Numéro d'ordre : 4161

Université des Sciences et Technologies de Lille
U.F.R. d'Informatique, Electronique, Electrotechnique et Automatique
Ecole Doctorale des Sciences Pour l'Ingénieur

THÈSE

présentée et soutenue publiquement
le 3 juillet 2008
par

Sébastien LEFEBVRE

pour l'obtention du grade de
Docteur de l'Université des Sciences et Technologies de Lille
en Automatique, Génie informatique, Traitement du Signal et des Images

Approche monodimensionnelle de la mise en correspondance stéréoscopique par corrélation - Application à la détection d'obstacles routiers

Composition du jury :

Rapporteurs :	Roland CHAPUIS, Professeur, LASMEA Fabrice MERIAUDEAU, Professeur, LE2I
Examinateurs :	Marion BERBINEAU, Directrice de recherche, INRETS-LEOST Michel DEVY, Directeur de recherche, LAAS-CNRS Jack-Gérard POSTAIRE, Professeur émérite, LAGIS
Invité :	Didier VAN DEN ABEELE, Ingénieur, Alstom Transport
Co-directeurs de thèse :	Sébastien AMBELLOUIS, Chargé de recherche, INRETS-LEOST François CABESTAING, Professeur, LAGIS

Thèse préparée au
Laboratoire d'**E**lectronique, **O**ndes et **S**ignaux pour les **T**ransports de l'**INRETS**

Remerciements

Ce travail de thèse a été réalisé au sein du Laboratoire d'Électronique, Ondes et Signaux pour les Transports de l'INRETS de Villeneuve d'Ascq. Il a été cofinancé par l'INRETS et la Région Nord-Pas de Calais que je remercie ici.

Tout d'abord, je tiens à remercier Monsieur Roland CHAPUIS, Professeur au LASMEA et Monsieur Fabrice MERIAUDEAU, Professeur au LE2I, qui ont accepté d'être rapporteurs de cette thèse et dont les remarques constructives m'ont permis d'améliorer ce mémoire.

Je remercie sincèrement Monsieur Michel DEVY, Directeur de recherche au LAAS et Monsieur Didier VAN DEN ABEELE, Ingénieur à Alstom Transport pour avoir accepté d'être examinateurs de ce travail de thèse et ce, malgré leurs emplois du temps chargés.

Aussi, je souhaite remercier Madame Marion BERBINEAU, Directrice de recherche et Directrice du LEOST pour m'avoir accueilli au sein de son unité de recherche et avoir accepté d'être examinatrice.

Je souhaite remercier très chaleureusement Monsieur Jack-Gérard POSTAIRE, Professeur émérite au LAGIS, pour avoir accepté la double mission d'être à la fois examinateur et président de ce jury. Sa présence fut pour moi un grand plaisir.

Et aussi, mes encadrants.. que dis-je... mes « Maîtres » ! !
Un immense merci à Sébastien AMBELLOUIS, Chargé de recherche au LEOST et François CABESTAING, Professeur au LAGIS, de m'avoir proposé ce sujet de thèse et pour leur encadrement exceptionnel. Leurs compétences techniques et leurs grandes qualités humaines m'ont permis de mener à bien ce projet : grâce à eux, ces trois dernières années ont été, professionnellement, un plaisir de chaque instant.

Un grand merci pour les relectures efficaces de papiers à la dernière minute...

Un merci tout particulier à Sébastien qui m'a convaincu de m'inscrire en DEA (actuel Master II recherche) ce sans quoi, je n'aurai jamais écrit ces quelques lignes ! !

Plus que des encadrants, ils sont devenus des amis. Qu'ils trouvent ici l'expression de mes plus sincères remerciements : si c'était à refaire, je reprendrais les mêmes !

Un grand merci à Amaury FLANCQUART, Ingénieur d'études au LEOST pour sa disponibilité, ses conseils et sa précieuse aide technique.

Un grand merci à toute l'équipe du LEOST pour son accueil sympathique et sa gentillesse en particulier, Lidwine et Olivier, nos deux gestionnaires de choc !

J'adresse toute ma sympathie à l'équipe Vision et Image du LAGIS. Un merci tout particulier à Ludovic MACAIRE pour son « regard extérieur » et ses précieux conseils lors de la préparation de ma soutenance de thèse.

Un clin d'œil à mes voisins de bureau successifs, Gérald et Julie, qui ont dû supporter au quotidien mon humour un peu lourd !

Merci beaucoup à l'ami Jérôme pour sa formation LaTeX sans qui mes différents exercices de rédaction auraient été bien laborieux.

Merci à Manu pour ses conseils et sa disponibilité !
Merci à Nanard pour son soutien technique pour l'étape primordiale du pot de thèse !
Merci à Daniel pour son appui logistique en reprographie !

Aussi, je tiens à remercier toutes les personnes qui ont amicalement partagé les moments de convivialité de mes pauses café et déjeuner : Emilie, Julie, Juliette, Amaury, Christophe, Cyril, François, Gérald, Jean, Jérôme, Jean-Luc, JPG, Manu, Philippe, Séb, Yann... et bien d'autres.

Je n'oublie pas la Team Excalibur qui m'a permis durant ces dernières années de passer des soirées fort sympathiques à me défouler : merci à vous !

Je tenais à remercier tous mes amis pour leur amitié, leur gentillesse et leur franchise : sans les citer un par un, qu'ils trouvent ici l'expression de ma plus sincère amitié !

Pour mes parents qui m'ont toujours laissé le choix de mes études, m'ont toujours soutenu et sans qui je ne serai pas là. Merci.

Merci à vous tous, qui êtes venu m'encourager lors de ma soutenance de thèse ; avec un merci tout particulier pour Chris et JF, tonton Marco et l'amie Mary !

Un immense merci à Aurélie pour m'avoir toujours soutenu et encouragé : pour tout ce qu'elle a fait pour moi et pour tout ce que qu'elle fera encore !

Et si avec après tout ce petit monde, il se trouve que j'ai oublié certaines personnes, qu'ils me pardonnent et trouvent ici toute ma sympathie.

Merci, merci, merci !!!

Table des matières

Introduction générale **1**
I.1 Les systèmes multi-capteur . . . 2
I.2 Les systèmes de vision . . . 3
I.3 La stéréovision . . . 4
I.4 Problématique . . . 5
I.5 Objectifs de la thèse . . . 7
I.6 Organisation du mémoire . . . 8
Références . . . 10

1 La stéréovision passive **11**
1.1 Le stéréoscope binoculaire . . . 12
1.2 Calibrage du stéréoscope . . . 13
1.3 Géométrie épipolaire . . . 14
1.4 Configuration parallèle . . . 14
1.5 La triangulation géométrique . . . 16
1.6 Les bases de la mise en correspondance . . . 17
1.6.1 Problèmes liés à la stéréovision passive . . . 18
1.6.2 Les contraintes de mise en correspondance . . . 18
1.6.2.1 La contrainte épipolaire . . . 19
1.6.2.2 La contrainte d'unicité . . . 20
1.6.2.3 La contrainte d'ordre . . . 20
1.6.2.4 La contrainte de position . . . 21
1.6.2.5 La contrainte de continuité de la disparité . . . 22
1.6.2.6 La contrainte de disparité maximale . . . 22
1.6.3 Primitives à apparier . . . 22
1.7 Techniques de mise en correspondance stéréoscopique . . . 23
1.7.1 Les méthodes locales . . . 23
1.7.1.1 La méthode de recherche exhaustive . . . 24
1.7.1.2 Les mesures de similarité . . . 24
1.7.1.3 La zone d'agrégation . . . 27
1.7.1.4 Présentation de quelques méthodes locales . . . 28

1.7.2 Les méthodes globales 30
1.7.2.1 Fonction de coût global 30
1.7.2.2 Les méthodes d'optimisation 31
1.7.3 Autres méthodes 33
1.8 Méthodes de raffinement des résultats de disparité 33
1.8.1 Propagation des disparités 33
1.8.2 Mise en correspondance au sous-pixel 34
1.8.3 Elimination des erreurs d'appariement 34
1.9 Conclusion 34
Références 36

2 La mise en correspondance stéréoscopique : un problème 1D ? 41
2.1 Méthodes linéaires 42
2.2 Calibration et rectification 46
2.3 Légitimité des fenêtres de similarité 2D 46
2.4 Méthode d'évaluation 47
2.4.1 Présentation 48
2.4.2 Les limites de la méthode d'évaluation 50
2.5 Comparaison 1D/2D 51
2.5.1 La méthode utilisée 52
2.5.2 Résultats obtenus sur le couple d'images *Head and Lamp* 53
2.5.2.1 Comparaison qualitative 54
2.5.2.2 Comparaison quantitative 60
2.5.3 Résultats obtenus sur le couple d'images *Venus* 63
2.5.4 Résultats obtenus sur une chaussée 64
2.5.4.1 Comparaison qualitative 65
2.5.4.2 Comparaison quantitative 67
2.6 Conclusion 67
Références 70

3 L'approche 1D proposée 71
3.1 Le principe de l'approche 1D 72
3.1.1 Calcul des scores de similarité 73
3.1.2 Modélisation de la confiance 75
3.1.3 La phase de décision 76
3.1.4 Validation directe/inverse 77
3.1.5 Raffinement des résultats 77

3.2 Etude de cas . . . 77
3.2.1 Mise en correspondance 1D multi-configurations basée sur la mesure de dissimilarité *SSD* . . . 77
3.2.2 Modélisation de la confiance par logique floue . . . 79
3.2.3 Décision . . . 83
3.3 Résultats . . . 86
3.3.1 Evaluation de Middlebury . . . 86
3.3.2 Méthodes comparables . . . 87
3.3.3 Comparaison des résultats obtenus par les sept méthodes . . . 88
3.3.4 Evaluation sur un plan incliné . . . 94
3.4 Conclusion . . . 100
Références . . . 102

4 La détection d'obstacles routiers 104
4.1 Un état de l'art . . . 105
4.2 Calcul de cartes des disparités quasi-denses et précises . . . 107
4.3 Projet RaViOLi . . . 109
4.3.1 Présentation générale . . . 110
4.3.2 Stéréoscope mono-caméra orientable . . . 110
4.3.3 Détection et extraction des obstacles . . . 113
4.3.3.1 Etat de l'art . . . 113
4.3.3.2 Comparaison des cartes des disparités obtenues par une méthode 2D et notre méthode 1D sur un couple d'images de route . . . 115
4.3.3.3 La détection des véhicules . . . 118
4.3.4 Comportement de la méthode 1D en situations critiques . . . 127
4.3.5 Quelques images réelles . . . 136
4.3.5.1 Environnement « transport guidé » . . . 136
4.3.5.2 Environnement « transport routier » . . . 138
4.4 Implantation matérielle . . . 139
4.5 Conclusion . . . 143
Références . . . 144

Conclusion générale 146

Perspectives 149
P.1 Pixels non affectés en zones texturées . . . 149
P.1.1 Décision au sous-pixel . . . 149
P.1.2 Ambiguïté d'une discontinuité . . . 150

P.2 Utilisation de l'information couleur 152
P.3 Augmentation de la densité de la carte 153

Liste des publications **154**

Table des figures **156**

Table des tableaux **156**

Résumé/Abstract **158**

Introduction générale

DEPUIS toujours, la sécurité des moyens de transport a fait l'objet d'une attention particulière, notamment la voiture qui est statistiquement le moyen de transport le moins sûr. Selon de nombreuses études en accidentologie, un accident apparaît comme un dysfonctionnement du système composé du conducteur, du véhicule et de l'environnement. Dans la grande majorité des cas, l'homme est à l'origine de l'accident, soit à cause d'une perception dégradée de l'environnement routier, soit à cause d'un contrôle inadapté de la dynamique du véhicule.

Afin de limiter les conséquences de ces défaillances humaines, de nombreuses équipes de recherche travaillent à la mise en place de systèmes embarqués d'aide à la conduite dont l'objectif est d'informer le conducteur du degré de dangerosité de la situation dans laquelle il se trouve, voire même d'agir directement sur le véhicule. Ces systèmes embarqués reconstruisent à chaque instant l'état de l'environnement routier autour du véhicule au moyen d'une perception multi-capteur et d'une étape de fusion de données. Certains indicateurs sont extraits de cette cartographie dynamique de l'environnement, puis exploités pour alerter le conducteur. Les moyens d'alerte peuvent être un ensemble de signes visuels et sonores transmis au conducteur. Ces mêmes indicateurs peuvent être également définis comme les entrées d'un module de contrôle/commande programmé pour agir sur le véhicule afin de réduire le niveau de dangerosité de la situation ou d'éviter l'accident.

Pour y parvenir, les constructeurs et autres professionnels du transport ont mis en place de nombreux programmes de recherche nationaux (ARCOS, LAVIA etc.) ou internationaux (PROMETHEUS, ESAFETY etc.) dans le domaine de la sécurité routière. L'objectif inscrit dans chacun de ces programmes est de développer une solution technologique tendant à réduire le risque encouru par les conducteurs et les passagers des véhicules routiers et plus récemment celui encouru par les piétons et les deux roues motorisées.

Les études les plus anciennes se sont conclues par le développement de systèmes composés d'un capteur et d'une unité de traitement dont la fonction est de surveiller l'environnement du véhicule à bord duquel ils sont embarqués (le véhicule porteur). L'objectif est de reconstruire l'environnement en suivant plusieurs étapes : la détection des objets dans la scène, leur positionnement par rapport au véhicule porteur, l'initialisation et le suivi de leur trajectoire. Plus récemment, les projets ont montré que la combinaison de plusieurs de ces capteurs constituait une réponse, partielle mais réaliste, à une partie des problèmes posés tels que le manque de robustesse, de fiabilité et de précision des résultats

obtenus. Aujourd'hui, les efforts de recherche sont donc portés sur la mise au point de systèmes multi-capteur.

I.1 Les systèmes multi-capteur

Un système de perception multi-capteur est défini comme un ensemble de plusieurs capteurs actifs ou passifs [1]. Les capteurs utilisés sont généralement de typologies différentes, c'est-à-dire fondés sur des technologies qui leur permettent, par exemple, de percevoir une scène dans différentes gammes de fréquences allant du domaine des micro-ondes et des ondes RADAR (longueur d'ondes de quelques millimètres) à celui de l'infrarouge et du visible (longueur d'ondes de quelques centaines de nanomètres). Ainsi, par l'intermédiaire d'un traitement spécifique des signaux fournis par chaque capteur, l'environnement surveillé est décrit par des données complémentaires et redondantes. Chacun de ces capteurs possède des modes de fonctionnement différents dépendant par exemple des conditions d'éclairage ou des conditions climatiques. Une complémentarité entre eux est possible [2] : ainsi, un objet qui n'est pas détecté par un capteur dans l'obscurité peut l'être par un autre. Par ailleurs, si deux capteurs fournissent à la fois la position et la trajectoire d'un même objet de la scène, la fusion des deux informations redondantes permet l'amélioration de la robustesse du résultat final, en particulier en cas de panne de l'un ou l'autre des capteurs.

Les capteurs actifs, tels que le RADAR et le LIDAR, émettent leur propre signal afin de mesurer leur distance aux objets. Cette mesure est réalisée à partir du temps de parcours d'une onde électromagnétique émise puis réfléchie par un objet. Ces capteurs comportent un seul faisceau ou peuvent être munis d'un système à balayage qui leur permet de détecter les objets situés dans un secteur angulaire donné. Le RADAR fonctionne dans la gamme des hyperfréquences. Il permet des mesures de distances à longues portées et ce, quelles que soient les conditions météorologiques. Cependant, ce capteur est très sensible aux caractéristiques de la scène observée telles que la pente, les virages. Par ailleurs, il ne permet pas la détection de petits obstacles notamment à cause d'une résolution angulaire faible. Le LIDAR opère quant à lui dans la gamme des ondes LASER. Il permet d'appréhender efficacement la navigation en terrain inconnu, notamment pour des applications en robotique mobile. Cependant, son encombrement, sa fragilité et son coût encore élevé sont des inconvénients majeurs à son utilisation.

