble des solutions admissibles $S = S_1 \cap S_2 \cap S_3$.

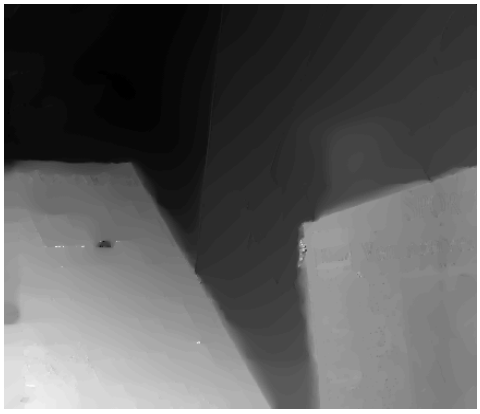
Dans le tableau 4.2, nous donnons, d'une part, les résultats quantitatifs correspondant aux cartes de disparité de la figure 4.4 (où $S = S_1 \cap S_2 \cap S_3$), suivant les deux critères que nous utilisons et, d'autre part, pour illustrer le bénéfice de chacune des contraintes S_1, S_2 et S_3, nous reportons les erreurs obtenues, lorsque l'une de ces contraintes n'est pas prise en compte, l'ensemble des solutions admissibles S se réduisant ainsi à l'intersection des deux autres contraintes. Dans ce tableau, nous notons en gras, pour chaque colonne, la meilleure valeur obtenue pour le critère considéré, évalué sur toute l'image (sauf pour la paire *Tsukuba* où il ne faut pas prendre en compte le bord de l'image, de dix-huit pixels, pour lequel la vérité terrain n'est pas fournie). Les résultats de ce tableau sont conformes à nos attentes puisque, pour toutes les paires d'images testées, le meilleur résultat est obtenu lorsque toutes les contraintes sont considérées. Par ailleurs, en comparant les valeurs de la deuxième et de la troisième colonne de ce tableau, nous pouvons constater que la régularisation par variation totale, en considérant $S = S_1 \cap S_2$, conduit à de meilleurs résultats que la régularisation de Nagel et Enkelmann où l'ensemble des solutions admissibles se réduit à $S_1 \cap S_3$, et ceci suivant les deux critères évalués.

Nous présentons sur la figure 4.5, pour la paire d'images *Venus*, les cartes de disparité estimées par minimisation de (4.2) sur les différents ensembles des solutions admissibles et nous indiquons également les valeurs, calculées à partir de ces cartes, de la variation totale, de la norme f_3 ainsi que de l'intervalle de disparité, que l'on note respectivement $\widetilde{\tau}$, $\widetilde{\delta}$ et $[\widetilde{u}_{\min}, \widetilde{u}_{\max}]$. Nous rappelons qu'une contrainte est satisfaite si la valeur de la fonction associée est inférieure à la borne imposée. Ainsi, en comparant les valeurs reportées sur cette figure à celles du tableau 4.1, nous pouvons remarquer que toutes les contraintes ne sont quasiment simultanément satisfaites que lorsque tous les ensembles sont utilisés. Ceci correspond à la carte de disparité de la figure 4.5(a) dont la qualité visuelle confirme qu'il est avantageux d'incorporer des contraintes multiples sur la solution. Notons également que la supériorité de la régularisation par variation totale par rapport à celle de Nagel-Enkelmann, déjà démontrée à partir des résultats du tableau 4.2, est confirmée par cette figure. En effet, les contours sur l'image de la figure 4.5(b) sont nettement mieux conservés que ceux de la figure 4.5(c) et l'effet régularisant dans les zones homogènes est beaucoup plus marqué. Cependant, la variation totale apparaît moins robuste aux grandes valeurs d'erreurs. En effet, comme on peut le voir sur la figure 4.5(c), le lissage des zones homogènes, avec une régularisation de Nagel-Enkelmann, a permis de mieux faire disparaître les pixels aberrants.

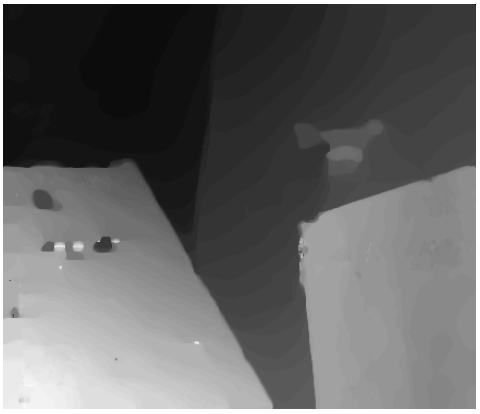
4.2.2.2 Robustesse vis-à-vis du choix des bornes des contraintes

En pratique, les bornes sur les ensembles des contraintes $(S_i)_{1 \leq i \leq 3}$ ne sont pas connues. C'est pourquoi, bien qu'il soit possible de les ajuster expérimentalement, avec une certaine précision, il nous paraît important d'étudier leurs influence sur les résultats obtenus. Dans ce qui suit, nous évaluons, en particulier, l'impact du choix des paramètres τ et δ sur la qualité de la solution. En effet, comme nous l'avons mentionné au paragraphe 4.1.2, le choix de $u_{\min}$ et $u_{\max}$ ne pose pas de problème et peut s'effectuer de manière précise.

Notons u_τ la carte de disparité obtenue par notre méthode pour une borne τ donnée sur l'ensemble S_2. Pour montrer l'influence de ce paramètre sur les performances de notre méthode, nous calculons, pour différentes valeurs de τ, la racine carrée de l'erreur quadratique moyenne normalisée $\|u_\tau - u_{\widehat{\tau}}\| / \|u_{\widehat{\tau}}\|$, où $u_{\widehat{\tau}}$ est la disparité obtenue en prenant la vraie valeur $\widehat{\tau}$. Le tracé de cette erreur en fonction de $\tau/\widehat{\tau}$, pour les paires d'images *Tsukuba*, *Sawtooth*



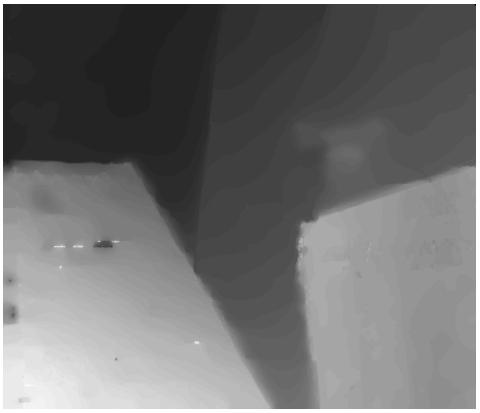
(a) $\widetilde{u}_{\min} = 2.95$, $\widetilde{u}_{\max} = 19.54$
$\widetilde{\tau} = 1.02 \times 10^4$, $\widetilde{\delta} = 1.75 \times 10^4$



(b) $\widetilde{u}_{\min} = 2.96$, $\widetilde{u}_{\max} = 19.26$
$\widetilde{\tau} = 1.02 \times 10^4$, $\widetilde{\delta} = 1.82 \times 10^5$



(c) $\widetilde{u}_{\min} = 2.89$, $\widetilde{u}_{\max} = 19.75$
$\widetilde{\tau} = 2.33 \times 10^4$, $\widetilde{\delta} = 1.72 \times 10^4$



(d) $\widetilde{u}_{\min} = 1.36$, $\widetilde{u}_{\max} = 21.55$
$\widetilde{\tau} = 1.02 \times 10^4$, $\widetilde{\delta} = 1.75 \times 10^4$

FIG. 4.5 – Cartes de disparité calculées pour la paire d'images *Venus* par projection sur les ensembles (a) $S_1 \cap S_2 \cap S_3$ (b) $S_1 \cap S_2$ (c) $S_1 \cap S_3$ (d) $S_2 \cap S_3$.

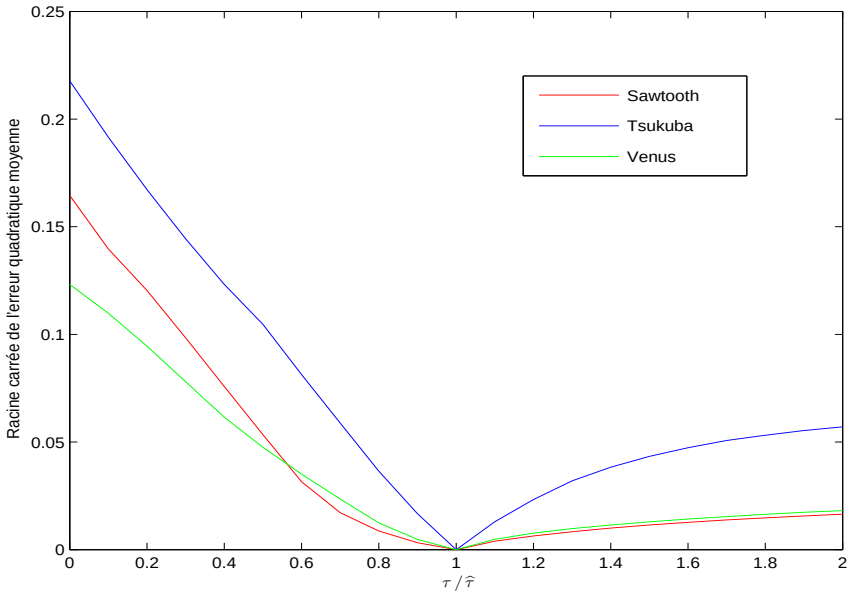


FIG. 4.6 – Racine carrée de l'erreur quadratique moyenne d'estimation en fonction de $\tau/\widehat{\tau}$.

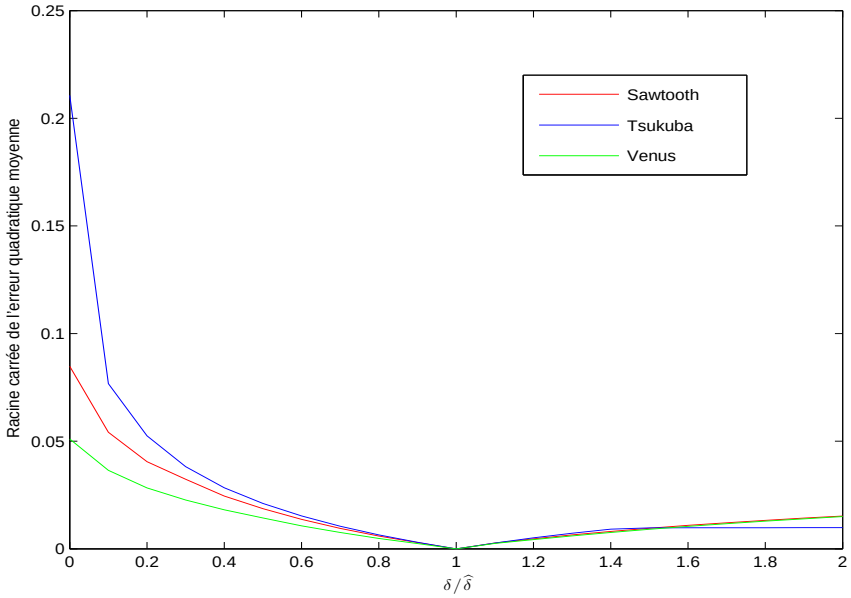


FIG. 4.7 – Racine carrée de l'erreur quadratique moyenne d'estimation en fonction de $\delta/\widehat{\delta}$.

et *Venus*, est représenté à la figure 4.6. Ces courbes montrent que notre méthode est robuste vis-à-vis du choix de τ dans la mesure où si $\tau/\widehat{\tau}$ varie entre 0.8 et 1.7, l'erreur ne dépasse pas 5%.

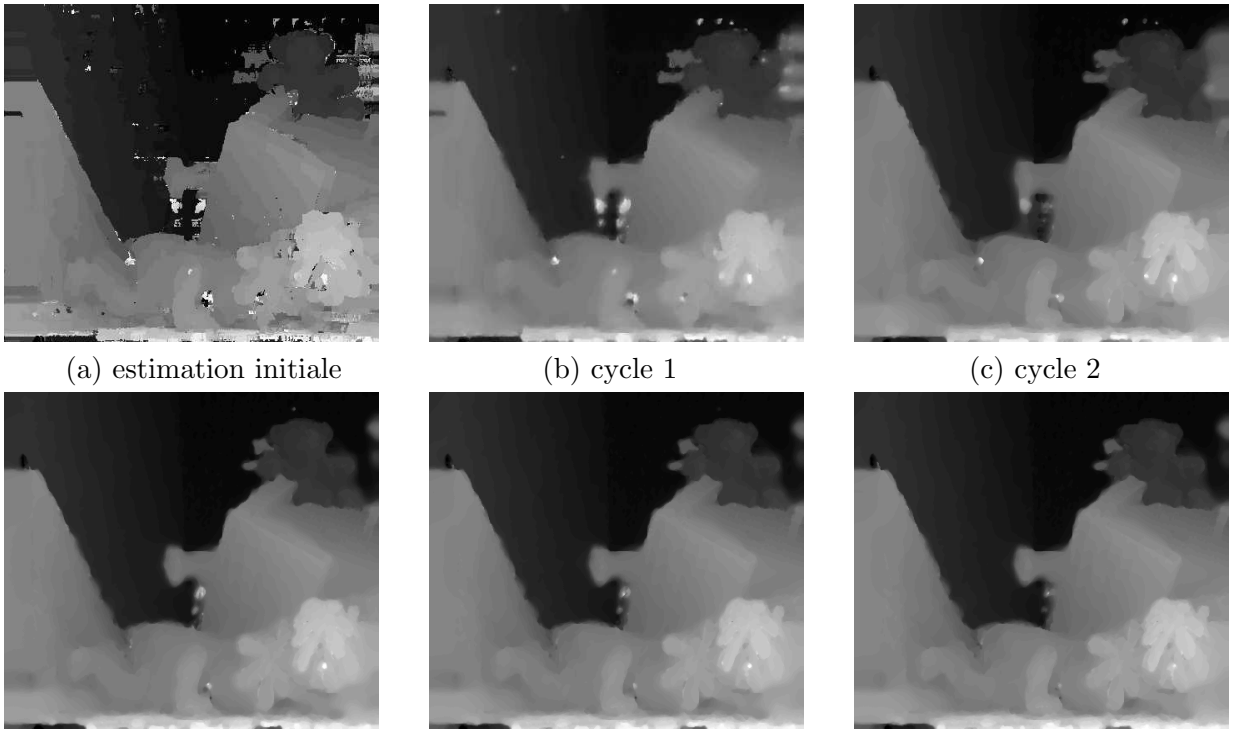
Nous procédons de même pour évaluer la robustesse de la méthode vis-à-vis du choix de la borne δ sur l'ensemble de contrainte S_3. La figure 4.7 montre l'erreur quadratique moyenne normalisée pour des valeurs différentes de cette borne. En comparant les courbes de cette figure avec celles de la figure 4.6, nous pouvons constater que notre approche est plus robuste vis-à-vis du choix de ce paramètre que vis-à-vis du choix de τ car, dans ce cas, l'erreur ne dépasse pas 2% pour des valeurs de $\delta/\widehat{\delta}$ variant entre 0.8 et 1.7.

4.2.2.3 Affinement des résultats par réinitialisation

Pour mettre en place notre méthode itérative d'estimation de la disparité, une étape préliminaire de calcul d'un champ de disparité initial est nécessaire. Comme nous l'avons mentionné au paragraphe 3.4.1.1, la méthode de mise en correspondance qui nous sert d'initialisation est la méthode par corrélation. Pour implanter cette méthode, nous avons choisi des fenêtres de corrélation carrées de taille 9×9, centrées sur le pixel considéré, et nous avons utilisé comme mesure de corrélation la somme des carrés des différences d'intensité SSD. L'intérêt de la méthode à base de corrélation, décrite plus en détails en 2.4.1, est qu'elle est simple à mettre en œuvre et surtout qu'elle est relativement rapide. Cependant, étant une approche locale, cette méthode est particulièrement sensible aux zones d'occultations et aux zones homogènes. Par conséquent, les erreurs d'appariement du champ initial générées dans ces zones difficiles seront importantes et auront ainsi tendance à se répercuter sur la solution. Nous pouvons observer ce phénomène sur les cartes de disparité de la figure 4.4. Notons, en effet, que les pics de disparité apparus en certains endroits de ces cartes proviennent des erreurs d'appariements commises à l'initialisation. Il faut donc affiner ces résultats par un processus de réinitialisation, qui consiste à itérer successivement les estimations de disparité en prenant la solution obtenue à un cycle d'itérations donné comme initialisation du cycle suivant et en linéarisant le critère (3.2) par rapport à ces valeurs. Cette procédure itérative a pour but d'améliorer l'estimation initiale à chaque cycle et donc d'éliminer les appariements de mauvaise qualité. Un autre avantage majeur de cette technique est qu'elle réduit la dépendance de la solution vis-à-vis du choix de l'estimation initiale.

La figure 4.8 montre les résultats intermédiaires pour la paire d'images *Teddy* où cinq cycles d'itérations ont été considérés. On voit sur ces images que ce processus itératif permet d'éloigner, progressivement, l'estimation courante du champ initial et de converger ainsi vers une solution plus cohérente et nettement meilleure que celle obtenue avec la première estimation de la disparité. En effet, les erreurs sur le champ initial n'apparaissent plus sur le champ final de la figure 4.8(f) et ceci malgré le fait que les amplitudes de disparité sur cette paire d'images sont importantes. Ceci confirme que notre méthode est également robuste vis-à-vis de grandes variations de la disparité. Nous pouvons aussi observer l'amélioration apportée par le processus de réinitialisation sur les erreurs EAM et Err indiquées dans le tableau 4.3. Par ailleurs, il ressort, de ce tableau et des images de la figure 4.8, que trois cycles d'itérations suffisent pour s'affranchir des erreurs d'initialisation et aboutir ainsi au meilleur résultat sans trop pénaliser le temps de calcul.

Nous avons reporté dans la dernière colonne du tableau 4.3 le nombre d'itérations nécessaires pour atteindre la convergence de l'algorithme à chaque étape de l'estimation. Ces



(a) estimation initiale (b) cycle 1 (c) cycle 2

(d) cycle 3 (e) cycle 4 (f) résultat final

FIG. 4.8 – Résultats intermédiaires pour l'estimation de la carte de disparité de la paire d'images *Teddy*.

Étape	EAM	`Err`	nombre d'itérations
estimation initiale	1.75	26	–
cycle 1	1.56	21	734
cycle 2	1.44	21	549
cycle 3	1.04	17	257
cycle 4	0.99	16	215
résultat final	0.98	17	201

TAB. 4.3 – Résultats intermédiaires pour l'estimation de la paire d'images *Teddy*.

Technique	Tsukuba			Venus			Teddy		
	$\mathcal{B}_{\text{img}}$	$\mathcal{B}_{\text{disc}}$	$\mathcal{B}_{\text{nontext}}$	$\mathcal{B}_{\text{img}}$	$\mathcal{B}_{\text{disc}}$	$\mathcal{B}_{\text{nontext}}$	$\mathcal{B}_{\text{img}}$	$\mathcal{B}_{\text{disc}}$	$\mathcal{B}_{\text{nontext}}$
AdaptingBP [Klaus *et al.*, 2006]	0.27_3	0.10_1	0.13_3	0.16_1	0.02_1	0.07_2	0.33_1	0.14_1	0.04_1
Gc+Occ [Kolmogorov, Zabih, 2001]	0.10_1	0.11_2	0.03_1	0.36_5	0.08_4	0.18_4	1.03_5	0.37_5	0.32_6
SegTreeDP [Deng, Lin, 2006]	0.31_4	0.13_3	0.17_4	0.19_2	0.02_1	0.05_1	0.66_4	0.27_3	0.15_4
AdaptWeight [Yoon, Kweon, 2006]	0.22_2	0.10_1	0.12_2	0.29_4	0.03_2	0.10_3	0.63_3	0.36_4	0.07_2
SSD+MF [Scharstein, Szeliski, 2002]	1.25_6	0.42_6	0.54_6	0.48_6	0.08_4	0.30_5	1.16_6	0.50_6	0.28_5
Corrélation	1.46_5	0.67_5	0.72_5	0.58_7	0.14_5	0.35_6	1.47_7	0.60_7	0.37_7
Notre approche	0.27_3	0.21_4	0.12_2	0.23_3	0.04_3	0.10_3	0.56_2	0.26_2	0.11_3

TAB. 4.4 – Résultats comparatifs pour les paires d'images *Tsukuba*, *Venus* et *Teddy* en utilisant le critère d'erreur absolue moyenne.

valeurs indiquent que le plus grand nombre d'itérations est requis au premier cycle du processus. Cela s'explique par le fait que les erreurs sur la première estimation initiale sont les plus importantes et que le champ de disparité initial est ainsi le plus éloigné de la solution finale recherchée.

4.2.2.4 Comparaison avec d'autres méthodes de mise en correspondance

Dans ce paragraphe, nous cherchons à comparer les performances de notre méthode à celles des approches développées dans la littérature. Nous utilisons, pour cela, les méthodes les plus récentes et les mieux classées d'après le protocole d'évaluation de Scharstein et Szeliski et dont il est possible de télécharger les résultats directement depuis le site du protocole[3]. Parmi ces méthodes, nous considérons cinq méthodes représentatives de différentes approches de mise en correspondance, en conservant les désignations données par Scharstein et Szeliski :

- la méthode de Klaus *et al.* [Klaus *et al.*, 2006] qui s'appuie sur une optimisation par propagation de croyance, notée AdaptingBP,
- l'algorithme de Kolmogorov et Zabih [Kolmogorov, Zabih, 2001] à base de coupures de graphes, noté Gc+Occ,
- la méthode de Deng et Lin [Deng, Lin, 2006] utilisant la programmation dynamique, notée SegTreeDP,
- la méthode de Yoon et Kweon [Yoon, Kweon, 2006] utilisant des fenêtres adaptatives, notée AdaptWeight,
- l'implémentation de Scharstein et Szeliski [Scharstein, Szeliski, 2002] utilisant des fenêtres multiples, notée SSD+MF .

Notons que les trois premières méthodes sont globales et que les deux dernières sont locales. Les résultats de notre méthode ont été obtenus en considérant trois cycles d'itérations, les ensembles de contraintes S_1, S_2 et S_3 et en ajustant les valeurs des bornes sur ces ensembles à partir des cartes de disparité théoriques disponibles. Pour comparer nos résultats à ceux obtenus par les différentes méthodes citées, nous utilisons le critère d'erreur absolue moyenne évalué en ne prenant en compte que les pixels visibles, dans les trois zones de la figure 4.3, à savoir toute l'image, notée $\mathcal{B}_{\text{img}}$, les zones des discontinuités, dont l'union est notée $\mathcal{B}_{\text{disc}}$ et les zones non texturées, dont l'union est notée $\mathcal{B}_{\text{nontext}}$. L'ensemble des résultats, pour les paires d'images *Tsukuba*, *Venus* et *Teddy*, est regroupé dans le tableau 4.4, où nous avons indiqué le rang de chaque méthode parmi toutes les méthodes évaluées. Nous y donnons également les résultats obtenus avec la méthode de corrélation nous servant d'initialisation. Grâce à ce tableau, nous pouvons remarquer que notre méthode, même si elle n'obtient pas le meilleur classement, soutient la comparaison avec les méthodes les plus performantes. Ceci confirme la fiabilité et l'efficacité de notre estimation. La méthode AdaptingBP basée sur la technique de propagation de croyance est en général la mieux classée. Notons que la propagation de croyance a été comparée à la méthode de coupures de graphes par Tappen et Freeman [Tappen, Freeman, 2003] qui ont montré que ces deux techniques d'optimisation récentes sont aussi performantes l'une que l'autre. Cependant, le comportement plus performant de la méthode AdaptingBP par rapport à celui de Gc+Occ est dû principalement à l'utilisation de la segmentation en couleur. En effet, ces dernières années ont vu se multiplier les publications

[3] http ://vision.middlebury.edu/stereo/eval/

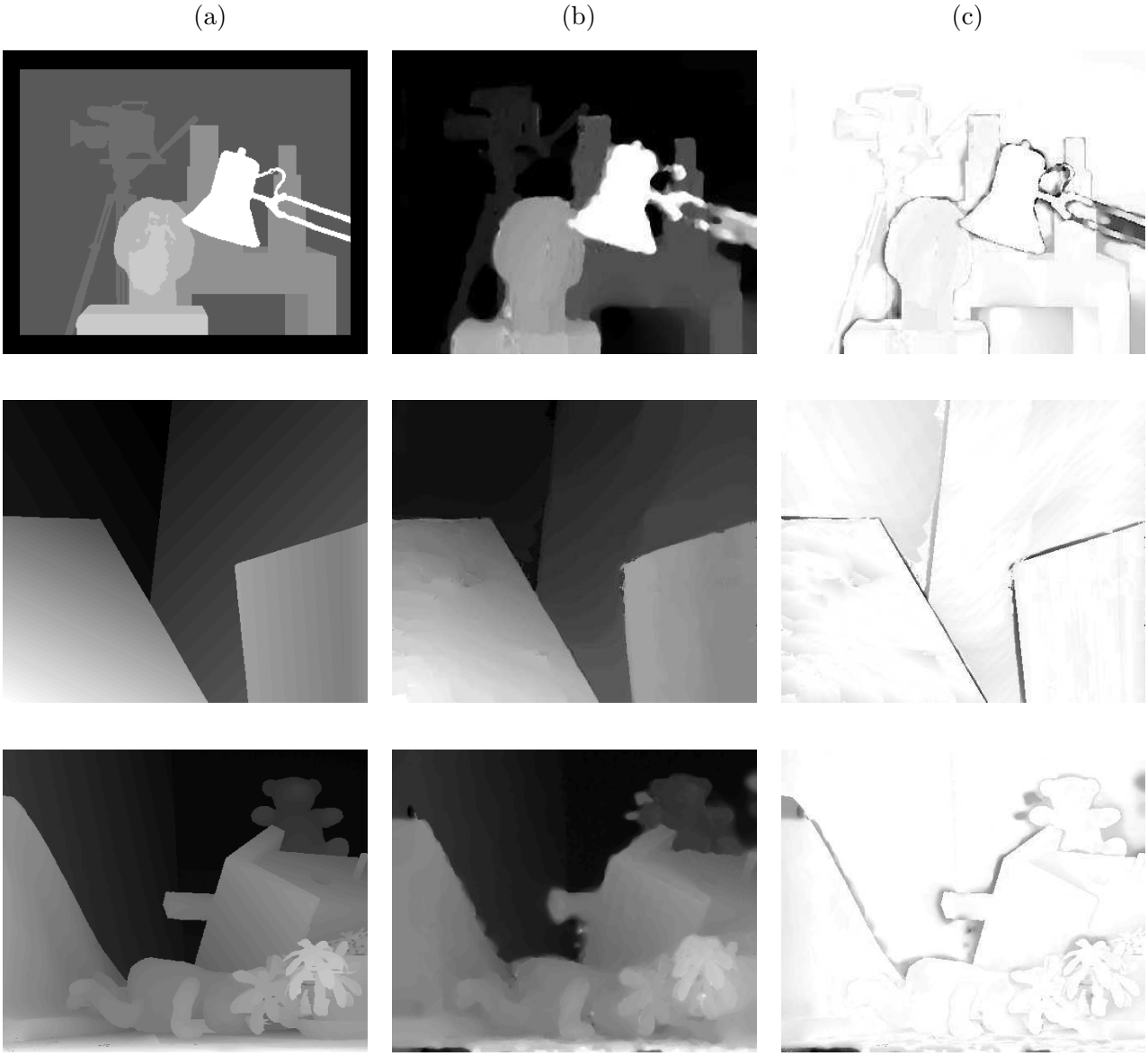


FIG. 4.9 – Résultas obtenus avec les paires d'images *Tsukuba*, *Venus* et *Teddy*. (a) vraie disparité (b) carte de disparité estimée (c) carte de pixels erronés : plus le pixel est foncé et plus l'appariement est mauvais. Ces résultats sont obtenus en considérant trois cycles de réinitialisations.

présentant des méthodes de mise en correspondance qui intègrent une technique de segmentation afin d'améliorer leurs performances, en particulier dans les zones non texturées et les zones des discontinuités. La méthode SegTreeDP consiste ainsi à effectuer une segmentation en régions de l'image de référence et à utiliser la programmation dynamique pour mettre en correspondance sur chaque droite épipolaire les segments appartenant à la même région. Enfin, notons que la méthode AdaptWeight, bien qu'étant une méthode locale, obtient de bons résultats mais, qu'en revanche, le résultat des autres approches locales, à savoir la méthode SSD+MF et celle par simple corrélation, est assez mauvais.

Les cartes de disparités estimées par notre méthode pour les paires d'images *Tsukuba*, *Venus* et *Teddy* sont présentées à la figure 4.9. Nous montrons sur cette figure, pour chaque paire d'images, l'erreur absolue sous la forme d'une carte de différence, que nous avons inversé pour avoir une meilleure visibilité. Les pixels foncés représentent ainsi les appariements erronés et les pixels blancs indiquent une ressemblance quasi-parfaite avec la carte de disparité théorique (cf. figure 4.9(c)). Cette figure montre que notre méthode obtient de bons résultats et permet de s'affranchir complètement des problèmes liés aux zones non texturées mais, qu'en revanche, elle se comporte moins bien dans les zones des discontinuités, qui néanmoins, sont globalement bien préservées (cf. figure 4.9(b)).

4.2.3 Introduction de contraintes définies dans le domaine des ondelettes

Nous introduisons dans cette partie les contraintes de régularisation définies dans le domaine des ondelettes, à savoir les contraintes S_4 et S_5, et nous comparons leurs performances à celles d'une régularisation par variation totale. En effet, bien qu'il soit avantageux d'introduire le plus grand nombre de contraintes possibles sur la solution, nous avons observé que l'utilisation simultanée des contraintes de régularisation avec préservation des discontinuités associées aux ensembles S_2, S_4 et S_5 augmente sensiblement le temps de calcul tout en améliorant de très peu les résultats. C'est pourquoi, nous avons choisi de ne considérer qu'une seule de ces contraintes que l'on ajoute aux contraintes associées aux ensembles S_1 et S_3.

4.2.3.1 Choix de la base d'ondelettes

Le choix d'une base d'ondelettes bien adaptée aux images de disparité constituées de zones homogènes séparées par des discontinuités, joue un rôle important dans la régularisation associée aux contraintes S_4 et S_5. Dans ce travail, où nous nous intéressons essentiellement aux coefficients de détails horizontaux et verticaux, nous avons choisi une base de Haar bien adaptée aux fortes discontinuités. Nous avons aussi évalué les ondelettes de Daubechies, les Symlets et les Coiflets. Les erreurs absolues moyennes, obtenues sur une portion 256×256 de la paire d'images *Sawtooth*, en considérant ces différentes bases d'ondelettes sont données dans le tableau 4.5. Notons que, dans le but de mieux évaluer l'impact du choix de l'ondelette sur la performance des contraintes S_4 et S_5, nous n'avons pas pris en compte la contrainte de régularisation de Nagel-Enkelmann traduite par l'ensemble S_3. Grâce au tableau 4.5, nous pouvons observer que le choix de l'ondelette n'a que peu d'influence sur le résultat de l'estimation. Toutefois, la simplicité de mise en œuvre et les résultats légèrement meilleurs de l'ondelette de Haar confirment notre choix de cette base. Par ailleurs, les gains observés avec une décomposition en ondelettes discrète redondante justifient la nécessité de recourir à une

		sans invariance par translation		avec invariance par translation	
Ondelettes	Niveau	$S_1 \cap S_4$	$S_1 \cap S_5$	$S_1 \cap S_4$	$S_1 \cap S_5$
Coiflets 2	$l = 1$	0.5754	0.5713	0.4548	0.4533
	$l = 2$	0.5746	0.5688	0.4536	0.4529
Symlets 4	$l = 1$	0.5430	0.5417	0.4399	0.4381
	$l = 2$	0.5426	0.5412	0.4387	0.5473
Daubechies 4	$l = 1$	0.5536	0.5532	0.4335	0.4329
	$l = 2$	0.5531	0.5529	0.4330	0.4327
Haar	$l = 1$	0.5411	0.5388	0.4296	0.4287
	$l = 2$	0.5398	0.5376	0.4292	0.4263

TAB. 4.5 – Erreurs absolues moyennes obtenues sur une portion 256×256 de la paire d'image *Sawtooth*.

transformation invariante par translation. Pour mettre en œuvre les contraintes S_4 et S_5, nous avons ainsi implémenté la décomposition invariante par translation, telle que nous l'avons décrite dans le chapitre précédent (cf. paragraphe 3.4.5.1), sur la base de Haar en considérant un seul niveau de résolution. En effet, en comparant les résultats obtenus avec une décomposition sur un puis deux niveaux de résolution (cf. tableau 4.5), nous n'observons qu'une légère amélioration pour un coût de calcul plus élevé. Il nous paraît donc justifié de ne faire la décomposition que sur un seul niveau de résolution. Ceci revient à effectuer les transformations pour quatre décalages circulaires de l'image, que nous avons alternés de manière à ce qu'un seul ensemble, associé à une transformation suivant un décalage $\mathrm{d} \in \{0,1\}^2$ donné, soit utilisé à chaque itération de l'algorithme.