Contrairement aux capteurs actifs, les capteurs passifs, tels que les caméras, n'émettent aucun signal. La caméra est capable de percevoir une scène à partir de la lumière réfléchie par les objets qui la compose. Elle fournit une image de la scène en la projetant sur un capteur photosensible 1D (caméra linaire) ou 2D (caméra matricielle). Cette image contient une information beaucoup plus riche que celle portée par le signal reçu par un RADAR ou un LIDAR. Certains systèmes de vision sont considérés comme actifs

lorsqu'ils sont associés à un système d'éclairage tel qu'un ensemble de LED infrarouges ou une source de lumière structurée. Les LED infrarouges sont souvent utilisées lorsque les conditions d'éclairage sont insuffisantes : dans ce cas, la luminosité et le contraste des images sont améliorés. Dans le cas d'une lumière structurée, un motif lumineux éclaire la scène à surveiller. Le système de vision enregistre et analyse alors les déformations de ce motif dues à la présence des objets dans la scène.

De nombreux projets ont proposé et proposent encore aujourd'hui de combiner plusieurs des capteurs cités précédemment. Le projet régional « RaViOLi[1] », pour « RADAR, Vision Orientable, LIDAR », soutenu par la région Nord-Pas De Calais, est un de ces projets. Son objectif consiste à surveiller l'environnement situé à longue distance à l'avant du véhicule sur autoroute. RaViOLi se décompose en plusieurs actions de recherche : un volet « vision », un volet « RADAR » et un volet « fusion ». Ce travail de thèse s'inscrit dans le premier volet de ce projet.

I.2 Les systèmes de vision

Dans le cadre d'une application transport, la vision permet de satisfaire un certain nombre de fonctions telles que la détection et l'identification des signalisations routières horizontales (les marquages au sol) ou verticales (les panneaux routiers), l'aide au stationnement [1], la détection et l'évitement d'obstacles [3]. Les deux dernières applications nécessitent la reconstruction de la structure de l'environnement dans lequel se déplace le véhicule, ainsi que la localisation et le suivi 3D de tous les objets qui s'y trouvent. Ces deux tâches peuvent être remplies grâce à la mise en place d'un système de vision composé d'une ou de plusieurs caméras.

Lorsqu'une seule caméra est utilisée, le système est dit mono-caméra. Lorsque deux caméras composent le système, il est appelé système stéréoscopique (ou stéréoscope) et on parle alors de *stéréovision*. Tous deux permettent le calcul des coordonnées tridimensionnelles des points d'une scène observée à partir de leur projection dans deux images de cette scène. Le mode de calcul est fondé sur le principe de la triangulation géométrique. Dans le cas mono-caméra, ces deux images sont acquises l'une après l'autre à deux instants différents. Il nécessite donc de connaître précisément le déplacement de la caméra entre les deux prises de vue. Dans le cas stéréoscopique, les deux images sont acquises au même instant mais à deux points de vue différents. Comme dans le cas précédent, la connaissance de la position relative des deux caméras est nécessaire. Elle est généralement obtenue de manière précise lors d'une étape de calibrage et varie peu dans des conditions raisonnables d'utilisation (pas ou peu de vibrations, de déformations ou de chocs).

Plusieurs études ont montré que, comparativement au système mono-caméra, un sté-

[1]RAdar, VIsion Orientable, LIdar : projet financé par l'Union Européenne, le gouvernement français et la région Nord-Pas de Calais sous le numéro de contrat OBJ2-2005/3-4.1-253-7820

réoscope fournissait des résultats plus précis sans faire appel à des hypothèses fortes sur le mouvement ou sur la géométrie des objets contenus dans la scène. Bien qu'il nécessite une procédure de calibrage initiale et qu'il demande la mise en place d'une procédure de recalibrage en ligne afin de corriger les effets des conditions de fonctionnement, le stéréoscope est considéré comme le moyen le plus adapté à la reconstruction 3D de l'environnement routier.

I.3 La stéréovision

L'objectif de la stéréovision est de calculer les coordonnées 3D de chaque point de la scène à partir des coordonnées de leur projection sur les deux images fournies par les deux caméras. L'étape essentielle des algorithmes de stéréovision est l'*appariement* de certaines *primitives* entre les deux images. Ces primitives, extraites des images, peuvent être des points de contour, des segments ou des sous-ensembles de pixels. L'appariement consiste alors à trouver, pour chaque primitive de l'une des deux images, son correspondant dans l'autre image aussi appelé *homologue*.

Quelle que soit la position relative des deux caméras, l'*homologue* d'une primitive de la première image se situe sur une droite dans la seconde image. Cette droite est appelée la *droite épipolaire* et constitue la région de l'image où il convient donc de mener la recherche. Ceci fait de l'étape de *mise en correspondance* un problème fondamentalement 1D.

Par ailleurs, lorsqu'une phase de calibrage fort est préalablement menée, ces droites épipolaires sont confondues avec les lignes des images. Dans ce dernier cas de figure, la position des primitives ne diffèrent que par leurs abscisses le long de ces droites épipolaires ; ce décalage est appelé *disparité*. Finalement, le résultat obtenu est une carte des disparités dont la densité dépend de la primitive appariée. Lorsque la primitive est un point de contour, la carte est *éparse* [4]. Lorsque la primitive est un ensemble de pixels centré autour de chaque pixel des images, la carte des disparités est *dense*. La valeur de disparité du pixel permet le calcul de la *profondeur* du point correspondant dans la scène réelle.

Les méthodes éparses atteignent généralement des performances « temps-réel », c'est-à-dire qu'elles fournissent des résultats à la même cadence que l'acquisition des images. Néanmoins, une carte éparse comportant moins d'information, il est plus difficile de lui appliquer des traitements de plus haut-niveau tels qu'une segmentation ou une reconnaissance d'objets. Les algorithmes de stéréovision dense constituent une solution à ces problèmes mais leur complexité algorithmique est plus importante, ce qui les rend difficilement compatibles avec la contrainte temps-réel tout particulièrement sur des calculateurs standards. Une solution est alors d'envisager leur implantation sur des matériels adaptés [5].

Dans ce travail, nous nous plaçons dans le cadre d'applications de modélisation, de visualisation et de « métrologie ». Ces applications nécessitent généralement le calcul d'une carte dense des disparités. Par ailleurs, la « métrologie » étant le cadre applicatif prioritaire que nous nous sommes fixés, cette carte des disparités doit être précise.

I.4 Problématique

La plupart du temps, les méthodes denses sont fondées sur la comparaison de *fenêtres* bidimensionnelles (2D) centrées ou non sur les pixels à apparier. La disparité d'un pixel de l'une des deux images est déterminée en recherchant, le long de la droite épipolaire de l'autre image, la fenêtre ressemblant le plus à celle centrée au pixel considéré et en mesurant le décalage de position entre ces deux fenêtres. Pour cela, ces méthodes posent les deux hypothèses suivantes :

1. la texture contenue dans la fenêtre doit être suffisante ;
2. tous les pixels contenus dans la fenêtre doivent avoir la même disparité.

Lorsque ces hypothèses sont vérifiées, la disparité est déterminée avec une grande certitude et une grande précision.

La première hypothèse se traduit par une distribution spatiale des niveaux de gris dont la variance est élevée comme le montre l'histogramme de la figure I.1(a) dans le cas de l'image *Head and Lamp*. La véracité de cette hypothèse dépend fortement des caractéristiques de l'environnement, des conditions d'acquisition et du matériel vidéo utilisé. La figure I.1(b) illustre, dans le cas d'une scène routière, les conséquences de la faible taille des grains du matériau composant la chaussée combinée aux caractéristiques de la caméra (distance focale de l'objectif et résolution du capteur). Dans ce cas, l'histogramme de la figure I.1(d) apparaît peu informatif et ne permet pas la discrimination de la fenêtre avec les fenêtres voisines.

La seconde hypothèse se traduit par le fait que la méthode mène sa recherche le long des droites épipolaires en appliquant le même *décalage* à tous les pixels de la fenêtre considérée. Elle signifie que tous les points de la scène se projetant sur cette fenêtre sont à la même distance du stéréoscope. Néanmoins, la figure I.2 présente deux cas rencontrés fréquemment pour lesquels l'hypothèse n'est pas vérifiée. Dans une carte des disparités, ces cas se manifestent par un taux d'erreur élevé dans les zones comprenant deux objets situés à des profondeurs différentes. Sur la figure I.2(a), les voisinages contiennent des points appartenant à plusieurs objets situés à des distances différentes (la lampe, la statue et la bibliothèque). Sur la figure I.2(b), il apparaît clairement que la première et la dernière ligne de la fenêtre 2D correspondent à des points de la route situés à des profondeurs différentes.

Plus la fenêtre est de taille importante, plus le risque de voir la première hypothèse non respectée et d'obtenir une carte des disparités peu dense est faible. A contrario, plus

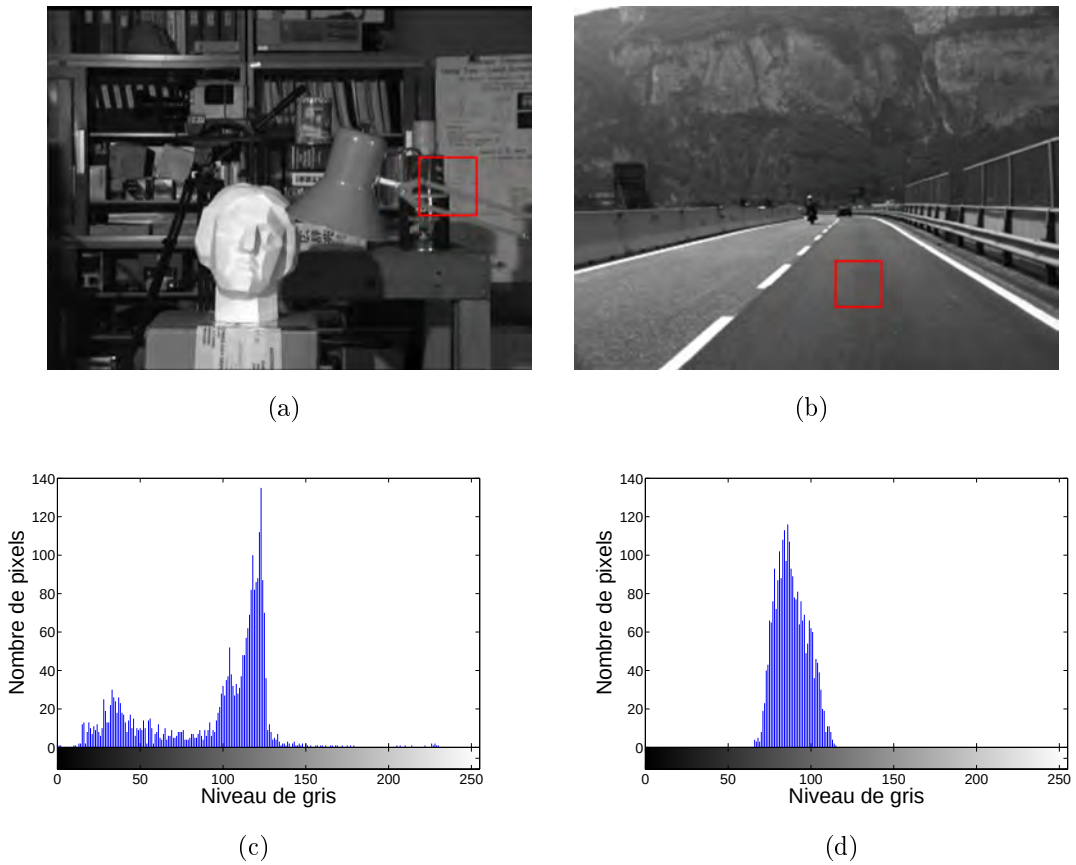


Fig. I.1 : Illustration de l'hypothèse 1 : (a)(c) vérifiée et (b)(d) non vérifiée avec respectivement l'image et l'histogramme associé.

la taille de la fenêtre est faible, plus la seconde hypothèse a de chances d'être vérifiée. Nous sommes donc en présence de deux phénomènes antagonistes qui engendrent un taux d'erreur élevé et une imprécision importante. Afin de limiter ce problème, certains auteurs ont proposé des méthodes fondées sur une analyse multi-fenêtres [6] et ont introduit la notion de carte des disparités *quasi-dense* [7].

Dans [7], l'auteur propose d'augmenter la densité d'une première carte des disparités éparse en lui appliquant un algorithme de croissance de régions. Cette méthode permet ainsi d'étendre le calcul des disparités des zones texturées aux zones non texturées tout en limitant l'apparition des erreurs. Dans un premier temps, la disparité est calculée pour l'ensemble des pixels contenus dans les régions texturées. La seconde étape consiste à augmenter la densité de la première carte obtenue en propageant les solutions les plus fiables selon un critère de confiance, calculé à partir de l'analyse de la distribution locale des niveaux de gris. Cette décomposition en deux étapes permet de respecter simultanément les deux hypothèses précédentes et donc de limiter l'imprécision de la carte des disparités finales. Cependant, cette méthode de propagation des disparités peut être assimilée à une méthode de post-traitement. Dans ce travail, nous proposons de calculer directement les valeurs précises des disparités.



(a) (b)

Fig. I.2 : Illustration de l'hypothèse 2 non vérifiée : (a) pour *Head and Lamp* et (b) pour une image de route.

I.5 Objectifs de la thèse

Dans la suite de ce mémoire, nous proposons une méthode générique de mise en correspondance stéréoscopique de fenêtres qui calcule une carte quasi-dense des disparités associée à une carte des confiances. Cette approche générique a pour finalité la mesure précise de distances et ce, quelle que soit l'application choisie. La carte des confiances est un moyen de mesurer, en chaque pixel, la véracité des deux hypothèses et donc d'accorder plus ou moins de crédit à chaque disparité calculée.

La confiance est calculée à partir de l'analyse de la forme d'un ensemble de courbes de similarité. Afin de limiter la contradiction entre les deux hypothèses, les courbes de similarité sont calculées en utilisant plusieurs fenêtres. Par ailleurs, comme nous l'avons présenté précédemment, les méthodes traditionnelles exploitent une ou plusieurs fenêtres 2D dont la taille est fixée, afin d'augmenter la densité de la carte tout en minimisant les erreurs dans les zones de discontinuité. Sachant que, comme nous l'avons rappelé, la recherche de deux pixels homologues est un problème 1D, la technique proposée utilise des fenêtres 1D. Toutefois, nous limitons notre étude à des couples d'images stéréoscopiques obtenus après un calibrage fort.

Nous souhaitons donc montrer que notre méthode générique 1D utilisant des fenêtres 1D est plus efficace que les méthodes utilisant des fenêtres 2D et déjà largement relatées dans la littérature. L'évaluation de la méthode générique 1D sera réalisée sur la base d'un protocole d'évaluation académique permettant une comparaison exhaustive avec d'autres méthodes déjà évaluées. De plus, de part les spécificités « transport » du laboratoire et du projet dans lequel s'intègre ce travail de thèse, la méthode générique sera évaluée dans un contexte de détection d'obstacles à l'avant d'un véhicule et ce, pour de longues distances. En effet, il est important de rappeler que dans le cadre d'une application liée à la sécurité des transports, la précision de la carte des disparités est très importante.

Dans ce cas, nous voulons montrer qu'il est préférable de ne pas affecter une disparité aux pixels contenus dans les zones peu texturées et de conserver ainsi une précision élevée.

Même s'il ne s'agit pas d'un des objectifs de ce travail, la méthode générique que nous proposons a été élaborée afin d'être implantée sur une architecture dédiée et de respecter les contraintes temps-réel imposées par toute application. Ce matériel est une carte électronique reconfigurable au format PCI et composée d'un composant logique programmable (FPGA[1]) [8]. Cette carte, appelée carte STREAM, a été développée dans le cadre d'un précédent projet de recherche soutenu par le Conseil Régional Nord-Pas de Calais, l'Etat et le FEDER[2].

I.6 Organisation du mémoire

Le présent mémoire s'articule autour de quatre chapitres.

Le chapitre 1 décrit les bases de la stéréovision passive telles le stéréoscope, le calibrage, le calcul de l'information tridimensionnelle, etc. Une attention toute particulière est portée à la description de l'étape de mise en correspondance.

Le chapitre 2 présente, dans un premier temps, les méthodes de mise en correspondance stéréoscopique basées sur l'utilisation de caméras linéaires. Dans un second temps, nous nous plaçons dans le contexte des méthodes d'appariement de fenêtres et nous comparons les résultats obtenus avec des fenêtres bidimensionnelles (fenêtres de forme carrée) et les résultats obtenus avec des fenêtre monodimensionnelles (fenêtres de forme ligne). La comparaison est réalisée selon des critères qualitatifs et quantitatifs sur la base de l'évaluation de Middlebury présentée dans [9] et introduite au préalable.

Dans le chapitre 3, nous débutons par une présentation du principe général de notre approche 1D. Nous poursuivons par la description d'une étude de cas de notre démarche et la présentation des résultats obtenus. Les cartes des disparités calculées grâce à notre méthode générique sont comparées à celles déterminées par d'autres méthodes similaires. Comme dans le chapitre précédent, cette comparaison est menée dans le cadre de l'évaluation académique proposée par Scharstein et Szelieski dans [9]. Nous compléterons cette comparaison en étudiant les résultats obtenus sur un couple d'images présentant un plan incliné.

[1]Field Programmable Gate Arrays

[2]Fonds Européen de DEveloppement Régional

Le chapitre 4 présente une application directe de notre méthode de mise en correspondance stéréoscopique à la détection d'obstacles en milieu routier. Les résultats de notre méthode sont obtenus sur des images introduites dans [10] par Wannes van der Mark *et al.* et comparés à ceux d'une méthode classique basée sur l'appariement de fenêtres 2D. Le projet RaViOLi, dans lequel s'intègre ce travail, est présenté et plus particulièrement le capteur stéréoscopique mono-caméra. Ensuite, une méthode simple de détection et d'extraction des obstacles, sur la base de nos cartes des disparités quasi-denses et précises, est détaillée ; les résultats associés sont présentés à partir d'images de synthèse et d'images réelles. A la fin de ce chapitre, la complexité ainsi que les possibilités d'implantation de notre méthode sur une architecture dédiée sont évoquées.

Enfin, la dernière partie de ce mémoire est consacrée aux conclusions de ce travail de thèse et permet d'amener différentes perspectives de poursuite de nos recherches.

Références

[1] V. LEMONDE : *Stéréovision embarquée sur véhicule : De l'auto-calibrage à la détection d'obstacles*. Thèse de doctorat, Institut National des Sciences Appliquées de Toulouse, 2005. 2, 3, 33, 46, 110, 115

[2] C. STILLER, J. HIPP, C. RÖSSIG et A. EWALD : Multisensor obstacle detection and tracking. *Image and vision computing*, 18(5):389–396, avril 2000. 2

[3] R. LABAYRADE : *Détection générique, robuste et rapide d'obstacles routiers par stéréovision embarquée*. Thèse de doctorat, Université Paris VI, 2004. 3, 105, 106

[4] S. NEDEVSCHI, R. DANESCU, D. FRENTIU, T. MARITA, F. ONIGA, C. POCOL, R. SCHMIDT et T. GRAF : High accuracy stereo vision system for far distance obstacle detection. *Proceedings of the IEEE Intelligent Vehicles Symposium 2004*, pages 292–297, 2004. 4

[5] M. PÉREZ, F. CABESTAING et J.G. POSTAIRE : A comparison of hardware resources required by real-time dense stereo algorithms. *IEEE International Workshop on Computer Architectures for Machine Perception*, New Orleans, LA, USA, mai 2003. 4

[6] A. FUSIELLO, V. ROBERTO et E. TRUCCO : Efficient stereo with multiple windowing. *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 858–863, San Juan, Porto Rico, juin 1997. 6

[7] M. LHUILLIER et L. QUAN : Match propagation for image-based modeling and rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8): 1140–1146, 2002. 6, 76, 106, 114

[8] O. FAUGERAS, B. HOTZ, H. MATHIEU, T. VIÉVILLE, Z. ZHANG, P. FUA, E. THÉRON, L. MOLL, G. BERRY, J. VUILLEMIN, P. BERTIN et C. PROY : Real time correlation-based stereo : Algorithm, implementation and applications. Rapport de recherche RR-2013, Institut National de Recherche en Informatique et en Automatique, INRIA, août 1993. 8, 141

[9] D. SCHARSTEIN et R. SZELISKI : A taxomomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1): 7–42, avril 2002. **www.middlebury.edu/stereo**. 8, 35, 46, 47, 54, 60, 63, 72, 77, 86, 88, 89, 90, 91, 92

[10] W. van der MARK et D.M. GAVRILA : Real-time dense stereo for intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):38–50, mars 2006. 9, 107, 108

Chapitre 1

La stéréovision passive

Sommaire

1.1 Le stéréoscope binoculaire 12
1.2 Calibrage du stéréoscope 13
1.3 Géométrie épipolaire 14
1.4 Configuration parallèle 14
1.5 La triangulation géométrique 16
1.6 Les bases de la mise en correspondance 17
1.6.1 Problèmes liés à la stéréovision passive 18
1.6.2 Les contraintes de mise en correspondance 18
1.6.3 Primitives à apparier 22
1.7 Techniques de mise en correspondance stéréoscopique 23
1.7.1 Les méthodes locales 23
1.7.2 Les méthodes globales 30
1.7.3 Autres méthodes 33
1.8 Méthodes de raffinement des résultats de disparité 33
1.8.1 Propagation des disparités 33
1.8.2 Mise en correspondance au sous-pixel 34
1.8.3 Elimination des erreurs d'appariement 34
1.9 Conclusion 34
Références 36

L'ACQUISITION d'une scène tridimensionnelle par une caméra se traduit par une image résultat bidimensionnelle. L'information de profondeur est perdue lors du processus d'acquisition. Ainsi, un pixel de l'image peut être associé à une infinité de points tridimensionnels situés le long de la droite passant par le dit pixel et le centre optique de la caméra. Pour lever cette ambiguïté et retrouver l'information de profondeur, il est nécessaire d'ajouter au moins un capteur supplémentaire au système d'acquisition : c'est le principe même de la stéréovision.

La principe de stéréovision est donc fondé sur l'utilisation d'un dispositif de prise de vues composé de plusieurs caméras. A l'instar de la vision animale, le système stéréoscopique le plus courant et le plus simple n'utilise que deux caméras, on parle alors de

stéréoscopie binoculaire. Certains systèmes stéréoscopiques exploitent les images issues de trois caméras [1] [2] alors que d'autres n'utilisent qu'une seule caméra associée à un jeu de miroirs [3] [4].

Le principe général de la stéréovision repose donc sur l'utilisation du *décalage* entre les images acquises par le stéréoscope pour déterminer l'information de *profondeur*. Dans ce chapitre, l'ensemble du processus de stéréovision sera décrit : de l'acquisition, au calcul de l'information de profondeur, en passant par l'étape la plus critique qui est la mise en correspondance.

1.1 Le stéréoscope binoculaire

Différents systèmes de prise de vues stéréoscopiques ont été l'objet de nombreux travaux. Le système classique, composé de deux caméras, permet l'acquisition de deux images au même instant mais à des points de vues légèrement différents.

Nous désignons par les indices g et d les grandeurs relatives respectivement aux caméras *gauche* et *droite*. Soit M_r un point de l'espace 3D réel, que nous désignons par la suite par « scène ». Ce point se projette sur le capteur droit en M_d et sur le capteur gauche en M_g comme présenté figure 1.1. Ceci est induit directement par le modèle *sténopé* constitué du plan image gauche (resp. droit) et du centre optique C_g (resp. C_d) pour la caméra gauche (resp. droite). Ainsi, le point M_r de coordonnées (X_r, Y_r, Z_r) dans la scène se projette sur le plan image gauche en un point M_g de coordonnées (x_g, y_g), intersection de la droite $(C_g M_g)$ avec le plan image gauche. Il en est de même pour le point M_d. Ces points M_g et M_d sont définis dans leurs repères respectifs $(R_g = (O_g, (X_g, Y_g))$ et $R_d = (O_d, (X_d, Y_d))$ et constituent un couple de points dits *homologues*.

Le point M_r de la scène et les deux projections M_g et M_d définissent le *plan épipolaire*. L'intersection de ce plan avec les deux plans image droit et gauche forme deux droites appelées *droite épipolaire droite* et *droite épipolaire gauche*. La droite épipolaire gauche constitue le lieu des points de l'image gauche pouvant correspondre à un point M_d de l'image droite. De même, la droite épipolaire droite constitue le lieu des points de l'image droite pouvant correspondre à un point M_g de l'image gauche. Ces deux droites épipolaires sont appelées *droites épipolaires conjuguées*.

L'ensemble des points tridimensionnels de la scène forme un ensemble de droites épipolaires. L'ensemble des droites épipolaires d'une même image (gauche ou droite) ont une intersection commune en un point appelé *épipole* (e_g ou e_d) comme le présente la figure 1.2.

Pour déterminer la profondeur du point M_r, il faut déterminer l'intersection des droites $(C_g M_g)$ et $(C_d M_d)$; pour ce faire, il faut exprimer les équations de ces deux droites dans un même repère, autrement dit, établir la géométrie du capteur stéréoscopique dans le même référentiel appelé *repère de référence*. La relation qui permet de passer d'un point

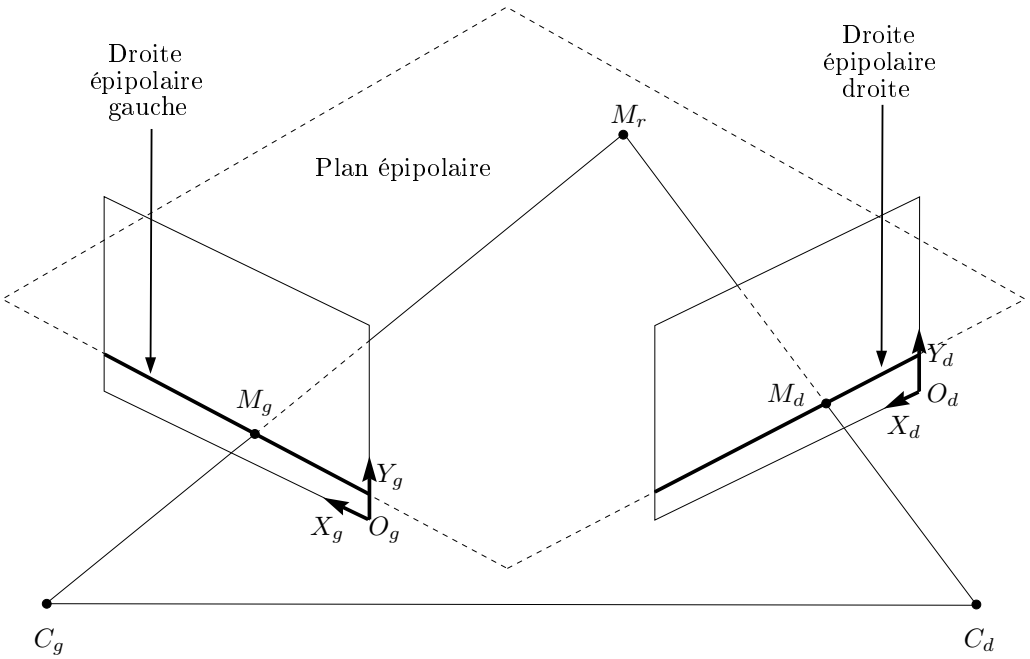


Fig. 1.1 : Le stéréoscope en configuration quelconque.

M_r de la scène dans le repère de référence à un point M_g ou M_d dans les repères 2D des capteurs gauche et droit, est déterminée lors de la phase de calibrage.

1.2 Calibrage du stéréoscope

Le calibrage consiste à établir les relations liant les repères 2D des caméras au repère de référence 3D de la scène observée : cette étape est primordiale.

Le stéréoscope étant positionné d'une manière quelconque, il est nécessaire de déterminer les paramètres intrinsèques et extrinsèques du système. Les paramètres intrinsèques constituent l'ensemble des caractéristiques propres à chacune des caméras indépendamment l'une de l'autre. Il s'agit de la résolution du capteur de la caméra, de la distance focale f de l'objectif de la caméra et de son modèle de déformation. Les paramètres extrinsèques constituent l'ensemble des paramètres liés à la position et l'orientation des caméras par rapport au repère de référence [5].

Ainsi, le calcul de ces paramètres permet de connaître la transformation à appliquer sur les images pour passer de leurs repères respectifs au repère de référence. Cela passe par l'estimation de la matrice fondamentale grâce à un ensemble de points de correspondance. Ces points sont extraits de mires (ou objets étalons) disposées dans la scène et dont les caractéristiques sont parfaitement connues. Le lecteur intéressé pourra trouver dans [6] un chapitre dédié à la géométrie et au calibrage des caméras, plus particulièrement dans le cas de systèmes stéréoscopiques.

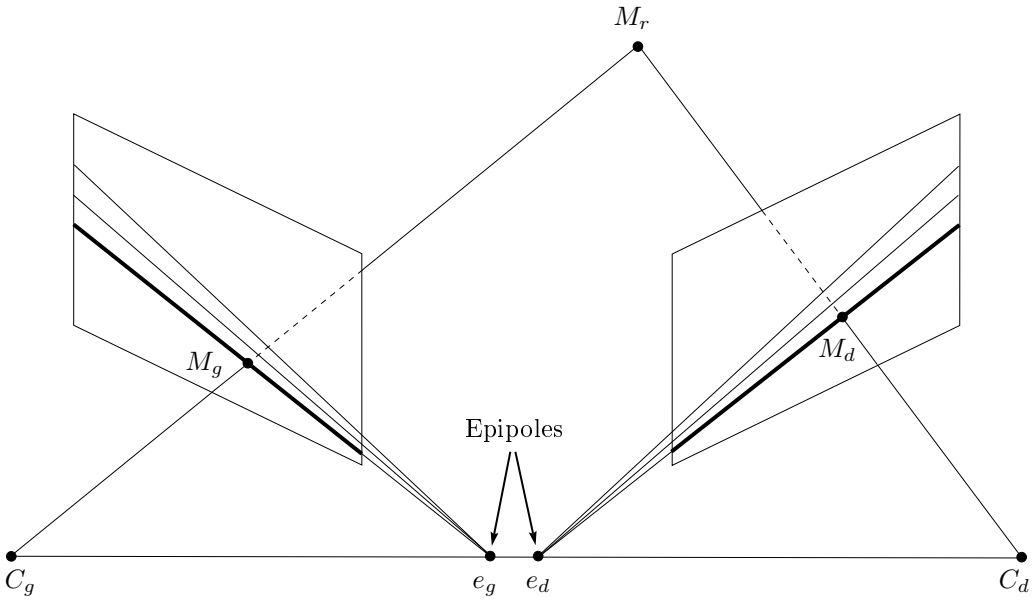


Fig. 1.2 : Illustration des épipoles.

1.3 Géométrie épipolaire

La notion de géométrie épipolaire est primordiale dans la phase de mise en correspondance (ou d'appariement) qui consiste à chercher dans l'une des deux images le correspondant d'une entité présente dans l'autre image. Chercher dans toute l'image droite le point homologue d'un point de l'image gauche peut s'avérer très coûteux en temps de calculs surtout si l'opération est réitérée pour tous les points de l'image.

La géométrie épipolaire permet d'affirmer que deux points homologues, dans les images gauche et droite, se situent sur les droites épipolaires conjuguées. Autrement dit, la géométrie épipolaire permet de réduire l'espace de recherche de deux points homologues à ces deux droites.

Pour ce faire, les équations de ces droites épipolaires doivent être déterminées avant l'étape de mise en correspondance. Cependant, la plupart des méthodes relatées dans la littérature passent outre ce calcul des équations des droites épipolaires en se plaçant dans le contexte de la *configuration parallèle*. La recherche des points homologues est donc grandement simplifiée puisque le calcul préalable des équations des droites épipolaires n'est plus nécessaire. Deux points homologues M_g et M_d se situent sur la même ligne y dans les images gauche et droite ; les coordonnées respectives de ces deux points ne différeront que de part leurs abscisses x_g et x_d.

1.4 Configuration parallèle

En configuration parallèle, les droites épipolaires des images gauche et droite, sont parallèles, horizontales et confondues. Dans cette situation, les épipoles se retrouvent

alors rejetés à l'infini. La recherche d'un couple de points homologues s'effectue, dans cette configuration, le long des différentes lignes horizontales des deux images. Deux approches permettent d'obtenir cette configuration intéressante.

La première approche consiste à positionner le stéréoscope dans une configuration toute particulière : c'est le *calibrage fort*. Celle-ci se définit par les caractéristiques suivantes :

- les axes optiques des caméras sont parallèles ;
- les distances focales des objectifs sont identiques ;
- les capteurs sont coplanaires ;
- la droite reliant les centres optiques C_g et C_d est parallèle au plan des capteurs.