Remarque 4.1 *Notons que l'expression* (3.73) *définissant la norme d'une image* $u \in \mathcal{H}$ *sur l'espace de Besov* $B^{\sigma}_{p,q}$ *en termes de ses coefficients en ondelettes, n'est valide que si l'ondelette choisie possède* $N+1$ *moments nuls avec* $N \geq \sigma$*. Or, l'ondelette de Haar possédant un seul moment nul (cf. Annexe A), la semi-norme associée à l'ensemble* S_4 *(et celle associée à l'ensemble* S_5 *étant équivalente), en choisissant cette ondelette, ne peut plus rigoureusement être considérée comme une semi-norme sur l'espace de Besov* $B^1_{1,1}$*.*

4.2.3.2 Comparaison des contraintes de régularisation

Pour illustrer les résultats des différentes contraintes de régularisation avec préservation des discontinuités, nous utilisons la paire d'images de synthèse *Corridor* de la figure 4.1. Cette paire, très utilisée pour l'évaluation des algorithmes de mise en correspondance, comporte peu d'occultations, une texture répétitive et des régions homogènes. La vérité terrain étant disponible, nous avons utilisé le critère EAM pour évaluer quantitativement la qualité des estimations. La figure 4.10 présente une comparaison entre la disparité estimée par notre méthode en utilisant les contraintes de régularisation introduites dans le domaine des ondelettes ($S = S_1 \cap S_3 \cap S_4$ ou $S = S_1 \cap S_3 \cap S_5$) et en utilisant la régularisation par variation totale ($S = S_1 \cap S_2 \cap S_3$). Nous avons choisi pour ces trois estimations de démarrer notre algorithme avec une estimée initiale donnée par la méthode de corrélation et de considérer par la suite trois cycles de réinitialisations. La carte de disparité initiale, pour le premier cycle, est

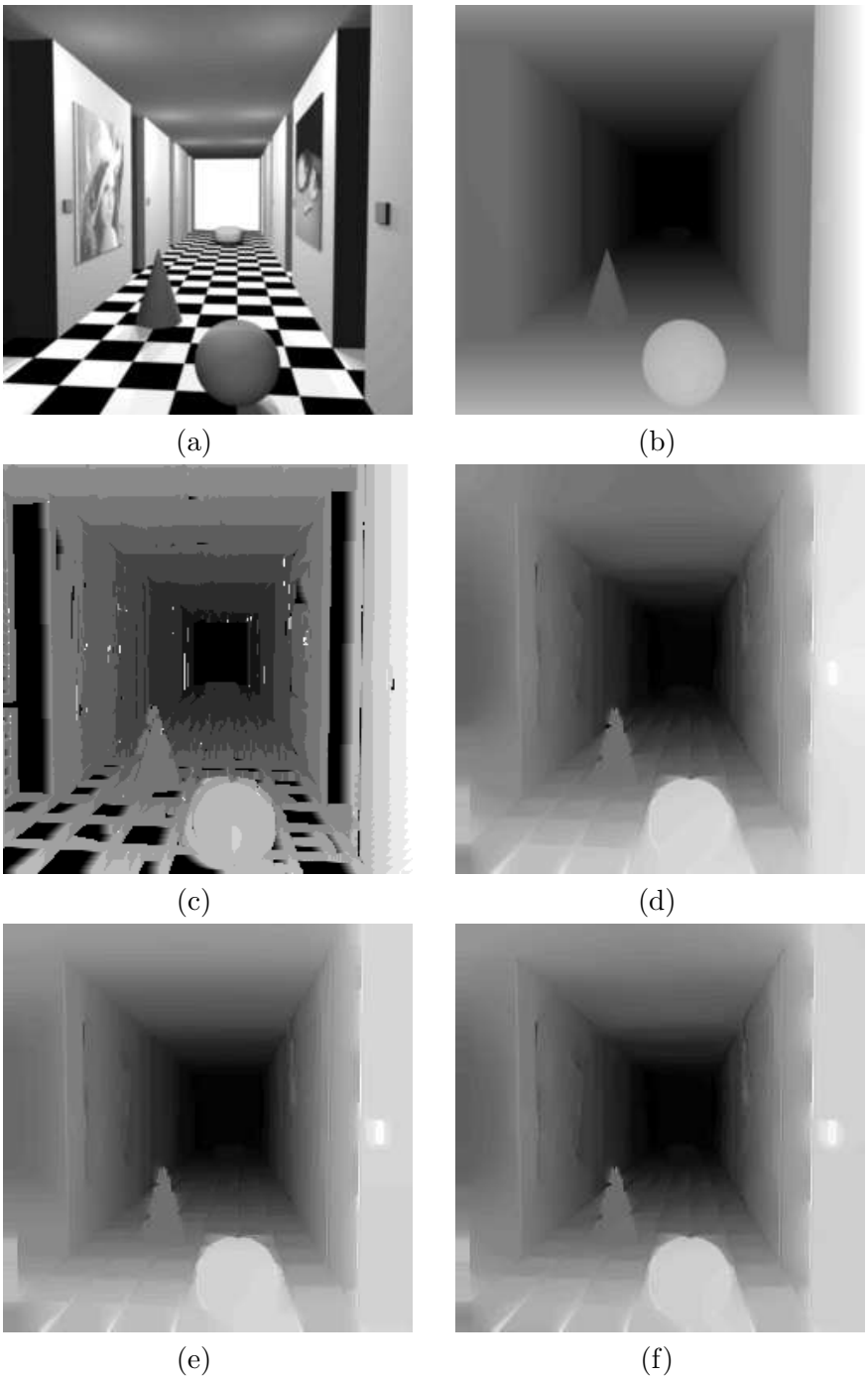


FIG. 4.10 – Résultats pour la paire stéréoscopique *Corridor*. (a) Image de gauche. (b) Disparité théorique. (c) Disparité initiale. Disparité estimée par notre méthode en considérant une régularisation par l'ensemble (d) S_2 (e) S_4 (f) S_5.

présentée à la figure 4.10(c). Celle-ci montre de nombreuses erreurs de mise en correspondance dûes aux motifs répétitifs du sol et au manque de texture sur les murs. En effet, deux points situés sur ces régions ont des attributs d'intensité identiques dans un voisinage donné, ce qui pose un problème pour le score de corrélation et engendre donc des erreurs d'appariement importantes. En observant l'ensemble de ces résultats, nous constatons la capacité de notre méthode, pour les différentes régularisations considérées, à fournir une carte de disparité précise où les larges zones homogènes tels que le plafond et les murs sont bien estimées et les contours des objets sont préservés. Notons qu'il peut parfois paraître difficile de distinguer visuellement les résultats des différentes régularisations mais, qu'en revanche, les erreurs absolues moyennes données dans le tableau 4.6 montrent clairement que la régularisation par la semi-norme quadratique, relative à l'ensemble S_5, conduit au meilleur résultat.

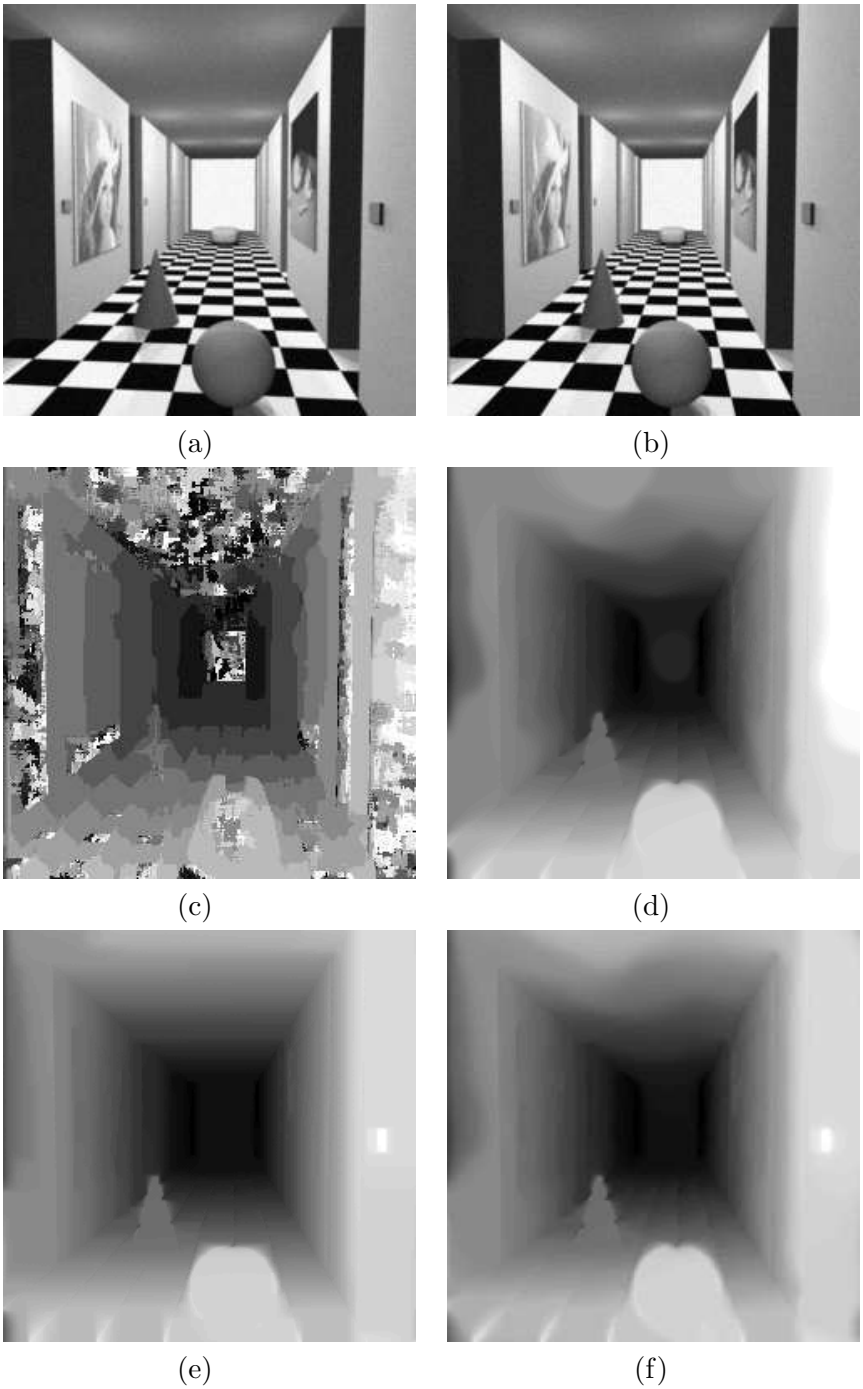
Technique	Sans ajout de bruit	niveaux de bruit $\sigma_n^2 = 1$	$\sigma_n^2 = 10$	$\sigma_n^2 = 100$
Corrélation	0.52	0.65	0.99	1.26
Régularisation par variation totale	0.24	0.29	0.33	0.50
Régularisation par l'ensemble S_4	0.21	0.26	0.30	0.46
Régularisation par l'ensemble S_5	**0.19**	**0.24**	**0.26**	**0.41**

TAB. 4.6 – Résultats comparatifs sur la paire d'images *Corridor* utilisant l'erreur absolue moyenne.

Par ailleurs, pour évaluer la robustesse de la méthode proposée en présence de bruit, nous avons ajouté aux images de la paire *Corridor* un bruit gaussien de moyenne nulle et d'écart type variable σ_n. Les erreurs obtenues pour trois niveaux différents de bruit sont indiqueés dans le tableau 4.6. Les figures 4.11 et 4.12 présentent la paire d'image *Corridor* à laquelle nous avons appliqué un bruit de variance σ_n^2 respectivement égale à 10 et 100, ainsi que les champs de disparité estimés. Ces résultats, avec ceux du tableau 4.6, mettent en évidence le bon comportement de notre méthode en présence de bruit additif gaussien. En effet, nous pouvons observer que le résultat de l'approche par corrélation se dégrade au fur et à mesure que l'amplitude de bruit augmente alors que le résultat de notre méthode reste précis et la qualité de l'image de disparité calculée demeure satisfaisante. Nous pouvons également remarquer que dans le cas de ces images bruitées, la régularisation dans le domaine des ondelettes apporte de nettes améliorations par rapport à la variation totale, tant visuellement que quantitativement, et permet de mieux s'affranchir des erreurs liées à l'initialisation.

4.2.3.3 Implémentation parallèle et coût de calcul

La parallélisation permet, lorsque plusieurs processeurs parallèles sont disponibles, de réduire le temps de calcul. Dans nos expériences précédentes, nous avons implémenté notre algorithme d'optimisation de manière cyclique en ignorant la possibilité de calcul parallèle. Toutefois, comme nous l'avons déjà indiqué dans le chapitre 3, grâce à sa structure itérative par bloc, cet algorithme peut être efficacement mis en œuvre sur une architecture à processeurs parallèles en adaptant, par le choix des blocs d'ensembles, la charge de calcul de chaque itération à la puissance des processeurs parallèles disponibles. Dans l'objectif de comparer la régularisation dans le domaine des ondelettes à celle dans le domaine spatial, nous avons



(a) (b) (c) (d) (e) (f)

FIG. 4.11 – Résultats pour la paire *Corridor* dégradée par un bruit gaussien de variance 10. (a) Image de gauche. (b) Image de droite. (c) Disparité initiale. Disparité estimée par notre méthode en considérant une régularisation par l'ensemble (d) S_2 (e) S_4 (f) S_5.

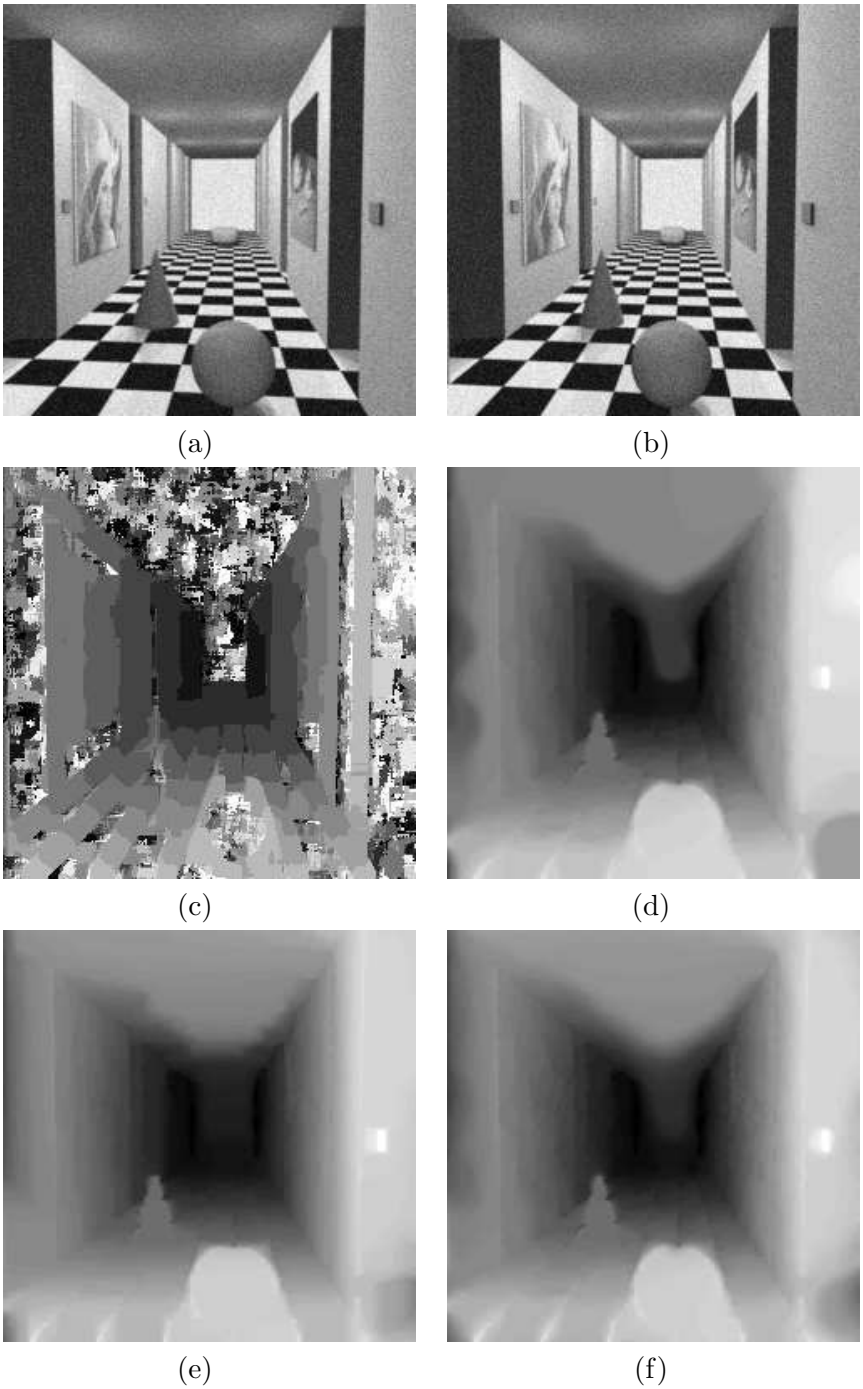


FIG. 4.12 – Résultats pour la paire *Corridor* dégradée par un bruit gaussien de variance 100. (a) Image de gauche. (b) Image de droite. (c) Disparité initiale. Disparité estimée par notre méthode en considérant une régularisation par l'ensemble (d) S_2 (e) S_4 (f) S_5.

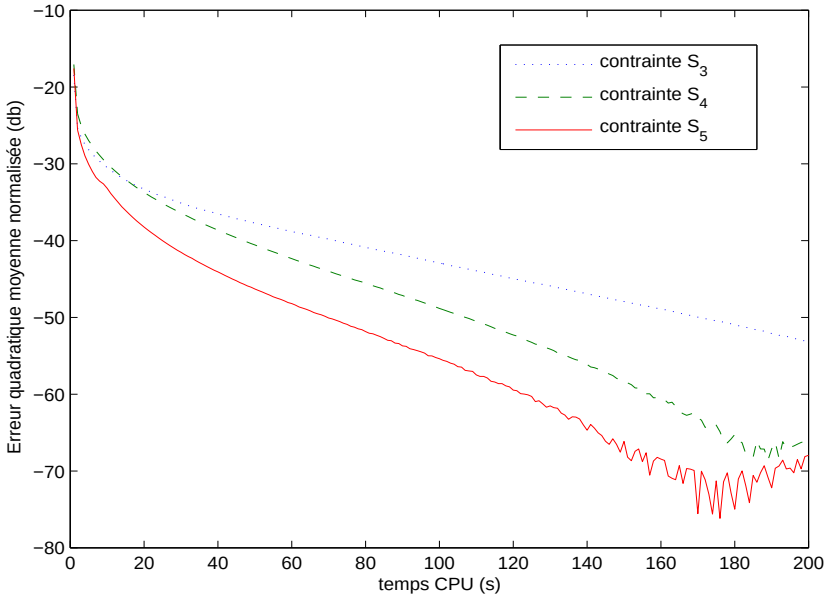


FIG. 4.13 – Erreur quadratique moyenne normalisée en fonction du temps CPU.

envisagé une telle stratégie en se restreignant à la parallélisation de la contrainte de régularisation. Il est évident que cette parallélisation est plus favorable à la régularisation dans le domaine des ondelettes qu'à la régularisation par variation totale. En effet, dans le cas des ensembles S_4 et S_5, au lieu d'alterner les quatre contraintes $S_4^{(\mathrm{d})}$ et $S_5^{(\mathrm{d})}$ suivant un décalage circulaire d donné, il est possible, dans le cadre d'une architecture parallèle, d'assigner chaque contrainte à un processeur. Cependant, ne disposant pas d'une telle architecture, nous avons simulé la mise en œuvre en parallèle de ces contraintes sous le logiciel de calcul Matlab, en supposant que quatre processeurs parallèles sont disponibles. La figure 4.13 illustre la convergence de l'algorithme, pour les différentes régularisations, en montrant l'évolution de l'erreur moyenne normalisée en fonction du temps CPU. Il est clair, d'après cette figure, que l'utilisation des contraintes définies dans le domaine des ondelettes, notamment la contrainte S_5, permet d'accélérer considérablement la convergence de l'algorithme.

Notons, par ailleurs, que les contraintes S_4 et S_5, définies dans le domaine des ondelettes, ont été introduites en employant les projections sous-différentielles sur les ensembles associés S_4^{d} et S_5^{d}, mais comme la projection exacte sur ces ensembles peut être calculée de manière quasi-explicite (cf. paragraphe 3.4.5.4), il peut être intéressant de comparer les performances des deux types de projections. Les comparaisons menées, dans ce but, nous ont ainsi permis de conclure que les cartes de disparité obtenues suivant les deux procédés de projection, sont proches mais, qu'en revanche, la projection exacte est plus coûteuse en temps de calcul. Celle-ci fait, en effet, appel, à chaque itération de l'algorithme, à une procédure de dichtomie qui est à l'origine de la complexité plus élevée de cette méthode.

4.3 Résultats sur des images réelles

Nous allons dans cette section présenter les résultats obtenus par notre méthode, en considérant la régularisation par l'ensemble S_5, sur des paires stéréoscopiques réelles. À titre de

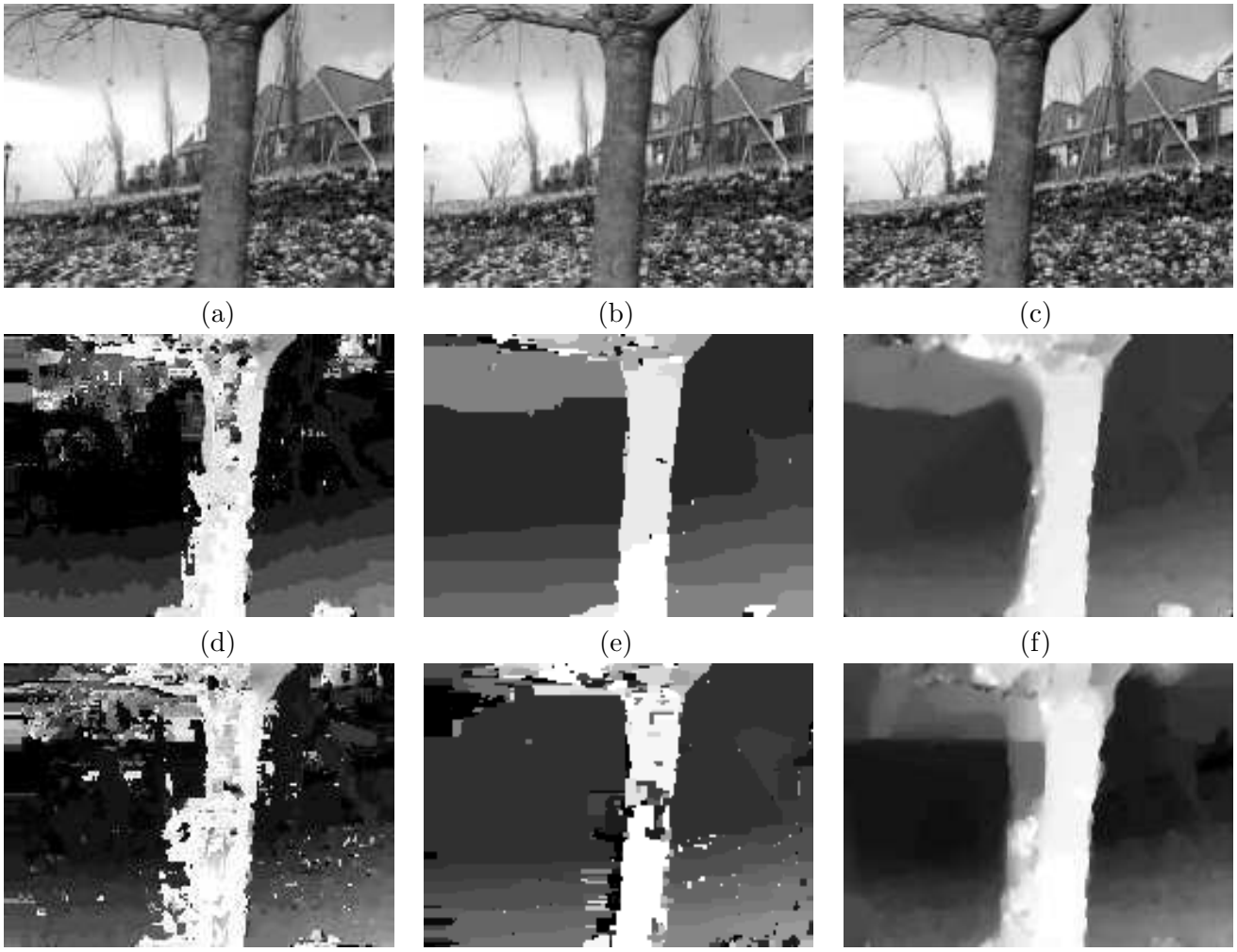
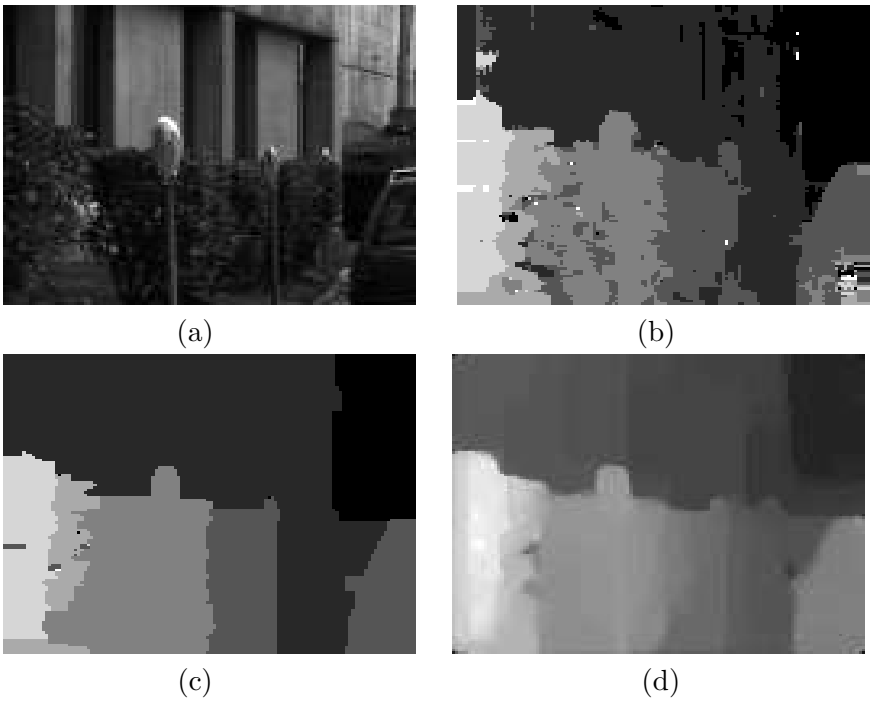


FIG. 4.14 – Séquence d'images *flower garden*. En haut : (a) image 1 (b) image 4 (c) image 10. Au milieu : Résultats obtenus sur les images 1 et 4 par (d) corrélation (e) Gc+Occ (f) notre méthode. En bas : Résultats obtenus sur les images 1 et 10 par (g) corrélation (h) Gc+Occ (i) notre méthode.



(a) (b)

(c) (d)

FIG. 4.15 – Résultats pour les images *parking meter*. (a) Image de gauche (b) méthode par corrélation (c) méthode de coupures de graphes (d) notre méthode.

comparaison, nous reportons également les résultats de la méthode par corrélation et ceux de la méthode Gc+Occ de Kolmogorov et Zabih [Kolmogorov, Zabih, 2001] utilisant les coupures de graphes. Cependant, la vérité terrain étant inconnue, seules des évaluations visuelles peuvent être effectuées.

4.3.1 Séquence *Flower garden*

La première expérimentation concerne les images extraites de la séquence vidéo *flower garden*, présentant un faible mouvement de translation entre deux images successives. La figure 4.14 montre les images 1, 4 et 10 de cette séquence. Nous avons, dans un premier temps, expérimenté notre méthode sur les images 1 et 4, ayant un déplacement maximal de 12 pixels entre les deux vues. Dans un second temps, pour évaluer la robustesse de notre méthode vis-à-vis de grands déplacements, nous l'avons testée sur les images 1 et 10 dont la disparité maximale est de 25 pixels. Les cartes de disparité estimées sont montrées à la figure 4.14. En comparant ces résultats, nous pouvons observer que la plage de variation de disparité n'a que peu d'impact sur la qualité de la solution. En effet, les résultats de notre estimation, aussi bien dans le cas d'images proches (images 1 et 4) que dans le cas d'images éloignées (images 1 et 10), sont satisfaisants et de meilleure qualité que ceux obtenus par les méthodes par corrélation et à base de coupures de graphe.

4.3.2 Paire d'images *Parking meter*

Nous présentons maintenant les résultats obtenus sur la paire d'images *Parking meter*, issue de la base de données JISCT[4]. L'image de gauche de cette paire stéréoscopique est présentée à la figure 4.15(a). Celle-ci montre trois parcmètres ainsi qu'une partie d'une voiture garée devant un immeuble. Les résultats de notre méthode sur cette paire d'images sont satisfaisants et cohérents avec la structuration des éléments de la scène observée. De plus, ces résultats sont nettement meilleurs que ceux obtenus avec la méthode de corrélation. En ajustant le paramètre de régularisation de la méthode par coupures de graphes, on parvient également à obtenir de bons résultats avec cette méthode sur la paire d'images considérée.

4.4 Conclusion

Dans ce chapitre, nous avons évalué les performances de notre méthode d'estimation de la disparité sur différentes paires d'images synthétiques et réelles. Pour cela, nous avons considéré une fonction objectif quadratique convexe et nous avons mis en place l'algorithme itératif par bloc décrit dans le chapitre précédent.
En exploitant, dans un premier temps, les contraintes définies dans le domaine spatial, nous avons mené une première série d'expérimentations grâce auxquelles, nous avons pu :

- mesurer l'impact de chaque contrainte sur la solution obtenue et confirmer ainsi l'avantage d'introduire des contraintes multiples sur la solution ;
- évaluer la robustesse de notre méthode vis-à-vis du choix des bornes sur les ensembles des contraintes ;
- montrer l'amélioration apportée par le processus de réinitialisation sur l'affinement des résultats ;
- comparer les performances de notre méthode à celles des approches concurrentes disponibles dans la littérature.

En introduisant, dans un second temps, les contraintes définies dans le domaine des ondelettes, nous nous sommes parvenus à :

- comparer les performances de la régularisation associée à ces contraintes à celle d'une régularisation par variation totale ;
- évaluer la robustesse de notre approche en présence d'un bruit additif gaussien ;
- illustrer la convergence de l'algorithme, notamment dans le cas d'une implémentation parallèle.

Les évaluations et les comparaisons établies confirment ainsi la robustesse et la fiabilité de la méthode proposée, qui permet notamment de s'affranchir des problèmes liés aux zones des discontinuités et aux zones non texturées de l'image. Dans le chapitre suivant, nous étendons notre méthode pour prendre en compte une autre difficulté de la mise en correspondance stéréoscopique, à savoir les variations d'illumination induites entre les images stéréoscopiques de la scène observée.

[4] http ://vasc.ri.cmu.edu/idb/html/jisct/

Chapitre 5

Prise en compte des variations d'illumination

5.1 Introduction

L'un des problèmes les plus délicats en vision par ordinateur, notamment dans le cadre de la mise en correspondance stéréoscopique, est le problème des changements d'illumination qui peuvent affecter deux vues d'une même scène. Les différentes méthodes proposées dans la littérature pour résoudre ce problème reposent essentiellement sur des techniques de prétraitement des images, des mesures de corrélation robustes aux variations d'illumination ou une modélisation affine de ces variations.

Dans ce chapitre, nous présentons une nouvelle méthode, qui étend l'approche décrite dans le chapitre 3, pour résoudre le problème d'estimation de la disparité en présence de changements d'illumination dans la scène observée. Nous proposons dans un premier temps un modèle d'illumination multiplicatif qui prend en compte les variations spatiales de luminosité entre les deux images de la paire stéréoscopique. À partir du modèle établi, nous formulons ensuite le problème de mise en correspondance comme un problème de programmation convexe où une fonction objectif quadratique convexe, intégrant un facteur d'illumination multiplicatif, est minimisée sur l'intersection d'ensembles convexes construits à partir des connaissances *a priori* et des observations. Pour résoudre le problème d'optimisation sous contraintes considéré, nous utilisons enfin l'algorithme itératif décrit au chapitre 3, que nous adaptons pour calculer simultanément les deux champs de variables qui représentent la disparité et les changements d'illumination.

Nous passons brièvement en revue dans la section 5.2 les principales méthodes de mise en correspondance existantes prenant en compte les changements d'illumination. La section 5.3 s'intéresse à la modélisation des variations spatiales de luminance à l'aide d'un modèle d'illumination multiplicatif. La section 5.4 présente notre approche qui permet d'intégrer dans un même formalisme le problème de mise en correspondance et celui de la prise en compte des changements d'illumination. Des résultats expérimentaux obtenus sur des images synthétiques et réelles montrent enfin l'efficacité et la robustesse de notre méthode.

5.2 Passage en revue des méthodes existantes

La plupart des méthodes de mise en correspondance reposent sur l'hypothèse de conservation de la luminance qui établit que les projections d'un même point 3D ont la même intensité dans les deux images stéréoscopiques. Cependant, cette contrainte physique, qui impose l'utilisation d'un modèle de réflectance lambertien (cf. paragraphe 5.3.1.1), n'est pas vérifiée en présence de surfaces non-lambertiennes [1], même si l'on ne prend pas en considération le bruit de mesure ou la différence de sensibilité entre les caméras utilisées. Par conséquent, bien que fréquemment utilisée, cette hypothèse si elle n'est pas respectée, peut entraîner de faux appariements et dégrader ainsi les résultats de la mise en correspondance.

Nous présentons dans ce qui suit les principales méthodes qui ont été proposées pour remédier à ce problème en les classant en trois catégories : les méthodes avec prétraitement, les méthodes utilisant des mesures de corrélation robustes et les méthodes avec modélisation affine.

5.2.1 Méthodes avec prétraitement

Pour compenser l'effet des variations d'illumination induites entre deux prises de vue d'une même scène, une approche simple, assez couramment employée, consiste à appliquer un prétraitement aux images avant de procéder à la mise en correspondance stéréoscopique. Ainsi, des opérations consistant à appliquer des transformations, souvent non linéaires, aux niveaux de gris des images permet de réduire le bruit ou de rehausser le contraste dans les images. Partant de l'idée que deux images d'une même paire stéréoscopique devraient se ressembler statistiquement, le principe de la méthode proposée dans [Konrad, Mancini, 1998] est d'égaliser leur moyenne et leur variance par une simple transformation linéaire appliquée à la luminance de l'une des deux images. En supposant que μ_g et σ_g sont respectivement la moyenne et l'écart type de l'image de gauche I_g et μ_d et σ_d sont ceux de l'image de droite I_d à ajuster, la transformation linéaire globale à appliquer à l'image I_d est alors :

$$I_d^{'} = \frac{\sigma_g}{\sigma_d} I_d + (\mu_g - \mu_d \frac{\sigma_g}{\sigma_d}), \tag{5.1}$$

où $I_d^{'}$ est la version corrigée de I_d.

En se basant sur des techniques d'analyse d'histogrammes, d'autres méthodes de prétraitement ont été conçues pour corriger les changements de luminosité sur une paire d'images stéréoscopique. Ces méthodes reposent sur l'observation que, lorsque la contrainte de conservation de la luminance est respectée, les histogrammes des intensités de gris dans les deux images sont similaires. La technique de spécification d'histogrammes [Gonzalez, Woods, 1993] consiste ainsi à ramener l'histogramme de l'une des deux images à l'histogramme de l'autre image laissé invariant. A l'issue de cette opération, on génère une nouvelle paire stéréoscopique constituée d'images d'histogrammes similaires. Une autre méthode basée sur la ressemblance d'histogrammes a été proposée dans [Cox *et al.*, 1995]. Elle consiste à apparier les histogrammes des deux images à l'aide d'une méthode d'optimisation globale basée sur la

[1] Une surface lambertienne est une surface diffusant la lumière de façon uniforme quelle que soit la direction d'observation.

programmation dynamique. Cette méthode s'avère meilleure que les méthodes de spécification d'histogrammes classiques, mais elle n'est appropriée que pour compenser des variations globales d'illumination.

5.2.2 Méthodes utilisant des mesures de corrélation robustes

D'autres ensembles de travaux adaptent des techniques de mise en correspondance classiques en utilisant des mesures de corrélation robustes aux changements d'illumination. Certains auteurs proposent d'appliquer les mesures de similarité habituellement utilisées avec des images de niveaux de gris, à des images dérivées comme par exemple l'image du gradient ou du laplacien. Une nouvelle mesure appelée corrélation des champs de gradients a été ainsi proposée par Crouzil *et al.* [Crouzil *et al.*, 1996] dans le cadre d'une mise en correspondance par corrélation. Les auteurs montrent que l'utilisation de la similarité de vecteurs gradients permet, en effet, d'obtenir une mesure invariante aux changements globaux d'illumination et de réduire, par là-même, la sensibilité des approches locales aux changements de luminosité.

Une autre mesure invariante aux transformations affines d'intensité des images est la corrélation croisée normalisée NCC (*Normalized Cross Correlation*), utilisée notamment dans [Zitnick, Kanade, 2000, Tsai *et al.*, 2003]. Dans le cas de changements d'intensité plus complexes que les changements affines, il existe la classe des mesures non paramétriques qui exploitent des outils de statistiques non paramétriques. Citons notamment les mesures de rang et de recensement proposées par Zabih et Woodfill [Zabih, Woodfill, 1994] et les mesures ordinales de Bhat et Nayar [Bhat, Nayar, 1998]. Ces mesures gèrent les transformations non linéaires, comme les reflets spéculaires, ainsi que le bruit des capteurs.

Récemment, l'information mutuelle, qui est une mesure de la dépendance statistique entre deux variables, a été aussi utilisée comme un critère de similarité [Egnal, 2000, Kim *et al.*, 2003, Fookes *et al.*, 2002, Hirschmüller, 2005]. Cette mesure, qui jusqu'à présent a été essentiellement utilisée pour mettre en correspondance deux images dans un cadre de recalage d'images [Viola, Wells, 1997], a l'avantage d'être insensible à différentes transformations complexes d'intensité des images. Utilisée dans le cadre d'une mise en correspondance par corrélation [Egnal, 2000], elle s'est montrée robuste vis-à-vis des changements d'illumination. Cependant, elle se comporte très médiocrement en présence d'un bruit gaussien en raison du faible nombre d'échantillons dans les fenêtres de corrélation. Pour pallier cet inconvénient, Hirschmüller propose de calculer l'information mutuelle globalement sur toute l'image. Sa méthode de mise en correspondance semi-globale, présentée dans [Hirschmüller, 2005], permet ainsi d'obtenir de bons résultats et s'est avérée particulièrement pertinente pour apparier des images avec des changements globaux d'illumination.

5.2.3 Méthodes avec modélisation affine

Le principe de ces méthodes est de compenser les changements affines d'illumination de la scène en exprimant la relation entre deux images d'une paire stéréoscopique par l'équation suivante :

$$I_d = \beta I_g + \gamma, \tag{5.2}$$

où I_g et I_d représentent respectivement l'intensité lumineuse de l'image de gauche et de droite et β et γ sont des paramètres à déterminer [Cox, 1994]. Notons que, contrairement à la transformation linéaire abordée au paragraphe 5.2.1, où les paramètres de la transformation sont exprimés en fonction des moyennes et des écarts types de l'image de gauche et de droite (cf. équation (5.1)), β et γ sont ici obtenus par minimisation d'un critère J_a associé au modèle affine (5.2). Ce critère s'écrit :

$$J_a(u(s), \beta, \gamma) = \sum_{(x,y)\in\mathcal{W}_s} \varphi_a(\beta I_g(x,y) - I_d(x - u(s), y) + \gamma), \tag{5.3}$$

où s est la position spatiale d'un point dans l'image, $u(s)$ est la disparité associée au point s et φ_a est une mesure d'écart. La région $\mathcal{W}_s \subset \mathbb{N}^2$ représente une fenêtre centrée sur le pixel s et, dans le cas d'une optimisation globale du critère (5.3), le support de toute l'image.

Le modèle affine remplace donc l'équation (3.1) traduisant l'hypothèse de conservation de la luminance pour prendre en compte les changements de luminosité existant entre deux images. Il suppose que β et γ sont constants sur $\mathcal{W}_s$ et stipule implicitement que les variations d'illumination de la scène sont faibles. L'étude menée dans [Cox *et al.*, 1995] sur quelques images de la base de données JISCT [2] a montré, cependant, que le modèle d'illumination affine ne permet pas de compenser la totalité des variations d'illumination susceptibles de se produire entre les deux images. C'est le cas notamment lorsque la scène présente des surfaces avec des propriétés de réflectance différentes.

Une autre alternative pour aborder le problème de mise en correspondance en présence de changements de luminosité dans la scène observée consiste à modéliser localement les variations d'illumination. Dans la section qui suit, nous justifions de manière physique un modèle simple d'illumination local plus approprié que le modèle affine, dans le sens où il ne requiert pas l'hypothèse de constance des paramètres du modèle sur une région et autorise ainsi une plus grande variabilité spatiale des changements d'illumination.

5.3 Modélisation des variations d'illumination

Dans le contexte du calcul du flot optique, plusieurs approches modélisant les changements spatiaux d'illumination dans une séquence d'images ont été proposées. Black *et al.* [Black *et al.*, 2000] considèrent les changements d'apparence dans une séquence d'images comme étant issus d'une combinaison complexe d'évènements et de processus et proposent un cadre statistique robuste pour modéliser ces variations. Dans [Hampson, Pesquet, 2000], un schéma récursif a été introduit pour estimer conjointement le champs de mouvement et les changements d'illumination en utilisant un modèle d'illumination multiplicatif. Kim *et al.* [Kim *et al.*, 2005b] proposent un modèle plus général qui autorise à la fois les variations spatiales d'illumination additives et multiplicatives.

Dans le cadre de l'estimation de la disparité, il existe très peu de travaux consacrés à la modélisation des variations d'illumination dans une scène. La plupart des approches proposées utilisent le modèle affine de l'équation (5.2). Cependant, comme nous l'avons souligné au paragraphe 5.2.3, ce modèle n'est pas très efficace et ne permet pas de prendre en compte

[2] http ://vasc.ri.cmu.edu/idb/html/jisct/

les variations locales d'illumination. Un modèle d'illumination qui permet de compenser au mieux les changements d'illumination a été proposé dans [Gennert, 1988] où Gennert a montré que les intensités des deux points homologues sont reliées par un terme multiplicatif qui varie spatialement. Ce modèle s'avère efficace mais l'inconvénient majeur de la méthode considérée est qu'elle requiert la minimisation d'une fonction de coût non convexe basée sur la résolution des équations d'Euler-Lagrange associées. Récemment, Zhang *et al.* ont adopté ce modèle d'illumination multiplicatif et ont modélisé le problème de mise en correspondance par des champs de Markov [Zhang *et al.*, 2006]. Ils ont ensuite employé une méthode d'optimisation itérative basée sur la technique de propagation de croyance (*belief propagation*) pour estimer conjointement l'image de disparité et les variations d'illumination. Les résultats présentés dans cet article montrent l'efficacité du modèle choisi. Cependant, la méthode d'optimisation utilisée est coûteuse en temps de calcul.

Nous nous proposons dans ce chapitre d'améliorer la robustesse de l'approche décrite dans le chapitre 3 vis-à-vis des variations locales d'illumination. Pour ce faire, le modèle d'illumination multiplicatif introduit par Gennert sera pris en compte dans la formulation du problème de mise en correspondance. Nous présentons dans ce qui suit ce modèle et nous justifions son utilisation au moyen de quelques notions de base en radiométrie.

5.3.1 Équation de formation des images

Pour caractériser les variations d'illumination existant entre deux images d'une paire stéréoscopique, il est nécessaire de comprendre le processus de formation de l'image sur un capteur de vision à partir d'une scène 3D éclairée. Ce processus revêt à la fois un aspect géométrique et un aspect radiométrique. Du point de vue de la géométrie, le modèle du processus peut être assimilé à celui d'une caméra sténopé, détaillé dans le chapitre 2 au paragraphe 2.2.1. Le modèle radiométrique relie les grandeurs réelles provenant de la scène, exprimées en termes d'énergie ou de puissance émise ou reçue, à celles mesurées par le système de prise de vue. Pour une caméra sténopé, l'information lumineuse enregistrée, exprimée en terme d'*intensité* quantifiée en niveau de gris, est proportionnelle à la radiance de la scène.

Pour définir ainsi le terme intensité, utilisé en photométrie pour mesurer la quantité de lumière qui parvient au capteur de vision, nous introduisons les grandeurs radiométriques suivantes :

- Le **flux radiant**, noté Φ_r, désigne l'énergie émise ou reçue par une surface donnée et s'exprime en Watts.

- La **radiance** $\mathcal{L}$ qui représente le flux radiant par unité de surface dA et par unité d'angle solide dS est donnée par :

$$\mathcal{L} = \frac{d^2\Phi_r}{dAdS} \quad (\mathrm{W.m^{-2}.sr^{-1}}). \tag{5.4}$$

- L'**irradiance** $\mathcal{E}$ est le flux radiant incident par unité de surface dA :

$$\mathcal{E} = \frac{d\Phi_r}{dA} \quad (\mathrm{W.m^{-2}}). \tag{5.5}$$

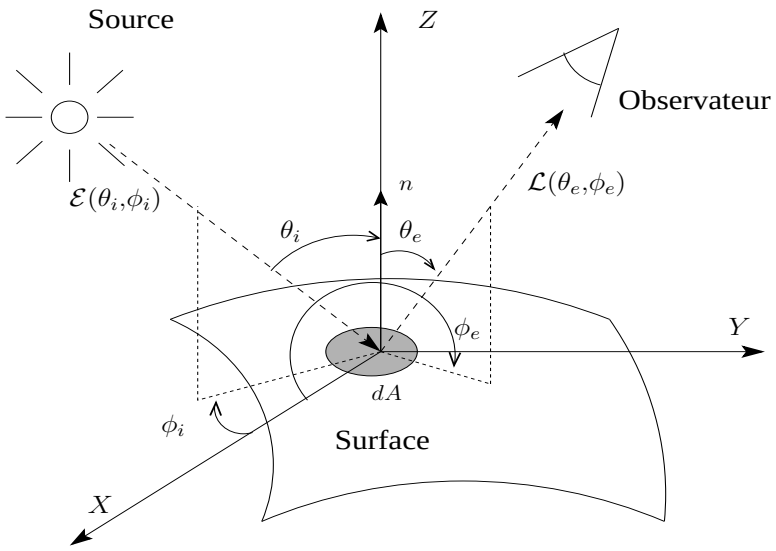


FIG. 5.1 – Le concept de la réflectance bidirectionnelle.

La radiance $\mathcal{L}$, qui mesure la quantité de lumière émise par une surface dans une direction donnée, dépend de l'illumination de la scène, de la direction d'observation du capteur et de la réflectance des surfaces. La réflectance caractérise la capacité d'une surface à réfléchir l'énergie incidente. Elle est fonction de

- la direction de la source par rapport à l'objet observé. Il s'agit de la direction de l'énergie incidente, caractérisée par les angles θ_i, ϕ_i et notée i (cf. figure 5.1).
- la direction suivant laquelle le capteur vise l'objet observé. Il s'agit de la direction de l'énergie réfléchie, caractérisée par les angles θ_e, ϕ_e et notée e (cf. figure 5.1).

En photométrie, pour définir la réflectance d'une surface on utilise habituellement la *fonction de réflectance bidirectionnelle* f_r ou BRDF[3] [Nicodemus *et al.*, 1977]. Elle est définie par le rapport suivant :

$$f_r(\theta_i, \phi_i, \theta_e, \phi_e) = \frac{d\mathcal{L}(\theta_e, \phi_e)}{d\mathcal{E}(\theta_i, \phi_i)}. \tag{5.6}$$

La BRDF représente le quotient du flux lumineux incident reçu par un élément de surface dA dans la direction (θ_i, ϕ_i) au flux rétrodiffusé dans la direction de l'observateur (θ_e, ϕ_e). Elle peut être décomposée en deux facteurs :

- La **réflectance bi-hémisphérique**, notée ρ_a et souvent appelée *albedo*, fournit le rapport entre l'énergie réfléchie par la surface dans tout l'hémisphère supérieur et l'énergie reçue par celle-ci de toutes les directions de l'hémisphère. L'albedo d'une surface ne dépend que des propriétés physiques de la surface et caractérise sa capacité à réémettre l'énergie reçue. Une surface opaque possède un albedo nul et une surface parfaitement réfléchissante un albedo unité.

[3] Bidirectional Reflectance Distribution Function

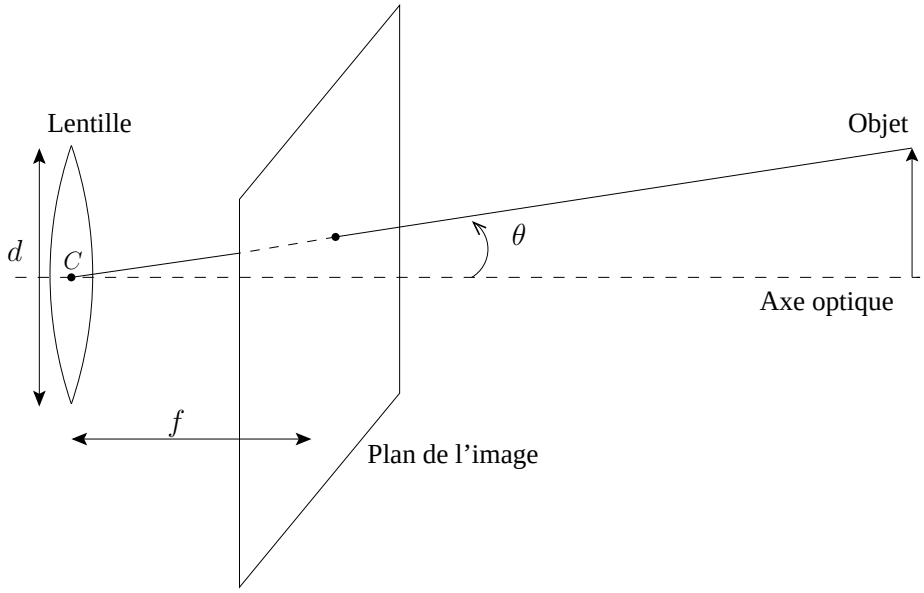


FIG. 5.2 – Géométrie de la formation des images.

- La **réflectance directionnelle**, notée $R(i, e, n)$ où n est le vecteur normal à la surface, dépend des directions d'illumination i et d'observation e. Elle est connue sous le nom anglais de *reflectance map* que nous traduisons par *carte de réflectance*.

À partir de ces deux grandeurs, l'équation (5.6) définissant la BRDF devient

$$f_r(\theta_i, \phi_i, \theta_e, \phi_e) = \rho_a R(i, e, n). \tag{5.7}$$

À présent, nous établissons la relation entre l'intensité lumineuse I mesurée par un système de formation d'images utilisant une lentille et la radiance de la scène $\mathcal{L}$. Cette relation est dûe à Horn [Horn, 1977, Horn, 1986] et se traduit par l'équation suivante :

$$I = \frac{\pi}{4} \left(\frac{d}{f}\right)^2 \cos^4 \theta \; \mathcal{L}, \tag{5.8}$$

où d est le diamètre de la lentille, f sa distance focale, et θ l'angle entre l'axe optique et la ligne d'observation (cf. figure 5.2). Par intégration de la réflectance bidirectionnelle f_r sur les angles solides définis par l'ouverture du système optique, on déduit l'expression de la radiance de la scène $\mathcal{L}$ [Nicodemus *et al.*, 1977], aboutissant ainsi à la forme simplifiée suivante de l'équation (5.8), souvent utilisée en pratique [Gennert, 1987] :

$$I = \rho \; R(i, e, n), \tag{5.9}$$

où le facteur ρ intègre tous les termes qui ne dépendent pas de la direction d'observation e.

5.3.1.1 Le modèle de Lambert

Le modèle de réflexion *lambertien* est le modèle le plus utilisé pour représenter la réflexion diffuse au sein des surfaces parfaitement mates ou diffuses (on parle aussi de surfaces lam-

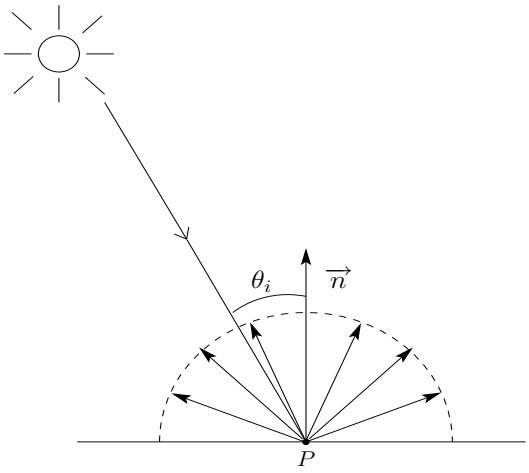


FIG. 5.3 – Modèle de réflexion lambertien : la diffusion est effectuée uniformément dans toutes les directions.

bertiennes). D'après ce modèle, l'intensité en un point P d'une surface dépend uniquement de l'angle entre la normale à la surface en ce point et la direction de la source lumineuse (cf. figure 5.3). En effet, l'équation d'illumination (5.9) s'écrit, pour une surface lambertienne, comme suit :

$$I = \rho \cos \theta_i, \tag{5.10}$$

où θ_i correspond à l'angle entre la normale à la surface et la direction de la lumière incidente. Remarquons que cette relation ne prend pas en compte la direction d'observation e, ce qui implique qu'un point sur un objet lambertien se projette sur une image avec toujours la même intensité quel que soit l'angle de vue de la caméra. Dans le cadre d'une mise en correspondance stéréoscopique, deux pixels homologues, projections d'un même point 3D de l'espace, auront donc la même intensité dans les deux images. Cependant, la majorité des surfaces qui nous entourent sont non-lambertiennes, d'où la nécessité de prendre en compte une modélisation plus adaptée du changement d'illumination.

5.3.2 Un modèle d'illumination multiplicatif

Soient I_g et I_d deux images gauche et droite d'une même scène prises de deux points de vue différents. Supposons que la scène soit représentée par une surface $\mathcal{C} = \{\mathcal{C}(s), \quad s \in \mathcal{D}\}$ où $\mathcal{D}$ représente le support de l'image, et que la direction de la source lumineuse soit la même en tout point de la scène. Par cette dernière hypothèse, la carte de réflectance, qu'on note désormais $R_e(n)$, est fonction uniquement de la direction d'observation $e \in \{g, d\}$ et de la normale à la surface n. D'après l'équation d'irradiance (5.9), l'intensité $I_e(s)$ correspondante à un point $\mathcal{C}(s)$ de la scène devient :

$$I_e(s) = \rho(s) R_e(n(s)), \tag{5.11}$$

où $n(s)$ est le vecteur normal à la surface au point $\mathcal{C}(s)$. Dans [Gennert, 1987], Gennert a montré qu'une portion de surface imagée par deux systèmes de prise de vue, de deux positions différentes, induit une relation de proportionnalité entre les intensités lumineuses dans les deux

images correspondantes :

$$\widetilde{I}_d(s) = v(s)\ I_g(s)\ , \tag{5.12}$$

où v est un facteur multiplicatif qui modélise les variations spatiales d'illumination de la scène et $\widetilde{I}_d$ est l'image de droite translatée du champ de disparité associé au couple d'images (I_g, I_d).

Pour justifier l'utilisation de ce modèle, nous considérons une portion de surface centrée autour d'un point $\mathcal{C}(s_0)$ avec pour normale à la surface le vecteur n_0. Pour de petits déplacements δs autour de $\mathcal{C}(s_0)$, l'équation (5.11) peut être développée en série de Taylor au premier ordre comme suit :

$$I_e(s) = \rho(s_0)R_e(n_0) + \left[\rho(s_0)\frac{\partial R_e(n(s))}{\partial s}\Big|_{s_0} + R_e(n_0)\frac{d\rho(s)}{ds}\Big|_{s_0}\right]\delta s. \tag{5.13}$$

En posant

$$\begin{cases} \rho_0 = \rho(s_0), \\ R_{e0} = R_e(n_0), \\ I_e(s_0) = \rho(s_0)R_e(n_0), \end{cases}$$

où $I_e(s_0)$ est l'intensité lumineuse au point s_0, les intensités $I_g(s)$ de l'image de gauche et $\widetilde{I}_d(s)$ de l'image de droite translatée peuvent être modélisées de la manière suivante :

$$I_g(s) = I_g(s_0)\ \Big(1 + \left[\frac{1}{R_{g0}}\frac{\partial R_g(n(s))}{\partial s}\Big|_{s_0} + \frac{1}{\rho_0}\frac{d\rho(s)}{ds}\Big|_{s_0}\right]\delta s\Big),$$

$$\widetilde{I}_d(s) = \widetilde{I}_d(s_0)\ \Big(1 + \left[\frac{1}{R_{d0}}\frac{\partial R_d(n(s))}{\partial s}\Big|_{s_0} + \frac{1}{\rho_0}\frac{d\rho(s)}{ds}\Big|_{s_0}\right]\delta s\Big). \tag{5.14}$$

En supposant que les variations de réflectance, appelées *ombrages*, sont les mêmes dans les deux images de la paire stéréoscopique, c'est à dire que

$$\frac{1}{R_{g0}}\frac{\partial R_g(n(s))}{\partial s}\Big|_{s_0} = \frac{1}{R_{d0}}\frac{\partial R_d(n(s))}{\partial s}\Big|_{s_0}, \tag{5.15}$$

on obtient la relation suivante :

$$\widetilde{I}_d(s) = \frac{\widetilde{I}_d(s_0)}{I_g(s_0)}\ I_g(s). \tag{5.16}$$

Le déplacement δs autour de s_0 étant supposé très petit, nous pouvons exprimer le facteur d'illumination en fonction de s en écrivant :

$$v(s) = \frac{\widetilde{I}_d(s_0)}{I_g(s_0)} = \frac{\widetilde{I}_d(s)}{I_g(s)}.$$

En remplaçant dans la relation (5.16), on aboutit au modèle d'illumination multiplicatif (5.12).

On s'est contenté ici d'analyser le cas où l'égalité de l'équation (5.15) est vérifiée. Toutefois, pour plus de détails sur la validité de ce modèle, on peut se référer à la thèse de Gennert [Gennert, 1987] où plusieurs autres cas justifiant son utilisation ont été étudiés. Notons, cependant, que le modèle d'illumination multiplicatif ne s'applique pas dans le cas de surfaces parfaitement spéculaires[4] [Gennert, 1987].

5.4 Mise en correspondance robuste aux changements de luminance

Dans la section précédente, nous avons défini un modèle d'illumination multiplicatif compensant les variations spatiales de luminosité dans une scène. L'utilisation de ce modèle dans le cadre de la mise en correspondance stéréoscopique conduit à une nouvelle formulation de ce problème, que nous détaillons ci-après.

5.4.1 Formulation ensembliste du problème

Soient I_g et I_d deux images d'une même scène acquises selon deux points de vue différents. Le modèle d'illumination établi dans le paragraphe 5.3.2 conduit à la relation suivante :

$$I_d(x - u(s)) = v(s)\ I_g(s), \tag{5.17}$$

où $s = (x, y)$ est la position spatiale dans chaque image et $u(s)$ est la disparité associée au point s. En considérant un critère de similitude quadratique, il est possible d'estimer les paramètres de ce modèle, à savoir la disparité u et l'illumination v, de manière conjointe en minimisant :

$$\tilde{J}(u, v) = \sum_{s \in \mathcal{D}} \left[v(s)\ I_g(s) - I_d(x - u(s), y)\right]^2 , \tag{5.18}$$

où $\mathcal{D} \subset \mathbb{N}^2$ est le support de l'image. La minimisation de ce critère est un problème non convexe, comme celui considéré à la section 3.2.2. Pour s'affranchir de cette limitation, on suit le même raisonnement que dans le chapitre 3 en approximant le terme non linéaire $I_d(x-u(s), y)$ par un développement de Taylor autour d'une estimée initiale $\bar{u}$ de la disparité. L'expression du critère (5.18) se réécrit ainsi de la manière suivante :

$$\widetilde{J}(u, v) \simeq \sum_{s \in \mathcal{D}} \left[L_1(s)\ u(s) + L_2(s)\ v(s) - r(s)\right]^2 , \tag{5.19}$$

où

$$\begin{aligned} L_1(s) &= \nabla I_d^x(x - \bar{u}(s), y), \\ L_2(s) &= I_g(s), \\ r(s) \quad &= I_d(x - \bar{u}(s), y) + \bar{u}(s)\ L_1(s). \end{aligned}$$

[4] Une surface parfaitement spéculaire est une surface qui réfléchit toute la lumière incidente selon une direction privilégiée.

En posant $w = (u, v)^\top$ et $L = [L_1, L_2]$, on aboutit finalement au critère quadratique convexe suivant :

$$J_{\mathcal{D}}(w) = \sum_{s \in \mathcal{D}} [L(s)\ w(s) - r(s)]^2 \ . \tag{5.20}$$

Minimiser ce critère est un problème inverse mal posé car les composantes de L peuvent s'annuler en certains points de l'image. Notons que ce caractère mal posé se trouve accentué par le fait que ce problème est sous-déterminé car on ne dispose que d'une unique équation pour trouver deux inconnues $u(s)$ et $v(s)$ en chaque point s. Pour obtenir des solutions uniques fiables, il faut donc incorporer autant d'information *a priori* que possible sur les deux champs à estimer.

Nous avons vu au cours du chapitre 3 que le problème d'estimation de la disparité initialement formulé comme un problème inverse mal posé peut être efficacement résolu par la minimisation d'une fonction objectif convexe sous des contraintes convexes. En représentant ces contraintes convexes, modélisant les connaissances a *priori* et les observations, par des ensembles convexes, le problème revient à sélectionner la solution qui minimise une certaine fonction convexe sur l'intersection de ces ensembles. En se plaçant dans ce cadre ensembliste, le problème de la détermination de w peut donc se formuler comme suit :

$$\text{trouver } \widehat{w} \in S = \bigcap_{i=1}^{m} S_i \text{ tel que } J(\widehat{w}) = \inf J(S) \ , \tag{5.21}$$

où J est un critère convexe approprié et les ensembles de contraintes $(S_i)_{1 \leq i \leq m}$ sont des convexes fermés d'un espace de Hilbert $\mathcal{H}$, modélisés sous la forme

$$\forall i \in \{1, \ldots, m\}, \qquad S_i = \{w \in \mathcal{H} \mid f_i(w) \leq \delta_i\} \ , \tag{5.22}$$

$(f_i)_{1 \leq i \leq m}$ étant une famille de fonctions convexes continues définies sur $\mathcal{H}$ et $(\delta_i)_{1 \leq i \leq m} \in \mathbb{R}$.

Nous avons également constaté au cours du chapitre 3 que l'algorithme itératif 3.2 initialement proposé par Combettes est un algorithme séduisant pour résoudre le problème (5.21) dans le cas de critères différentiables. Dans le cas particulier d'un critère J quadratique, la variante 3.3 de cet algorithme offre un moyen de résolution simple et efficace tout en conservant tous les avantages de l'algorithme original.

Ce constat nous conduit naturellement à adopter l'approche ensembliste pour formuler le problème d'estimation de la disparité en présence de changements d'illumination et à utiliser l'algorithme 3.3 pour le résoudre numériquement.

Dans ce qui suit, nous commençons dans un premier temps par définir la fonction objectif globale à minimiser que nous considérons, dans ce chapitre, quadratique. Nous introduisons dans un deuxième temps les ensembles convexes traduisant les propriétés connues *a priori* sur les champs à estimer. Pour cela, nous reprenons les contraintes définies dans le chapitre 3 sur le champ de disparité et nous proposons de nouvelles contraintes pour le champ d'illumination. Nous détaillons enfin l'algorithme que nous employons pour calculer simultanément le champ de disparité et les changements d'illumination.

5.4.2 Définition d'une fonction objectif quadratique

La fonction objectif à minimiser s'appuie sur le modèle d'observation (5.17) linéarisé autour d'une estimée initiale $\bar{u}$ de la disparité. Pour que cette linéarisation soit valide, il est important que l'estimée initiale soit consistante avec la solution. Comme alternative à l'initialisation classique consistant à choisir un champ nul au niveau le plus grossier d'une pyramide d'images, nous utilisons ici une approche par corrélation combinée à un modèle affine par bloc de changements d'illumination. Ce choix nous permet de calculer à la fois une estimée initiale pour le champ de disparité et une estimée initiale pour le champ d'illumination, dont nous aurons besoin par la suite. Pour être conforme avec le modèle d'illumination que nous utilisons, on propose de ne considérer que le terme multiplicatif du modèle affine[5] (5.2) et de prendre ainsi $\gamma = 0$ dans l'équation (5.2). Plus précisément, nous calculons

$$\bar{u}(x,y) = \arg\min_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{B}} \left[\beta_{x,y}(u)\ I_g(x+i,y+j) - I_d(x+i-u,y+j)\right]^2 , \tag{5.23}$$

où $\mathcal{U} \subset \mathbb{N}$ correspond à la zone de recherche, $\mathcal{B}$ est le bloc de corrélation centré sur le pixel (x,y) et $\beta_{x,y}(u)$ est l'estimée au sens des moindres carrés du facteur d'illumination, supposé constant en tout point (i,j) du bloc $\mathcal{B}$. Ce facteur est donné par :

$$\beta_{x,y}(u) = \frac{\sum_{(i,j)\in\mathcal{B}} I_g(x+i,y+j) I_d(x+i-u,y+j)}{\sum_{(i,j)\in\mathcal{B}} I_g(x+i,y+j)^2}. \tag{5.24}$$

L'estimée initiale du champ d'illumination s'obtient donc en substituant $\bar{u}$ dans l'équation (5.24), soit :

$$\bar{v}(x,y) = \beta_{x,y}\big(\bar{u}(x,y)\big). \tag{5.25}$$

Par ailleurs, pour s'affranchir des problèmes liés aux zones d'occultations, qui sont les portions de la scène visibles depuis un seul point de vue, nous proposons de les exclure du processus de minimisation. En notant $\mathcal{O}$ l'union de ces zones, le critère à minimiser devient

$$J_{\mathcal{D}\setminus\mathcal{O}}(w) = \sum_{s\in\mathcal{D}\setminus\mathcal{O}} \left[L(s)\ w(s) - r(s)\right]^2. \tag{5.26}$$

Cependant, d'après les conditions assurant la convergence de l'algorithme que nous utilisons (cf. section 5.4.4), la fonction objectif J doit être strictement convexe. Étant donné que $L(s)^\top L(s)$ est de rang 1, que les composantes de l'opérateur L peuvent s'annuler et que le champ d'occultations $\mathcal{O}$ est toujours non vide, le critère $J_{\mathcal{D}\setminus\mathcal{O}}$ n'est pas strictement convexe. Nous introduisons donc un terme additif strictement convexe au critère, qui devient :

$$J(w) = \sum_{s\in\mathcal{D}\setminus\mathcal{O}} \left[L(s)\, w(s) - r(s)\right]^2 + \alpha \sum_{s\in\mathcal{D}} |w(s) - \bar{w}(s)|_2^2 , \tag{5.27}$$

où $\bar{w} = (\bar{u},\bar{v})$ est l'estimée initiale calculée comme décrit précédemment et α une constante positive qui pondère l'influence du deuxième terme vis-à-vis du premier. Nous soulignons ici

[5] D'après Gennert [Gennert, 1987], le terme additif s'avère être inutile en théorie comme en pratique.

que le rôle premier de ce terme additif n'est pas de régulariser la solution mais d'imposer la stricte convexité du critère J, conformément à l'hypothèse 3.3.3.1 de l'algorithme employé. Enfin, il est important de noter que l'estimée initiale $\bar{w}$ est affinée par un processus itératif, où le champ estimé à un cycle d'itérations donné sert d'initialisation pour le calcul du champ au cycle suivant. Ce processus réduit la dépendance de la solution vis-à-vis du choix de l'estimation initiale.

5.4.3 Introduction de contraintes convexes

Le problème d'estimation conjointe de la disparité et des changements d'illumination appartenant à la classe des problèmes inverses mal posés, c'est par l'introduction de certaines contraintes que le problème devient bien posé. Ces contraintes s'interprètent comme l'information *a priori* dont nous disposons sur les variables à estimer, à savoir la disparité u et le champ d'illumination v. Dans le cadre d'une approche ensembliste, il convient de modéliser ces contraintes comme des sous-ensembles convexes de la forme (5.22).

Dans le chapitre 3, des sous-ensembles de contraintes convexes ont été construits à partir des connaissances *a priori* sur le champ de disparité u. Nous proposons ici de retenir quelques unes de ces contraintes sur u et de formuler de nouvelles contraintes sur le champ d'illumination v.