Notons que si les caméras sont situées à la même hauteur, les points homologues M_g et M_d se trouvent alors obligatoirement sur la ligne de même ordonnée y dans les deux images (figure 1.3). Pour cette approche, notons que les déformations dues aux objectifs ne sont pas prises en compte.

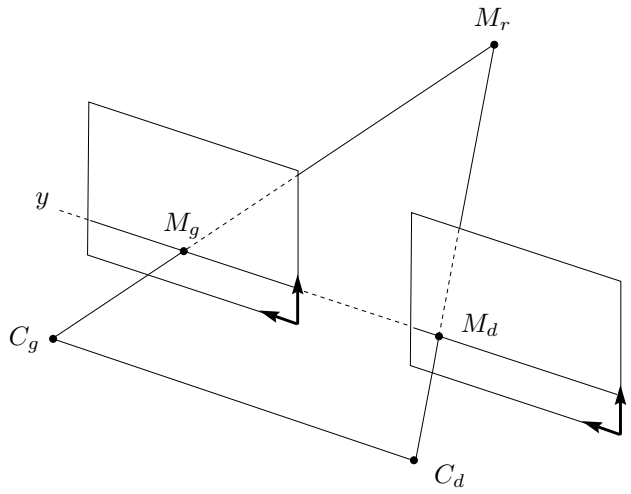


Fig. 1.3 : Configuration parallèle d'un stéréoscope binoculaire.

La seconde approche est appelée technique de *rectification* et vise aussi à confondre les droites épipolaires avec les lignes des images. Le système est d'abord calibré *faiblement* de manière à ce que les caméras observent la même scène. Les images sont ensuite re-projetées dans un plan parallèle à la ligne reliant les centres optiques, de telle manière que les droites épipolaires soient parallèles et horizontales [7] [8].

Apparier deux points M_g et M_d permet de déterminer les coordonnées 3D du point correspondant M_r dans la scène observée par simple triangulation géométrique (comme dans le cas d'une configuration quelconque).

1.5 La triangulation géométrique

La triangulation géométrique permet de déterminer les coordonnées tridimensionnelles d'un point de la scène à partir de ses projections M_g et M_d sur les capteurs gauche et droit. Ces projections ont respectivement les coordonnées (x_g, y_g) et (x_d, y_d). La figure 1.4 représente un système stéréoscopique calibré fortement. Dans ce cas, les droites épipolaires sont confondues, autrement dit, les ordonnées y_g et y_d sont identiques. f représente la distance focale des objectifs et E l'entraxe du stéréoscope (c'est-à-dire la distance entre ses deux axes optiques parallèles).

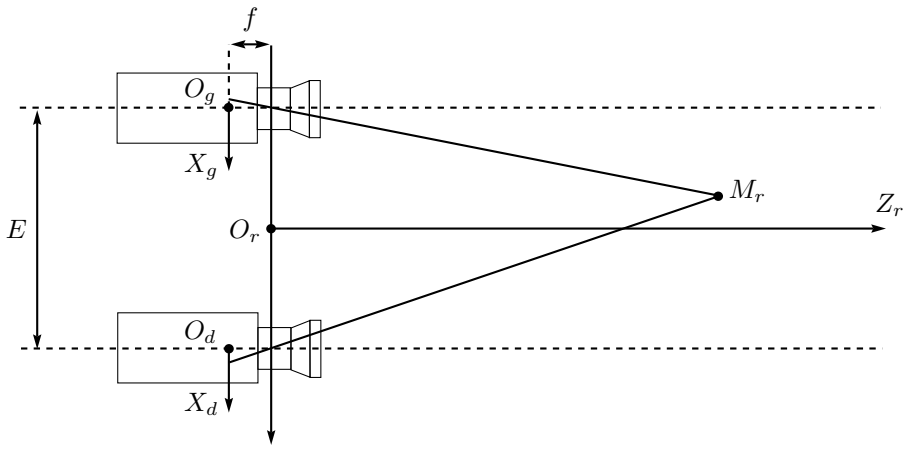


Fig. 1.4 : Triangulation géométrique dans le plan $(O_r, (X_r, Z_r))$.

Les positions x_g et x_d des points image se déduisent de la position du point de la scène $M_r = (X_r, Y_r, Z_r)$ par les relations détaillées dans [6] :

$$x_g = f \cdot (X_r + \frac{E}{2}) \cdot \frac{1}{Z_r} \ . \tag{1.1}$$

et

$$x_d = f \cdot (X_r - \frac{E}{2}) \cdot \frac{1}{Z_r} \ . \tag{1.2}$$

En configuration parallèle, la disparité d_{gd} correspond à la différence des abscisses des points M_g et M_d (l'image gauche étant l'image de référence) :

$$d_{gd} = x_g - x_d. \tag{1.3}$$

En remplaçant dans l'équation (1.3) les valeurs de x_g et x_d obtenues en (1.1) et (1.2), on obtient :

$$d_{gd} = \frac{E \cdot f}{Z_r} \ . \tag{1.4}$$

Quelle que soit l'abscisse du point M_r, on peut calculer sa coordonnée Z_r caractérisant la profondeur par :

$$Z_r = \frac{E \cdot f}{d_{gd}} \ . \tag{1.5}$$

De même, les deux autres coordonnées du point M_r dans le repère de référence (O_r, X_r, Y_r, Z_r) peuvent être déterminées par les équations suivantes :

$$X_r = \frac{x_g \cdot E}{d_{gd}} \ , \tag{1.6}$$

$$Y_r = \frac{y_g \cdot E}{d_{gd}} \ . \tag{1.7}$$

Ainsi, à distance focale constante, un entraxe faible induira une précision moindre des valeurs des disparités. A contrario, pour une estimation précise de la profondeur, il est préférable d'utiliser un entraxe important. Cependant, dans ce cas, le nombre de valeurs de disparité à tester devient important ce qui augmente considérablement les possibilités de faux appariements [9]. Aussi, avec un entraxe important, la distance minimale de vision du stéréoscope augmente puisque le champ commun de vision entre les deux caméras se retrouve éloigné. La mise en œuvre du stéréoscope et la détermination de ses paramètres seront donc directement liées à l'application visée et constitue donc un compromis.

Pour ce début de chapitre, la dénomination « point » était employée. Dans la suite de ce mémoire, nous détaillons l'étape de mise en correspondance de « pixels ». Ceci est lié au processus de discrétisation de l'image qui définit des pixels ayant des attributs plus nombreux que pour le « point » (i.e. plutôt employé au sens géométrique dans l'espace continu). Ces différents attributs, tels la position, le niveau de gris, la couleur, permettent la mise en correspondance. La discrétisation du pixel entraîne une discrétisation des valeurs de distance associées aux valeurs entières de disparité : ainsi, plusieurs points peuvent être projetés sur le même pixel dans l'image. L'utilisation de mesures au sous-pixel présentées en fin de ce chapitre s'avère alors indispensable.

Comme décrit précédemment, la mise en correspondance de points homologues détermine la disparité au pixel considéré qui, par triangulation, permet le calcul de la profondeur du point de la scène. Dans la suite de ce chapitre, l'étape cruciale de mise en correspondance est détaillée.

1.6 Les bases de la mise en correspondance

La mise en correspondance est une étape critique du processus de stéréovision. L'objectif est d'apparier chaque pixel de l'image de référence avec son homologue dans l'autre

image. Cependant, cela n'est pas toujours possible. En effet, contrairement aux techniques de mesures de distances basées sur des capteurs actifs, la stéréovision passive ne permet pas toujours la mesure, ce qui rend le problème très complexe.

1.6.1 Problèmes liés à la stéréovision passive

Le principe de la stéréovision passive est de constituer des couples de pixels homologues en comparant les pixels extraits de deux images. L'hypothèse des surfaces *Lambertiennes* est souvent utilisée : celle-ci stipule que l'intensité des projections, dans les deux images, d'un point 3D doit être indépendante du point de vue, autrement dit, deux pixels homologues ont la même intensité lumineuse. Cependant, cette contrainte n'est plus valable lorsque les deux caméras possèdent des caractéristiques différentes (gain, offset) ou pour des variations de l'illuminant.

D'autres cas de figures entraînent des difficultés lors de la phase d'appariement. Un pixel de l'image de référence n'a pas toujours d'homologue dans l'autre image, on parle alors d'*occultations*. Mettre en correspondance un pixel qui n'a pas d'homologue s'avère être, par définition, une tâche impossible. Certains auteurs proposent de gérer ce problème dans leurs algorithmes ([10] [11] [12]).

Aussi, certaines zones de l'image ne contiennent que très peu d'information. Certains objets peu texturés apparaissent comme très bruités dans ces zones. Dans ce cas, mettre en correspondance deux pixels homologues s'avère être une tâche très ardue. En pratique, il n'y a pas toujours de solution au problème de mise en correspondance.

Cependant, les méthodes de mise en correspondance stéréoscopique sont souvent utilisées pour des applications de reconstruction 3D pour lesquelles la densité et le rendu sont très importants. Seulement, comme nous l'avons écrit, la stéréovision passive ne permet pas toujours la mesure.

La mise en correspondance constitue donc le processus qui va permettre l'appariement de deux pixels homologues dans les images gauche et droite. Si pour un pixel de l'image de référence, chaque pixel de l'autre image est sollicité dans la recherche d'homologues, cette étape peut devenir très gourmande en calculs ; l'application de *contraintes* de mise en correspondance permet de réduire l'espace de recherche mais aussi de lever des ambiguïtés [13]. Les principales contraintes de mise en correspondance sont présentées dans la section suivante.

1.6.2 Les contraintes de mise en correspondance

Les contraintes de mise en correspondance peuvent être classées en plusieurs catégories. Les contraintes d'unicité, épipolaire, d'ordre et de position sont liées à la géométrie du

capteur ou encore de la scène observée. Les contraintes de continuité de la disparité ou de disparité maximale se rapprochent plus de « techniques » pour la mise en correspondance.

1.6.2.1 La contrainte épipolaire

La contrainte épipolaire, liée à la géométrie du capteur, constitue la contrainte principale en stéréovision. Elle est directement liée à la géométrie épipolaire présentée précédemment. L'espace de recherche est ainsi limité aux droites épipolaires puisque chaque pixel situé sur une droite épipolaire dans l'image de référence trouve son homologue sur la droite épipolaire conjuguée dans l'autre image.

En se référant à la représentation graphique introduite par Otha et Kanade dans [14], la *grille de correspondance* est présentée figure 1.5. Celle-ci représente les niveaux de gris respectifs des pixels $I_g(i)$ et $I_d(j)$ (tels que $i \in [0..N]$ et $j \in [0..N]$) le long des droites épipolaires dans l'image gauche et l'image droite pour la configuration parallèle ; les lignes verticales et horizontales indiquent respectivement les positions des pixels dans les images gauche et droite. L'intersection de ces lignes, appelée *nœud*, est définie comme un couple de pixels $\{I_g(i), I_d(j)\}$. Chaque *nœud à tester* (disque blanc) représente un couple potentiel de pixels homologues alors qu'un *nœud valide* (disque noir) représente un couple de pixels mis en correspondance.

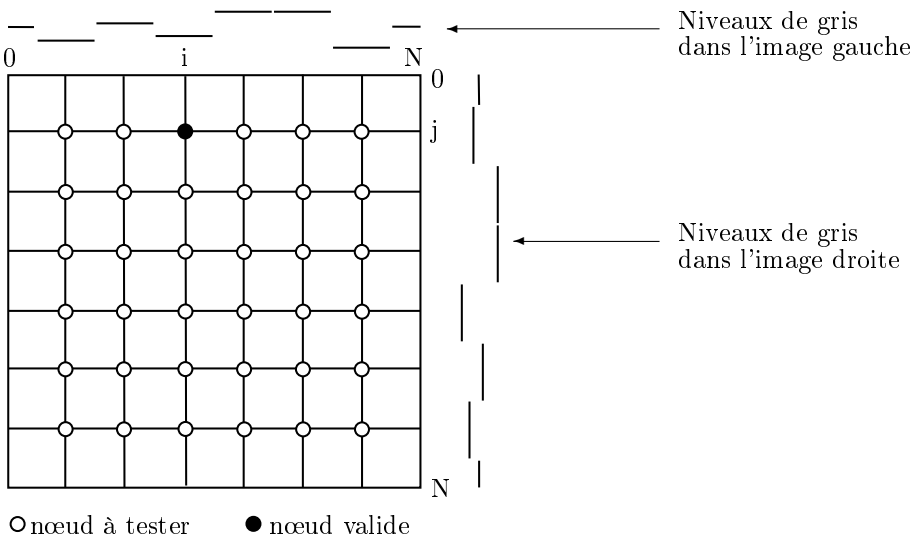


Fig. 1.5 : Grille de mise en correspondance basée sur la contrainte épipolaire.

Si les droites épipolaires des images gauche et droite sont toutes deux composées de 500 pixels (soit $i \in [0..499]$ et $j \in [0..499]$), 250 000 nœuds seront donc à examiner. Le nombre de nœuds à tester peut être considérablement diminué par l'application de contraintes de mise en correspondance supplémentaires, liées à la géométrie de la scène.

1.6.2.2 La contrainte d'unicité

La contrainte d'unicité impose qu'un pixel de l'image de référence ne peut avoir qu'un unique homologue dans l'autre image. Cette contrainte est illustrée figure 1.6. Sur cette figure, pour laquelle l'image gauche est prise comme référence, on observe que la composition d'un couple de pixels homologues (ou nœud valide) entraîne l'élimination des autres nœuds à tester pour le pixel considéré dans l'image gauche. Graphiquement, les disques blancs correspondant aux nœuds à tester sont supprimés et la droite verticale du pixel apparié dans l'image gauche est mise en pointillés.

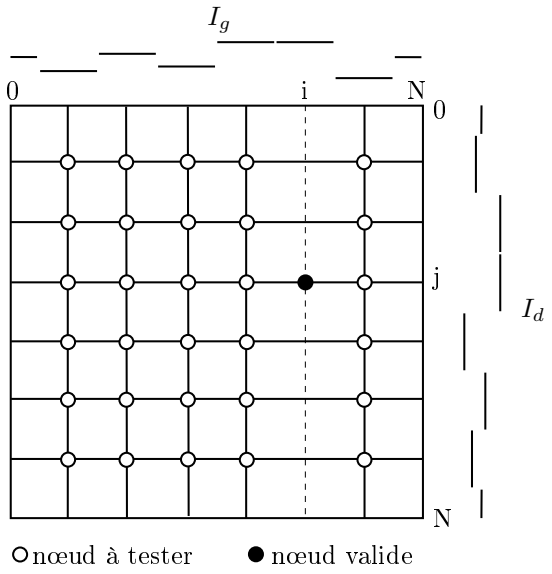


Fig. 1.6 : Application de la contrainte d'unicité, l'image gauche étant l'image de référence.

1.6.2.3 La contrainte d'ordre

La contrainte d'ordre (figure 1.7) implique la conservation de l'ordre des pixels homologues le long des deux droites épipolaires. La contrainte est respectée si la position entre deux pixels d'une droite épipolaire est identique à celle de leurs homologues sur la droite épipolaire conjuguée.

Les objets transparents violent cette contrainte. De cette manière, la plupart des algorithmes de mise en correspondance supposent que la scène observée est dépourvue d'objets transparents. De la même façon, lorsque l'entraxe entre les deux caméras est important et que la scène est constituée de deux objets situés l'un derrière l'autre, l'ordre entre ces deux objets peut ne pas être conservé.

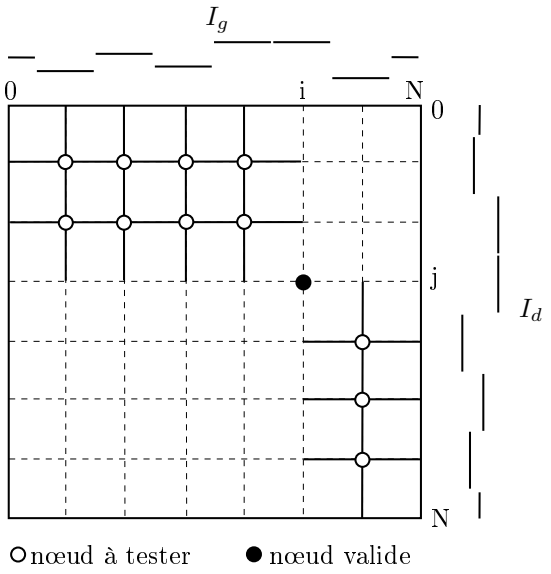


Fig. 1.7 : Application de la contrainte d'ordre.