5.4.3.1 Contraintes sur le champ de disparité

Comme nous venons de le mentionner, dans le chapitre 3 ont été imposées des contraintes convexes $(S_i)_{1\leq i\leq 5}$ sur le champ de disparité. À partir d'expérimentations décrites dans ce même chapitre, l'ensemble des solutions admissibles donnant les meilleurs résultats est l'ensemble $S = S_1 \cap S_3 \cap S_5$. Ceci nous incite naturellement à retenir les contraintes S_1, S_3 et S_5 sur u que nous rappelons brièvement dans ce qui suit. Notons que les ensembles S_1 et S_3 sont définis dans le domaine spatial et l'ensemble S_5 dans le domaine des ondelettes. Il est également important de souligner la capacité de l'algorithme 3.3 à pouvoir combiner efficacement des contraintes définies dans les deux domaines.

En nous plaçant dans le cas discret, l'espace des champs 2D considérés est l'espace de Hilbert produit $\mathcal{H} = \mathbb{R}^{N_\mathrm{l}\times N_\mathrm{c}} \times \mathbb{R}^{N_\mathrm{l}\times N_\mathrm{c}}$, où $N_\mathrm{l} \times N_\mathrm{c}$ est la taille des champs u et v qui doivent être tous les deux estimés.

Contrainte de plage de valeurs

Le premier *a priori* que l'on peut imposer sur l'image de disparité est l'intervalle de recherche. L'ensemble correspondant à cette information est

$$S_1 = \{(u, v) \in \mathcal{H} \mid u_{\min} \leq u \leq u_{\max}\} \,. \tag{5.28}$$

Rappelons que les amplitudes minimale $u_{\min} \geq 0$ et maximale $u_{\max}$ de la disparité sont souvent connues en pratique.

Régularisation suivant les contours de l'image

Le second *a priori* est basé sur une estimation en accord avec les discontinuités en niveau de gris présentes dans l'image de référence I_g. L'idée fondamentale de cette régularisation, comme nous l'avons déjà dit, est de réduire l'effet de lissage en présence de frontières où le gradient ∇I_g est élevé et d'encourager ce lissage le long de ces mêmes frontières. L'ensemble associé à cette contrainte est :

$$S_3 = \{(u, v) \in \mathcal{H} \mid f_3(u) \leq \delta\} \,, \tag{5.29}$$

où δ est une constante positive qui peut être estimée à partir d'expérimentations et la fonction f_3 est donnée par l'équation (3.67).

Régularisation dans le domaine des ondelettes

Le principe de la régularisation dans le domaine des ondelettes est de supposer que la norme de l'image de disparité exprimée en fonction des coefficients en ondelettes de l'image, est bornée. La contrainte de régularisation associée à la semi-norme quadratique que nous avons proposée dans le chapitre 3 se formule comme suit :

$$S_5 = \bigcap_{\mathrm{d} \in \{0,\ldots,2^l-1\}^2} S_5^{(\mathrm{d})}, \tag{5.30}$$

avec

$$S_5^{(\mathrm{d})} = \{(u, v) \in \mathcal{H} \mid \sum_{j=1}^{l} \sum_{k \in \mathbb{Z}^2} \Big(\sum_{o \in \{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u^{(\mathrm{d})})|^2 \Big)^{1/2} \leq \kappa_5\}, \tag{5.31}$$

où $u^{(\mathrm{d})}$ est la version translatée de u par d, $\big(c^{\mathcal{B}}_{j,k,o}(u^{(\mathrm{d})})\big)_{j,k,o}$ sont les coefficients en ondelettes de $u^{(\mathrm{d})}$ issus d'une décomposition dans une base séparable $\mathcal{B}$ d'ondelettes 2D, l est le niveau de résolution le plus grossier et κ_5 est une constante positive.

5.4.3.2 Contraintes sur le champ d'illumination

Afin de pallier le caractère mal-posé du problème de mise en correspondance en présence des changements d'illumination, il est utile d'introduire, en plus des contraintes sur le champ de disparité u, des contraintes sur le facteur d'illumination v qui traduit les variations locales d'illumination de la scène.

Contrainte de régularisation

Le modèle d'illumination multiplicatif, que nous proposons dans ce travail, suppose implicitement que les changements d'illumination ont des variations douces et isotropes dans l'espace. Pour modéliser cette information *a priori* de régularité du champ v, une régularisation quadratique de type Tikhonov [Tikhonov, Arsenin, 1977] semble particulièrement appropriée. Formuler cette contrainte de régularisation comme un sous-ensemble convexe de la

forme (5.22) revient à imposer une borne $\zeta > 0$ sur la norme quadratique du gradient discret $\widehat{\nabla} v$ de v :

$$S_6 = \{(u, v) \in \mathcal{H} \mid |\widehat{\nabla} v|_2^2 \leq \zeta\} \ . \tag{5.32}$$

Un autre type de régularisation qui peut être envisagé pour modéliser un champ d'illumination variant de manière douce est celui basé sur le laplacien de l'image. La régularisation que nous avons testée consiste plus précisément à imposer une borne sur la valeur absolue du laplacien, confinant la solution à l'ensemble convexe suivant :

$$S_7 = \{(u, v) \in \mathcal{H} \mid |\widehat{\Delta} v| \leq \eta\} \ , \tag{5.33}$$

où $\eta > 0$ et $\widehat{\Delta} v$ est le laplacien discret de v défini par

$$(\widehat{\Delta} u)(x, y) = u(x+1, y) + u(x-1, y) + u(x, y+1) + u(x, y-1) - 4u(x, y), \tag{5.34}$$

si $1 < x < N_\mathrm{l}$ et $1 < y < N_\mathrm{c}$.

Notons que les contraintes (5.32) et (5.33) ont pour vocation de traduire la même information *a priori* de régularité du champ d'illumination v. Il nous paraît donc justifié de n'introduire qu'une seule de ces deux contraintes. Une comparaison de l'influence des deux régularisations qu'elles induisent sur le champ d'illumination (cf. paragraphe 5.5) montre que la régularisation de Tikhonov obtient les meilleurs résultats.

Contrainte d'intervalle

Le modèle d'illumination (5.12) autorise des changements spatiaux de luminosité au travers d'un coefficient d'illumination multiplicatif v. Au lieu de supposer que ce facteur d'illumination est constant et égal à 1, nous considérons dans ce travail un terme v qui varie spatialement en chaque point, tout en supposant, néanmoins, que ses valeurs sont proches de 1. Ceci a été prouvé par Gennert [Gennert, 1987] en démontrant, à partir d'expérimentation sur des images réelles, que v est, de manière usuelle, comprise entre $v_{\min} = 0.8$ et $v_{\max} = 1.2$. L'ensemble convexe associé à cette information est

$$S_8 = \{(u, v) \in \mathcal{H} \mid v_{\min} \leq v \leq v_{\max}\} \ . \tag{5.35}$$

5.4.4 Méthode d'optimisation

Ce paragraphe a pour objectif de développer une solution numérique au problème de mise en correspondance qui soit peu sensible aux changements d'illumination. Ce problème qui a été formulé comme un problème de programmation convexe sera résolu à l'aide de l'algorithme itératif par bloc détaillé ci-après. Cet algorithme permet, en effet, de minimiser la fonction objectif quadratique (5.27) sur l'ensemble des solutions admissibles défini par les contraintes multiples introduites précédemment.

5.4.4.1 Projection sous-différentielle

L'algorithme que nous utilisons pour résoudre à l'aide d'un même formalisme le problème de mise en correspondance et celui des changements d'illumination emploie les projections sous-différentielles associées aux diverses contraintes introduites sur la solution. Notons que seules les projections exactes sur les ensembles S_1 et S_8 peuvent être calculées explicitement. La projection sous-différentielle $P_i w$ de $w = (u, v) \in \mathcal{H}$ sur S_i est définie, de manière similaire à (3.38), en prenant un sous-gradient quelconque $t_i \in \partial f_i(w)$ et en posant :

$$p_i w = \begin{cases} w + \dfrac{\delta_i - f_i(w)}{\|t_i\|^2} t_i, & \text{si } f_i(w) > \delta_i; \\ w, & \text{si } f_i(w) \leq \delta_i. \end{cases} \tag{5.36}$$

Il est important de souligner ici que les contraintes que nous considérons dans ce chapitre sont séparables dans le sens où elles portent chacune soit sur le champ de disparité u, soit sur le champ d'illumination v. Ceci équivaut à dire que les projections sous-différentielles associées à des contraintes sur u (resp. v) laissent v (resp. u) invariant. Concrètement, la projection sous-différentielle de $w = (u, v)$ sur l'ensemble S_i s'écrit :

$$p_i w = p_i(u, v) = (\widetilde{p}_i u, v), \tag{5.37}$$

si l'ensemble S_i porte sur le champ de disparité u. En revanche, si l'ensemble S_i est associé au champ d'illumination v, $p_i w$ s'écrit :

$$p_i w = p_i(u, v) = (u, \widetilde{p}_i v). \tag{5.38}$$

En combinant les équations (5.37) et (3.69), il est désormais facile de calculer la projection sous-différentielle de $w \in \mathcal{H}$ sur l'ensemble S_3. De la même manière, on peut calculer la projection sous-différentielle sur S_5 en exploitant le résultat du paragraphe 3.4.5.3 et celui de l'équation (5.37).

Nous allons à présent calculer la projection sous-différentielle sur l'ensemble de contrainte S_6 relatif au champ d'illumination v. Soit f_6 la semi-norme quadratique du gradient de v donnée par :

$$f_6(v) = \sum_{x=1}^{N_\mathrm{l}} \sum_{y=1}^{N_\mathrm{c}} \left((\widehat{\nabla} v)^1(x, y)\right)^2 + \left((\widehat{\nabla} u)^2(x, y)\right)^2. \tag{5.39}$$

En remarquant que la semi-norme f_6 est égale à la semi-norme f_3, définie par (3.67), dans le cas où l'opérateur D donné par l'équation (3.66) correspond à la matrice identité $I_{2,2}$, le gradient de la fonction quadratique f_6 se déduit facilement de l'équation (3.68). On tire, ensuite, de (5.36) et (5.38) l'expression de la projection sous-différentielle associée à cette contrainte. Enfin, notons que la projection sous-différentielle sur l'ensemble S_7 se calcule de manière similaire, en écrivant la fonctionnelle associée à cette contrainte et en dérivant un sous-gradient correspondant.

5.4.4.2 Algorithme proposé

Nous allons à présent décrire l'algorithme itératif par bloc qui permet de calculer simultanément le champ de disparité u et le champ d'illumination v.

Algorithme 5.1

Etape 0 Poser $n = 0$. Calculer w_0 comme suit

$$w_0(s) = \begin{cases} (L(s)^\top L(s) + \alpha\ I_{2,2})^{-1}(L(s)^\top r(s) + \alpha\ \bar{w}(s)) & \text{si } s \in \mathcal{D} \setminus \mathcal{O}, \\ \bar{w}(s) & \text{sinon.} \end{cases} \tag{5.40}$$

Etape 1 Prendre un ensemble d'indices non vide $I_n \subseteq \{1, \ldots, m\}$.

Etape 2 Pour chaque indice $i \in I_n$, poser $a_{i,n} = P_{i,n} - w_n$ où $P_{i,n}$ est une projection sous-différentielle de w_n sur S_i (cf. (3.38)).

Etape 3 Poser $z_n = |I_n|^{-1} \sum_{i \in I_n} a_{i,n}$ et $\kappa_n = |I_n|^{-1} \sum_{i \in I_n} \|a_{i,n}\|^2$, où $|I_n|$ est le nombre d'éléments dans I_n.

Etape 4 Si $\kappa_n = 0$, sortir. Sinon, poser

- $b_n = w_0 - w_n$,
- c_n tel que
$$c_n(s) = \begin{cases} (L(s)^\top L(s) + \alpha\ I_{2,2})\ b_n(s) & \text{si } s \in \mathcal{D} \setminus \mathcal{O}, \\ \alpha\ b_n(s) & \text{sinon,} \end{cases} \tag{5.41}$$
- d_n tel que
$$d_n(s) = \begin{cases} (L(s)^\top L(s) + \alpha\ I_{2,2})^{-1} z_n(s) & \text{si } s \in \mathcal{D} \setminus \mathcal{O}, \\ \alpha^{-1} z_n(s) & \text{sinon,} \end{cases} \tag{5.42}$$
- $\lambda_n = \kappa_n / \langle d_n, z_n \rangle$.

Etape 5 Poser $\tilde{d}_n = \lambda_n d_n$, $\pi_n = -\langle c_n, \tilde{d}_n \rangle$, $\mu_n = \langle b_n, c_n \rangle$, $\nu_n = \lambda_n \langle \tilde{d}_n, z_n \rangle$ et $\varrho_n = \mu_n \nu_n - \pi_n^2$.

Etape 6 Poser

$$w_{n+1} = \begin{cases} w_n + \tilde{d}_n, & \text{si } \varrho_n = 0,\ \pi_n \geq 0; \\ w_0 + (1 + \frac{\pi_n}{\nu_n})\tilde{d}_n, & \text{si } \rho_n > 0,\ \pi_n \nu_n \geq \varrho_n; \\ w_n + \frac{\nu_n}{\rho_n}(\pi_n b_n + \mu_n \tilde{d}_n), & \text{si } \varrho_n > 0,\ \pi_n \nu_n < \rho_n. \end{cases}$$

Etape 7 Incrémenter n et aller à l'étape 1.

Le bon comportement asymptotique de cet algorithme est garanti par le résultat suivant :

Théorème 5.1 *Supposons qu'il existe un entier positif K tel que*

$$\forall\ n \in \mathbb{N}, \qquad \bigcup_{l=n}^{n+K-1} I_l = \{1, \ldots, m\}, \tag{5.43}$$

alors la suite $(w_n)_n$ engendrée par l'algorithme 5.1 converge vers l'unique solution du problème (5.21).

Démonstration.
L'algorithme 5.1 peut être vu comme un cas particulier de l'algorithme 3.3 décrit dans le chapitre 3 et développé dans [Combettes, 2003] pour la minimisation de critères quadratiques sous des contraintes convexes. En effet, les poids $(\omega_{i,n})_{i \in I_n}$, à l'étape 3 de l'algorithme, sont ici choisis égaux et une valeur maximale du paramètre de relaxation λ_n est utilisée à chaque itération.
Pour montrer la convergence de l'algorithme, nous devons donc nous assurer que les hypothèses 3.3.3.1 sont vérifiées :

1. Étant donné que $\alpha > 0$, la fonction objectif J est une fonction quadratique strictement convexe.
2. L'ensemble des solutions admissibles $S = \bigcap_{i=1}^{m} S_i$, intersection des ensembles de contraintes convexes, est non vide. En effet, si u et v sont tous les deux des champs constants ayant des valeurs dans $[u_{\min}, u_{\max}] \times [v_{\min}, v_{\max}]$, alors (u, v) satisfait de manière évidente les contraintes considérées.

■

5.5 Résultats expérimentaux

Nous présentons dans ce paragraphe quelques résultats expérimentaux permettant d'évaluer la robustesse de notre approche vis-à-vis de changements d'illumination. Dans une première partie, nous montrons des résultats obtenus sur un exemple synthétique où un changement de luminosité artificiel a été simulé. Nous illustrons ensuite les résultats de la méthode sur des paires d'images réelles ayant subi de réels changements d'illumination. Dans le cadre des images réelles, nous distinguons les images avec vérité terrain et les images sans vérité terrain. Tous ces résultats ont été obtenus avec des ondelettes de Haar invariantes par translation, sur un niveau de résolution, comme préconisé au paragraphe 4.2.3.1. À des fins de comparaison, nous reportons également les résultats obtenus avec

- une approche par corrélation combinée à un modèle affine des changements d'illumination, que l'on notera SSD+MA ;
- une approche par corrélation utilisant une mesure de corrélation normalisée, que l'on notera NCC ;
- l'approche semi-globale développée par Hirschmüller [Hirschmüller, 2005] et basée sur l'information mutuelle, que l'on notera SGM (pour *semi-global matching*).

Notons que les méthodes utilisées dans l'évaluation de Scharstein et Szeliski ne prennent pas en compte le problème des changements d'illumination [Scharstein, Szeliski, 2002]. Pour pouvoir

établir une comparaison avec ces méthodes, il paraît judicieux d'effectuer un prétraitement des images du couple stéréoscopique pour d'atténuer les effets des changements d'illumination. Nous avons choisi ici d'appliquer une transformation d'histogrammes aux images stéréoscopiques pour ensuite procéder à leur mise en correspondance en utilisant l'algorithme de Kolmogorov et Zabih [Kolmogorov, Zabih, 2001] à base de coupures de graphes. Les résultats de la combinaison de ces deux techniques, que l'on notera TH+CG, sont aussi inclus dans notre étude comparative. Notons que l'algorithme de Kolmogorov et Zabih, disponible sur le Web[6], utilise un paramètre de régularisation que nous avons ajusté de manière à obtenir les meilleurs résultats avec cette méthode.

5.5.1 Exemple synthétique avec des variations d'illumination simulées

Le premier exemple que nous avons utilisé pour valider quantitativement la méthode proposée est la paire d'images synthétique *corridor* montrée à la figure 4.1. Pour simuler des variations de luminosité sur cette paire stéréoscopique, nous avons multiplié l'image de droite par un profil gaussien, tandis que l'image de gauche est restée inchangée. Les deux images obtenues sont représentées en figure 5.4. Les dimensions de ces images étant égales à 256×256 pixels, le profil de changement d'illumination gaussien est défini par :

$$g(x, y) = c_1 e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\varsigma^2}} + c_2, \tag{5.44}$$

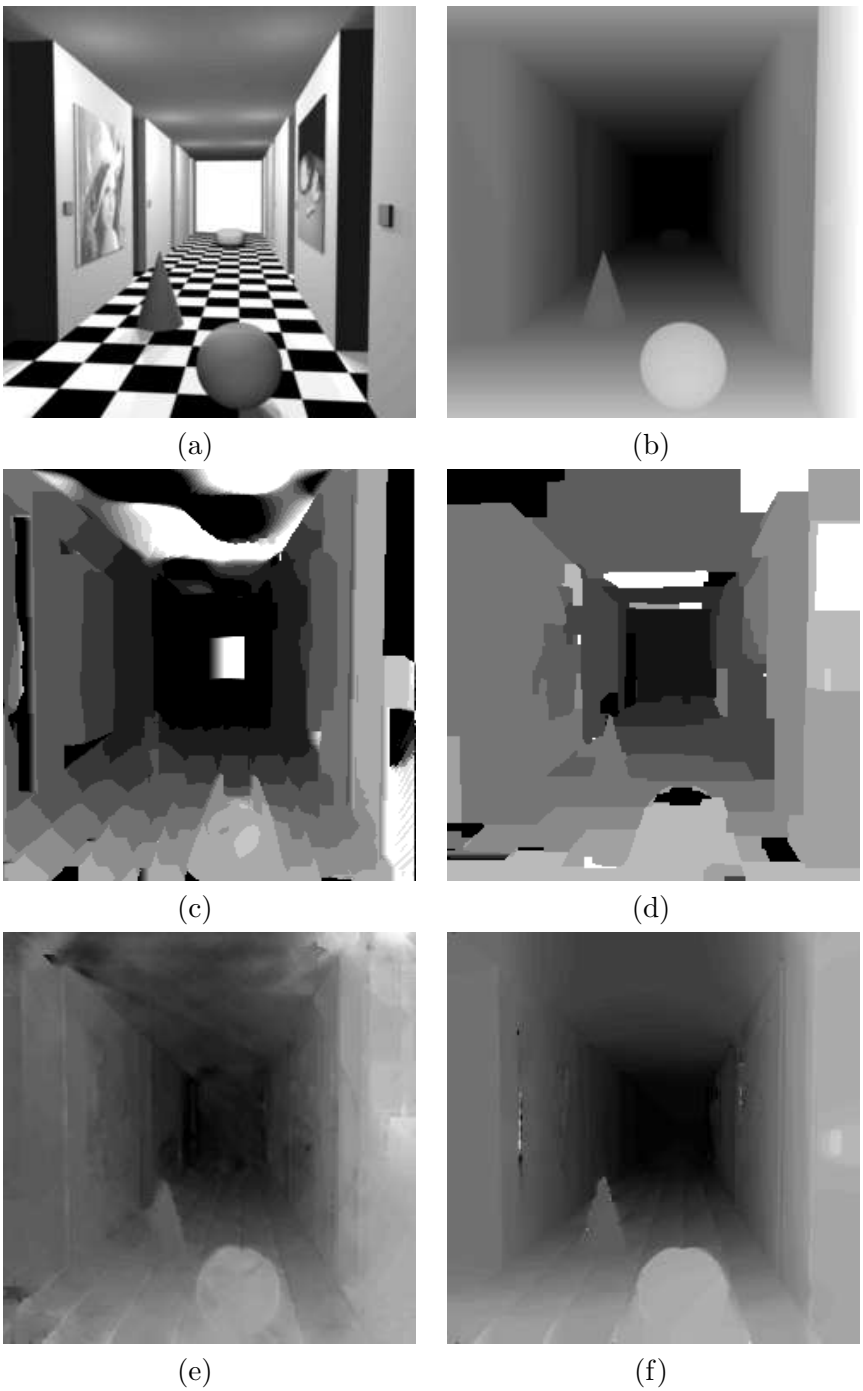
où $(x_0, y_0) = (128, 128)$ est le centre, $\varsigma = 256$ est l'écart type de l'effet d'illumination et c_1 et c_2 sont deux constantes que nous avons fixées à 1.8 et -0.6 respectivement, de manière à ce que le profil g vérifie la contrainte de plage de valeurs (5.35).



FIG. 5.4 – Paire d'images *Corridor* modifiée pour avoir un changement d'illumination artficiel. L'image de droite est multipliée par un profil gaussien dont la formule est donnée en (5.44).

La carte de disparité théorique et celle calculée par notre approche pour ce couple d'images sont montrées à la figure 5.5. Comme nous pouvons le constater, l'utilisation du modèle d'illumination proposé a permis de compenser les changements de luminosité entre l'image de gauche et l'image de droite et d'obtenir ainsi une estimation de la carte de disparité très robuste vis-à-vis de ces changements. En effet, les larges zones homogènes, comme le plafond

[6]http ://www.adastral.ucl.ac.uk/ vladkolm/software.html



(a) (b) (c) (d) (e) (f)

FIG. 5.5 – Résultats pour la paire stéréoscopique *Corridor* ayant subi des variations d'illuminations artificielles. (a) Image de gauche (b) Vraie disparité (c) NCC (d) GC+HT (e) SGM (f) Notre approche.

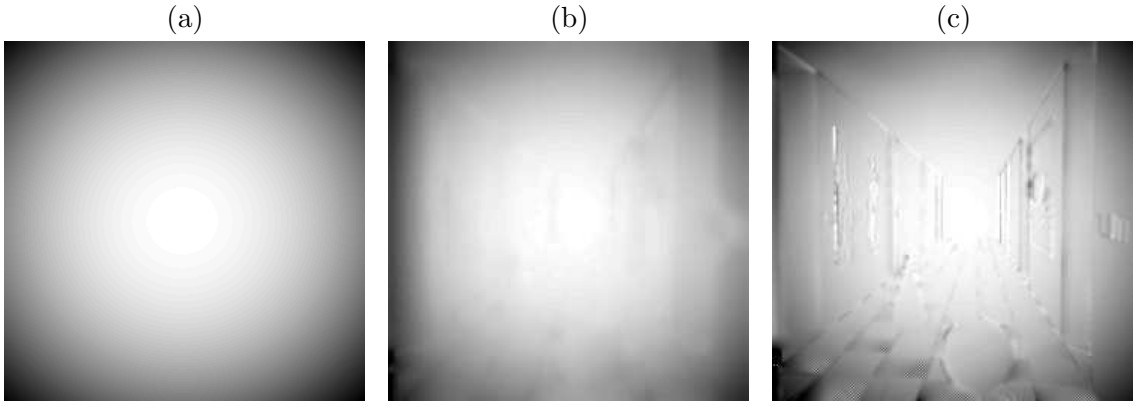


FIG. 5.6 – Cartes de changements d'illumination pour la paire d'images *Corridor* modifiée : (a) la carte synthétique (b) la carte estimée avec une régularisation de Tikhonov (c) la carte estimée avec une régularisation basée sur le laplacien.

ou les murs, sont bien lissées et les contours autour des objets sont préservés. Nous reportons également sur cette figure les cartes de disparité obtenues par les différentes méthodes comparées. L'ensemble de ces résultats met en évidence la supériorité de notre méthode sur celle par corrélation normalisée et sur la technique CG+TH. La carte de disparité de la figure 5.5(e) souligne le bon comportement de la méthode SGM. Toutefois, comparée à notre méthode, elle se comporte moins bien, notamment pour le lissage des zones homogènes de l'image qui apparaissent, en effet, comme les zones les plus marquées par l'effet des variations d'illumination.

Pour une analyse quantitative des performances des différentes méthodes, nous proposons d'utiliser les critères d'évaluation du chapitre 4, à savoir l'erreur absolue moyenne et le pourcentage d'appariements erronés. Les résultats obtenus sont indiqués dans le tableau 5.1(a). La supériorité de notre méthode, déjà appréciée visuellement sur la figure 5.5, est clairement confirmée par ces valeurs. Notons que l'approche de Hirschmüller [Hirschmüller, 2005] obtient également de bons résultats. Cependant, contrairement à SGM, notre approche permet de mesurer les changements locaux d'illumination. Pour modéliser la régularité du champ d'illumination mesuré, deux expériences ont été ainsi menées : 1) en considérant la régularisation quadratique de Tikhonov traduite par l'ensemble S_6 et 2) en considérant une régularisation basée sur le laplacien et traduite par l'ensemble S_7. La figure 5.6 présente le champ d'illumination théorique et celui estimé suivant les deux régularisations. La solution obtenue avec une régularisation de Tikhonov est nettement meilleure que celle obtenue avec une régularisation basée sur le laplacien. En effet, tandis que le lissage induit par l'utilisation de l'ensemble S_6 est bien effectué de manière conforme sur toute l'image, celui obtenu en prenant en compte la contrainte S_7 est influencé par les discontinuités de l'image de référence. Cela s'explique par le fait que la dérivée seconde est plus sensible aux petites perturbations. En outre, l'erreur absolue moyenne d'estimation du champ d'illumination avec une régularisation de Tikhonov est de 0.016 tandis que celle avec une régularisation basée sur le laplacien est de 0.021. Ceci confirme que la régularisation quadratique de Tikhonov convient mieux à ce type de champs.

Par ailleurs, pour évaluer la robustesse de notre approche en présence d'un bruit additif, nous l'avons testée sur trois variantes de la paire *Corridor*, modifiée par les variations d'illumination et dégradée par un bruit additif gaussien centré de variance σ_n^2. À l'aide du tableau 5.1(b), nous pouvons constater que notre méthode est très robuste au bruit et fournit de

(a) Performance globale

Technique	EAM	Err
SSD+MA	1.58	28
NCC	1.54	26
GC [Kolmogorov, Zabih, 2001] +HT	0.80	18
SGM [Hirschmüller, 2005]	0.47	13
Notre approche	**0.37**	**10**

(b) Impact du bruit utilisant le critère EAM

Technique	niveaux de bruit $\sigma_n^2 = 1$	$\sigma_n^2 = 10$	$\sigma_n^2 = 100$
SSD+MA	1.18	1.20	1.27
NCC	1.10	1.14	1.16
GC [Kolmogorov, Zabih, 2001]+HT	0.72	0.85	0.89
SGM [Hirschmüller, 2005]	0.38	0.41	0.48
Notre approche	**0.34**	**0.38**	**0.44**

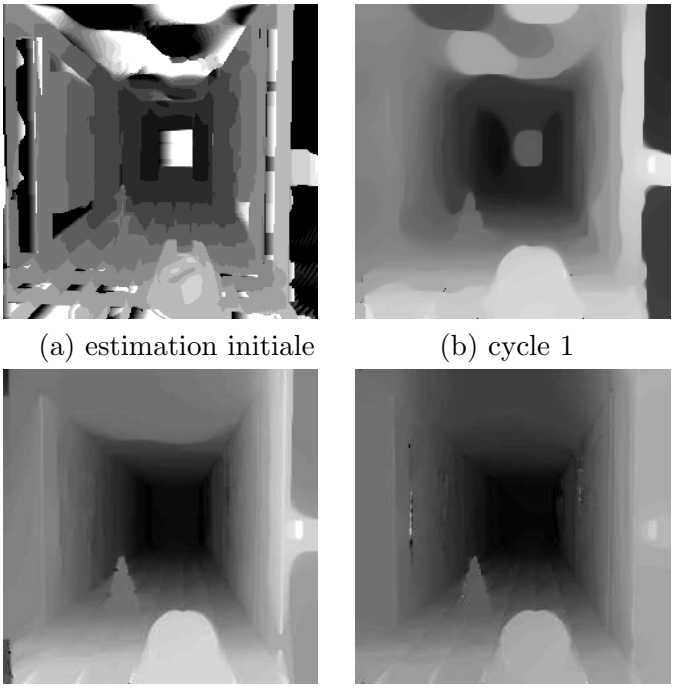
TAB. 5.1 – Résultats comparatifs sur la paire d'images *Corridor* modifiée : (a) sans ajout de bruit (b) avec bruit additif gaussien.

bons résultats même si le niveau de bruit est important. En effet, le résultat de notre méthode pour un niveau de bruit égal à 100 est meilleur que celui obtenu par les autres méthodes, sur la paire d'image non bruitée. Les résultats de l'algorithme SGM sur la paire d'image bruitée confirment que le calcul global de l'information mutuelle sur toute l'image permet de pallier la sensibilité de cette mesure à un bruit additif gaussien.

Il est important de noter qu'afin d'améliorer la qualité de nos résultats et réduire la dépendance de la solution vis-à-vis du choix de l'estimation initiale, le processus d'estimation conjointe de la disparité et des changements d'illumination a été réitéré, comme dans le cas d'une estimation sans variations d'illumination, où les champs estimés à un cycle d'itérations donné servent d'initialisation pour le cycle suivant. Les figures 5.7 et 5.8 ainsi que le tableau 5.2 donnent les résultats intermédiaires pour la paire *Corridor*, en considérant trois cycles d'itérations. Notons que, pour évaluer les champs d'illumination estimés suivant le critère Err, nous avons porté la valeur du seuil δ_s, qui intervient dans ce critère, à 0.1 en raison de la dynamique de ce champ. Les résultats obtenus confirment visuellement et quantitativement l'intérêt d'utiliser une procédure itérative pour affiner les estimations initiales et par conséquent les résultats.

Étape	Champ de disparité		Champ d'illumination	
	EAM	Err	EAM	Err
estimation initiale	1.58	28	0.029	9
cycle 1	0.69	17	0.025	6
cycle 2	0.53	13	0.017	4
résultat final	0.37	10	0.016	4

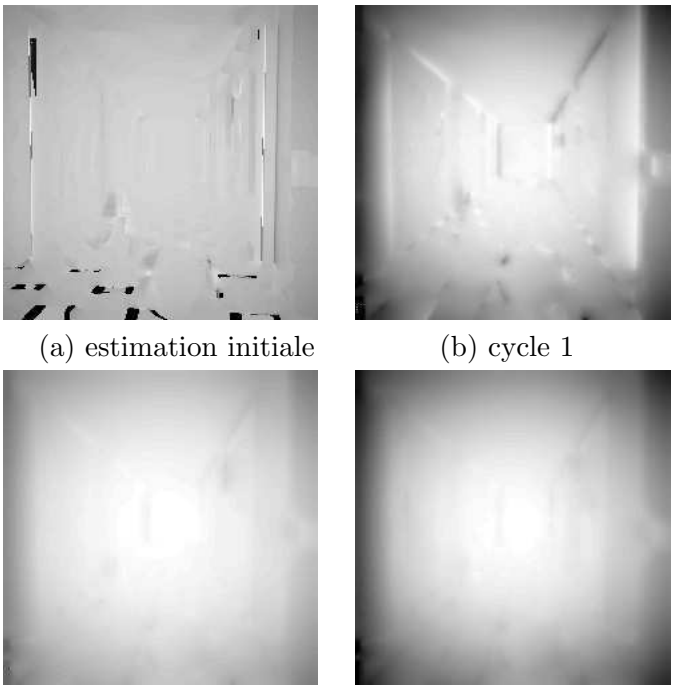
TAB. 5.2 – Résultats intermédiaires pour la paire d'images *Corridor* modifiée.



(a) estimation initiale (b) cycle 1

(c) cycle 2 (d) résultat final

FIG. 5.7 – Résultats intermédiaires pour l'estimation des champs de disparité de la paire d'images *Corridor* modifiée.



(a) estimation initiale (b) cycle 1

(c) cycle 2 (d) résultat final

FIG. 5.8 – Résultats intermédiaires pour l'estimation des champs d'illumination de la paire d'images *Corridor* modifiée.

5.5.2 Images réelles avec des changements d'illumination réels

5.5.2.1 Images réelles avec vérité terrain

Nous présentons dans ce paragraphe les résultats obtenus avec des images de scènes réelles soumises à des changements d'illuminations réels. Ces bases d'images ont été créées et mises à disposition sur le web [7] par Daniel Scharstein. Chaque base d'images contient 7 vues rectifiées prises de points de vue équidistants le long d'une même droite et soumises à trois changements d'illumination différents. La vérité terrain, obtenue par des méthodes utilisant des sources de lumière structurée [Scharstein, Szeliski, 2003], est fournie pour les vues 2 et 6 de chaque base d'images. Nous donnons ici les résultats pour les deux paires d'images présentées à la figure 5.9, à savoir *Dolls* et *Books*. Ces images sont difficiles à apparier car, d'une part, les changements d'illuminations sont significatifs et, d'autre part, la plage de variation de disparité est très large. Les valeurs de la disparité vont, en effet, de 20 à 75 pixels.

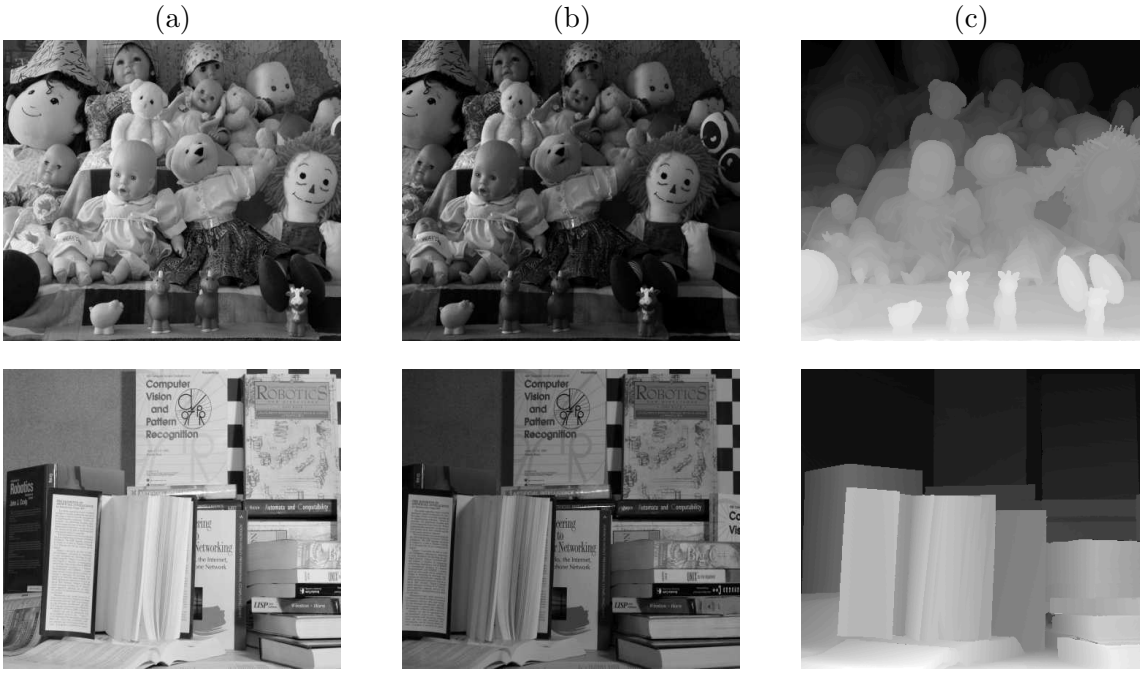
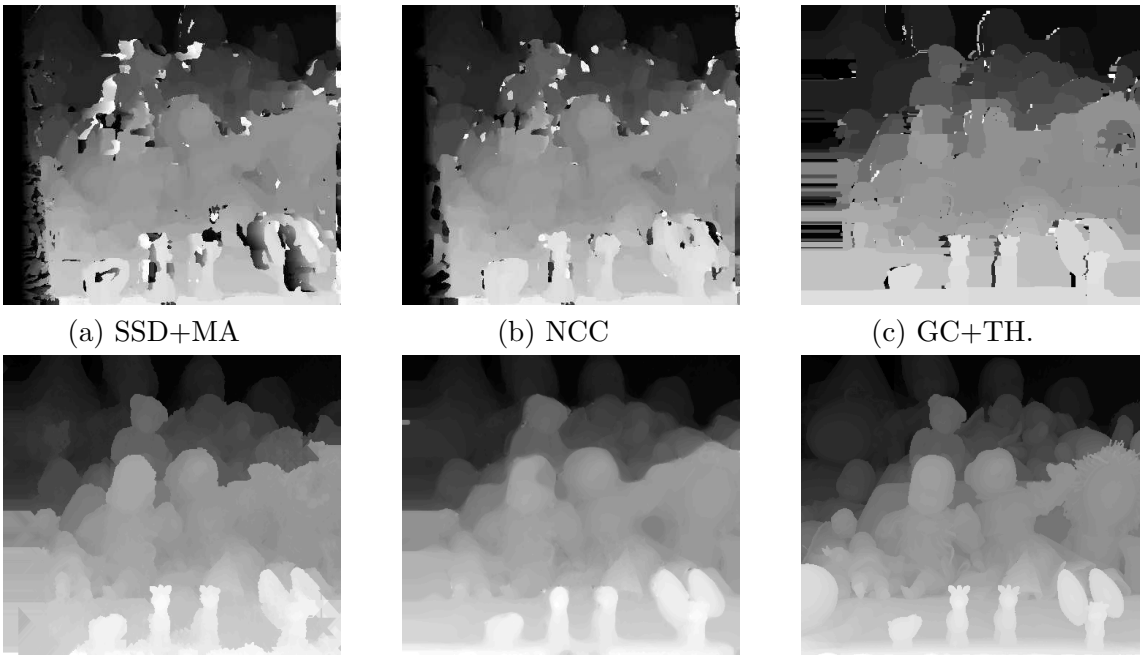


FIG. 5.9 – Paires d'images *Dolls* (en haut) et *Books* (en bas) avec vérité terrain. (a) Image de gauche. (b) Image de droite. (c) Vérité terrain.

Les résultats obtenus pour ces paires d'images avec les différentes méthodes considérées sont exposés sur les figures 5.10 et 5.11. Nous pouvons constater que les approches SSD+MA et NCC sont toutes les deux très sensibles aux changements d'illumination et entraînent des erreurs d'appariement significatives dans les zones de l'image où l'effet des variations de luminosité est important. L'algorithme CG combiné à une transformation d'histogramme maintient globalement un lissage des zones homogènes de l'image mais ne permet pas de reconstruire les détails fins, notamment au niveau des discontinuités des objets. Le résultat fourni par notre approche traduit la pertinence du modèle d'illumination choisi en dépit de la

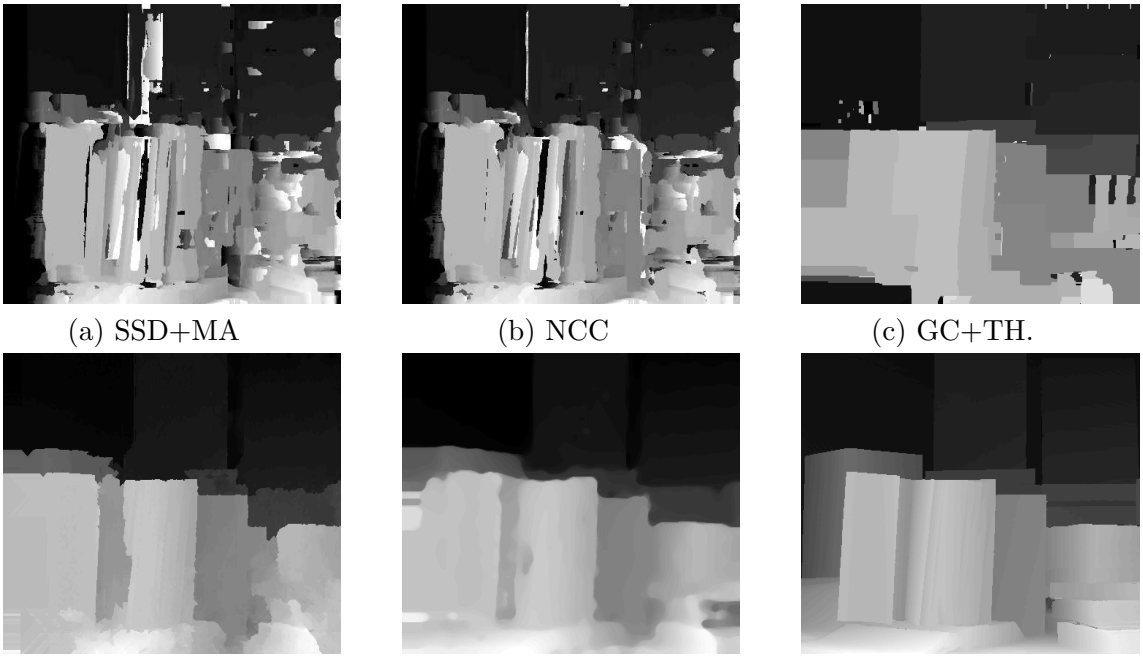
[7] http ://cat.middlebury.edu/stereo/scenes2005/



(a) SSD+MA (b) NCC (c) GC+TH.

(d) SGM (e) Notre approche (f) Verité terrain

FIG. 5.10 – Images de disparité pour la paire stéréoscopique *Dolls*.



(a) SSD+MA (b) NCC (c) GC+TH.

(d) SGM (e) Notre approche (f) Verité terrain

FIG. 5.11 – Images de disparité pour la paire stéréoscopique *Books*.

complexité des couples d'images considérés. Pour l'analyse quantitative, les mesures d'erreurs sont calculées comme précédemment. Les résultats présentés dans le tableau 5.3 montrent que notre méthode permet d'obtenir les meilleurs résultats parmi toutes les méthodes, en ce qui concerne l'erreur absolue moyenne et le pourcentage de mauvais appariements. Enfin, notons que les performances de notre approche et celles de l'algorithme SGM sont très proches, non seulement visuellement, mais aussi par rapport aux erreurs mesurées.

	Dolls		*Books*	
Technique	EAM	`Err`	EAM	`Err`
SSD+MA	2.65	23	4.74	34
NCC	1.73	21	3.35	27
GC [Kolmogorov, Zabih, 2001] +TH	1.26	25	2.64	26
SGM [Hirschmüller, 2005]	0.67	11	0.93	**16**
Notre approche	**0.51**	**9**	**0.85**	**16**

TAB. 5.3 – Résultats comparatifs pour les paires stéréoscopiques *Dolls* et *Books*.

5.5.2.2 Image réelle sans vérité terrain

Nous montrons enfin les résultats obtenus par les différentes méthodes comparées sur une paire d'images réelle ayant subi de réels changements d'illumination. La figure 5.12 présente la paire stéréoscopique *Shrub* issue de la base de données JISCT[8]. En examinant cette paire d'images, nous avons remarqué une différence entre les histogrammes des intensités de gris dans les images de gauche et de droite. Ceci implique que la contrainte de conservation de la luminance n'est pas respectée pour cette paire stéréoscopique, ce qui peut s'expliquer par un éventuel changement d'illumination entre les deux images.

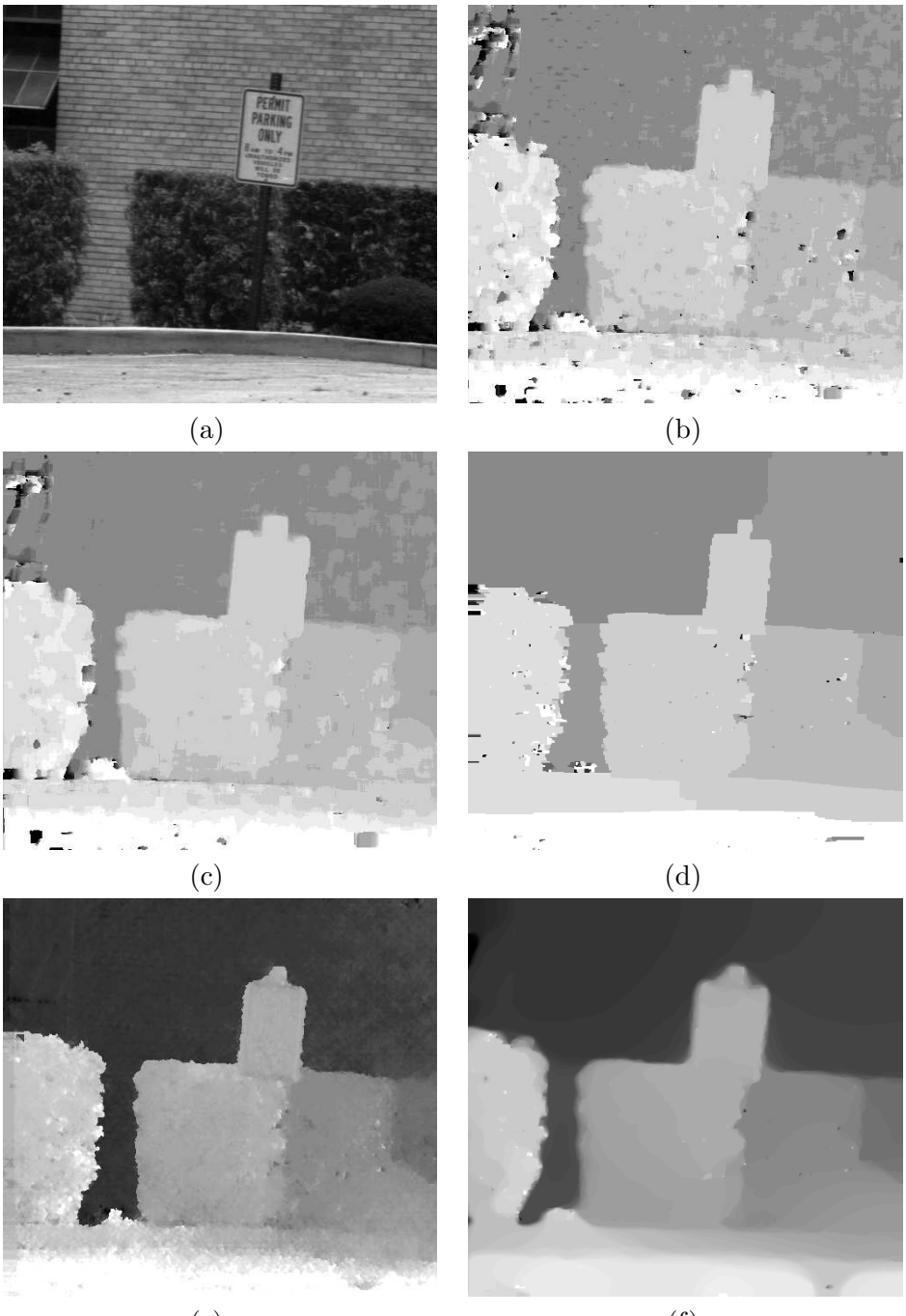


(a) (b)

FIG. 5.12 – Paire d'images *Shrub*. (a) Image de gauche. (b) Image de droite.

Nous reportons sur la figure 5.13 le résultat obtenu par la méthode proposée, que nous comparons avec les résultats des autres méthodes. Cependant, la vérité terrain étant inconnue, la comparaison ne peut être établie que visuellement. Nous observons ainsi que les méthodes

[8]http ://vasc.ri.cmu.edu/idb/html/jisct/



(a) (b)

(c) (d)

(e) (f)

FIG. 5.13 – Résultats pour la paire stéréoscopique réelle *Shrub*. (a) Image de gauche. (b) SSD+MA. (c) NCC. (d) GC+HT (e) SGM. (f) Notre approche.

locales sont, comme cela a été constaté précédemment, très sensibles aux changements d'illumination. À l'opposé, la technique GC+TH montre une nette amélioration pour ce couple d'images, due à la présence de larges zones homogènes. Enfin, nous pouvons signaler que notre approche ainsi que l'algorithme SGM permettent de mieux s'affranchir des problèmes liés aux variations de luminosité.

5.6 Conclusion

La plupart des algorithmes de vision par ordinateur supposent que les objets sont lambertiens et qu'aucun changement d'illumination ne se produit entre les images d'une même scène. Dans le cadre de la mise en correspondance stéréoscopique, ceci se traduit par une hypothèse de conservation de la luminance entre les projections, dans les images gauche et droite, d'un même point physique de la scène. Il s'agit bien sûr d'une hypothèse qui n'est pas toujours réaliste.

Les approches classiques proposées dans la littérature pour pallier ce problème ne permettent pas de prendre en compte la totalité des variations d'illuminations susceptibles de se produire entre les deux vues d'une même scène, dans la mesure où elles supposent que ces variations sont globales (au moins sur certaines régions). Dans ce chapitre, nous avons proposé un modèle d'illumination multiplicatif qui varie spatialement. Ce modèle permet de compenser localement les changements d'illumination.

À partir du modèle proposé, nous avons formulé le problème de mise en correspondance comme un problème de programmation convexe où une fonction objectif quadratique convexe est minimisée sur l'intersection d'ensembles convexes construits à partir des connaissances *a priori* sur les champs de disparité et d'illumination. Les résultats obtenus sur un exemple synthétique, où des variations de luminosité ont été simulées, montrent que notre approche est robuste vis-à-vis de ces changements. Nous l'avons par la suite testée sur des images ayant subi de vrais changements d'illumination et les résultats se sont également avérés satisfaisants. Par ailleurs, nous avons comparé notre approche aux approches de mise en correspondance utilisées dans la littérature pour prendre en compte le problème des changements d'illumination. Parmi les méthodes passées en revue dans ce chapitre, nous avons considéré les suivantes : une méthode avec un prétraitement (CG+TH), une méthode utilisant une mesure de corrélation robuste (NCC), une modélisation affine des changements d'illumination (SSD+MA) et une méthode récente (SGM) basée sur l'information mutuelle. Au vu des comparaisons visuelles et quantitatives établies, il s'avère que l'approche proposée est plus robuste et permet de mieux s'affranchir des problèmes liés aux variations de luminosité.

Chapitre 6

Application à la détection d'obstacles routiers

6.1 Introduction

La détection d'obstacles dans un milieu routier est une tâche primordiale pour la navigation autonome des systèmes de transport intelligents. Cette tâche était essentiellement fondée sur l'utilisation de capteurs actifs tels que les radars ou les télémètres lasers. Cependant, l'inconvénient majeur de ce type de capteurs est leur coût prohibitif. À cela, s'ajoute une résolution angulaire faible ne permettant pas de détecter tous les types d'obstacles. Avec l'amélioration des techniques d'analyse d'images, la vision est devenue un moyen de détection d'obstacles plus attractif, car elle fournit des solutions peu coûteuses et offre une perception plus homogène des éléments qui structurent la scène routière.

Dans ce chapitre, nous proposons une méthode de détection d'obstacles à partir d'un système de vision embarqué à bord d'un véhicule en mouvement. Grâce aux travaux présentés dans les chapitres précédents, nous disposons désormais d'une méthode robuste et efficace pour l'estimation de cartes de disparité denses et précises. Le principe de l'approche décrite dans ce chapitre est d'exploiter ces cartes pour la détection d'obstacles routiers. Cette approche se décompose en trois étapes fondamentales. Dans un premier temps, l'image de disparité est estimée à partir de la méthode présentée dans le chapitre 5 en prenant en compte les changements d'illumination qui peuvent se produire entre les deux images stéréoscopiques. La détection d'obstacles revient ensuite à classifier les pixels de cette image, à l'aide d'un critère approprié d'orientation de surface, en deux catégories : les pixels appartenant à la route et ceux appartenant aux obstacles. Enfin, à partir de cette détection, une étape de caractérisation d'obstacles est réalisée.

Dans ce qui suit, nous proposons tout d'abord un survol non exhaustif des méthodes de détection d'obstacles fondées sur la vision. La section suivante décrit notre méthode de détection d'obstacles par stéréovision. Les résultats montrés sur des images réelles de scènes routières permettent de valider la méthode proposée.

6.2 Un survol des méthodes de détection d'obstacles par vision

Le développement de méthodes d'aide à la conduite ou d'automatisation complète de cette tâche fait ressortir un verrou majeur : celui de la détection d'obstacles et de leur évitement. Les techniques de vision par ordinateur se prêtent particulièrement bien à cette tâche du fait de leurs aptitudes en matière de segmentation et de perception. De nombreuses méthodes de détection d'obstacles fondées sur la vision ont été ainsi développées ces dernières années. Nous pouvons distinguer deux types d'approches : utilisant une seule caméra (vision monoculaire) ou exploitant deux caméras (vision stéréoscopique).

6.2.1 Les approches basées sur la vision monoculaire

L'approche monoculaire se limite à l'analyse d'une seule image en se basant sur des connaissances *a priori* des obstacles que le véhicule est susceptible de rencontrer. Cette analyse repose sur la recherche d'attributs ou de caractéristiques qui peuvent être de différents types : photométriques (niveaux de gris, texture, histogramme), morphologiques (segments, éléments de contours, symétries) et cinématiques (mouvement apparent).

Caractérisation photométrique

D'une manière générale, les méthodes fondées sur cette caractérisation se décomposent en deux étapes : une étape d'extraction d'attributs propres aux objets à détecter (principalement des piétons ou des véhicules) et une étape d'apprentissage ou de reconnaissance de formes. La première étape permet de trouver l'information pertinente dans les images. L'étape d'apprentissage consiste ensuite à analyser cette information en mettant en correspondance les primitives extraites avec un modèle d'obstacle ou une classe d'objets. L'élaboration d'un modèle nécessite la création au préalable d'une base de données composée d'images d'obstacles prises pour différentes poses et différents angles de vue. La classification obstacle/non-obstacle est générée par l'utilisation de réseaux de neurones [Zhao, Thorpe, 2000] ou par SVM (*Support Vector Machine*) [Shashua *et al.*, 2004, Papageorgiou, Poggio, 1999].

Ces méthodes ont l'avantage d'être simples et rapides mais elles sont très sensibles au bruit et aux variations d'illumination de la scène. Le principe de la méthode proposée dans [Hotter *et al.*, 1996], pour s'affranchir de celle limitation, est de segmenter l'image en régions selon un critère basé sur des informations statistiques, comme l'espérance et la variance des niveaux de gris. En effet, le fait d'utiliser des caractéristiques globales (des moments) et non plus simplement des attributs locaux (niveaux de gris) permet d'améliorer la robustesse de ces méthodes en présence de bruit. Cependant, ne permettant pas de conclure de façon certaine quant à la présence d'un obstacle, elles sont surtout utilisées pour localiser des régions d'intérêt contenant des obstacles potentiels ou pour confirmer la détection d'un obstacle et réduire ainsi les fausses alarmes d'autres méthodes.

Caractérisation morphologique

Les méthodes fondées sur cette caractérisation s'appuient sur les mêmes principes que celles décrites précédemment. Néanmoins, du fait des primitives choisies, elles bénéficient d'une

plus grande robustesse au bruit et aux changements d'illumination. Parmi ces méthodes, nous pouvons citer les nombreux travaux de Broggi et al. [Broggi *et al.*, 2000, Bertozzi *et al.*, 2002, Bertozzi *et al.*, 2005] qui caractérisent les piétons par la symétrie ou la taille de leurs contours verticaux. Papageorgiou et al. [Papageorgiou, Poggio, 1999] proposent de définir le modèle de piéton à partir des coefficients de détails horizontaux, verticaux et diagonaux, issus d'une décomposition en ondelettes 2D d'une base d'images de piétons.

Caractérisation cinématique

Dans cette catégorie, un obstacle correspond à une région de l'image différente, au sens du mouvement apparent, du reste de la scène [Demonceaux *et al.*, 2004, Giachetti *et al.*, 1998, Enkelmann, 1991]. Les méthodes reposant sur cette caractérisation d'obstacles nécessitent l'estimation ou la connaissance *a priori* du mouvement de la caméra fixée sur le véhicule. Les obstacles correspondent ainsi aux points de l'image dont le mouvement apparent, calculé entre deux images consécutives, n'est pas conforme à celui induit par le mouvement de la caméra. Lors d'un mouvement de translation de la caméra, le champ de mouvement convergeant vers un point appelé *foyer d'expansion*, les régions de l'image dont le vecteur de mouvement ne converge pas vers ce point peuvent être interprétées comme appartenant à des obstacles [Takeda *et al.*, 1996].

Notons que les méthodes d'analyse du mouvement permettent non seulement la détection d'obstacles mais également une estimation de leur vitesse et par conséquent une analyse dynamique de la scène. Cependant, ces méthodes sont sensibles aux vibrations de la caméra et sont généralement complexes à mettre en œuvre.

6.2.2 Les approches basées sur la vision stéréoscopique

On peut classer les approches de détection d'obstacles par stéréovision en deux catégories : celles mettant en œuvre une rectification homographique et celles basées sur une mise en correspondance entre les deux images stéréoscopiques.

Les méthodes dans la première catégorie appliquent une transformation homographique sur une des deux images de manière à ce que le plan de la route soit identique dans les deux images stéréoscopiques. Les régions de l'image qui ne vérifient pas l'homographie induite par ce plan indiquent alors la présence d'objets sur la route. L'estimation de la matrice d'homographie est obtenue suite à une calibration ou par minimisation de la différence des niveaux de gris entre les deux images [Okutomi *et al.*, 2002]. L'avantage de ces méthodes est ne pas réaliser d'appariements, d'où des gains importants en temps de calcul. Cependant, elles sont assez sensibles aux variations d'illumination qui peuvent se produire entre les deux images.

Les méthodes par mise en correspondance se basent sur la carte de disparité issue de l'appariement de primitives entre les deux images de la paire stéréoscopique [Badal *et al.*, 1994, van der Mark, Gavrila, 2006, Yu *et al.*, 2003, Andersen *et al.*, 2005, Toulminet *et al.*, 2004]. La plupart des approches dans cette catégorie exploitent cette carte pour estimer le plan de la route et localiser les objets situés au-dessus de la surface de la route.

La méthode développée par Labayrade *et al.* [Labayrade *et al.*, 2002] est basée sur la

construction d'une image, appelée *V-disparité*[1], fournissant une bonne représentation géométrique de la scène routière. Une carte de disparité éparse est tout d'abord calculée par mise en correspondance des maxima locaux des gradients horizontaux. À partir de cette carte, l'image *V-disparité* est construite en accumulant les pixels d'une même ligne ayant une disparité commune. Elle présente ainsi pour chaque ligne de l'image la fonction de répartition en niveaux de gris des disparités mesurées, reportée en abscisse. Dans l'espace *V-disparité*, le plan de la route est transformé en une droite oblique, et les obstacles en segments verticaux. L'extraction de ces droites, par une transformée de Hough, permet enfin d'estimer le profil longitudinal de la route et de détecter les obstacles sur la surface de la route.

L'approche *V-disparité*, permettant de segmenter de manière fiable et rapide les surfaces planaires (route, obstacles) observées dans la scène, a été exploitée et étendue par de nombreux auteurs dans le cadre de la détection d'obstacles routiers. Lemonde *et al.* [Lemonde, Devy, 2004] reprennent cette approche pour prendre en compte le roulis, le dévers ou les variations de pente de la route en travaillant dans l'espace tridimensionnel et non dans l'espace *V-disparité*. En outre, pour permettre l'utilisation de la *V-disparité* sur des routes non structurées, Soquet *et al.* [Soquet *et al.*, 2007] proposent d'améliorer la répartition des primitives appariées en adaptant le seuil sur le gradient de l'image en fonction de la complexité de l'environnement routier. Ceci permet, en effet, l'obtention d'un profil de route précis, prenant en compte des géométries plus complexes de la route. Hu *et al.* [Hu *et al.*, 2005] utilisent un nouveau concept, appelé *U-V disparité*, qui consiste à calculer, en plus de l'image *V-disparité*, une image *U-disparité* [Labayrade, 2004]. Celle-ci est construite en accumulant les pixels de même disparité le long de chaque colonne de l'image. L'exploitation de ces deux images permet d'extraire la surface de la route, les obstacles ainsi que les structures latérales telles que les façades des bâtiments ou les parties latérales des véhicules.
D'autres approches, comme [Bertozzi *et al.*, 2005], utilisent la vision stéréoscopique seulement pour mesurer la distance d'un objet au véhicule après l'avoir détecté par une approche monoculaire. En effet, seules les méthodes qui exploitent la disparité entre deux images stéréoscopiques permettent d'obtenir des informations tridimensionnelles sur la scène observée.

En conclusion, la vision monoculaire, bien qu'elle permette dans la majorité des cas de détecter les obstacles routiers, ne permet pas d'estimer leurs distances sans ajouter un autre capteur, des hypothèses supplémentaires sur la scène ou une connaissance *a priori* sur les objets à détecter. L'utilisation de la stéréovision permet, en revanche, de mesurer précisément la profondeur de chaque point de la scène par un simple procédé de triangulation. L'information de distance étant primordiale pour localiser les obstacles détectés, nous avons opté, dans le cadre de cette thèse, pour une méthode de détection d'obstacles par stéréovision que nous décrivons dans ce qui suit.

6.3 Notre méthode de détection d'obstacles par stéréovision

Dans ce paragraphe, nous proposons une méthode de détection d'obstacles dans un milieu routier à partir d'un système de stéréovision embarqué à bord d'un véhicule. Cette méthode repose essentiellement sur l'estimation et l'exploitation de cartes de disparité denses et précises (cf. chapitres 3 et 5).

[1] Selon les notations des auteurs, V correspond à l'ordonnée d'un pixel dans un repère image (U, V).

6.3.1 Matériel d'expérimentation

Le Cycab

La plate-forme sur laquelle ont été réalisées les expérimentations de ce chapitre est le Cycab (voir figure 6.1). C'est un véhicule électrique de petite taille (1.90 m de long sur 1.20 m de large), qui peut transporter deux personnes à une vitesse maximale de 30 km/h. Il est conçu pour être utilisé de façon sûre et simple avec une conduite manuelle simplifiée et sécurisée ou une conduite totalement automatique. La conduite manuelle repose sur l'utilisation d'une manette positionnée entre les deux sièges. Le véhicule dispose également d'un écran tactile pour permettre à l'utilisateur de récupérer des informations telle que la localisation du véhicule dans l'environnement ou, en cas de conduite automatique, d'indiquer le trajet qu'il veut effectuer.



FIG. 6.1 – Le Cycab.

Le Cycab constitue actuellement une plate-forme de recherche dédiée pour expérimenter le concept du véhicule intelligent et son intégration dans la ville de demain, avec l'objectif à terme d'équiper ces villes de flotte de véhicules en libre-service [Parent, Fortelle, 2005]. Il intègre ainsi dans sa conception matérielle et logicielle, les contraintes de bas coût, de robustesse et de facilité de maintenance.

Le système de stéréovision Bumblebee

Le système d'acquisition vidéo adopté sur la plate-forme Cycab est le système de stéréovision Bumblebee[2] (cf. figure 6.2). Ce système, qui se trouve à l'avant du véhicule à une hauteur $h = 80$ cm au dessus du sol, contient deux caméras numériques CCD (*Charge-Coupled Device*) et fournit des images à une fréquence de 15 images par seconde. Les images peuvent être

[2]http ://www.erasme.org/IMG/bumblebee.pdf

acquises en couleur ou en niveau de gris et avoir une taille de 480×640 ou 240×320 pixels. Les paramètres intrinsèques des caméras sont :

- la distance focale $f = 4$ mm ;
- la distance entre les centres optiques $B = 12$ cm ;
- la longueur focale $\alpha_f = \alpha_x = \alpha_y = 269$ pixels.



FIG. 6.2 – Système de vision Bumblebee.

Les deux caméras sont pré-calibrées et placées dans une configuration parallèle où leurs axes optiques sont parallèles entre eux et perpendiculaires à la droite joignant leurs centres optiques (cf. figure 6.3). Ceci implique que la géométrie épipolaire du capteur stéréoscopique utilisé est rectifiée. En outre, un logiciel permettant la correction des distorsions optiques et la rectification des images, dans le cas d'un éventuel désalignement des caméras, est fourni avec ce système de stéréovision.

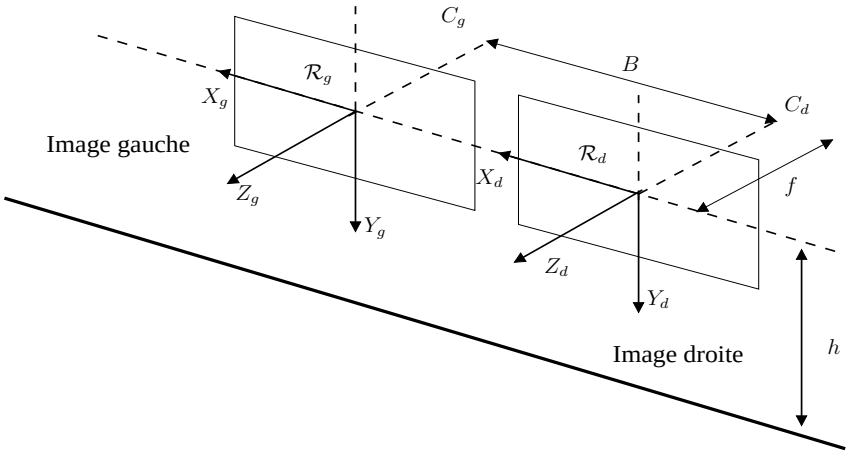
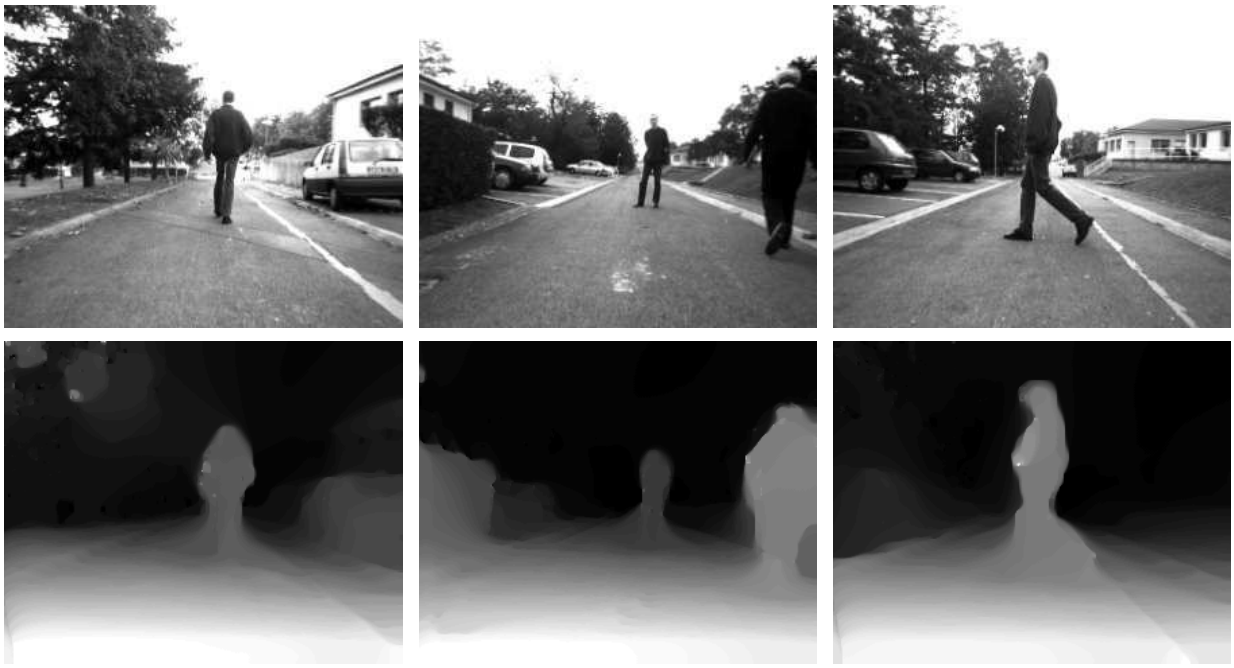


FIG. 6.3 – Configuration des caméras du système de vision stéréoscopique.