1.6.2.4 La contrainte de position

La contrainte de position impose que les abscisses x_g et x_d des pixels homologues dans les images gauche et droite vérifient la relation $x_g > x_d$. Cette contrainte est présentée figure 1.8 pour laquelle seuls les nœuds respectant cette contrainte sont conservés.

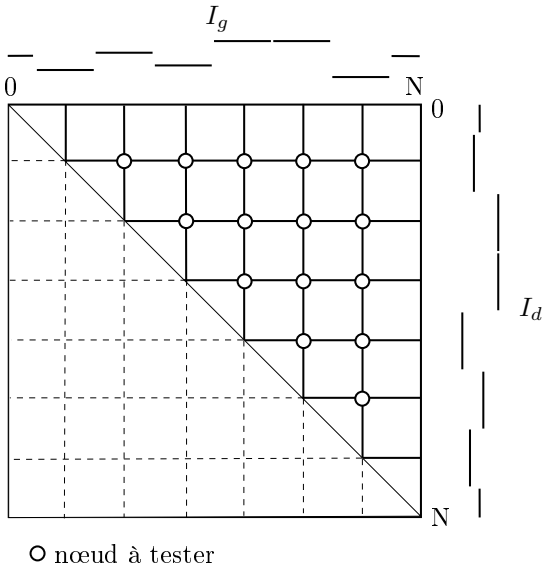


Fig. 1.8 : Application de la contrainte de position.

1.6.2.5 La contrainte de continuité de la disparité

La contrainte de continuité de la disparité considère que l'évolution des valeurs de la disparité doit être continue le long d'un même contour sous l'hypothèse qu'il appartient à un seul et même objet. De nombreux auteurs utilisent ainsi cette contrainte et prennent en compte une partie de ce contour pour le calcul de la disparité au pixel courant.

1.6.2.6 La contrainte de disparité maximale

En fonction de la configuration du système stéréoscopique, on détermine la distance minimale pouvant être détectée et par conséquent le décalage maximal d_{max} pouvant séparer deux pixels homologues. Cela permet de limiter l'intervalle des disparités à traiter.

1.6.3 Primitives à apparier

La mise en correspondance permet l'appariement de pixels homologues sur la base d'indices visuels les caractérisant. Ces indices visuels sont appelés *primitives*. Le processus d'appariement consiste donc à mettre en correspondance ces primitives qui peuvent être de différentes natures.

Les primitives les plus simples sont les pixels de l'image pour lesquels le niveau de gris ou la couleur sont étudiés. Cette primitive permet d'obtenir des cartes des disparités dites *denses* c'est-à-dire pour lesquelles une valeur de disparité est affectée en chaque pixel de l'image. Dans ce cas, les temps de calculs sont importants.

D'autres primitives dont l'utilisation s'avère moins coûteuse en temps de calcul ont également été proposées. Celles-ci se basent sur l'utilisation de points caractéristiques de la scène observée.

Certaines méthodes utilisent des *points d'intérêt* : ceux-ci sont définis par des propriétés les rendant plus discriminants par rapport à leurs voisins. L'un des détecteurs de points d'intérêt des plus populaires est celui de Harris [15] ; Ishikawa et Geiger utilisent quant à eux un détecteur de coins [16].

Les primitives de type *points de contour* font partie des plus populaires puisqu'utilisées dans de nombreux travaux. Ces points de contour peuvent être extraits par l'opérateur de Canny [17].

D'autres primitives, comme les segments de droite, ont été utilisées [18].

Ces méthodes, fondées sur l'utilisation de points caractéristiques de la scène, mènent à la constitution de cartes éparses des disparités. Dans la suite de ce chapitre, différentes techniques pour la mise en correspondance stéréoscopique sont présentées.

1.7 Techniques de mise en correspondance stéréoscopique

De manière générale, les algorithmes de mise en correspondance se décomposent en deux étapes. Dans un premier temps, une *fonction de coût* évalue, pour un pixel donné, toutes ou un sous-ensemble des solutions possibles de l'espace des disparités.

Dans un second temps, l'extremum de cette fonction de coût est recherché grâce à une *technique d'optimisation* de manière à fournir la valeur de disparité à associer au pixel considéré.

Les techniques de mise en correspondance stéréoscopique peuvent être classées en deux principales catégories. Tout d'abord, les méthodes dites *globales* qui appliquent les contraintes stéréoscopiques sur la globalité de l'image. L'autre approche dite *locale* applique ces contraintes sur un ensemble réduit de pixels situés dans le voisinage du pixel à apparier. Les sections suivantes présentent les principales méthodes de mise en correspondance utilisées en stéréovision ; le lecteur pourra se référer à [19] et [20] pour deux études bibliographiques complémentaires.

1.7.1 Les méthodes locales

Les méthodes locales, aussi appelées méthodes par corrélation, permettent de mesurer la ressemblance entre deux ensembles de pixels (dans les images gauche et droite) par des mesures de corrélation. De manière plus générique, nous emploierons les termes de *mesures* de *similarité* (ou de *dissimilarité*).

Ces méthodes locales de mise en correspondance par corrélation calculent des cartes éparses des disparités, si les primitives employées sont des points caractéristiques de l'image, mais aussi de cartes denses. Ces dernières sont obtenues en effectuant l'analyse des niveaux de gris (ou couleur) des pixels contenus dans une *zone d'agrégation* centrée ou non sur chaque pixel à apparier. L'objectif est de mesurer la similarité (ou la dissimilarité) des pixels contenus dans deux zones de forme identique extraites des images gauche et droite. Selon les caractéristiques de la mesure de similarité utilisée, une ressemblance forte entre deux zones situées autour de deux pixels caractérisera un couple potentiel de pixels homologues. C'est l'obtention de cartes denses des disparités qui nous intéresse dans ce travail.

Ces méthodes locales se décomposent généralement en trois étapes qui sont :

- le choix de la mesure de similarité entre pixels ;
- la détermination de la zone d'agrégation des pixels considérés ;
- la méthode de recherche exhaustive qui affectera la valeur finale de disparité au pixel à apparier.

1.7.1.1 La méthode de recherche exhaustive

La méthode de recherche exhaustive va permettre la détermination de la valeur de disparité à affecter au pixel à apparier. Comme elle présente le principe général des méthodes locales, nous détaillons tout d'abord cette étape.

Cette méthode repose sur la constatation que deux régions des images, situées autour de deux pixels homologues, sont localement semblables. La zone d'agrégation, située autour du pixel à apparier dans l'image de référence, reste fixe. Une zone d'agrégation identique est définie dans l'autre image à la même position et sera décalée le long de la droite épipolaire (ici une ligne de l'image). Pour chaque valeur de décalage dans l'intervalle des disparités, les scores de similarité sont calculés. Le calcul de la similarité pour les pixels de la zone d'agrégation constitue le *coût de correspondance* $E_{correspondance}$. La valeur de décalage qui maximise (ou minimise) le coût de correspondance sera considérée comme la valeur de disparité à affecter au pixel à apparier. Cette méthode de recherche exhaustive constitue la méthode d'optimisation des méthodes locales. Elle est appelée méthode « Winner-Take-All[1] » et notée WTA.

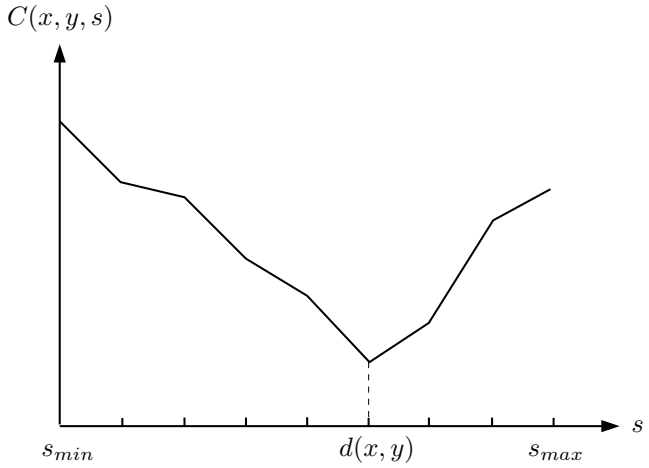


Fig. 1.9 : Exemple d'une courbe de dissimilarité.

Ainsi, le coût de correspondance $E_{correspondance}$ peut être calculé de différentes manières en fonction de la mesure de similarité et de la zone d'agrégation utilisées ; ceci est détaillé dans la suite de ce chapitre.

1.7.1.2 Les mesures de similarité

Plusieurs mesures de similarité sont disponibles dans la littérature ([21] et [22]) ; elles fournissent des résultats de qualité inégale et sont, en termes de traitement, de complexité variable. Ces expressions s'articulent pour la plupart autour de l'étude de l'intensité lumineuse locale dans une zone d'agrégation donnée. Les mesures de similarité peuvent être classées en trois catégories : différence d'intensité lumineuse, corrélation et rang. Dans la

[1] le gagnant prend tout

suite de ce chapitre, une présentation non exhaustive de différentes mesures de similarité est réalisée. Les équations sont exprimées dans le cadre de la configuration parallèle introduite précédemment.

Les mesures de différences d'intensité lumineuse :

Les mesures de différences d'intensité lumineuse les plus classiques sont le *SAD* (pour *Sum of Absolute Differences*) et le *SSD* (pour *Sum of Squared Differences*). Apparier, sur la base du critère *SAD*, revient à minimiser la somme des différences absolues des intensités contenues dans les zones d'agrégation positionnées sur les images gauche et droite. Ces deux mesures de similarité peuvent être formalisées comme suit :

$$SAD(x,y,s) = \sum_{i=-w}^{i=w} \sum_{j=-w}^{j=w} |I_g(x+i,y+j) - I_d(x-s+i,y+j)| \,, \tag{1.8}$$

$$SSD(x,y,s) = \sum_{i=-w}^{i=w} \sum_{j=-w}^{j=w} (I_g(x+i,y+j) - I_d(x-s+i,y+j))^2 \,, \tag{1.9}$$

où I_g et I_d sont respectivement les niveaux de gris des pixels dans les images gauche et droite ; i et j représentent les translations horizontales et verticales dans la zone d'agrégation de dimension $(2w+1)^2$ autour du pixel à apparier de coordonnées (x,y) ; s est le décalage de la zone d'agrégation le long de le droite épipolaire (ici une ligne de l'image).

Il existe différentes variantes de ces deux mesures de base : centrée, normalisée et centrée/normalisée. La mesure centrée du *SSD*, notée *ZSSD* pour *Zero-mean Sum of Squared Differences*, permet de mettre en correspondance des zones de l'image dont l'illumination est légèrement différente :

$$ZSSD(x,y,s) = \sum_{ij} ((I_g(x+i,y+j) - \overline{I_g}(x,y)) - (I_d(x-s+i,y+j) - \overline{I_d}(x,y)))^2 \,. \tag{1.10}$$

Ici, $\overline{I_g}$ et $\overline{I_d}$ sont les moyennes des intensités sur chaque fenêtre de similarité centrée en (x,y).

La mesure normalisée du *SSD*, notée *NSSD* pour *Normalised Sum of Squared Differences*, qui normalise par la variance locale des intensités, est définie par l'équation suivante :

$$NSSD(x,y,s) = \frac{\sum_{ij}(I_g(x+i,y+j) - I_d(x-s+i,y+j))^2}{\sqrt{(\sum_{ij} I_g(x+i,y+j))^2} \times \sqrt{(\sum_{ij} I_d(x-s+i,y+j))^2}} \,. \tag{1.11}$$

La mesure centrée/normalisée du *SSD*, notée *ZNSSD* pour *Zero-mean Normalised Sum of Squared Differences*, allie le centrage à la normalisation. De cette manière, on s'affranchit en partie des différences de gain et d'offset des deux caméras.

Les mesures de corrélation :

Les mesures de corrélation exploitent la corrélation croisée notée CC (pour *Cross-Correlation*) et calculent le produit des intensités lumineuses comme suit :

$$CC(x, y, s) = \sum_{ij} I_g(x + i, y + j) \times I_d(x - s + i, y + j) \ . \tag{1.12}$$

De la même manière que pour la mesure de similarité SSD, la mesure CC peut être centrée (NCC) et centrée/normalisée ($ZNCC$). La mesure NCC est assez gourmande en calculs : il s'agit de la somme des produits entre les termes correspondants aux zones d'agrégation, normalisée par le produit des moyennes quadratiques calculé pour chacune de ces zones d'agrégation. De même, le $ZNCC$ est identique au NCC sauf qu'il s'agit ici de la corrélation croisée entre les écarts et la moyenne.

Les mesures de rang :

Dans [23], Zabih et Woodfill ont proposé une méthode de mise en correspondance basée sur une transformation non-paramétrique de l'image afin de réduire la sensibilité aux imperfections de luminosité. Cette section, intitulée « mesures de rang » par abus de langage, présente les transformations appelées *transformée de Rank* et *transformée de Census*.

Rank définit, dans une fenêtre donnée, le nombre de pixels ayant une valeur de niveau de gris inférieure à celle du pixel à apparier et l'affecte à ce dernier. Le tableau 1.1 présente à gauche les niveaux de gris d'une fenêtre de taille 5×5 et à droite la transformée *Rank* pour le pixel central ; on constate que 11 pixels de la fenêtre ont un niveau de gris inférieur à celui du pixel à apparier. Ensuite, une mesure de similarité de type SSD est utilisée pour déterminer les appariements, les images de rang remplaçant les images originales.

12	14	26	32	31
13	16	24	30	33
11	15	***23***	31	33
10	14	22	29	32
12	17	25	34	36

...	...	...	...	...
...	...	...	...	...
...	...	***11***	...	...
...	...	...	...	...
...	...	...	...	...

Tab. 1.1 : Illustration de la Transformée *Rank*.

La *transformée de Census* est, quant à elle, basée sur la structure spatiale locale des niveaux de gris dans le voisinage. Elle conserve la distribution des niveaux de gris en les codant en une chaîne de bits. Le tableau 1.2 présente à gauche les niveaux de gris d'une fenêtre de taille 5×5 et au centre la transformée *Census* pour le pixel central. Chaque pixel de niveau de gris supérieur à celui du pixel central est codé par la valeur 1 ; si la valeur est inférieure, la valeur 0 sera affectée au pixel. Un mot binaire (à droite) est

ainsi déterminé pour caractériser les pixels voisins du pixel à mettre en correspondance. L'appariement est effectué en calculant la distance de Hamming, autrement dit le nombre de bits qui diffèrent entre les deux mots binaires des pixels à apparier. L'appariement sera effectué pour le couple de pixels minimisant cette distance de Hamming.

12	14	26	32	31
13	16	24	30	33
11	15	23	31	33
10	14	22	29	32
12	17	25	34	36

0	0	1	1	1
0	0	1	1	1
0	0	×	1	1
0	0	0	1	1
0	0	1	1	1

00111001110011000110011 1

Tab. 1.2 : Illustration de la Transformée *Census*.

Pour conclure, on constate dans la littérature que les mesures de similarité sont nombreuses et fournissent des résultats assez différents. Dans [24], Hirschmüller et Scharstein ont comparé plusieurs mesures de similarité associées à différentes méthodes d'optimisation (globales, locales) ; la différence des résultats obtenus est liée aux différentes associations de la mesure de similarité à la méthode d'optimisation utilisée. Cependant, le choix de la zone d'agrégation utilisée compte pour beaucoup dans le processus de mise en correspondance.

1.7.1.3 La zone d'agrégation

Pour les méthodes par corrélation, on définit deux types de zones d'agrégation. Le premier consiste à appliquer la mesure de similarité sur l'ensemble des pixels contenus dans une fenêtre donnée, appelé *fenêtre de similarité*. Le deuxième consiste à n'utiliser que les pixels contenus dans une fenêtre donnée répondant à un certain critère, appelé *voisinage de similarité* (dans ce cas, tous les pixels de la fenêtre ne sont pas utilisés).

La *fenêtre de similarité* est définie comme une fenêtre bidimensionnelle (2D), de forme rectangulaire, contenant le pixel à apparier. Les performances des méthodes de stéréovision par corrélation sont particulièrement sensibles au choix de ces fenêtres de similarité [25]. Trop petites, elles ne contiennent pas suffisamment d'informations. Trop grandes, elles peuvent contenir des éléments correspondant à des points de la scène situés à des profondeurs différentes. De nombreuses méthodes d'appariement de fenêtres ont été décrites dans la littérature. Celles-ci se fondent exclusivement sur l'utilisation de fenêtres 2D et font l'hypothèse que tous les pixels contenus dans cette fenêtre ont la même disparité.

Dans [26], les auteurs ont déterminé le nombre de valeurs de disparité différentes présentes dans une fenêtre de similarité en fonction de la taille de cette fenêtre et ce, pour une image donnée. Naturellement, plus la taille de la fenêtre est importante et plus le nombre de disparités différentes contenues dans la fenêtre augmente. Par exemple, pour

une fenêtre de taille 21×21, plus de 50% des fenêtres contiennent au moins deux valeurs de disparité différentes. De nombreux auteurs ont tenté de définir la zone d'agrégation optimale comme nous allons le présenter dans la section suivante.

1.7.1.4 Présentation de quelques méthodes locales

Les méthodes locales peuvent être très efficaces mais elles restent cependant sensibles au contenu de l'image comme pour le cas des zones peu texturées. Comme Nishihara le précise dans [27], en chaque pixel (x,y) d'une image, la qualité de la mise en correspondance dépendra de l'adéquation entre la taille du voisinage et de l'information présente dans cette région de l'image. Certaines méthodes proposent de trouver, en chaque pixel des images gauche et droite, la taille et la géométrie optimales de la zone d'agrégation afin de lever les ambiguïtés qui apparaissent lorsqu'une fenêtre de taille fixe est utilisée. Plusieurs d'entre-elles sont introduites ci-après.

Méthode à taille de fenêtre variable :

Kanade et Okutomi ont proposé une méthode à voisinage adaptatif [25], pour laquelle la taille est modifiée en fonction des variations locales de luminosité et des disparités locales estimées.

Méthode à position variable du pixel à apparier :

Certains auteurs calculent les scores de similarité pour plusieurs fenêtres dont le pixel à apparier est situé en différentes positions. Fusiello *et al.* ont présenté leur algorithme appelé « Symmetric Multi-Window » (SMW) dans [28]. Ils calculent les indices de dissimilarité sur la base du SSD pour neuf fenêtres différentes (figure 1.10) : pour chacune d'elles, le pixel à apparier (en noir sur la figure) se situe en une position différente. La fenêtre qui minimise le score de dissimilarité est sélectionnée.

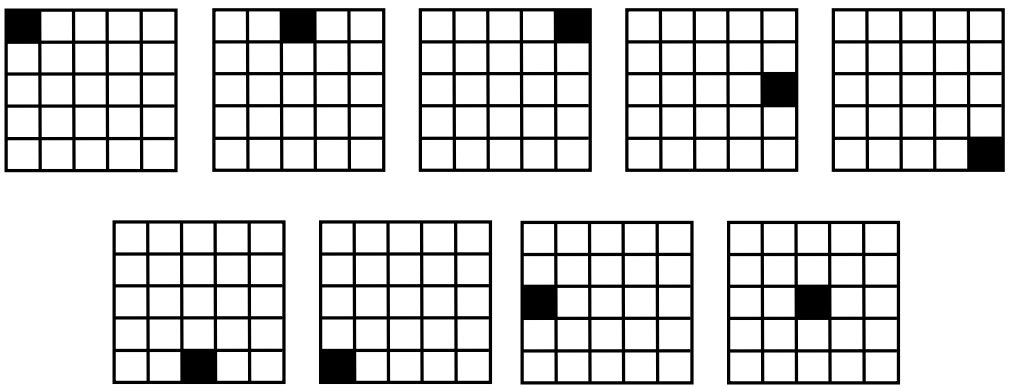


Fig. 1.10 : Illustration des fenêtres utilisées par Fusiello *et al.* dans [28].

Méthode multi-fenêtres :

Comme précédemment, plusieurs fenêtres sont utilisées. La différence réside dans le fait qu'un score de similarité est calculé en combinant plusieurs fenêtres. Par exemple dans [29], Hirschmüller combine les scores de similarité calculés pour une fenêtre C_0 centrée sur le pixel à apparier, avec les indices calculés pour plusieurs autres fenêtres appelées « fenêtres support ». Comme présenté figure 1.11, l'auteur propose trois configurations différentes de fenêtres. L'algorithme exploite l'une de ces trois configurations. Pour la configuration choisie, plusieurs combinaisons de fenêtres sont possibles : la fenêtre centrale C_0 (grisée) associée à plusieurs autres « fenêtres support » (C_1, C_2,..). En chaque décalage, la combinaison qui minimise le score de similarité est retenue.

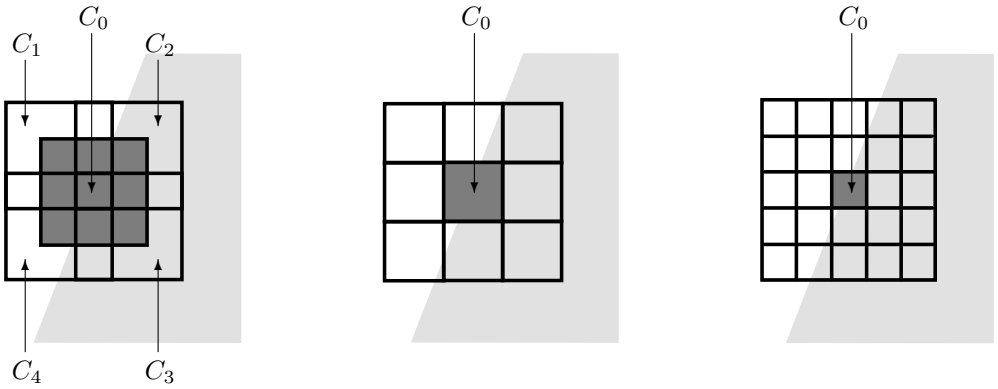


Fig. 1.11 : Illustration des fenêtres utilisées par Hirschmüller dans [29] (le pixel noir représente la position du pixel à apparier dans la fenêtre).

Méthode à voisinage variable :

La majorité de ces méthodes combine une étape de segmentation et une étape classique de stéréovision par corrélation ; les deux étapes peuvent être réalisées simultanément ou séquentiellement. La segmentation consiste à extraire certaines primitives selon un critère de similarité d'un ou plusieurs attributs. Par exemple, il s'agit de constituer des régions homogènes et cohérentes, si possible, liées à un même objet ou à une partie d'un même objet de la scène.

Pérez *et al.* utilisent un voisinage adaptatif dont la forme variable est déterminée par l'analyse de la similarité entre pixels. Dans [30], la segmentation des images exploite l'hypothèse que les pixels connexes de même teinte appartiennent à un même objet. Le voisinage utilisé lors du calcul du score de similarité est composé des pixels voisins de teinte « semblable ». Grâce à cette méthode, la carte des disparités est améliorée puisque seuls les pixels de la fenêtre appartenant à un même objet (au sens de l'hypothèse retenue) interviennent dans le calcul de similarité.

D'autres travaux, tels que [31] et [32], ont été menés en utilisant d'autres attributs que la teinte. Cependant, la segmentation ne fournit pas nécessairement des primitives de

mêmes caractéristiques géométriques et morphologiques pour les images gauche et droite. De plus, le nombre des primitives extraites peut être différent.

Pour conclure, les différentes méthodes locales présentées modifient la taille et la forme de la zone d'agrégation selon certains critères dans le but d'obtenir le meilleur appariement possible notamment dans des zones de l'image peu texturées ou contenant plusieurs valeurs de disparité.

1.7.2 Les méthodes globales

Les méthodes globales font généralement intervenir l'ensemble des pixels des images dans le processus d'optimisation [33]. D'autres auteurs proposent de réduire l'espace de recherche à la droite épipolaire [34] ou plusieurs droites épipolaires adjacentes [35]. La suite de cette section présente le principe des méthodes globales selon deux axes : l'élaboration de la fonction de coût global et la méthode d'optimisation.

1.7.2.1 Fonction de coût global

La fonction de coût global E_{global} permet de mesurer la ressemblance entre deux primitives. Cette fonction se décompose en deux fonctions de coût [36] :

- le coût de correspondance $E_{correspondance}$;
- le coût des contraintes $E_{contraintes}$.

Le coût de correspondance s'exprime, comme pour les méthodes locales, en fonction de la mesure de similarité et de la zone d'agrégation retenue. Généralement, les méthodes globales utilisent des zones d'agrégation limitées au seul pixel courant. Le principe reste le même que pour les méthodes locales.

Le coût des contraintes est exprimé en fonction des contraintes de mise en correspondance exposées précédemment. Dans [37], la fonction coût intègre les contraintes de signe de gradient, d'unicité, d'ordre et de continuité. Celle-ci passe par un extremum lorsque toutes les contraintes sont satisfaites : cet extremum représente le bon appariement.

Le travail sur la fonction de coût est primordial puisque c'est elle qui servira de base au processus d'optimisation. Si celle-ci n'est pas construite efficacement, les résultats obtenus ne seront pas optimaux et ce, quelle que soit la méthode d'optimisation mise en œuvre.

Devant la difficulté à résoudre ces problèmes d'optimisation qui comportent un grand nombre de solutions sous-optimales, il est courant d'avoir recours à des métaheuristiques. Celles-ci se comportent comme des algorithmes de recherche, tentant d'apprendre les caractéristiques d'un problème afin d'en trouver une approximation de la meilleure solution. Parmi ces méthodes d'optimisation, nous pouvons citer les algorithmes génétiques, les ré-

seaux de neurones, le recuit simulé, etc. Nous en décrivons quelques unes dans la suite de ce chapitre.

1.7.2.2 Les méthodes d'optimisation

Les méthodes d'optimisation peuvent être de nature non-stochastique ou stochastique. Dans cette partie, ces deux types de méthodes vont être présentés : la programmation dynamique et les réseaux de neurones en tant que méthodes non stochastiques et les algorithmes génétiques et le recuit simulé en tant que méthodes stochastiques.

Programmation dynamique :

Les méthodes basées sur la technique de programmation dynamique modélisent le problème de mise en correspondance stéréoscopique comme la détermination d'un parcours optimal reliant le premier pixel au dernier pixel de la droite épipolaire [38]. La méthode consiste à déterminer la meilleure trajectoire parmi les trajectoires possibles en passant par différents nœuds de la grille de mise en correspondance. Un coût de correspondance est assigné à chaque nœud ; le coût global d'une trajectoire équivaut à la somme des coûts individuels de correspondance des différents nœuds disposés sur cette trajectoire. La trajectoire minimisant la fonction de coût global est retenue [39].

Les contraintes d'unicité et d'ordre peuvent être utilisées dans le but de limiter l'espace de recherche et déterminer le chemin optimal [20].

Un des principaux inconvénients de cette méthode est la propagation à toute la ligne d'une erreur locale pouvant ainsi entraîner un grand nombre de faux appariements [19].

Les réseaux de neurones :

Un réseau de neurones est un modèle de calcul dont la conception est schématiquement inspirée du fonctionnement des neurones biologiques. Comme pour toutes les méthodes d'optimisation, une fonction coût, modélisant le problème, est définie : celle-ci est aussi appelée fonction d'énergie du réseau. Le réseau va ensuite évoluer vers un état stable qui correspond à la valeur minimisant la fonction d'énergie [40].

Dans [37], les auteurs utilisent un réseau de Hopfield dans le but d'apparier des pixels de contour. Le réseau de Hopfield est un réseau de neurones entièrement interconnecté. Cette technique présente deux inconvénients majeurs ; la minimisation de la fonction d'énergie est très gourmande en temps de calcul et ce type de fonction converge souvent vers le premier minimum local rencontré, ce qui est gênant lorsque la fonction coût présente plusieurs minima.

Les algorithmes génétiques :

Inspiré des mécanismes de la sélection naturelle, les algorithmes génétiques permettent la résolution de problèmes d'optimisation complexes par une approche stochastique. Chaque

solution potentielle du problème est représentée par un chromosome. Le contenu de la population initiale des chromosomes est généré aléatoirement. Ensuite, ces chromosomes évoluent vers la solution du problème par le biais d'une opération de sélection d'individus sur une population affectée par des mutations et des croisements. Cette méthode est itérative. La solution est représentée par le chromosome dont la valeur minimise la fonction d'évaluation.

Dans [41], les auteurs utilisent des mesures qui caractérisent la similarité entre pixels sur une fenêtre donnée et calculent plusieurs cartes des disparités pour différentes vues de la scène observée. Les algorithmes génétiques sont employés à la recherche de la carte des disparités qui minimise la fonction d'évaluation sur la base des contraintes de similarité et de continuité.

La méthode d'optimisation par algorithmes génétiques comporte cependant certaines limites. La taille de la population ou le taux de mutation sont des paramètres difficiles à déterminer. Aussi, le succès de l'évolution en dépend et plusieurs essais sont donc nécessaires, ce qui limite l'efficacité de l'algorithme.

Le recuit simulé :

Le recuit simulé est une méthode d'optimisation inspirée d'un processus utilisé en métallurgie. Celui-ci alterne des cycles de refroidissement lent et de réchauffage (recuit) qui tendent à minimiser l'énergie du matériau en partant d'une température élevée pour finir à une température proche de zéro [42].

Le processus est initialisé avec une solution aléatoire associée à une température élevée ; si cette solution tend à diminuer la fonction d'énergie à minimiser, celle-ci est acceptée. Dans le cas contraire, la solution est acceptée mais pondérée par une température élevée. De manière itérative, le processus tend à minimiser la fonction d'énergie en diminuant la température associée à la solution et ce, jusqu'à convergence. Cette méthode a été utilisée pour la mise en correspondance d'images stéréoscopiques [43].

Les principaux inconvénients du recuit simulé résident dans le choix des nombreux paramètres, tels que la température initiale, la loi de décroissance de la température, les critères d'arrêt ou la longueur des paliers de température. Ces paramètres sont souvent choisis de manière empirique. De plus, la méthode du recuit simulé est très gourmande en temps de calcul.

Différentes méthodes de minimisation de fonction d'énergie ont été étudiées et comparées dans [44]. Dans la suite de ce chapitre, nous allons présenter le principe de méthodes de mise en correspondance en marge des méthodes locales et globales.

1.7.3 Autres méthodes

Les méthodes globales et locales forment la grande majorité des algorithmes de mise en correspondance stéréoscopique. Cependant, d'autres méthodes de mise en correspondance ont été présentées dans la littérature. Certaines d'entre-elles sont introduites ci-après.

Dans [45], l'auteur propose une méthode hiérarchique (ou multirésolution) qui permet de réduire l'intervalle des disparités à traiter. Le processus de mise en correspondance s'effectue d'un niveau de résolution faible des images vers des résolutions plus élevées. Les résultats d'appariement des niveaux de résolution faible sont conservés comme base pour les niveaux plus élevés. Cette technique permet un gain en temps de calcul et semble plus robuste par rapport au décalibrage et aux imperfections de rectification [46]. Cependant, la perte des détails pour les résolutions les plus faibles peut engendrer un effet de lissage des valeurs de disparité ; de plus, des valeurs erronées de disparité peuvent se propager aux niveaux de résolution supérieure.

Dans [47], les auteurs présentent une méthode hybride qui applique deux passages, l'un exploitant la méthode de programmation dynamique et l'autre utilisant celle du recuit simulé.

Dans [48], les auteurs proposent une méthode par croissance de germes. Celle-ci débute par la mise en correspondance de points d'intérêt appelés « germes ». Ensuite, l'algorithme de croissance de régions propage les appariements obtenus par certains germes des zones les plus texturées (a priori pour lesquelles la mise en correspondance est la plus fiable) aux zones les moins texturées.

1.8 Méthodes de raffinement des résultats de disparité

Différentes méthodes permettent le raffinement des cartes des disparités obtenues : cela consiste à améliorer ces cartes en supprimant les erreurs ou encore en donnant aux valeurs des disparités une résolution au sous-pixel. Cette étape est utilisée dans l'immense majorité des méthodes de mise en correspondance dans le but d'obtenir des cartes moins brutes notamment pour des applications de reconstruction 3D. A noter qu'il ne s'agit pas ici de mesures mais bien d'une étape de post-traitement supplémentaire.