6.3.2 Estimation de la carte de disparité

La première étape de la méthode de détection d'obstacles proposée consiste à estimer le champ de disparité à partir de deux images stéréoscopiques acquises par le système de stéréovision embarqué à bord du véhicule Cycab. Pour des contraintes de temps réel, la plupart des approches de détection existantes utilisent une carte de disparité éparse, c'est-à-dire une carte où l'appariement n'est réalisé qu'en certains points d'intérêt dans l'image. L'utilisation d'une carte de disparité éparse ne permet cependant pas de localiser et de caractériser de manière très fiable les obstacles détectés. De plus, la méthode d'appariement, souvent utilisée, est locale et est donc nécessairement sensible aux faux appariements. Ceci empêche d'exploiter directement la carte de disparité générée comme carte de profondeur de l'environnement.



(a) (b) (c)

FIG. 6.4 – Résultats de l'estimation de la disparité sur des paires d'images de scènes routières. En haut : les images de gauche. En bas : les cartes de disparité estimées.

Dans ce travail, nous adoptons la méthode d'estimation de la disparité présentée au chapitre 5. En effet, celle-ci fournit des cartes de disparité denses, précises, robustes au bruit et aux variations d'illumination de la scène observée. Elle semble donc la mieux adaptée. Les images que nous avons considérées sont en niveaux de gris et de taille 240×320 pixels. Les paramètres qui interviennent dans notre algorithme, pour ces images testées, sont donnés dans le tableau 6.1. La figure 6.4 représente les images de gauche de trois paires stéréoscopiques de scènes routières ainsi que les cartes de disparité calculées par notre méthode par rapport à ces images de référence. Ces cartes sont cohérentes avec la structuration des éléments constituant les différentes scènes filmées. En effet, on y identifie clairement le plan de la route, les piétons ainsi que les voitures stationnées sur le parking. En outre, sur les trois cartes, le plan de la route montre un dégradé cohérent avec la profondeur de la scène et les disparités associées aux piétons sont en adéquation avec leurs profondeurs réelles. Par exemple, sur la figure 6.4(b), le piéton à droite de l'image a une disparité plus claire et est donc plus proche du véhicule que le piéton au centre, ce qui correspond bien à la réalité.

α	δ	κ_5	ζ	intervalle de disparité	intervalle d'illumination
50	6×10^5	4×10^3	2	$[2, 20]$	$[0.8, 1.2]$

TAB. 6.1 – Paramètres de l'algorithme d'estimation de la disparité pour les images considérées dans ce chapitre.

Il s'agit à présent d'exploiter ces cartes de disparité pour distinguer les obstacles du plan de la route.

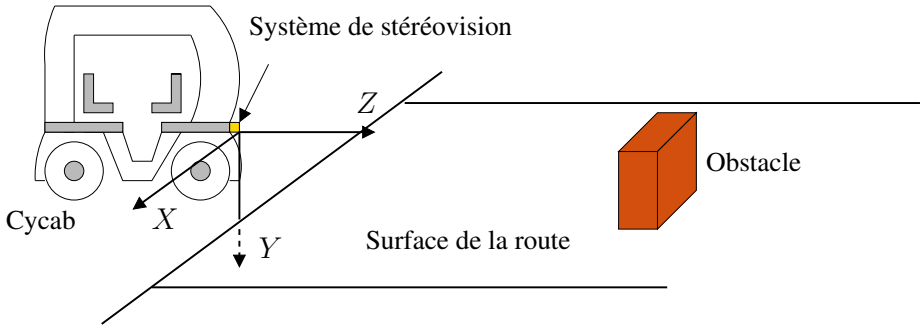


FIG. 6.5 – Système de coordonnées du véhicule.

6.3.3 Classification des pixels de la carte de disparité

En analysant les cartes de disparité de la figure 6.4, nous pouvons observer qu'une scène de la route est constituée essentiellement d'une large surface horizontale représentant le plan de la route et d'objets, ayant des surfaces verticales, situés au-dessus de ce plan. L'environnement routier peut ainsi être modélisé simplement par des surfaces horizontales et verticales. Pour identifier les éléments constituant cet environnement, il s'agit donc d'extraire, à partir de l'image de disparité, les surfaces globales qui correspondent soit à la surface de la route, soit à des obstacles dans la scène.

La modélisation de l'environnement routier par des surfaces planes a été adoptée dans plusieurs travaux. La technique *V-disparité* [Labayrade *et al.*, 2005] et sa version étendue *U-V disparité* [Hu *et al.*, 2005, Soquet *et al.*, 2007] déjà évoquées font partie des plus célèbres. Travaillant dans l'espace *V-disparité* et/ou *U-disparité*, l'estimation des plans de la scène, suivant ces approches, revient à détecter des droites projetées dans ces espaces. L'approche que nous proposons ici est différente dans la mesure où elle travaille directement dans l'espace tri-dimensionnel, en tirant profit de la densité et de la précision des cartes de disparité estimées par notre méthode, décrite dans le chapitre précédent. Pour caractériser et séparer les surfaces contenues dans ces cartes, nous utilisons un critère de segmentation en plans. Le critère, que nous avons adopté est le critère d'orientation de surface.

Il existe différentes façons de spécifier l'orientation de la surface d'un plan. Nous pouvons, par exemple, donner l'équation définissant le plan ou, de manière équivalente, la direction du vecteur normal à la surface de ce plan. Dans le repère de la caméra de gauche, ramené au repère de l'espace, l'équation d'un plan ne passant pas par l'origine de l'espace (centre optique de la caméra de gauche) peut s'écrire comme suit :

$$aX + bY + cZ = 1, \tag{6.1}$$

où (X, Y, Z) sont les coordonnées spatiales. Comme nous l'avons déjà mentionné au paragraphe 2.2.3.7, la disparité est inversement proportionnelle à la distance des objets aux caméras. Nous avons exprimé, dans ce paragraphe, l'équation liant la profondeur d'un point M de

coordonnées (X, Y, Z) à la disparité u de ses projections dans les deux images comme suit :

$$u = \frac{B\ \alpha_f}{Z}, \tag{6.2}$$

où B est la distance entre les centres optiques des deux caméras et α_f est la longueur focale exprimée en pixels. Il est également intéressant de rappeler ici l'expression des coordonnées, exprimées dans le repère image, de la projection du point M sur l'image de gauche. Celle-ci est donnée par (voir paragraphe 2.2) :

$$\begin{aligned} x &= \alpha_f \frac{X}{Z}, \\ y &= \alpha_f \frac{Y}{Z}. \end{aligned} \tag{6.3}$$

En combinant les équations (6.1), (6.2) et (6.3), nous pouvons définir l'équation d'un plan dans l'espace tri-dimensionnel (x, y, u), que nous apppelons *espace de disparité*, de la manière suivante :

$$Ba\ x + Bb\ y - u = -B\alpha_f c. \tag{6.4}$$

À partir de cette équation, nous pouvons conclure qu'un plan de l'espace de vecteur normal (a, b, c) est aussi un plan dans l'espace de disparité (x, y, u) de vecteur normal $(Ba, Bb, -1)$. Par ailleurs, il est intéressant de noter que le gradient de l'image de disparité s'écrit en fonction des composantes a et b du vecteur (a, b, c) normal au plan (6.1). En effet, on vérifie aisément que :

$$(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}) = B(a, b). \tag{6.5}$$

Nous allons dans ce qui suit caractériser les plans de l'espace que l'on cherche à déterminer en s'appuyant uniquement sur les vecteurs normaux à la surface générés à partir de l'image de disparité. En considérant des vecteurs unitaires, ceci revient plus précisément à calculer en chaque point de la carte de disparité un vecteur normal unitaire (a_n, b_n, c_n) tel que :

$$(a_n, b_n, c_n) = (\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, -1) / \sqrt{\Big((\frac{\partial u}{\partial x})^2 + (\frac{\partial u}{\partial y})^2 + 1\Big)}. \tag{6.6}$$

En raisonnant toujours dans l'espace de disparité, nous définissons à présent les différents plans de cet espace en nous intéressant plus particulièrement aux plans horizontaux et verticaux.

Les plans horizontaux

En supposant que dans le système de coordonnées du véhicule, représenté à la figure 6.5, l'axe Y est toujours perpendiculaire au plan du sol, nous pouvons considérer que les valeurs

absolues de la composante b_n du vecteur normal unitaire (a_n, b_n, c_n) proches de 1 caractérisent les plans horizontaux et définir ainsi les projections de ces plans par :

$$P_h = \{(x, y) \in \mathbb{R}^2 \ \mid \ |b_n| \geq s_y\}, \tag{6.7}$$

où $s_y \leq 1$ est un seuil à fixer. Il est évident que ce seuil doit être choisi proche de 1.

Les plans verticaux

De même, l'axe Z du repère du véhicule étant supposé parallèle au plan horizontal du sol, un seuillage approprié sur les valeurs absolues de la composante c_n permet d'extraire les plans verticaux, qui peuvent être donc définis par :

$$P_v = \{(x, y) \in \mathbb{R}^2 \ \mid \ |c_n| \geq s_z\}, \tag{6.8}$$

où $s_z \leq 1$ est un seuil à prendre proche de 1.

Notons que nous pouvons définir de manière similaire des plans transverses en considérant la composante a_n des vecteurs normaux unitaires. Soit, pour un seuil s_x donné,

$$P_t = \{(x, y) \in \mathbb{R}^2 \ \mid \ |a_n| \geq s_x\}. \tag{6.9}$$

Dans ce travail, nous ne prenons pas en compte explicitement ces plans au cours de la segmentation, mais, en revanche, on les classe comme appartenant à des plans horizontaux ou verticaux suivant leurs angles d'inclinaison. Pour cela, nous adoptons une définition étendue des plans P_h et P_v, prenant ainsi en compte les plans inclinés. Nous posons alors

$$P_{hi} = \{(x, y) \in \mathbb{R}^2 \ \mid \ |b_n| \geq s_y \text{ ou } |c_n| \leq (1 - s_y)\}, \tag{6.10}$$

et

$$P_{vi} = \{(x, y) \in \mathbb{R}^2 \ \mid \ |c_n| \geq s_z \text{ ou } |b_n| \leq (1 - s_z)\}. \tag{6.11}$$

6.3.3.1 Expérimentations sur la paire d'images synthétique Corridor

Pour illustrer la méthode proposée, nous allons, dans un premier temps, la valider sur les images de la scène *Corridor*, constituée de plans horizontaux, verticaux et inclinés. Cette scène de synthèse est, en effet, représentative d'un environnement routier : le sol peut être assimilé à un plan de la route et les objets situés sur le sol, tels que le cône et la balle, peuvent être considérés comme des obstacles. Nous avons calculé en chaque point de la carte de disparité associée à cette paire, présentée à la figure 6.6, un vecteur normal à la surface comme cela a été décrit précédemment. Les valeurs absolues des composantes de ces vecteurs suivant les axes X, Y et Z sont montrées sur les figures 6.6(d), (e) et (f), respectivement. Sur ces figures, plus le pixel est foncé, plus sa valeur est proche de 0 et plus il est clair, plus sa valeur est proche de 1. Nous pouvons ainsi observer que les valeurs absolues proches de 1 suivant chaque composante caractérisent un plan particulier. La composante a_n permet, en effet, d'indiquer la présence de surfaces transverses. De même, la composante b_n met en évidence les surfaces horizontales et la composante c_n les surfaces verticaux. Pour segmenter l'image de disparité suivant ces différents plans, nous avons dans un premier temps appliqué

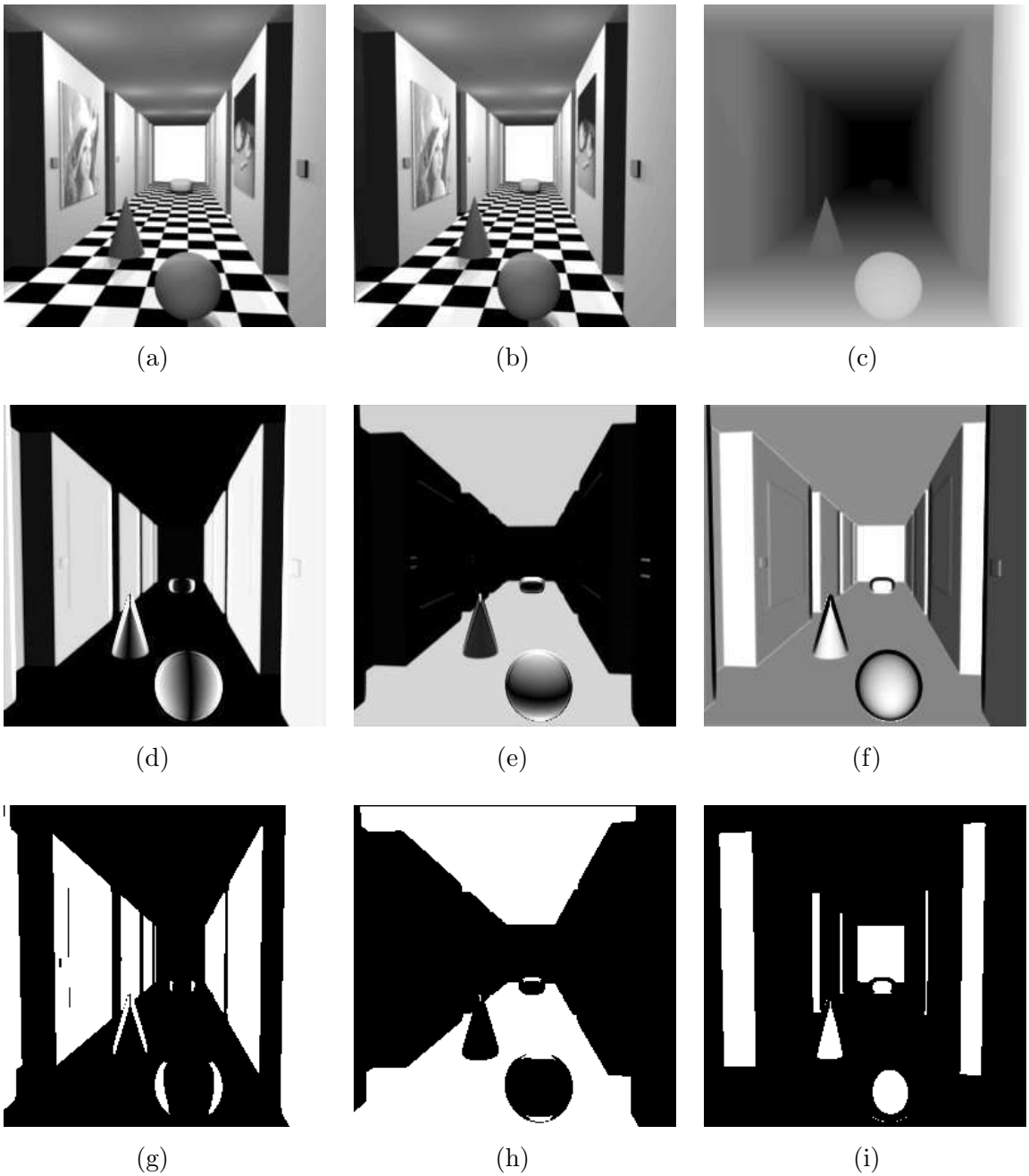


FIG. 6.6 – Résultats sur la paire de synthèse *Corridor*. En haut : les images stéréoscopiques et la carte de disparité associée. Au milieu : les valeurs absolues des composantes des vecteurs normaux à la surface suivant les axes (d) X, (e) Y et (f) Z. Résultats du seuillage sur les composantes (g) a_n, (h) b_n et (i) c_n.

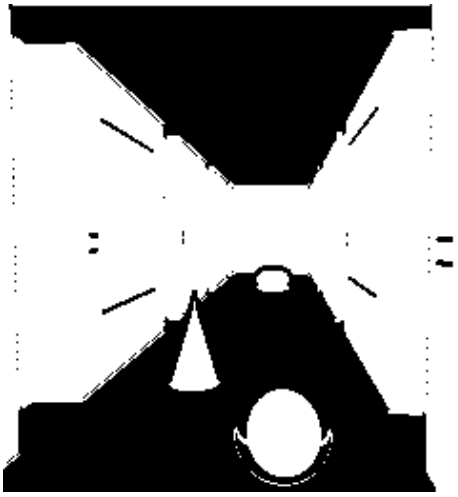


FIG. 6.7 – Résultat du seuillage sur la composante c_n de la figure 6.6(f) mettant en évidence les plans verticaux et obliques.

un seuillage sur les différentes composantes a_n, b_n et c_n en adaptant les définitions (6.7), (6.8) et (6.9) des plans correspondants et en prenant un seuil identique ($s_x = s_y = s_z$) égal à 0.8. Le résultat de cette segmentation parvient clairement à séparer les plans transverses (cf. figure 6.6(g)), horizontaux (cf. figure 6.6(h)) et verticaux (cf. figure 6.6(i)).

Nous allons à présent segmenter l'image de disparité uniquement en plans horizontaux et verticaux. Pour cela, nous utilisons les définitions (6.10) et (6.11) permettant de prendre en compte les plans inclinés. Nous montrons, en particulier, à la figure 6.7 le résultat de l'extraction des plans verticaux selon la définition (6.11) et nous observons que les plans transverses des murs verticaux ressortent bien dans cette image et sont donc désormais considérés comme des plans verticaux.

Le résultat final de notre segmentation superposé à l'image de gauche de la paire *Corridor* est présenté à la figure 6.8. Les pixels bleus appartiennent ainsi à des surfaces horizontales et les pixels rouges appartiennent à des objets verticaux. Ce résultat confirme que notre méthode permet une segmentation efficace de la carte de disparité en surfaces horizontales et verticales. Nous pouvons donc, à présent, l'appliquer sur des images de scènes routières pour la détection des obstacles routiers.

6.3.3.2 Expérimentations sur des images réelles de scènes routières

Nous présentons dans cette section les résultats de notre méthode sur des images réelles de scènes routières acquises par le système de stéréovision Bumblebee embarqué à bord du Cycab. La première séquence, composée de 60 images, montre un piéton immobile au milieu de la route et un autre piéton avançant en face du Cycab à la même vitesse. Nous donnons figure 6.9 (lignes 2 et 4) les résultats de notre méthode de détection d'obstacles sur six images de cette séquence. Ces résultats sont présentés, comme précédemment, en superposant les résultats de la classification des cartes de disparité aux images de référence. Les zones bleues correspondent ainsi au plan de la route et les zones rouges représentent les obstacles. Les



FIG. 6.8 – Classification des pixels de la carte de disparité : les pixels bleus appartiennent à des surfaces horizontales et les pixels rouges appartiennent à des objets verticaux.

cartes de disparité estimées par notre méthode sont également présentées sur la figure 6.9 (lignes 1 et 3). Notons que dans le cas de ces images, nous n'avons pas pris en compte, lors de la segmentation en plans de l'image de disparité, les points ayant une disparité inférieure à un. Ces points qui correspondent aux pixels noirs de l'image de disparité sont, en effet, très éloignés et ne présentent pas de danger pour le véhicule. La segmentation en plans de l'image de disparité est effectuée en considérant les définitions étendues des plans verticaux et horizontaux. En effet, comme on peut clairement le voir sur les premières images de cette séquence, le plan transverse correspondant au buisson est considéré comme un plan vertical. Pour éliminer les pixels isolés mal classifiés, en raison de mauvais appariements dans l'image de disparité, nous avons appliqué des opérations élémentaires de morphologie mathématique [Serra, 1988]. Les différents obstacles et le plan de la route sont ainsi correctement détectés.
La deuxième séquence montre un piéton traversant la route face au Cycab. Les cartes de disparités associées à six images de cette séquence sont montrées sur la figure 6.10 (lignes 1 et 3). Les résultats de la segmentation en plans de ces cartes de disparité superposés aux images de gauche sont présentés figure 6.10 (lignes 2 et 4). Sur les différentes images de la séquence, les piétons ainsi que les voitures stationnées sont correctement détectés comme des obstacles et le plan de la route, représenté en bleu, est bien estimé.

6.3.4 Caractérisation des obstacles détectés

Ayant désormais à disposition une carte de disparité de l'ensemble de l'environnement du véhicule segmentée en plans horizontaux et verticaux, nous cherchons, dans ce paragraphe, à caractériser les obstacles par leur distance au véhicule. Ceci permet de localiser les obstacles potentiels par rapport au véhicule et d'évaluer leur degré de dangerosité. Compte tenu de ses qualités, la carte de disparité estimée par notre méthode peut être utilisée directement comme carte de profondeur de l'environnement du véhicule.
Notre méthode de caractérisation consiste, dans un premier temps, à positionner des cadres englobants autour des objets verticaux détectés lors de la phase de segmentation. Les extré-

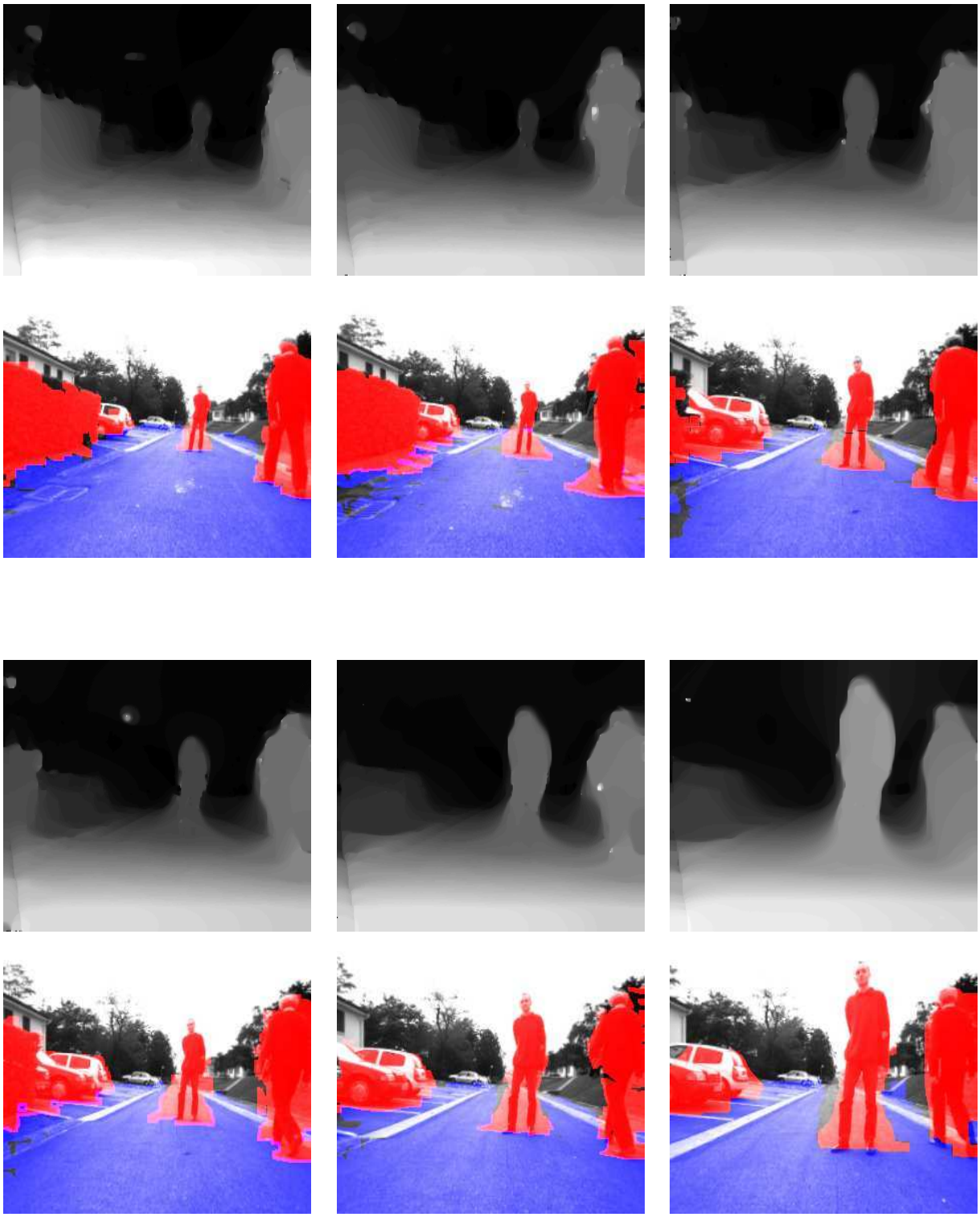


FIG. 6.9 – Résultats sur les images d'une scène routière. Première et troisième lignes : les cartes de disparité estimées par notre méthode. Deuxième et quatrième lignes : classification des pixels de la carte de disparité : les zones bleues correspondent au plan de la route et les zones rouges représentent les obstacles.

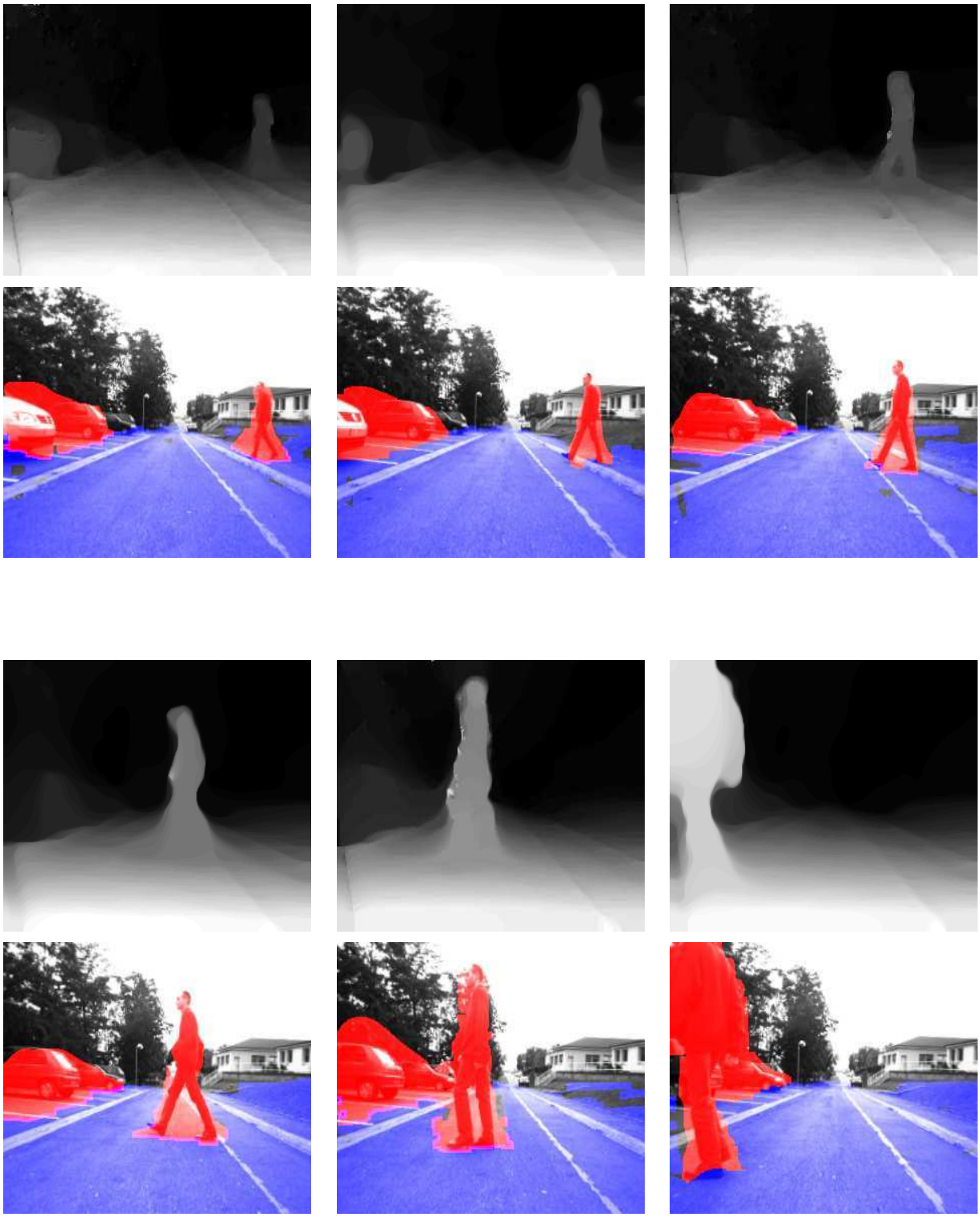
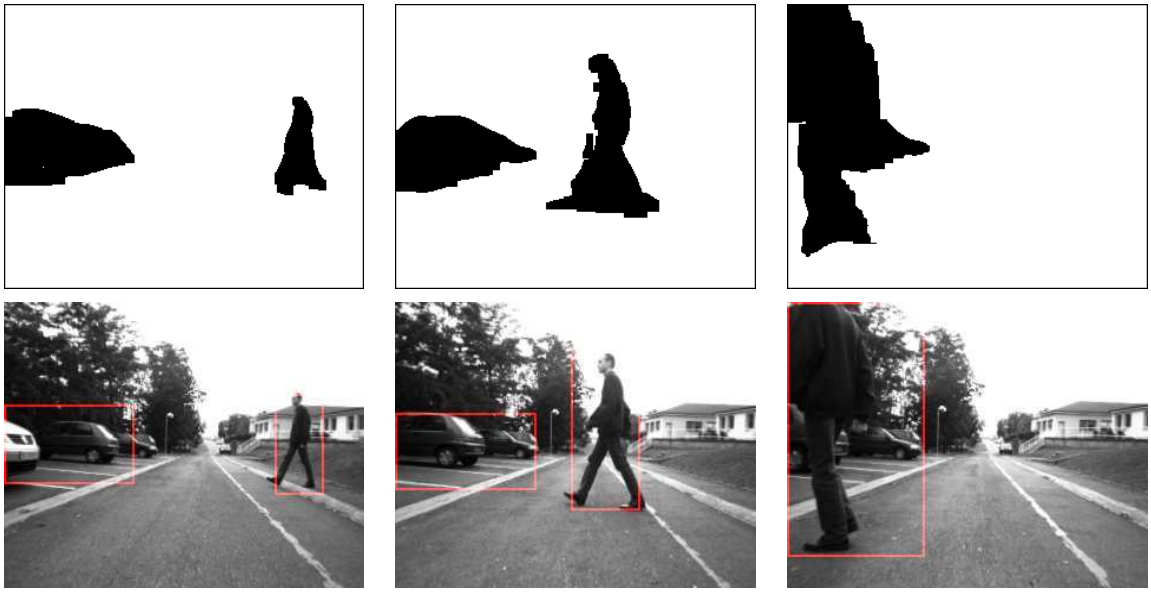


FIG. 6.10 – Résultats sur les images d'une scène routière. Première et troisième lignes : les cartes de disparité estimées par notre méthode. Deuxième et quatrième lignes : classification des pixels de la carte de disparité : les zones bleues correspondent au plan de la route et les zones rouges représentent les obstacles.



Obstacle à 3.59 m Obstacle à 3.64 m Obstacle à 3.04 m

FIG. 6.11 – Résultats de la détection d'obstacles. En haut : obstacles détectés. En bas : cadres englobants construits autour des obstacles détectés.



Obstacle à 4.61 m Obstacle à 3.88 m Obstacle à 2.14 m

FIG. 6.12 – Résultats de la détection d'obstacles. En haut : obstacles détectés. En bas : cadres englobants construits autour des obstacles détectés.

mités des cadres sont obtenues en récupérant les valeurs minimales et maximales des bords de chaque partie connexe classifiée comme un obstacle. Pour chaque cadre ainsi positionné, on calcule la profondeur des pixels appartenant à ce cadre en utilisant les paramètres de calibration des caméras et la relation (6.2). La moyenne des distances de ces points est enfin attribuée à la distance de l'obstacle encadré. Sur les figures 6.11 et 6.12, les rectangles rouges représentent les boites englobantes autour des obstacles détectés dans les images 6.9 et 6.10, respectivement. Nous indiquons sur chacune de ces figures la distance de l'obstacle le plus proche du véhicule.

6.4 Conclusion

Dans ce chapitre, nous avons introduit une méthode de détection d'obstacles dans un environnement routier à partir d'un système de stéréovision fixé sur un véhicule. Nous avons dans un premier temps estimé les cartes de disparité à l'aide de notre méthode d'estimation de la disparité. En supposant que la route est plane et horizontale et que les obstacles ont une structure verticale, nous avons modélisé l'environnement routier, dans un espace tridimensionnel de disparité, par des plans horizontaux et verticaux. Pour extraire ces plans, nous avons utilisé un critère d'orientation de surface classifiant ainsi les pixels des cartes de disparité en deux catégories : la route et les obstacles. Pour évaluer la dangerosité des obstacles détectés, nous avons calculé la distance de chaque obstacle au véhicule par un procédé de triangulation, exploitant ainsi à nouveau les cartes de disparité. Sur les séquences testées, les différents obstacles rencontrés ont été correctement détectés par notre méthode.

Chapitre 7

Conclusion et perspectives

Les recherches dans le domaine de la route intelligente, portant sur le développement de véhicules de nouvelles générations et de systèmes d'aide à la conduite, ont pour objectif l'amélioration de la sécurité et de l'efficacité du transport routier. Dans le cadre de ce travail de thèse, nous avons tenté de contribuer à cette entreprise en proposant une nouvelle approche de détection d'obstacles routiers à partir d'un système de vision stéréoscopique embarqué à bord d'un véhicule en mouvement. L'utilisation de la vision stéréoscopique en vue de déterminer la profondeur des objets et de reconstruire la structure tridimensionnelle de la scène observée, nous a amené à aborder le problème de la mise en correspondance stéréoscopique. La résolution de ce problème de grande taille, qui se ramène à l'estimation d'un champ de disparité, est rendue difficile par la présence inévitable de variations d'illumination de la scène, d'occultations et de surfaces non texturées. L'apport de cette thèse réside dans le développement d'une nouvelle méthode d'estimation de la disparité palliant ces difficultés et se prêtant bien à la détection d'obstacles routiers.

Nous résumons maintenant les principaux points que nous avons abordés dans ce mémoire ainsi que les contributions apportées.