1.8.1 Propagation des disparités

De manière générale, les auteurs tendent à fournir des cartes denses des disparités et ce, même si ces cartes ne sont pas complètement denses en fin de traitement. Pour ce faire, les auteurs peuvent utiliser une méthode pour propager les disparités calculées aux pixels non affectés par une technique d'interpolation linéaire ou encore une technique de croissance de germes [48].

1.8.2 Mise en correspondance au sous-pixel

L'estimation de la disparité avec une précision au sous-pixel peut être réalisée en interpolant la courbe de dissimilarité par une courbe du second degré (ou parabole) autour du minimum de la courbe [45] [49]. La valeur non entière, minimisant cette courbe approchée, est considérée comme la valeur de disparité au sous-pixel. Cette technique ne demande pas beaucoup de calculs supplémentaires. Les auteurs l'utilisent notamment pour obtenir un rendu plus agréable à l'œil pour des applications liées à la reconstruction 3D.

La mise en correspondance peut aussi mener à un calcul de disparité au sous-pixel. Pour cela, il est nécessaire d'interpoler les niveaux de gris de l'image originale de manière à obtenir une image de plus grande résolution [34].

1.8.3 Elimination des erreurs d'appariement

Certains auteurs concluent leur méthode par une étape de filtrage de manière à éliminer des disparités erronées soit par des opérateurs morphologiques simples ou encore un filtre médian [31]. Cette étape permet la suppression dans un voisinage donné de valeurs de disparité très éloignées de leurs voisins.

La *validation directe/inverse* est une méthode classique [50] largement relatée dans la littérature qui permet la suppression de certaines erreurs d'appariement : ici, ce sont les erreurs dues aux occultations qui sont principalement visées. Cette technique s'applique à un couple de cartes des disparités obtenues en prenant respectivement comme image de référence l'image gauche puis l'image droite.

De façon formelle, on définit la relation suivante :

$$d_{gd}(x, y) = d_{dg}(x - d_{gd}(x, y), y) \ , \tag{1.13}$$

où d_{gd} et d_{dg} sont les valeurs de disparité calculées en prenant respectivement comme image de référence l'image gauche et l'image droite. Lorsque cette relation n'est pas vérifiée, la validation directe/inverse n'est pas réalisée et aucune valeur de disparité n'est associée au pixel (x, y).

1.9 Conclusion

La stéréovision passive permet donc de retrouver, sous certaines conditions, l'information de profondeur perdue lors du processus d'acquisition. Dans ce premier chapitre, les différentes étapes de la stéréovision ont été détaillées, du capteur stéréoscopique classique, au calcul des distances en détaillant l'étape primordiale de mise en correspondance ; des méthodes classiques, largement relatées dans la littérature, ont été présentées.

L'objectif de ces méthodes est souvent lié à la reconstruction 3D dont le but est d'obtenir un résultat dense et agréable à l'œil. En effet, ces méthodes calculent des valeurs de disparité dans des zones de l'image dépourvues d'information lumineuse locale (non texturées) mais aussi en des pixels qui n'ont pas d'homologue dans l'une ou l'autre image (occultés). A posteriori, ces méthodes utilisent des techniques de raffinement des cartes des disparités pour en améliorer le rendu.

Aussi, les performances de ces différentes méthodes ont été comparées dans la littérature [51]. Bien que celles basées sur des techniques d'optimisation globale fassent partie des meilleures dans cette évaluation [52] [53], la recherche de la solution optimale est très coûteuse en temps de calculs ; des travaux [54] [55] [56], liés aux méthodes globales, ont montré qu'il était possible d'atteindre des traitements à la cadence vidéo en implantant les algorithmes directement dans le processeur graphique d'une architecture PC classique. Cependant, pour ce type de méthode d'optimisation, les points de convergence associés ne sont pas prévisibles ce qui les rend difficilement utilisables pour nos systèmes de vision embarqués (de type FPGA[1]). A contrario, les algorithmes basés sur la recherche exhaustive WTA ont des propriétés qui permettent leur implantation dans ces architectures parallèles dans le but d'atteindre la cadence imposée par la vidéo. En effet, avec la méthode de recherche exhaustive WTA, le nombre d'opérations nécessaires pour l'obtention de la solution est fixe. Ainsi, même si les méthodes exploitant la similarité présentent des temps de calculs assez importants sur des architectures classiques, leur implantation dans une architecture de calcul spécialisée peut considérablement diminuer ces temps de calculs [13] [57].

Pour ce travail de thèse, nous nous intéressons donc aux méthodes de mise en correspondance par corrélation (basées sur la méthode WTA) permettant le calcul de cartes denses des disparités. Ces méthodes sont exclusivement basées sur l'appariement de fenêtres ou de voisinages 2D.

Cependant, la grande majorité des méthodes se placent dans le cas particulier de la configuration parallèle pour laquelle les droites épipolaires sont confondues avec les lignes des images gauche et droite. Autrement dit, deux pixels homologues dans les images gauche et droite se trouvent nécessairement sur la même ligne. L'information se trouvant sur la ligne courante, il semble que le problème de mise en correspondance soit un problème 1D. Aussi pourrait-on se demander si l'information située sur cette unique ligne ne serait pas suffisante pour permettre des appariements efficaces.

Dans le chapitre suivant, nous allons présenter des travaux ayant en commun une approche linéaire de la stéréovision. Ensuite, une comparaison sur la base d'un algorithme classique de mise en correspondance basé sur des fenêtres de similarité 1D et 2D, sera détaillée.

[1]Field Programmable Gate Arrays

Références

[1] T. WILLIAMSON et C. THORPE : A trinocular stereo system for highway obstacle detection. *IEEE International Conference on Robotics and Automation*, volume 3, pages 2267–2273, Detroit, MI, USA, juin 1999. 12

[2] N. AYACHE et F. LUSTMAN : Trinocular stereo vision for robotics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):73–85, 1991. 12

[3] H. MATHIEU et F. DEVERNAY : Systèmes de miroirs pour la stéréoscopie. Rapport technique RR-0172, Institut National de Recherche en Informatique et en Automatique, INRIA, juin 1995. 12, 111

[4] J. GLUCKMAN et S.K. NAYAR : Rectified catadioptric stereo sensors. *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 380–387, Hilton Head Island, SC, USA, juin 2000. 12

[5] Y. RUICHEK : *Stéréovision linéaire par réseau de neurones de Hopfield.* Thèse de doctorat, Université des Sciences et Technologies de Lille, 1997. 13, 44

[6] R. HORAUD et O. MONGA : *Vision par ordinateur : outils fondamentaux.* Editions Hermès, Paris, 1995. 13, 16

[7] A. FUSIELLO, E. TRUCCO et V. ROBERTO : A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000. 15, 46, 143

[8] N. AYACHE : *Vision stéréoscopique et perception multisensorielle : application à la robotique mobile.* Inter-Editions (MASSON), 1989. 15

[9] M. OKUTOMI et T. KANADE : A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, avril 1993. 17

[10] S. CHAMBON et A. CROUZIL : Towards correlation-based matching algorithms that are robust near occlusions. *International Conference on Pattern Recognition*, pages 20–23, Cambridge, UK, août 2004. 18

[11] A.F. BOBICK et S.S. INTILLE : Large occlusion stereo. *International Journal of Computer Vision*, 33(3):181–200, novembre 1999. 18

[12] G. EGNAL et P. WILDES : Detecting binocular half-occlusions : Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1127–1133, août 2002. 18

[13] M. PÉREZ : *Stéréovision dense par traitement adaptatif temps réel : Algorithmes et implantation.* Thèse de doctorat, Université des Sciences et Technologies de Lille, France, février 2005. 18, 35, 141, 142

[14] Y.-I. OHTA et T. KANADE : Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154, mars 1985. 19

[15] C. HARRIS et M. STEPHENS : A combined corner and edge detector. *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, Manchester, UK, janvier 1988. 22

[16] H. ISHIKAWA et D. GEIGER : Occlusions, discontinuities, and epipolar lines in stereo. *European Conference on Computer Vision*, volume 1, pages 232–248, Freiburg, Germany, juin 1998. 22

[17] J. CANNY : A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, novembre 1986. 22

[18] J. M. De La CRUZ, G. PAJARES et J. A. ALMANSA : A neural network model in stereovision matching. *Neural Networks*, 8(5):805–813, 1995. 22

[19] M. BROWN, D. BURSCHKA et G. HAGER : Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, août 2003. 23, 31

[20] Z. ZHANG : Le problème de la mise en correspondance : L'état de l'art. Rapport de recherche RR-2146, Institut National de Recherche en Informatique et en Automatique, INRIA, décembre 1993. 23, 31

[21] P. ASCHWANDEN et W. GUGGENBÜL : Experimental results from a comparative study on correlation type registration algorithms. *Robust computer vision : Quality of Vision Algorithms*, pages 268–282, 1992. 24

[22] F. DEVERNAY : *Vision stéréoscopique et propriétés différentielles des surfaces.* Thèse de doctorat, Ecole Polytechnique, France, janvier 1997. 24

[23] R. ZABIH et J. WOODFILL : Non-parametric local transforms for computing visual correspondence. *European Conference on Computer Vision*, pages 151–158, Stockholm, Suède, mai 1994. 26

[24] H. HIRSCHMÜLLER et D. SCHARSTEIN : Evaluation of cost functions for stereo matching. *IEEE International Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, juin 2007. 27

[25] T. KANADE et M. OKUTOMI : A stereo matching algorithm with an adaptive window : theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, septembre 1994. 27, 28

[26] M. AGRAWAL et L. DAVIS : Window-based, discontinuity preserving stereo. *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 66–73, Washington, USA, juillet 2004. 27

[27] H. K. NISHIHARA : Prism, a pratical real-time imaging stereo matcher. Rapport technique A. I. Memo 780, Institut de Technologie du Massachusetts, MIT, MA, USA, 1984. 28

[28] A. FUSIELLO, V. ROBERTO et E. TRUCCO : Symmetric stereo with multiple windowing. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(8):1053–1066, 2000. 28, 75

[29] H. HIRSCHMÜLLER, P. INNOCENT et J. GARIBALDI : Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, 47(1):229–246, 2002. 29

[30] M. PÉREZ, F. CABESTAING et O. COLOT : A SBAN stereovision algorithm using Hue as pixel similarity criterion. *International Conference on Computer Vision and Graphics*, Warsaw, Poland, 2004. 29

[31] K. MÜHLMANN, D. MAIER, J. HESSER et R. MÄNNER : Calculating dense disparity maps from color stereo images, an efficient implementation. *IEEE Workshop on Stereo and Multi-Baseline Vision*, pages 30–36, Kauai, HI, USA, juin 2001. 29, 34, 75, 76, 77, 79, 88, 89, 90, 91, 92

[32] M. BLEYER et M. GELAUTZ : A layered stereo algorithm using image segmentation and global visibility constraints. *IEEE International Conference on Image Processing*, volume 5, pages 2997–3000, Singapore, octobre 2004. 29

[33] O. VEKSLER : Fast variable window for stereo correspondence using integral images. *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 565–570, Madison, WI, USA, juin 2003. 30

[34] S. FORSTMANN, Y. KANOU, J. OHYA, S. THUERING et A. SCHMITT : Real-time stereo by using dynamic programming. *Workshop on real-time 3D sensors and their use, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 3, pages 29–36, Washington, DC, USA, juin 2004. 30, 34

[35] M. GONG et Y.H. YANG : Near real-time reliable stereo matching using programmable graphics hardware. *IEEE International Conference on Computer Vision and Pattern Recognition*, San diego, CA, USA, juin 2005. 30

[36] S. CHAMBON : *Mise en correspondance stéréoscopique d'images couleur en présence d'occultations*. Thèse de doctorat, Université Paul Sabatier - Toulouse III, 2005. 30, 47

[37] Y. RUICHEK, H. ISSA et J.G. POSTAIRE : Genetic approach for obstacle detection using linear stereo vision. *IEEE International Vehicles Symposium*, pages 261–266, Dearborn, MI, USA, octobre 2000. 30, 31, 44

[38] O. VEKSLER : Stereo correspondence by dynamic programming on a tree. *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 384–390, San Diego, CA, USA, juin 2005. 31

[39] S. BIRCHFIELD et C. TOMASI : Depth discontinuities by pixel-to-pixel stereo. *International Journal of Computer Vision*, 35(3):269–293, 1999. 31

[40] J. HU et P. SIY : Stereo correspondence through multiple constraint neural networks. *Proceedings of the IEEE International Conference on Neural Networks*, pages 126–131, San Francisco, USA, mars 1993. 31

[41] M. GONG et Y.H. YANG : Genetic-based stereo algorithm and disparity map evaluation. *International Journal of Computer Vision*, 47(1):63–77, avril 2002. 32

[42] P. Van LAARHOVEN et E. AARTS : *Simulated Annealing : Theory and Applications*. Reidel, 1988. 32

[43] J. P. Pascual STARINK et E. BACKER : Finding point correspondences using simulated annealing. *Pattern Recognition*, 28(2):231–240, février 1995. 32

[44] R. SZELISKI, R. ZABIH, D. SCHARSTEIN, O. VEKSLER, V. KOLMOGOROV, A. AGARWALA, M. F. TAPPEN et C. ROTHER : A comparative study of energy minimization methods for markov random fields. *European Conference on Computer Vision*, volume 2, pages 16–29, Graz, Austria, mai 2006. 32

[45] C. SUN : Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques. *International Journal of Computer Vision*, 47(1-3):99–117, 2002. 33, 34, 77

[46] V. LEMONDE : *Stéréovision embarquée sur véhicule : De l'auto-calibrage à la détection d'obstacles*. Thèse de doctorat, Institut National des Sciences Appliquées de Toulouse, 2005. 2, 3, 33, 46, 110, 115

[47] P. N. BELHUMEUR et D. MUMFORD : A bayesian treatment of the stereo correspondence problem using half-occluded regions. *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 5, pages 506–512, Champaign, USA, juin 1992. 33

[48] M. LHUILLIER et L. QUAN : Quasi-dense reconstruction from image sequence. pages 125–139, London, UK, 2002. 33

[49] H. HIRSCHMÜLLER : Accurate and efficient stereo processing by semi-global matching and mutual information. *IEEE International Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, juin 2005. 34

[50] P. FUA : A parallel stereo algorithm that produces dense depth maps and preserves image features. Rapport de recherche RR-1369, Institut National de Recherche en Informatique et en Automatique, INRIA, janvier 1991. 34, 77

[51] D. SCHARSTEIN et R. SZELISKI : A taxomomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1): 7–42, avril 2002. **www.middlebury.edu/stereo**. 8, 35, 46, 47, 54, 60, 63, 72, 77, 86, 88, 89, 90, 91, 92

[52] V. KOLMOGOROV et R. ZABIH : Multi-camera scene reconstruction via graph cuts. *European Conference on Computer Vision*, pages 82–96, London, UK, 2002. 35

[53] A. KLAUS, M. SORMANN et K. KARNER : Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. *IEEE International Conference on Pattern Recognition*, pages 15–18, Washington, DC, USA, 2006. 35

[54] L. WANG, M. LIAO, M. GONG, R. YANG et D. NISTER : High-quality real-time stereo using adaptive cost aggregation and dynamic programming. *3DPVT '06 : Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission*, pages 798–805, Washington, DC, USA, 2006. 35

[55] Q. YÁNG, L. WANG, R. YANG, H. STEWÉNIUS et D. NISTÉR : Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2347–2354, New York, NY, USA, juin 2006. 35, 51

[56] M. GONG et Y.H. YANG : Real-time stereo matching using orthogonal reliability-based dynamic programming. *IEEE Transactions on Image Processing*, 16(3):879–884, mars 2007. 35

[57] J.L. BOIZARD, M. DEVY, P. FILLATREAU, J.Y. FOURNIOLS, P. LACROIX, N. NASREDDINNE, F.X. BERNARD et T. SENTENAC : Real-time stereovision by an integrated sensor. *Proceedings of the 6th IFAC Symposium on Intelligent Autonomous Vehicles (IAV 2007)*, Toulouse, France, septembre 2007. 35

Chapitre 2

La mise en correspondance stéréoscopique : un problème 1D ?