L'application visée dans cette thèse a mené à une formulation variationnelle du problème de mise en correspondance pour l'obtention d'un champ de disparité dense et précis. La résolution de ce problème passe par la mise en œuvre de contraintes qui ont trait à la géométrie du capteur stéréoscopique, à la structure de la scène et aux objets observés. La formulation, au chapitre 3, de l'hypothèse de conservation de la luminance entre les deux projections d'un même point de l'espace a conduit à la minimisation d'une fonction objectif qui, de par la nature du problème, est non convexe. La minimisation d'un critère non convexe ne permet cependant pas d'exclure la présence de minima locaux et, par conséquent, ne garantit pas l'obtention de la solution globale. Pour contourner cette difficulté, nous avons approximé ce critère par un développement de Taylor autour d'une estimée initiale de la disparité, aboutissant ainsi à un critère convexe. La limitation, que peut constituer cette linéarisation, est compensée par le profit qu'on peut tirer des propriétés importantes liées à la convexité et elle a été, par ailleurs, atténuée par l'utilisation d'une procédure itérative de réinitialisation. Le problème d'estimation de la disparité formulé ainsi comme un problème de minimisation d'une fonctionnelle convexe est un problème inverse. Pour pallier le caractère mal posé de ce problème, l'une de nos contributions a porté sur l'introduction de contraintes convexes

traduisant les propriétés connues *a priori* sur le champ de disparité à estimer. Nous avons cherché, en particulier, à traduire la régularité de ce champ, isotrope dans les zones homogène et discontinu aux frontières des objets. Des contraintes de régularisation avec préservation des discontinuités ont alors été proposées, tant dans le domaine spatial que dans le domaine des ondelettes. Nous étant placés dans un cadre ensembliste, nous avons modélisé les contraintes introduites, comme des ensembles de niveau, faisant ainsi intervenir des bornes sur les fonctions convexes associées. Le problème de réglage de ces paramètres se trouve dès lors posé. Ce problème se présente dans la plupart des méthodes d'optimisation, notamment celles minimisant des critères pénalisés. Bien qu'il puisse constituer un frein à ces méthodes, il n'en est rien dans le cas de notre approche qui s'est montrée robuste vis-à-vis du choix des bornes sur les ensembles de contraintes. Le problème d'estimation de la disparité revient finalement à minimiser une fonction objectif convexe sur l'intersection des ensembles de contraintes convexes. La résolution numérique de ce problème nous a amené à nous intéresser de près au domaine de l'optimisation de critères convexes différentiables. Nous avons alors considéré des algorithmes parallèles itératifs, utilisant des développement récents d'analyse convexe et mettant en œuvre des projections sous-différentielles sur des ensembles convexes. Ces algorithmes offrent des méthodes de résolution puissantes et efficaces, tout en permettant de s'affranchir des limitations de grande taille des problèmes considérés.

La deuxième étape de notre travail a consisté à traiter le problème de l'estimation de la disparité en présence de variations d'illumination dans la scène observée. Nous avons pour cela adopté, au chapitre 5, un modèle d'illumination multiplicatif qui varie spatialement. Ce modèle s'avère performant et a permis de compenser les variations locales d'illumination. Il se traduit par l'ajout, dans la fonction objectif globale à minimiser, d'un facteur multiplicatif variant spatialement, que nous nous somme proposés d'estimer conjointement avec la disparité. Des contraintes convexes, portant notamment sur la douceur de ce champ d'illumination, ont été introduites en plus des contraintes sur le champ de disparité. La résolution du problème de l'estimation conjointe des deux champs de variables a été appréhendée dans le cadre d'une approche ensembliste. Les résultats obtenus sur des images synthétiques et réelles, soumises à des variations d'illumination, démontrent l'efficacité et la robustesse de cette approche.

La précision et la fiabilité de notre estimation a conduit à la mise en place, dans le chapitre 6, d'une méthode de détection d'obstacles routiers. Notre objectif a été de mettre à profit dans cette application le maximum d'informations extraites des cartes de disparités denses fournies. Nous avons, dans un premier temps, segmenté dans l'espace tridimensionnel les principaux plans qui structurent la scène routière, à l'aide d'un critère d'orientation de surface approprié. Nous sommes ainsi parvenus à identifier le plan de la route modélisé par une surface horizontale, et à détecter les obstacles routiers correspondant aux surfaces verticales. Une fois les obstacles détectés, nous avons cherché à les caractériser, notamment par leur distance au véhicule.

Des expérimentations réalisées hors-ligne ont permis de valider la méthode proposée. Cependant, le temps d'exécution relativement long de cette méthode ne permet pas son intégration immédiate dans un système embarqué. Le temps de calcul le plus important étant associé à l'étape de mise en correspondance, une première perspective des travaux présentés, pour accélérer ces méthodes, est l'implantation des algorithmes utilisés sur des architectures à processeurs parallèles. En effet, grâce à leurs structures itératives par bloc, ces algorithmes peuvent être efficacement mis en œuvre sur une configuration de *clusters* d'ordinateurs ou de cartes numériques multiprocesseurs, pour le traitement en parallèle de chaque contrainte introduite sur la solution. Par ailleurs, dans le cadre de la détection d'obstacles, ces algorithmes

devraient être évalués sur un grand nombre de séquences de scènes routières pour déterminer le taux de détection et de fausses alarmes.

De nombreuses autres perspectives sont ouvertes par notre travail tant sur l'aspect théorique que sur l'aspect applicatif.

➤ **Contraintes convexes supplémentaires**

Les algorithmes que nous utilisons ont pour avantage de pouvoir incorporer, avec une grande flexibilité, un grand nombre de contraintes convexes sur la solution. Dans nos travaux, nous avons exploité la contrainte d'intervalle qui découle de la géométrie du capteur stéréoscopique et des contraintes de régularisation avec préservation des discontinuités. Il pourrait être intéressant d'introduire d'autres contraintes permettant de réduire encore plus l'ensemble des solutions admissibles. Une contrainte qui peut être envisagée et qui s'exprime facilement comme l'appartenance à un ensemble convexe est la contrainte de limite de gradient de disparité évoquée au chapitre 2. Toutefois, les premiers essais qui ont été realisés, à partir de notre approche, semblent montrer que cette contrainte est toujours satisfaite.

➤ **Relaxation de la contrainte épipolaire rectifiée**

La méthode d'estimation de la disparité proposée se limite au cas des images rectifiées. Une direction de recherche de grand intérêt serait de généraliser cette méthode pour prendre en compte une géométrie quelconque du capteur stéréoscopique. Les travaux présentés dans [Oisel *et al.*, 2003] peuvent constituer un point de départ pour cette extension. L'utilisation, dans ces travaux, de la contrainte épipolaire calculée de façon robuste, a permis le passage d'un problème d'estimation 2D de champs de vecteurs à un problème d'estimation 1D, le long de droites épipolaires. Ce problème, formulé comme un problème de minimisation d'un critère convexe, se prête ainsi à une résolution par l'approche proposée dans ce travail.

➤ **Prise en compte de la couleur**

La plupart des capteurs de vision actuels permettent l'acquisition d'images en couleur. La couleur apporte une information supplémentaire plus riche et plus discriminante que les niveaux de gris. Il serait donc intéressant d'exploiter cette information pour améliorer les performances de notre méthode. Une première tentative menée dans cette direction a donné des résultats encourageants, ce d'autant plus que notre méthode s'adapte facilement à la couleur. Le problème de mise en correspondance, prenant en compte la couleur, revient à minimiser une fonction objectif convexe qui devient simplement la somme de trois critères de similarité suivant chaque composante couleur. Les contraintes convexes portant sur le champ de disparité restent valides, à l'exception de la contrainte de Nagel-Enkelmann qui fait intervenir l'image de gauche du couple stéréoscopique. Toutefois, une extension à la couleur du terme de régularisation associé à cette contrainte a été proposée dans [Alvarez, Sánchez, 2000]. Notons enfin que l'étude que nous avons menée s'est limitée à l'utilisation d'un seul système de couleur, à savoir le système RGB. Il pourrait donc être important d'étudier les différents systèmes de représentation de la couleur pour déterminer celui qui permet d'obtenir les meilleurs résultats.

➤ **Estimation des paramètres**

Comme nous l'avons souligné, l'algorithme d'optimisation considéré dans ce travail présente des paramètres de réglage. Bien qu'il se soit montré robuste vis-à-vis du choix de ces paramètres, notamment des bornes sur les ensembles de contraintes, une perspective serait de

développer des méthodes automatiques d'estimation de ces paramètres. Cela permettrait d'accroître l'adaptativité de l'algorithme et d'améliorer les performances de la méthode proposée.

➤ **Optimisation d'un critère non différentiable**

Dans ce travail, nous avons formulé le problème de mise en correspondance comme un problème de programmation convexe consistant à minimiser un critère convexe et différentiable sur l'intersection d'ensembles convexes. Dans certaines situations, la fonction objectif convexe peut ne pas être différentiable. On parle alors de problème d'optimisation convexe non différentiable. Certains critères non-différentiables (norme ℓ^1) sont plus robustes aux fortes erreurs que les critères différentiables mais ils sont plus difficiles à manipuler car les algorithmes d'optimisation posent souvent des problèmes de lenteur de convergence. Si des algorithmes efficaces existent dans le cas différentiable, des difficultés théoriques surgissent, en effet, lorsque l'on se place dans le cadre plus général de l'optimisation convexe non différentiable. Toutefois, les algorithmes issus de recherches récentes dans ce domaine se sont révélés très efficaces dans diverses applications en traitement du signal et des images [Combettes, Luo, 2002a, Combettes, Pesquet, 2007]. Une perspective intéressante serait donc d'explorer les possibilités qu'offrent ces algorithmes récents pour le problème de mise en correspondance.

➤ **Estimation conjointe disparité-mouvement**

Une séquence d'images stéréoscopiques montre l'évolution temporelle d'une scène tridimensionnelle à partir de deux points de vue distincts. Pour procéder à l'analyse dynamique d'une telle séquence, il est nécessaire de s'appuyer sur les liens spatio-temporels qui existent entre les images de la séquence. Ces liens se résument à un champ de déplacement spatial (disparité) et un champ de déplacement temporel (mouvement). Grâce à la contrainte stéréo-cinétique qui relie ces deux champs, une technique d'estimation conjointe peut être envisagée. Cette estimation conjointe peut profiter à de nombreuses applications. Dans le cadre de la détection d'obstacles, le mouvement apparent peut fournir de précieuses informations permettant de détecter la présence, voire d'estimer le mouvement, des obstacles mobiles. Une autre application concerne le codage spatio-temporel (3D+t) de scènes observées.

➤ **Extension au cas multi-vues et multispectral**

L'extension de notre méthode au cas multi-vues semble être une piste de recherche intéressante. En effet, l'utilisation de vues multiples de la même scène, issues d'un réseau de caméras, contribue fortement à réduire les parties occultées de la scène, sur lesquelles aucune information 3D ne peut être obtenue. Cette extension pourrait ainsi s'avérer particulièrement utile dans le cas d'images satellitaires, où les zones d'occultations sont très larges et posent de nombreuses difficultés. Par ailleurs, la plupart des dispositifs d'imagerie satellitaire sont aujourd'hui capables de fournir des images de la même scène observée dans plusieurs bandes spectrales. La mise en correspondance d'images multispectrales est un problème qui a été peu traité dans la littérature. L'adaptation de notre méthode au cas de ces images, tout en prenant en compte les interactions entre les différents canaux, est une piste de recherche qui mérite d'être explorée.

Annexe A

Un bref rappel sur les ondelettes

La théorie des ondelettes est devenue un outil mathématique très utilisé depuis les années 1980. En effet, beaucoup d'applications y font appel dans des domaines aussi variés que la mécanique des fluides, les statistiques, le traitement du signal et des images, la vision par ordinateur... Cet outil, de par sa capacité à représenter un signal à différentes résolutions, s'est révélé particulièrement puissant en traitement d'images pour résoudre de nombreux problèmes tels que la restauration d'images, la compression et le débruitage.

Nous exposons ici brièvement quelques rappels sur la théorie des ondelettes. Le lecteur intéressé pourra se référer aux ouvrages de [Mallat, 1998, Truchetet, 1998, Cohen, 1992, Meyer, 1992a] pour une description plus précise.

Transformée en ondelettes continue

La première définition de la transformée en ondelettes a été établie par Morlet [Morlet, 1981]. L'idée est d'utiliser une famille de fonctions $\{\psi_{a,b}(t),\ a \in \mathbb{R}_+^*,\ b \in \mathbb{R}\} \subset L^2(\mathbb{R})$ bien localisées en temps et en fréquence pour analyser un signal. Ces fonctions, appelées *ondelettes*, sont générées par dilatations (ou contractions) et translations d'une ondelette mère ψ :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \tag{A.1}$$

où a représente le facteur d'échelle et b le paramètre de localisation. On définit la transformée en ondelettes continue d'un signal f par la formule suivante :

$$W_f(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-b}{a}\right)dt. \tag{A.2}$$

Le signal f peut être reconstruit à partir de ses coefficients en ondelettes $W_f(a,b)$ sous réserve que l'ondelette ψ vérifie la condition d'admissibilité suivante :

$$C_\psi = \int_0^{+\infty} \frac{|\widehat{\psi}(\nu)|^2}{\nu} d\nu < +\infty, \tag{A.3}$$

où $\widehat{\psi}$ désigne la transformée de Fourier de ψ. La condition (A.3) implique que l'ondelette mère ψ soit de moyenne nulle ($\int_{-\infty}^{+\infty} \psi(t)dt = 0$).

La formule de reconstruction du signal f s'écrit donc

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_f(a,b) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) db \frac{da}{a^2} \tag{A.4}$$

Cette représentation faisant intervenir des ondelettes correspondant à toutes les localisations temporelles et à tous les facteurs d'échelles est difficle à mettre en œuvre du point de vue pratique. En effet, l'inconvénient majeur de la transformée en ondelettes continue est la redondance de l'information contenue dans cette transformée. On préfère souvent utiliser la transformée en ondelettes discrète qui permet d'implémenter des algorithmes rapides de calcul des coefficients en ondelettes.

Transformée en ondelettes discrète

L'idée de la discrétisation de la transformée en ondelettes est de pouvoir décomposer le signal sur un ensemble dénombrable d'ondelettes sans perdre les propriétés de la transformée en ondelettes continue. La discrétisation la plus couramment utilisée consiste à faire varier le facteur d'échelle de façon dyadique, en choisissant $a = 2^j,\ j \in \mathbb{Z}$ et en prenant $b = k2^j,\ k \in \mathbb{Z}$. Dans ce cas, la famille d'ondelettes générée par l'ondelette mère ψ est donnée par :

$$\{\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k), \quad (j,k) \in \mathbb{Z}^2\}, \tag{A.5}$$

et est construite de façon à constituer une base de $L^2(\mathbb{R})$.

La transformée en ondelettes discrète se formalise comme suit :

$$W_{j,k}(f) = \int_{-\infty}^{+\infty} f(t)\psi_{j,k}^*(t)dt. \tag{A.6}$$

Si la famille d'ondelettes $(\psi_{j,k})_{(j,k)\in\mathbb{Z}^2}$ forme une base orthonormale de $L^2(\mathbb{R})$, on a la formule de reconstruction suivante

$$f(t) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} W_{j,k}(f)\psi_{j,k}(t). \tag{A.7}$$

L'analyse multirésolution de $L^2(\mathbb{R})$

L'analyse multirésolution, introduite par Mallat en 1989 [Mallat, 1989], est un formalisme mathématique qui permet de construire aisément des bases orthonormées d'ondelettes. Une analyse multirésolution est définie comme une famille de sous-espaces vectoriels fermés emboîtés $\{V_j\}_{j\in\mathbb{Z}}$ de $L^2(\mathbb{R})$ vérifiant les propriétés suivantes :

1.
$$\forall j \in \mathbb{Z}, \quad V_{j+1} \subset V_j, \tag{A.8}$$

2.
$$\forall j \in \mathbb{Z}, \ \ \forall t \in \mathbb{R}, \quad f(t) \in V_{j+1} \Leftrightarrow f(2t) \in V_j, \tag{A.9}$$

3.
$$\lim_{j \to +\infty} V_j = \bigcap_{j \in \mathbb{Z}} V_j = 0, \tag{A.10}$$

et
$$\lim_{j \to -\infty} V_j = \bigcup_{j \in \mathbb{Z}} V_j \text{ est dense dans } L^2(\mathbb{R}). \tag{A.11}$$

4. Il existe une fonction $\phi(t) \in V_0$, appelée *fonction d'échelle*, telle que la famille $\{\phi(t-k),\ k \in \mathbb{Z}\}$ forme une base orthonormale de V_0.

On vérifie à partir des conditions 2 et 4 précédentes que la famille $\{2^{-j/2}\phi(t/2^j - k),\ k \in \mathbb{Z}\}$ constitue une base orthonormale de l'espace V_j.
Les espaces V_j constituent des espaces d'approximation. L'analyse multirésolution d'un signal f de $L^2(\mathbb{R})$, qui consiste à calculer les projections orthogonales de ce signal sur les espaces V_j, permet donc d'obtenir des *approximations* de f de plus en plus grossières au fur et à mesure que le niveau de résolution j croît. La différence entre deux approximations consécutives représente l'information de *détails*. Cette information est contenue dans le sous-espace W_j, complémentaire orthogonal de V_j dans V_{j-1}, tel que :

$$V_{j-1} = V_j \oplus W_j \tag{A.12}$$

On montre alors qu'il existe une ondelette mère $\psi(t)$ telle que la famille $\{2^{-j/2}\psi(t/2^j - k),\ k \in \mathbb{Z}\}$ soit une base orthonormale de W_j. Il est clair, ainsi, que la famille $\{2^{-j/2}\psi(t/2^j - k),\ (j,k) \in \mathbb{Z}^2\}$ est une base orthonormale de $L^2(\mathbb{R})$.

Si l'on note A_j (resp. D_j) l'opérateur de projection orthogonale sur l'espace V_j (resp. W_j), on a

$$A_j f(t) = \sum_{k=-\infty}^{+\infty} a_{j,k} \frac{1}{2^{j/2}} \phi(\frac{t}{2^j} - k),$$
$$D_j f(t) = \sum_{k=-\infty}^{+\infty} c_{j,k} \frac{1}{2^{j/2}} \psi(\frac{t}{2^j} - k),$$

où les coefficient d'approximations $(a_{j,k})_{k \in \mathbb{Z}}$ et de détails $(c_{j,k})_{k \in \mathbb{Z}}$ sont définis au niveau de résolution j par :

$$\forall k \in \mathbb{Z}, \quad a_{j,k} = \langle f(t) \ \mid \ \frac{1}{2^{j/2}} \phi(\frac{t}{2^j} - k) \rangle, \tag{A.13}$$
$$c_{j,k} = \langle f(t) \ \mid \ \frac{1}{2^{j/2}} \psi(\frac{t}{2^j} - k) \rangle. \tag{A.14}$$

En conclusion, l'analyse multirésolution permet à partir d'une fonction d'échelle $\phi(t)$ et d'une ondelette mère $\psi(t)$, de décomposer un signal en coefficients d'approximations et de

détail. Considérons une fonction f de $L^2(\mathbb{R})$, sa décomposition en ondelettes s'écrit :

$$f = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} c_{j,k}\psi_{j,k}. \tag{A.15}$$

Si l'on limite la sommation aux niveaux de résolutions $j \leq l$, où l est le niveau de résolution le plus grossier, f est alors décomposée ainsi :

$$f = \sum_{k\in\mathbb{Z}} a_{l,k}\phi_{l,k} + \sum_{j\leq l} \sum_{k\in\mathbb{Z}} c_{j,k}\psi_{j,k}. \tag{A.16}$$

Extension au cas 2D

Le modèle d'ondelettes défini précédemment peut se généraliser à n'importe quelle dimension $n > 0$. Nous étudierons ici le cas 2D pour des applications en traitement d'images.

Dans le cas 2D, on peut définir de façon semblable au cas 1D, une analyse multirésolution comme une séquence de sous-espaces emboîtés $\{V_j^2\}_{j\in\mathbb{Z}}$ de $L^2(\mathbb{R}^2)$ où chaque espace V_j^2 est défini par le produit tensoriel de deux espaces identiques V_j de $L^2(\mathbb{R})$. On peut ainsi construire la fonction d'échelle associée à $\{V_j^2\}_{j\in\mathbb{Z}}$ par

$$\phi(x, y) = \phi(x)\phi(y). \tag{A.17}$$

où $\phi(t)$ est la fonction d'échelle 1D de $\{V_j\}_{j\in\mathbb{Z}}$. Les ondelettes 2D sont définies comme suit

$$\begin{aligned} \psi^1(x, y) &= \phi(x)\psi(y), \\ \psi^2(x, y) &= \psi(x)\phi(y), \\ \psi^3(x, y) &= \psi(x)\psi(y), \end{aligned} \tag{A.18}$$

ψ étant l'ondelette mère associée à $\{V_j\}_{j\in\mathbb{Z}}$.

Il es usuel de représenter la transformée en ondelettes 2D d'une image $f \in L^2(\mathbb{R}^2)$ en partageant le plan image en imagettes de façon récursive comme le montre la figure A.1. Sur l'image A.1, On peut interpréter $D_j^1 f$ comme les détails horizontaux de f à la résolution j, $D_j^2 f$ comme les détails verticaux, et $D_j^3 f$ comme les détails diagonaux. $A_2(f)$ représente l'approximation de l'image f au niveau de résolution le plus grossier $l = 2$.

En notant par $(c_{j,k,1})_{k\in\mathbb{Z}^2}$, $(c_{j,k,2})_{k\in\mathbb{Z}^2}$ et $(c_{j,k,3})_{k\in\mathbb{Z}^2}$ les coefficients de détails horizontaux, verticaux et diagonaux respectivement, à un niveau de résolution j donné, une fonction f de $L^2(\mathbb{R}^2)$ se décompose comme suit :

$$f = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}^2, o\in\{1,2,3\}} c_{j,k,o}\psi_{j,k,o}. \tag{A.19}$$

La base d'ondelettes de Haar

Si on considère une analyse multirésolution telle que les sous-espaces $\{V_j\}_{j\in\mathbb{Z}}$ soient constitués des fonctions constantes par morceaux, alors nous pouvons prendre pour $\phi(t)$ la fonction définie par :

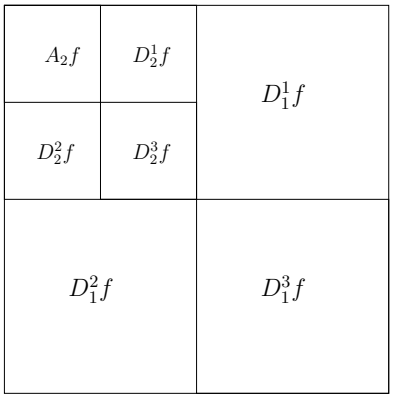


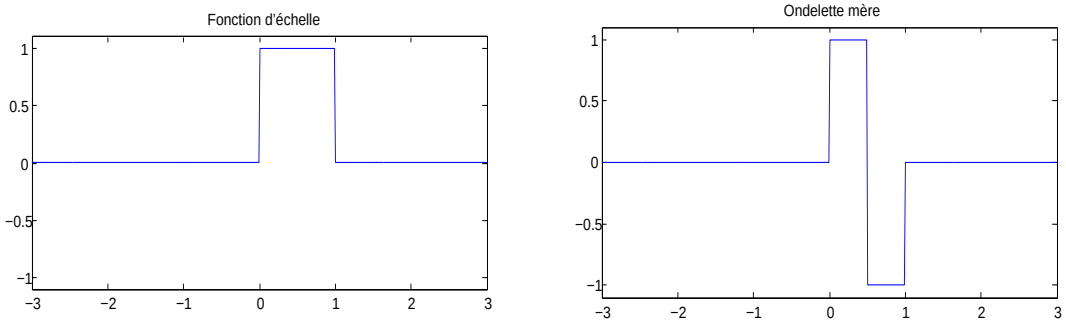
FIG. A.1 – Disposition des images d'approximations et de détails issues d'une décomposition en ondelettes 2D pour $l = 2$ niveaux de résolutions.

$$\phi(t) = \begin{cases} 1 & \text{si } t \in [0,1[, \\ 0 & \text{ailleurs.} \end{cases} \tag{A.20}$$

L'ondelette associée à cette fonction d'échelle est l'ondelette de Haar définie par :

$$\psi(t) = \begin{cases} 1 & \text{si } t \in [0,1/2[\\ -1 & \text{si } t \in [1/2,1[\\ 0 & \text{ailleurs.} \end{cases} \tag{A.21}$$

Les graphes de la fonction d'échelle ϕ et de l'ondelette mère ψ sont donnés par les figures A.2(a) et (b).



(a) (b)

FIG. A.2 – (a) Allure des fonctions de Haar (a) $\phi(t)$ (b) $\psi(t)$.

L'ondelette de Haar est très simple et est donc facile à mettre en œuvre. De plus, son support est compact, elle est donc bien localisée en espace. Cependant, elle n'a qu'un seul moment nul et est discontinue ce qui implique une piètre localisation en fréquence.

Table des figures

2.1 Le modèle de caméra sténopé. 22
2.2 La projection perspective. 23
2.3 Le modèle géométrique du capteur stéréoscopique. 27
2.4 La géométrie épipolaire. 28
2.5 La rectification épipolaire. 33
2.6 Couple d'images Cones. (a) Image gauche. (b) Image droite. (c) Carte de disparité. 34
2.7 Relation entre la disparité et la profondeur. 35
2.8 Le problème d'occultation : le point m est visible dans l'image de gauche mais pas dans l'image de droite. 37
2.9 Cas de non respect de la contrainte d'unicité : tous les pixels qui se trouvent sur le segment de droite $[M_1, M_2]$ se projettent sur le pixel m' dans l'image de droite. 40
2.10 La contrainte d'ordre. 40
2.11 (a) Zone interdite associée au point M. (b) Cas d'un objet transparent fortement incliné. 41
2.12 Mise en correspondance par corrélation. 44
2.13 Programmation dynamique. 47

3.1 Le principe d'une approche ensembliste. 58
3.2 Projection sur l'hyperplan de séparation H_u délimitant Q_u. 63

4.1 Paire d'images *Corridor* et carte de disparité associée. 90
4.2 Exemples de paires stéréoscopiques proposées par Scharstein et Szeliski. De haut en bas : *Tsukuba*, *Sawtooth*, *Venus*, *Map*, *Teddy*. (a) Image de gauche (b) Image de droite (c) Carte de disparité théorique. 91
4.3 Zones d'évaluation considérées dans le protocole de Scharstein et Szeliski. Les pixels noirs correspondent aux pixels occultés et les pixels blancs correspondent aux (a) pixels de l'image (b) pixels proches d'une discontinuité (c) pixels dans une zone non texturée. 92
4.4 Cartes de disparité calculées par projection sur l'ensemble des solutions admissibles $S = S_1 \cap S_2 \cap S_3$. 93
4.5 Cartes de disparité calculées pour la paire d'images *Venus* par projection sur les ensembles (a) $S_1 \cap S_2 \cap S_3$ (b) $S_1 \cap S_2$ (c) $S_1 \cap S_3$ (d) $S_2 \cap S_3$. 95
4.6 Racine carrée de l'erreur quadratique moyenne d'estimation en fonction de $\tau/\widehat{\tau}$. 96
4.7 Racine carrée de l'erreur quadratique moyenne d'estimation en fonction de $\delta/\widehat{\delta}$. 96

4.8 Résultats intermédiaires pour l'estimation de la carte de disparité de la paire d'images *Teddy*. 98
4.9 Résultas obtenus avec les paires d'images *Tsukuba*, *Venus* et *Teddy*. (a) vraie disparité (b) carte de disparité estimée (c) carte de pixels erronés : plus le pixel est foncé et plus l'appariement est mauvais. Ces résultats sont obtenus en considérant trois cycles de réinitialisations. 100
4.10 Résultats pour la paire stéréoscopique *Corridor*. (a) Image de gauche. (b) Disparité théorique. (c) Disparité initiale. Disparité estimée par notre méthode en considérant une régularisation par l'ensemble (d) S_2 (e) S_4 (f) S_5. 103
4.11 Résultats pour la paire *Corridor* dégradée par un bruit gaussien de variance 10. (a) Image de gauche. (b) Image de droite. (c) Disparité initiale. Disparité estimée par notre méthode en considérant une régularisation par l'ensemble (d) S_2 (e) S_4 (f) S_5. 105
4.12 Résultats pour la paire *Corridor* dégradée par un bruit gaussien de variance 100. (a) Image de gauche. (b) Image de droite. (c) Disparité initiale. Disparité estimée par notre méthode en considérant une régularisation par l'ensemble (d) S_2 (e) S_4 (f) S_5. 106
4.13 Erreur quadratique moyenne normalisée en fonction du temps CPU. 107
4.14 Séquence d'images *flower garden*. En haut : (a) image 1 (b) image 4 (c) image 10. Au milieu : Résultats obtenus sur les images 1 et 4 par (d) corrélation (e) Gc+Occ (f) notre méthode. En bas : Résultats obtenus sur les images 1 et 10 par (g) corrélation (h) Gc+Occ (i) notre méthode. 108
4.15 Résultats pour les images *parking meter*. (a) Image de gauche (b) méthode par corrélation (c) méthode de coupures de graphes (d) notre méthode. 109

5.1 Le concept de la réflectance bidirectionnelle. 116
5.2 Géométrie de la formation des images. 117
5.3 Modèle de réflexion lambertien : la diffusion est effectuée uniformément dans toutes les directions. 118
5.4 Paire d'images *Corridor* modifiée pour avoir un changement d'illumination artficiel. L'image de droite est multipliée par un profil gaussien dont la formule est donnée en (5.44). 129
5.5 Résultats pour la paire stéréoscopique *Corridor* ayant subi des variations d'illuminations artificielles. (a) Image de gauche (b) Vraie disparité (c) NCC (d) GC+HT (e) SGM (f) Notre approche. 130
5.6 Cartes de changements d'illumination pour la paire d'images *Corridor* modifiée : (a) la carte synthétique (b) la carte estimée avec une régularisation de Tikhonov (c) la carte estimée avec une régularisation basée sur le laplacien. . . 131
5.7 Résultats intermédiaires pour l'estimation des champs de disparité de la paire d'images *Corridor* modifiée. 133
5.8 Résultats intermédiaires pour l'estimation des champs d'illumination de la paire d'images *Corridor* modifiée. 133
5.9 Paires d'images *Dolls* (en haut) et *Books* (en bas) avec vérité terrain. (a) Image de gauche. (b) Image de droite. (c) Vérité terrain. 134
5.10 Images de disparité pour la paire stéréoscopique *Dolls*. 135
5.11 Images de disparité pour la paire stéréoscopique *Books*. 135
5.12 Paire d'images *Shrub*. (a) Image de gauche. (b) Image de droite. 136

5.13 Résultats pour la paire stéréoscopique réelle *Shrub*. (a) Image de gauche. (b) SSD+MA. (c) NCC. (d) GC+HT (e) SGM. (f) Notre approche. 137

6.1 Le Cycab. 143
6.2 Système de vision Bumblebee. 144
6.3 Configuration des caméras du système de vision stéréoscopique. 144
6.4 Résultats de l'estimation de la disparité sur des paires d'images de scènes routières. En haut : les images de gauche. En bas : les cartes de disparité estimées. 145
6.5 Système de coordonnées du véhicule. 146
6.6 Résultats sur la paire de synthèse *Corridor*. En haut : les images stéréoscopiques et la carte de disparité associée. Au milieu : les valeurs absolues des composantes des vecteurs normaux à la surface suivant les axes (d) X, (e) Y et (f) Z. Résultats du seuillage sur les composantes (g) a_n, (h) b_n et (i) c_n. 149
6.7 Résultat du seuillage sur la composante c_n de la figure 6.6(f) mettant en évidence les plans verticaux et obliques. 150
6.8 Classification des pixels de la carte de disparité : les pixels bleus appartiennent à des surfaces horizontales et les pixels rouges appartiennent à des objets verticaux. 151
6.9 Résultats sur les images d'une scène routière. Première et troisième lignes : les cartes de disparité estimées par notre méthode. Deuxième et quatrième lignes : classification des pixels de la carte de disparité : les zones bleues correspondent au plan de la route et les zones rouges représentent les obstacles. 152
6.10 Résultats sur les images d'une scène routière. Première et troisième lignes : les cartes de disparité estimées par notre méthode. Deuxième et quatrième lignes : classification des pixels de la carte de disparité : les zones bleues correspondent au plan de la route et les zones rouges représentent les obstacles. 153
6.11 Résultats de la détection d'obstacles. En haut : obstacles détectés. En bas : cadres englobants construits autour des obstacles détectés. 154
6.12 Résultats de la détection d'obstacles. En haut : obstacles détectés. En bas : cadres englobants construits autour des obstacles détectés. 154

A.1 Disposition des images d'approximations et de détails issues d'une décomposition en ondelettes 2D pour $l = 2$ niveaux de résolutions. 165
A.2 (a) Allure des fonctions de Haar (a) $\phi(t)$ (b) $\psi(t)$. 165

Liste des tableaux

2.1 Quelques termes de régularisation utilisés en traitement d'images. 49

4.1 Vraies valeurs des bornes sur les ensembles de contraintes S_1, S_2 et S_3 pour les paires d'images de la figure 4.2. 92
4.2 Impact des contraintes sur les résultats obtenus par notre méthode. Le meilleur résultat est obtenu lorsque toutes les contraintes sont prises en compte. 93
4.3 Résultats intermédiaires pour l'estimation de la paire d'images *Teddy*. 98
4.4 Résultats comparatifs pour les paires d'images *Tsukuba*, *Venus* et *Teddy* en utilisant le critère d'erreur absolue moyenne. 98
4.5 Erreurs absolues moyennes obtenues sur une portion 256×256 de la paire d'image *Sawtooth*. 102
4.6 Résultats comparatifs sur la paire d'images *Corridor* utilisant l'erreur absolue moyenne. 104

5.1 Résultats comparatifs sur la paire d'images *Corridor* modifiée : (a) sans ajout de bruit (b) avec bruit additif gaussien. 132
5.2 Résultats intermédiaires pour la paire d'images *Corridor* modifiée. 132
5.3 Résultats comparatifs pour les paires stéréoscopiques *Dolls* et *Books*. 136

6.1 Paramètres de l'algorithme d'estimation de la disparité pour les images considérées dans ce chapitre. 145

Bibliographie

[Agrawal, Davis, 2004] Agrawal, M., Davis, L. (2004). Window-based, discontinuity preserving stereo. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 1, pages 66–73, Washington, États-Unis.

[Alvarez *et al.*, 2002] Alvarez, L., Deriche, R., Papadopoulos, T., Sánchez, J. (2002). Symmetrical dense optical flow estimation with occlusions detection. *In European Conference on Computer Vision*, vol. 1, pages 721–735, Copenhague, Danemark.

[Alvarez *et al.*, 2000a] Alvarez, L., Deriche, R., Sánchez, J., Weickert, J. (2000a). Dense disparity map estimation respecting image discontinuities : A PDE and scale-space based approach. Rapport de recherche RR-3874, Institut National de Recherche en Informatique et en Automatique, INRIA.

[Alvarez, Sánchez, 2000] Alvarez, L., Sánchez, J. (2000). 3-D geometry reconstruction using a color image stereo pair and partial differential equations. *Cuadernos del Instituto Universitario de Ciencias y Tecnologas Ciberneticas*, 6:1–26.

[Alvarez *et al.*, 2000b] Alvarez, L., Weickert, J., Sánchez, J. (2000b). Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39(1):41–56.

[Ambrosio *et al.*, 2000] Ambrosio, L., Fusco, N., Pallara, D. (2000). Functions of bounded variations and free discontinuity problems. *Oxford Mathematical Monographs*.

[Andersen, Kak, 1984] Andersen, A. H., Kak, A. C. (1984). Simultaneous algebraic reconstruction technique SART : a superior implementation of the art algorithm. *Ultrasonic imaging*, 6:81–94.

[Andersen *et al.*, 2005] Andersen, H., Kirk, K., Dideriksen, T., Madsen, C., Holte, M., Bak, T. (2005). Obstacle detection by stereo vision, introducing the PQ method. *In IEEE Conference Proceedings of the 2nd International Conference on Informatics in Control, Automation and Robotics*, vol. 2, pages 250–257, Barcelone, Espagne.

[Aschwanden, Guggenbül, 1992] Aschwanden, P., Guggenbül, W. (1992). Experimental results from a comparative study on correlation type registration algorithms. *In* Förstner, W., Ruwiedel, S., éditeurs : *Robust computer vision : Quality of Vision Algorithms*, pages 268–282. Wichmann, Karlsruhe, Allemagne.

[Aubert *et al.*, 1999] Aubert, G., Deriche, R., Kornprobst, P. (1999). Computing optical flow via variational techniques. *SIAM Journal on Numerical Analysis*, pages 156–182.

[Aubert, Vese, 1997] Aubert, G., Vese, L. (1997). A variational method in image recovery. *SIAM Journal on Numerical Analysis*, 34(5):1948–1979.

[Badal *et al.*, 1994] Badal, S., Ravela, S., Draper, B., Hanson, A. (1994). A practical obstacle detection and avoidance system. *In IEEE Conference Proceedings of the 2nd Workshop on Applications of Computer Vision*, pages 97–104, Sarasota, FL, USA.

[Bect *et al.*, 2004] Bect, J., Blanc-Féraud, L., G., Aubert, Chambolle, A. (2004). A l^1-unified variational framework for image restoration. *In proceedings of European Conference on Computer Vision, ECCV*, vol. LNCS 3024, pages 1–13, Prague, Czech Republic.

[Bertero *et al.*, 1988] Bertero, M., Poggio, T. A., Torre, V. (1988). Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889.

[Bertozzi *et al.*, 2005] Bertozzi, M., Binelli, E., Broggi, A., Rose, M. D. (2005). Stereo vision-based approaches for pedestrian detection. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 3, page 16, San Diego, USA.

[Bertozzi *et al.*, 2002] Bertozzi, M., Broggi, A., Fascioli, A., Lombardi, P. (2002). Vision-based pedestrian detection : will ants help ? *In IEEE Conference Proceedings of Intelligent Vehicle Symposium*, vol. 1, pages 17–21.

[Bhat, Nayar, 1998] Bhat, D., Nayar, S. (1998). Ordinal measures for image correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):415–423.

[Birchfield, Tomasi, 1998] Birchfield, S., Tomasi, C. (January 1998). Depth discontinuities by pixel-to-pixel stereo. *In IEEE Conference Proceedings of the Sixth International Conference on Computer Vision, ICCV*, vol. 2, pages 1073–1080, Bombay, India.

[Black *et al.*, 2000] Black, M., Fleet, D., Yacoob, Y. (2000). Robustly estimating changes in image appearance. *Computer Vision and Image Understanding*, 78(1):8–31.

[Bleyer, Gelautz, 2004] Bleyer, M., Gelautz, M. (2004). A layered stereo algorithm using image segmentation and global visibility constraints. *In IEEE Conference Proceedings of International Conference on Image Processing, ICIP*, vol. 5, pages 2997–3000, Singapour.

[Bleyer, Gelautz, 2005] Bleyer, M., Gelautz, M. (2005). Graph-based surface reconstruction from stereo pairs using image segmentation. *In Videometrics VIII*, vol. SPIE–5665, pages 288–299, San José, États-Unis.

[Boukir *et al.*, 1992] Boukir, S., Bouthemy, P., Chaumette, F., Juvin, D. (1992). Mise en correspondance de segments dans une séquence d'images par une approche locale. Rapport de recherche 1792, Institut National de Recherche en Informatique et en Automatique, INRIA.

[Boyle, Dykstra, 1986] Boyle, J., Dykstra, R. (1986). A method for finding projections onto the intersection of convex sets in Hilbert spaces. *Lecture Notes Statist.*, 37:28–47.

[Broggi *et al.*, 2000] Broggi, A., Bertozzi, M., Fascioli, A., Sechi, M. (2000). Shape-based pedestrian detection. *In IEEE Conference Proceedings of Intelligent Vehicles Symposium*, pages 215 – 220, Dearborn, MI.

[Burt, Julesz, 1980] Burt, P., Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science*, 208(9):615–617.

[Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 8(6):679–698.

[Capricelli, Combettes, 2005] Capricelli, T., Combettes, P. (2005). Parallel block-iterative reconstruction algorithms for binary tomography. *Electronic Notes in Discrete Mathematics*, 20:263–280.

[Chambolle *et al.*, 1998] Chambolle, A., DeVore, R. A., Lee, N., Lucier, B. J. (1998). Nonlinear wavelet image processing : variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–335.

[Charbonnier *et al.*, 1997] Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M. (1997). Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311.

[Choi, Baraniuk, 1999a] Choi, H., Baraniuk, R. (1999a). Interpolation and denoising of nonuniformly sampled data using wavelet-domain processing. *In IEEE Conference Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, vol. 3, pages 1645–1648, Phoenix, USA.

[Choi, Baraniuk, 1999b] Choi, H., Baraniuk, R. (1999b). Multiple basis wavelet denoising using besov projections. *In IEEE Conference Proceedings of International Conference on Image Processing, ICIP*, vol. 1, pages 595–599, Kobe, Japan.

[Cochran, Médioni, 1992] Cochran, S., Médioni, G. (1992). 3-D surface description from binocular stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 14(10):981–994.

[Cohen, 1992] Cohen, A. (1992). *Ondelettes et traitement numérique du signal*. Masson.

[Coifman, D.Donoho, 1995] Coifman, R., D.Donoho (1995). Translation invariant denoising. *In A. Antoniadis and G. Oppenheim, editors : Wavelets and Statistics*, 103:125–150.

[Combettes, 1996] Combettes, P. (1996). The convex feasibility problem in image recovery. *In* P. Hawkes, E., éditeur : *Advances in Imaging and Electron Physics*, vol. 95, pages 155–270. New York : Academic Press.

[Combettes, 1997] Combettes, P. (1997). Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections. *IEEE Transactions on Image Processing*, 6(4):493–506.

[Combettes, 2000] Combettes, P. (2000). Strong convergence of block-iterative outer approximation methods for convex optimization. *SIAM J. Control Optim.*, 38:538–565.

[Combettes, 2003] Combettes, P. (2003). A block iterative surrogate constraint splitting method for quadratic signal recovery. *IEEE Transactions on Signal Processing*, 51(7):1771–1782.

[Combettes, Luo, 2002a] Combettes, P., Luo, J. (2002a). An adaptive level set method for nondifferentiable constrained image recovery. *IEEE Transactions on Image Processing*, 11(11):1295–1304.

[Combettes, Pesquet, 1993] Combettes, P., Pesquet, J. (1993). The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208.

[Combettes, Pesquet, 2004a] Combettes, P., Pesquet, J. (2004a). Image restoration subject to a total variation constraint. *IEEE Transactions on Image Processing*, 13(9):1213–1222.

[Combettes, Pesquet, 2004b] Combettes, P., Pesquet, J. (2004b). Wavelet-constrained image restoration. *International Journal of Wavelets, Multiresolution, and Information Processing*, 2(4):371–389.

[Combettes, Pesquet, 2007] Combettes, P., Pesquet, J. (2007). A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *International Journal of Selected Topics in Signal Processing*, to appear.

[Combettes, Luo, 2002b] Combettes, P. L., Luo, J. (2002b). An adaptive level set method for nondifferentiable constrained image recovery. *IEEE Transactions on Image Processing*, 11(11):1295–1304.

[Combettes, Pesquet, 2003] Combettes, P. L., Pesquet, J. C. (2003). Incorporating total variation information in image recovery. *In IEEE Conference Proceedings of International Conference on Image Processing, ICIP*, vol. 3, pages 373–376.

[Cox, 1994] Cox, I. (1994). A maximum likelihood N-camera stereo algorithm. *In IEEE Conference Proceedings of International Conference on Computer Vision and Pattern Recognition, CVPR*, pages 733–739, Seattle, WA.

[Cox *et al.*, 1996] Cox, I., Hingorani, S., Rao, S. (1996). A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3):542–567.

[Cox *et al.*, 1995] Cox, I., Roy, S., Hingorani, S. (1995). Dynamic histogram warping of image pairs for constant image brightness. *In IEEE Conference Proceedings of International Conference on Image Processing, ICIP*, vol. 2, pages 366–369, Washignton, USA.

[Crouzil *et al.*, 1996] Crouzil, A., Massip-Pailhes, L., Castan, S. (1996). A new correlation criterion based on gradient fields similarity. *In IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR*, vol. 1, pages 632–636, Vienne, Autriche.

[Demonceaux *et al.*, 2004] Demonceaux, C., A.Potelle, Kachi-Akkouche, D. (2004). Obstacle detection in road scene based on motion analysis. *IEEE Transactions on Vehicular Technology*, 53(6):1649–1656.

[Deng, Lin, 2006] Deng, Y., Lin, X. (2006). A fast line segment based dense stereo algorithm using tree dynamic programming. *In proceedings of European Conference on Computer Vision, ECCV*, pages 201–212, Graz, Austria.

[Deriche, Faugeras, 1995] Deriche, R., Faugeras, O. (1995). Les EDP en traitement d'images et vision par ordinateur. Rapport de recherche RR-2697, Institut National de Recherche en Informatique et en Automatique, INRIA.

[Deriche *et al.*, 1996] Deriche, R., Kornprobst, P., Aubert, G. (1996). Optical-flow estimation while preserving its discontinuities : A variational approach. *In ACCV '95 : Invited Session Papers from the Second Asian Conference on Computer Vision*, pages 71–80, London, UK. Springer-Verlag.

[DeVore, Lucier, 1992] DeVore, R., Lucier, B. (1992). Fast wavelet techniques for near-optimal image processing. *In IEEE Military Communications Conference Record*, vol. 3, pages 1129–1135, San Diego, USA.

[Dey *et al.*, 2004] Dey, N., Blanc-Féraud, L. B., Zimmer, C., Roux, P., Kam, Z., Olivo-Marin, J. C., Zerubia, J. (2004). 3D microscopy deconvolution using Richardson-Lucy algorithm with total variation regularization. Rapport de recherche RR-5272, Institut National de Recherche en Informatique et en Automatique, INRIA.

[Dobson, Vogel, 1997] Dobson, D. C., Vogel, C. R. (1997). Convergence of an iterative method for total variation denoising. *SIAM Journal on Numerical Analysis*, 34(5):1779–1791.

[Durand, Froment, 2003] Durand, S., Froment, J. (2003). Reconstruction of wavelet coefficients using total variation minimization. *SIAM Journal on Scientific Computing*, 24(5):1754–1767.

[Egnal, 2000] Egnal, G. (2000). Mutual information as a stereo correspondence measure. Technical report ms-cis-00-20, computer and information science, University of Pennsylvania, Philadelphia, USA.

[Egnal, Wildes, 2002] Egnal, G., Wildes, R. P. (2002). Detecting binocular half-occlusions : empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1127–1133.

[Ekeland, Temam, 1974] Ekeland, I., Temam, R. (1974). *Analyse convexe et problèmes variationnels*. Dunod, Paris.

[Enkelmann, 1991] Enkelmann, W. (1991). Obstacle detection by evaluation of optical flow fields from image sequences. *Image Vision Computing*, 9(3):160–168.

[Evans, Gariepy, 1992] Evans, L. C., Gariepy, R. F. (1992). Measure theory and fine properties of functions. *Studies in Advanced Mathematics*, 19.

[Faugeras, 1993] Faugeras, O. (1993). *Three-Dimensional Computer Vision : a Geometric Viewpoint*. MIT Press.

[Feraud, 2000] Feraud, L. (2000). *Sur quelques problèmes inverses en traitement d'image*. Habilitation à diriger des recherches, Université de Nice-Sophia Antipolis, France.

[Fookes *et al.*, 2002] Fookes, C., Bennamoun, M., Lamanna, A. (2002). Improved stereo image matching using mutual information and hierarchical prior probabilities. *In 16th International Conference on Pattern Recognition*, vol. 2, pages 937–940, Quebec, Canada.

[Fua, 1993] Fua, P. (1993). A parallel stereo algorithm that produces dense depth maps and preserves image features. *International Journal of Machine Vision and Applications*, 6(1):35–49.

[Fusiello *et al.*, 2000a] Fusiello, A., Roberto, V., Trucco, E. (2000a). Symmetric stereo with multiple windowing. *International Journal of Pattern Recognition and Artificial Intelligence,*, 14(8):1053–1066.

[Fusiello *et al.*, 2000b] Fusiello, A., Trucco, E., Verri, A. (2000b). A compact algorithm for rectification of stereo pairs. *International Journal of Machine Vision and Applications, MVA*, 12(1):16–22.

[G. E. Mailloux, Bertrand, 1989] G. E. Mailloux, F. Langlois, P. Y. S., Bertrand, M. (1989). Restoration of the velocity field of the heart from two-dimensional echocardiograms. *IEEE Transactions on Medical Imaging*, 8(2):143–153.

[Geman, Reynolds, 1992] Geman, D., Reynolds, G. (1992). Constrained restoration and the recovery of discontinuities. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 14(3):367–383.

[Gennert, 1987] Gennert, M. A. (1987). *A computational Framework for Understanding Problems in Stereo Vision*. Thèse de doctorat, MIT.

[Gennert, 1988] Gennert, M. A. (1988). Brightness-based stereo matching. *In IEEE Conference Proceedings of International Conference on Computer Vision, ICCV*, pages 139–143.

[Giachetti *et al.*, 1998] Giachetti, A., Campani, M., Torre, V. (1998). The use of optical flow for road navigation. *IEEE transactions on Robotics and Automation*, 14(1):34–48.

[Giusti, 1984] Giusti, E. (1984). *Minimal surfaces and functions of bounded variation*. Boston : Birkhäuser.

[Gonzalez, Woods, 1993] Gonzalez, R. C., Woods, R. E. (1993). *Digital Image Processing.* Addison-Wesley.

[Hadamard, 1902] Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. pages 49–52.

[Hampson, Pesquet, 2000] Hampson, F., Pesquet, J. (2000). Motion estimation in the presence of illumination variations. *Signal Processing Image Communication*, 16(4):373–381.

[Harris, Stephens, 1988] Harris, C., Stephens, M. (1988). A combined corner and edge detector. *In Alvey Vision Conference*, pages 147–151, Manchester, Royaume-Uni.

[Hartley, Gupta, 1993] Hartley, R., Gupta, R. (1993). Computing matched-epipolar projections. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, pages 549–555, New York, États-Unis.

[Hartley, Zisserman, 2004] Hartley, R., Zisserman, A. (2004). *Multiple View Geometry in Computer Vision.* Cambridge University Press, deuxième édition.

[Haugazeau, 1968] Haugazeau, Y. (1968). *Sur les inéquations variationnelles et la minimisation de fonctionnelles convexes.* Thèse de doctorat, Université de Paris, Paris, France.

[Hirschmüller, 2005] Hirschmüller, H. (2005). Accurate and efficient stereo processing by semi-global matching and mutual information. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, pages 807–814, Washington, DC, USA.

[Hong, Chen, 2004] Hong, L., Chen, G. (2004). Segment-based stereo matching using graph cuts. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 1, pages 74–81, Washington, États-Unis.

[Horaud, Monga, 1993] Horaud, R., Monga, O. (1993). *Vision par ordinateur, outils fondamentaux.* Traité des nouvelles technologies, série Informatique. Hermès, Paris, France.

[Horn, 1977] Horn, B. K. P. (1977). Image intensity understanding. *Artificial Intelligence*, 8(2):201–231.

[Horn, 1986] Horn, B. K. P. (1986). *Robot vision.* The MIT Press, Cambridge, Massaachusetts and Mc-Graw-Hill Book, New York.

[Horn, Schunck, 1981] Horn, K. P., Schunck, G. (1981). Determining optical flow. *Artificial Intelligence*, (7):185–203.

[Hotter *et al.*, 1996] Hotter, M., Mester, R., Müller, F. (1996). Detection and description of moving objects by stochastic modelling and analysis of complex scenes. *Signal Processing : Image Communication*, 8:281–293.

[Hu *et al.*, 2005] Hu, Z., Lamosa, F., Uchimura, K. (2005). A complete U-V-disparity study for stereovision based 3D driving environment analysis. *In Proceedings of the Fifth International Conference on 3D Digital Imaging and Modeling, 3DIM*, pages 204–211, Ontario, Canada.

[Ishikawa, Geiger, 1998] Ishikawa, H., Geiger, D. (1998). Occlusions, discontinuities, and epipolar lines in stereo. *In European Conference on Computer Vision*, vol. 1, pages 232–248, Freibourg, Allemagne.

[Kanade, Okutomi, 1994] Kanade, T., Okutomi, M. (1994). A stereo matching algorithm with an adaptive window : Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932.

[Kang *et al.*, 2001] Kang, S., Szeliski, R., Chai, J. (2001). Handling occlusions in dense multi-view stereo. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 1, pages 103–110, Kauai, États-Unis.

[Kim *et al.*, 2005a] Kim, C., Lee, K., Choi, B., Lee, S. (2005a). A dense stereo matching using two-pass dynamic programming with generalized ground control points. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 2, pages 1075–1082, San Diego, États-Unis.

[Kim *et al.*, 2003] Kim, J., Kolmogorov, V., Zabih, R. (2003). Visual correspondence using energy minimization and mutual information. *In IEEE Conference Proceedings of International Conference on Computer Vision, ICCV*, pages 1033–1040, Nice, France.

[Kim *et al.*, 2005b] Kim, Y., Martinez, A., Kak, A. C. (2005b). Robust motion estimation under varying illumination. *Image and Vision Computing*, 23(4):365–375.

[Kim *et al.*, 1997] Kim, Y. S., Lee, J. J., Ha, Y. H. (1997). Stereo matching algorithm based on modified wavelet decomposition process. *Pattern Recognition*, 30(6):929–952.

[Klaus *et al.*, 2006] Klaus, A., Sormann, M., Karner, K. (2006). A segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. *In IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR*, vol. 3, pages 15–18, Hong Kong.

[Kolmogorov, Zabih, 2001] Kolmogorov, V., Zabih, R. (2001). Computing visual correspondence with occlusions using graph cuts. *In International Conference on Computer Vision*, vol. 2, pages 508–515, Vancouver, BC, Canada.

[Kolmogorov, Zabih, 2005] Kolmogorov, V., Zabih, R. (2005). Graph cut algorithms for binocular stereo with occlusions. *Mathematical Models in Computer Vision : The Handbook, Springer-Verlag*, pages 423–437.

[Konrad, Lan, 2000] Konrad, J., Lan, Z. D. (2000). Dense disparity estimation from feature correspondences. *In Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 3957, pages 90–101, San Jose, CA, USA.

[Konrad, Mancini, 1998] Konrad, J., Mancini, A. (1998). Robust quadtree-based disparity estimation for the reconstruction of intermediate stereoscopic images. *In Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 3995, pages 53–64.

[Labayrade, 2004] Labayrade, R. (2004). *Détection générique, robuste et rapide d'obstacles routiers par stéréovision embarquée.* Thèse de doctorat, Université Pierre et Marie Curie Paris VI, Jussieu, France.

[Labayrade *et al.*, 2002] Labayrade, R., Aubert, D., Tarel, J. P. (2002). Real time obstacle detection in stereovision on non flat road geometry through "V-disparity" representation. *In IEEE Conference Proceedings of Intelligent Vehicle Symposium*, vol. 2, pages 646–651, Versailles, France.

[Labayrade *et al.*, 2005] Labayrade, R., Royere, C., Gruyer, D., Aubert, D. (2005). Cooperative fusion for multi-obstacles detection with use of stereovision and laser scanner. *Autonomous Robots*, 19:117–140.

[Lemonde, Devy, 2004] Lemonde, V., Devy, M. (2004). Obstacle detection with stereovision. *In Mechatronics and Robotics*, Aachen, Germany.

[Longuet-Higgins, 1981] Longuet-Higgins, H. C. (1981). A computer program for reconstructing a scene from two projections. *Nature*, 293:133–135.

[Malgouyres, 2002] Malgouyres, F. (2002). Minimizing the total variation under a general convex constraint for image restoration. *IEEE Transactions on Image Processing*, 11(12): 1450–1456.

[Mallat, 1989] Mallat, S. G. (1989). A theory for multiresolution signal decomposition : The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

[Mallat, 1998] Mallat, S. G. (1998). *A Wavelet Tour of Signal Processing*. Academic Press.

[Marr, Poggio, 1976] Marr, D., Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, 194(4262):283–287.

[Marr, Poggio, 1979] Marr, D., Poggio, T. (1979). A theory of human stereo vision. *In Proceedings of the Royal Society of London*, vol. 204, pages 301–328.

[Meyer, 1992a] Meyer, Y. (1992a). *Ondelettes et algorithmes concurrents*. Hermann, Paris, France.

[Meyer, 1992b] Meyer, Y. (1992b). *Wavelets and operators*, vol. 37 *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

[Morlet, 1981] Morlet, J. (1981). Sampling theory and wave propagation. *In Proceedings of 51st Annual Meeting of the Society of Exploration Geophysicists*, Los Angeles.

[Nagel, Enkelmann, 1986] Nagel, H., Enkelmann, W. (1986). An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593.

[Nicodemus *et al.*, 1977] Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., Limperis, T. (1977). Geometric considerations and nomenclature for reflectance. *Final Report National Bureau of Standards (US)*.

[Ohta, Kanade, 1985] Ohta, Y., Kanade, T. (1985). Stereo by intra and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154.

[Oisel *et al.*, 2003] Oisel, L., Memin, E., Morin, L., Galpin, F. (2003). 1D dense disparity estimation for 3D reconstruction. *IEEE Transactions on Image Processing*, 12(9):1107–1119.

[Okutomi *et al.*, 2002] Okutomi, M., Nakano, K., Maruyama, J., Hara, T. (2002). Robust estimation of planar regions for visual navigation using sequential stereo images. *In IEEE International Conference on Robotics and Automation*, vol. 4, pages 3321–3327, Washington DC, USA.

[Osher *et al.*, 2003] Osher, S., Sole, A., Vese, L. (2003). Image decomposition and restoration using total variation minimization and the h^{-1} norm. *Multiscale Modeling and Simulation : A SIAM Interdisciplinary Journa*, 1(3):349–370.

[Papageorgiou, Poggio, 1999] Papageorgiou, C., Poggio, T. (1999). Trainable pedestrian detection. *In Proceedings of the International Conference on Image Processing, ICIP*, vol. 4, pages 35–39, Kobe, Japan.

[Parent, Fortelle, 2005] Parent, M., Fortelle, A. D. L. (2005). Cybercars : Past, present and future of the technology. *In ITS World Congress*, San Francisco, USA.

[Perona, Malik, 1990] Perona, P., Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639.

[Pesquet *et al.*, 1996] Pesquet, J., Krim, H., Carfantan, H. (1996). Time invariant orthonormal wavelet representations. *IEEE Transactions on Signal Processing*, 44:1964–1970.

[Pollard *et al.*, 1985] Pollard, S. B., Mayhew, J. E. W., Frisby, J. P. (1985). Pmf : A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14(4):449–470.

[Pons *et al.*, 2003] Pons, J., Keriven, R., Faugeras, O., Hermosillo Valadez, G. (2003). Variational and 3D scene flow estimation with statistical similarity measures. *In IEEE International Conference on Computer Vision*, vol. 1, pages 597–602, Nice, France.

[Robert, Deriche, 1996] Robert, L., Deriche, R. (1996). Dense depth map reconstruction : A minimization and regularization approach which preserves discontinuities. *In proceedings of European Conference on Computer Vision, ECCV*, vol. A, pages 439–451, Cambridge, Royaume-Uni.

[Rockafellar, 1970] Rockafellar, R. T. (1970). *Convex analysis.* NJ : Princeton Univ. Press, Princeton.

[Roma *et al.*, 2002] Roma, N., Santos-Victor, J., Tome, J. (2002). A comparative analysis of cross-correlation matching algorithms using a pyramidal resolution approach. *In* Christensen, H. I., Phillips, P. J., éditeurs : *Empirical Evaluation Methods in Computer Vision*, pages 117–142. World Scientific Press.

[Roy, Cox, 1998] Roy, S., Cox, I. (1998). A maximum-flow formulation of the N-camera stereo correspondence problem. *In IEEE International Conference on Computer Vision*, pages 492–499, Bombai, Inde.

[Rudin *et al.*, 1992] Rudin, L., Osher, S., Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268.

[Rziza *et al.*, 2001] Rziza, M., Aboutajdine, D., Morin, L., Tamtaoui, A. (2001). New disparity map estimation using higher order statistics. *In IEEE Conference Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, vol. 3, pages 1685–1688, Salt Lake City, USA.

[Scharstein, Szeliski, 2002] Scharstein, D., Szeliski, R. (2002). A taxomomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision, IJCV*, 47(1):7–42.

[Scharstein, Szeliski, 2003] Scharstein, D., Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 1, pages 195–202, Madison, WI.

[Serra, 1988] Serra, J. (1988). *Image Analysis and Mathematical Morphology.* Ac. Press.

[Shashua *et al.*, 2004] Shashua, A., Gdalyahu, Y., Hayon, G. (2004). Pedestrian detection for driving assistance systems : single-frame classification and system level performance. *In IEEE Conference Proceedings of Intelligent Vehicles Symposium*, pages 1–6, Parme, Italie.

[Shi *et al.*, 2001] Shi, F., Huges, N. R., Roberts, G. (2001). SSD matching using shift-invariant wavelet transform. *In Proceedings of the British Machine Vision Conference, BMVC*, Manchester, UK.

[Simard, Mailloux, 1988] Simard, P., Mailloux, G. (1988). A projection operator for the restoration of divergence-free vector fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 10(2):248–256.

[Simard, Mailloux, 1990] Simard, P., Mailloux, G. (1990). Vector field restoration by the method of convex projections. *Computer Vision, Graphics, and Image Processing*, 52(3): 360–385.

[Slesareva *et al.*, 2005] Slesareva, N., Bruhn, A., Weickert, J. (2005). Optic flow goes stereo : A variational method for estimating discontinuity preserving dense disparity maps. *In DAGM-Symposium*, pages 33–40.

[Soquet *et al.*, 2007] Soquet, N., Aubert, D., Hautiere, N. (2007). Road segmentation supervised by an extended V-Disparity algorithm for autonomous navigation. *In IEEE Conference Proceedings of Intelligent Vehicles Symposium*, pages 160–165, Istanbul, Turkey.

[Sun, 2002] Sun, C. (2002). Fast stereo matching using rectangular subregioning and 3D maximum surface techniques. *International Journal of Computer Vision*, 47(1):99–117.

[Takeda *et al.*, 1996] Takeda, N., Watanabe, M., Onoguchi, K. (1996). Moving obstacle detection using residual error of FOE estimation. *In Proceedings of the International Conference on Robots and System, IROS*, vol. 3, pages 1642–1647, Osaka, Japan.

[Tappen, Freeman, 2003] Tappen, M. F., Freeman, W. T. (2003). Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. *In IEEE Conference Proceedings of International Conference on Computer Vision, ICCV*, vol. 2, pages 900–907, Nice, France.

[Tikhonov, Arsenin, 1977] Tikhonov, A., Arsenin, A. (1977). Solution of ill-posed problems. *John Wiley and Sons*.

[Tikhonov, Arsénine, 1976] Tikhonov, A., Arsénine, V. (1976). *Méthodes de résolution de problèmes mal posés*. Editions MIR, Moscow.

[Toulminet *et al.*, 2004] Toulminet, G., Mousseot, S., bensrhair, A. (2004). Fast and accurate stereo vision-based estimation of 3D position and axial motion of road obstacles. *International Journal of Image and Graphics*, 4(1):99–126.

[Truchetet, 1998] Truchetet, F. (1998). *Ondelettes pour le signal numérique*. Hermès, Paris, France.

[Trussell, Civanlar, 1984] Trussell, H., Civanlar, M. (1984). The feasible solution in signal restoration. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 32(2):201–212.

[Tsai *et al.*, 2003] Tsai, D., Lin, C., Chen, J. (2003). The evaluation of normalized cross correlations for defect detection. *Pattern Recognition Letters*, 24(15):2525–2535.

[van der Mark, Gavrila, 2006] van der Mark, W., Gavrila, D. M. (2006). Real-time dense stereo for intelligent vehicles. *IEEE Transactions on intelligent transportation systems*, 7(2):38–50.

[Veksler, 1999] Veksler, O. (1999). *Efficient graph-based energy minimization methods in computer vision*. Thèse de doctorat, Université de Cornell, Ithaca, États-Unis.

[Veksler, 2001] Veksler, O. (2001). Stereo matching by compact windows via minimum ratio cycle. *In IEEE Conference Proceedings of International Conference on Computer Vision, ICCV*, vol. 1, pages 540–547, Vancouver, BC, Canada.

[Veksler, 2005] Veksler, O. (2005). Stereo correspondence by dynamic programming on a tree. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 2, pages 384–390, San Diego, États-Unis.

[Viola, Wells, 1997] Viola, P. A., Wells, W. M. (1997). Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154.

[Wei, Quan, 2004] Wei, Y., Quan, L. (2004). Region-based progressive stereo matching. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 1, pages 106–113, Washington, États-Unis.

[Weng *et al.*, 1992] Weng, J., Ahuja, N., Huang, T. (1992). Matching two perspective views. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 14(8):806–825.

[Yamada *et al.*, 1998] Yamada, I., Ogura, N., Yamashita, Y., Sakaniwa, K. (1998). Quadratic optimization of fixed points of nonexpansive mappings in Hilbert space. *Numerical Functional Analysis and Optimization*, 19:165–190.

[Yoon, Kweon, 2006] Yoon, K. J., Kweon, I. S. (2006). Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 28(4):650–656.

[Yoon, Kweon, 2005] Yoon, K. J., Kweon, I. S. (June 2005). Locally adaptive support-weight approach for visual correspondence search. *In IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 2, pages 924– 931, San Diego, CA, USA.

[Youla, Webb, 1982] Youla, D. C., Webb, H. (1982). Image restoration by the method of convex projections : Part 1, theory. *IEEE Transactions on Medical Imaging*, 1:81–94.

[Yu *et al.*, 2003] Yu, Q., Araújo, H., Wang, H. (2003). Stereo-vision based real time obstacle detection for urban environments. *In IEEE Conference Proceedings of the 11th International Conference on Advanced Robotics, ICAR*, vol. 3, pages 1671–1676, Coimbra, Portugal.

[Yuille, Poggio, 1984] Yuille, A. L., Poggio, T. (1984). A generalized ordering constraint for stereo correspondence. *Technical Report AIM-777*.

[Zabih, Woodfill, 1994] Zabih, R., Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. *In proceedings of European Conference on Computer Vision, ECCV*, pages 151–158, Stockholm, Suède.

[Zhang *et al.*, 2006] Zhang, J., McMillan, L., Yu, J. (2006). Robust tracking and stereo matching under variable illumination. *In IEEE Conference Proceedings of International Conference on Computer Vision and Pattern Recognition, CVPR*, vol. 1, pages 871–878, New York, USA.

[Zhang, Kambhamettu, 2002] Zhang, Y., Kambhamettu, C. (2002). Stereo matching with segmentation-based cooperation. *In European Conference on Computer Vision*, vol. 2, pages 556–571, Copenhague, Danemark.

[Zhang, 1996] Zhang, Z. (1996). Determining the epipolar geometry and its uncertainty : A review. Rapport de recherche RR-2927, Institut National de Recherche en Informatique et en Automatique, INRIA.

[Zhang *et al.*, 1994] Zhang, Z., Deriche, R., Faugeras, O., Luong, Q. (1994). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Rapport de recherche RR-2273, Institut National de Recherche en Informatique et en Automatique, INRIA.

[Zhang, Shan, 2000] Zhang, Z., Shan, Y. (2000). A progressive scheme for stereo matching. *In European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, vol. 2018 *Lecture Notes in Computer Science*, pages 68–85, Dublin, Irlande.

[Zhang, Xu, 1997] Zhang, Z., Xu, G. (1997). A general expression of the fundamental matrix for both projective and affine cameras. *In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, (IJCAI'97)*, pages 1502–1507, Nagoya, Japan.

[Zhao, Thorpe, 2000] Zhao, L., Thorpe, C. E. (2000). Stereo and neural network-based pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 1(3):148–154.

[Zitnick, Kanade, 2000] Zitnick, C. L., Kanade, T. (2000). A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 22(7):675–684.