Sommaire

2.1 Méthodes linéaires 42
2.2 Calibration et rectification 46
2.3 Légitimité des fenêtres de similarité 2D 46
2.4 Méthode d'évaluation 47
2.4.1 Présentation 48
2.4.2 Les limites de la méthode d'évaluation 50
2.5 Comparaison 1D/2D 51
2.5.1 La méthode utilisée 52
2.5.2 Résultats obtenus sur le couple d'images *Head and Lamp* 53
2.5.3 Résultats obtenus sur le couple d'images *Venus* 63
2.5.4 Résultats obtenus sur une chaussée 64
2.6 Conclusion 67
Références 70

DANS la grande majorité des travaux, la mise en correspondance stéréoscopique par corrélation s'appuie sur l'utilisation de fenêtres bidimensionnelles (2D). De ce fait, la mise en correspondance stéréoscopique semble être un problème 2D. Notre objectif est de montrer que l'approche 1D n'est pas à exclure a priori. Dans ce chapitre, quelques travaux (assez rares) basés sur l'utilisation de caméras linéaires, pour des applications de stéréovision, sont présentés. L'utilisation de telles caméras engendre des contraintes non négligeables sur la calibration du système. Ensuite, nous présenterons une méthode d'évaluation largement utilisée dans le domaine de la stéréovision, de manière à effectuer une première comparaison entre les résultats obtenus sur la base de fenêtres 2D et 1D, par un algorithme classique de mise en correspondance stéréoscopique.

2.1 Méthodes linéaires

Réaliser les calculs très rapidement, pour des applications à la cadence vidéo, a toujours été une contrainte forte en stéréovision. Afin de diminuer le volume des calculs, dans les années 90, quelques chercheurs ont été conduits à travailler sur des systèmes stéréoscopiques composés de caméras linéaires.

Dans [1], à la place de caméras matricielles (fournissant une image à deux dimensions), Bruyelle utilise des caméras linéaires (l'image ne comporte qu'une seule ligne comme présenté figure 2.1) pour son dispositif stéréoscopique de détection d'obstacles sur la chaussée. Ces caméras linéaires possèdent des caractéristiques très intéressantes. La structure très simple des capteurs permet d'augmenter leur largeur : ils disposent encore aujourd'hui d'une résolution horizontale bien supérieure à celle d'un capteur matriciel. Malgré tout, le capteur linéaire délivre une quantité d'information faible puisqu'une seule ligne (voire quelques lignes) est disponible. Cependant, de part leur cadence d'acquisition très élevée, les caméras linéaires permettent d'enregistrer des mouvements avec une grande précision.

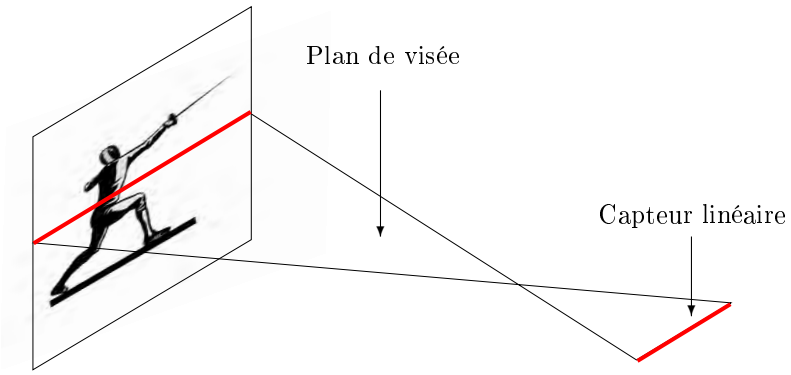


Fig. 2.1 : Principe d'une caméra linéaire.

L'objectif principal de ce travail était le développement d'un capteur stéréoscopique, sa mise en œuvre et sa validation. Le dispositif stéréoscopique de prise de vue, basé sur deux caméras linéaires, n'est pas différent dans son principe d'un système basé sur deux caméras matricielles. Néanmoins, une étape de calibrage très rigoureuse est nécessaire afin que les deux caméras linéaires observent la même scène, sans quoi la mise en correspondance s'avère impossible (cette étape est décrite de façon très complète dans [1]).

Avec ce type de caméra, l'étape de rectification ne peut pas être appliquée : soit les images gauche et droite observent strictement la même scène et la rectification n'est pas nécessaire ; soit elles observent des scènes différentes et l'étape de rectification ne pourra rien améliorer. Dans cette approche, certains paramètres, comme la distorsion, sont négligés lors de la phase de calibration.

L'auteur a mis en place une procédure de calibration précise de manière à ce que les plans de visée des deux caméras linéaires soient confondus (cf figure 2.2).

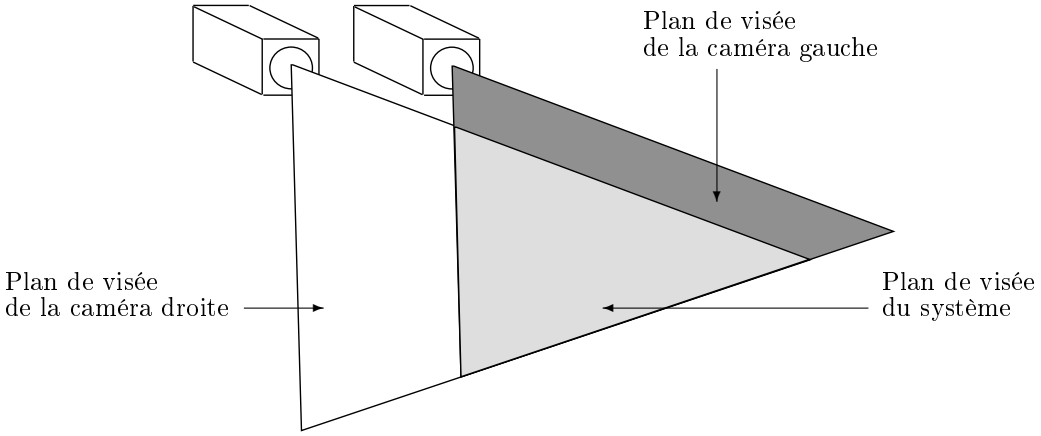


Fig. 2.2 : Plan de visée du système.

Avec l'image linéaire, il n'y a pas vraiment de notion de contour comme on peut la rencontrer avec des images matricielles. Ici, l'auteur définit des « points de contour ». Au vu de l'extrême pauvreté des informations géométriques en vision linéaire, les points d'intérêt se limitent en pratique à ces points de contour. Aussi, les primitives de types région sont ramenées à des segments de droite, définis comme des portions de l'image linéaire, pour lesquels le niveau de gris est uniforme, appelés « régions homogènes ». Chaque région est donc limitée par deux points de contour ; autrement dit, chaque point de contour sert de frontière entre deux régions. Ces deux primitives sont donc dépendantes : il en résulte que l'extraction de primitives passe par la détection des points de contour présents dans l'image.

Pour déterminer les points de contour, le gradient de l'image est calculé ; il correspond ici à la dérivée première de l'image linéaire dont les maxima correspondent aux points de contour de l'image. Dans cette approche, seuls les maxima supérieurs à un seuil donné sont retenus de manière à éliminer ceux caractérisant le bruit. La mise en correspondance est effectuée sur ces points de contour par le biais de fenêtres de similarité de même largeur, centrées sur le point de contour à apparier.

Comme les images linéaires contiennent relativement peu d'informations, quelques erreurs ou ambiguïtés peuvent apparaître lors du processus de mise en correspondance.

Dans [2], Burie reprend la suite des travaux de Bruyelle sur le stéréoscope linéaire afin d'améliorer la partie algorithmique et réduire le nombre d'erreurs d'appariement. La partie extraction des primitives a été soignée puisqu'elle constitue une étape critique dans le processus d'appariement. Une opération de lissage permet l'élimination du bruit présent dans l'image linéaire avant de lui appliquer l'opérateur de dérivation. Une opération

d'extraction des extrema locaux significatifs permet d'extraire les contours pertinents de l'image.

L'étape de mise en correspondance se base sur un ensemble de contraintes : épipolaire, signe du gradient, unicité et ordre. Les deux premières sont des contraintes locales (figure 2.3) qui interrompent la suite du processus de mise en correspondance lorsqu'elles ne sont pas vérifiées simultanément.

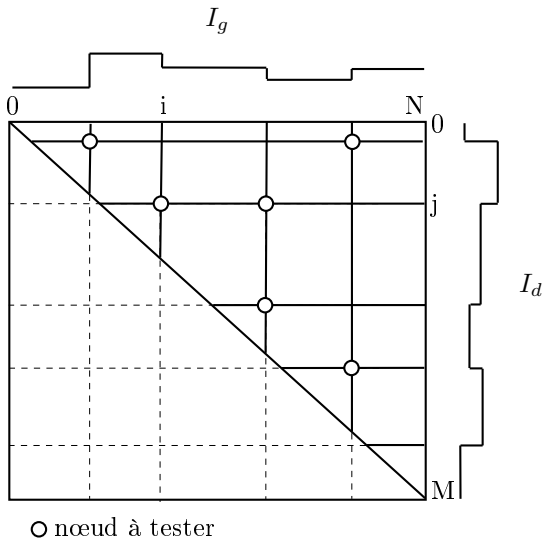


Fig. 2.3 : Application des contraintes épipolaire et de signe du gradient afin d'éliminer des couples candidats à l'appariement.

Lorsque le premier couple de points de contour homologues est validé, l'algorithme cherche le couple suivant sur la base du premier couple obtenu. La mise en correspondance de ces primitives peut être vue comme la recherche d'un chemin dont chaque étape est constituée par l'appariement d'un couple de points de contour. Les primitives sont appariées hiérarchiquement en commençant par les plus significatives, c'est-à-dire celles pour lesquelles la valeur absolue du gradient est la plus élevée.

Cette mise en correspondance est effectuée de deux manières : en partant du début de la ligne (analyse progressive) ou de la fin de ligne (analyse rétrograde). La comparaison des résultats obtenus pour ces deux analyses permet une confirmation de certains appariements et une élimination des appariements les plus ambigus.

Dans [3], Ruichek utilise aussi un stéréoscope linéaire associé à une méthode d'appariement basée sur les réseaux de neurones pour une application de détection d'obstacles à l'avant de véhicules [4]. Le problème d'appariement est ici ramené à un problème d'optimisation. Une fonction coût, représentant les contraintes du problème, est construite de telle sorte que les neurones correspondent à des appariements corrects. Les contraintes utilisées sont les contraintes de position, de signe du gradient, d'unicité, d'ordre et de

continuité. Cette fonction coût est ensuite associée à la fonction d'énergie d'un réseau de neurones de Hopfield mettant en œuvre l'étape de minimisation de la fonction.

La suite de ces travaux a conduit à l'utilisation de capteurs matriciels mais avec une approche bi-modale pour une application de détection d'obstacles à l'avant de véhicules guidés. Dans [5], l'auteur conserve le principe de la stéréovision linéaire en n'utilisant que quelques lignes de l'image issue du capteur matriciel. Ceci permet de détecter d'éventuels obstacles sur les rails à l'avant du véhicule et ce, à différentes distances (différentes lignes de l'image). Une droite reliant les points centraux des rails est définie ; tout point d'intérêt se trouvant au dessus de cette droite est considéré comme appartenant à un obstacle potentiel. Si des points d'intérêt sont détectés sur la ligne, le stéréoscope change de mode de fonctionnement.

Ce deuxième mode, bidimensionnel cette fois-ci, utilise une partie de l'image matricielle pour confirmer la présence de l'objet et le caractériser. En effet, l'information contenue dans une image matricielle est bien plus riche que celle issue d'une image linéaire (ou d'une ligne de l'image). L'auteur définit, sous hypothèses, une fenêtre de recherche située au dessus de la ligne courante. Ensuite, les points de contour situés à l'intérieur de cette fenêtre sont mis en correspondance de manière à définir les caractéristiques de l'obstacle potentiel.

Toutes ces techniques, basées sur l'utilisation de caméras linéaires, permettent d'obtenir de bons résultats de mise en correspondance alors que seule l'information située sur la ligne de l'image est prise en compte (sous la condition d'une calibration forte) ; les cartes des disparités obtenues sont éparses. Cette expérience forte du laboratoire LAGIS, dans lequel ont été menés tous ces travaux, peut nous servir de caution quant à l'utilisation de la seule information située sur une unique ligne. Ces techniques ont été mises en œuvre dans le but d'augmenter la rapidité des traitements avec les ordinateurs de l'époque. Aujourd'hui, nous pouvons raisonnablement dire que l'utilisation d'images 2D n'est plus un frein à l'exécution de traitements à la cadence vidéo d'autant qu'il existe des composants électroniques dont l'architecture est particulièrement adaptée à la structure des algorithmes en jeu. De plus, ces images 2D apporteront bien plus d'information sur la scène observée qu'une image composée d'une seule ligne.

Nous verrons dans la suite de ce mémoire que l'utilisation de fenêtres 1D est aussi pertinente que l'utilisation systématique de voisinages 2D pour le processus de mise en correspondance stéréoscopique. Autrement dit, il est possible de calculer un champ dense des disparités par la mise en correspondance de fenêtres 1D. Avant tout, la stricte utilisation de l'information située sur la ligne ne peut se faire que si et seulement si les lignes des capteurs gauche et droit caractérisent strictement la même scène : les étapes de calibration et rectification s'avèrent donc cruciales.

2.2 Calibration et rectification

Depuis de nombreuses années, la calibration d'un système de prise de vues stéréoscopiques est le sujet de nombreux travaux. Les techniques fondées sur l'utilisation de mires ou autres objets de géométrie connue sont maintenant bien maîtrisées. Les mires utilisées sont, en fonction du système, de plus ou moins grande taille, ce qui rend assez laborieuse leur utilisation notamment pour des applications embarquées.

Dans [1], l'auteur a mis en place une procédure de calibration manuelle basée sur une mire adaptée à l'utilisation de caméras linéaires.

D'autres méthodes dites d'autocalibration ne nécessitent aucune connaissance de l'espace et permettent de s'affranchir des objets de calibration. Celles-ci utilisent des points caractéristiques des images comme base pour la calibration.

Dans [6], l'auteur se base sur une calibration exacte exécutée hors ligne de manière à effectuer une recalibration du système en embarqué en considérant que le système n'est pas complètement décalibré. Le système est donc initialement calibré. Une tâche de fond exploite la géométrie épipolaire de manière à évaluer la qualité de calibration courante ; enfin, le système se recalibre si besoin. Dans [7], les auteurs proposent également une méthode de calibration en ligne. Certains paramètres extrinsèques et intrinsèques sont considérés comme connus après une étape de calibration précise hors ligne. Ici, aucune mire de calibration n'est requise. Cependant, la calibration en ligne doit se faire dans un environnement transport bien particulier qui se compose d'une route plane et de lignes blanches parfaitement identifiables.

Dans [8], les auteurs ont développé une méthode de rectification très légère et très rapide puisque celle-ci tient sur une vingtaine de lignes de code. Ceci permet d'obtenir les droites épipolaires parallèles et confondues avec les lignes de l'image.

On peut aujourd'hui considérer que la rectification quasi-parfaite est opérationnelle.

2.3 Légitimité des fenêtres de similarité 2D

Comme explicité dans la section précédente, il existe de nombreux travaux portant sur les étapes de calibration et rectification. Ces étapes sont cruciales car elles permettent de confondre les droites épipolaires aux lignes horizontales de l'image. Cette propriété est largement utilisée puisque la plupart des algorithmes décrits dans la littérature [9] sont appliqués à des couples d'images rectifiées. De cette façon, les indices de similarité sont calculés le long des lignes de l'image. Les travaux associés exploitent le contenu de fenêtres bidimensionnelles de manière à pallier le problème lié au manque de texture et rendre la carte des disparités plus dense.

L'utilisation de fenêtres bidimensionnelles se fonde sur l'hypothèse que tous les pixels contenus dans ces fenêtres de similarité sont caractérisés par la même valeur de disparité.

Or, il s'avère que cette hypothèse est loin d'être toujours respectée. En effet, les images naturelles sont souvent caractérisées par une perspective convergeant vers un point de fuite. Dans le cas d'un plan incliné, on constate aisément que deux lignes situées aux extrémités haute et basse d'une fenêtre de similarité 2D seront caractérisées par deux profondeurs différentes, l'hypothèse de départ étant ainsi invalidée. On retrouve le problème au niveau des discontinuités puisque la fenêtre de similarité contient des pixels appartenant à des objets différents situés à des profondeurs différentes. Ainsi, les scènes réelles contenant de nombreux objets situés à des profondeurs différentes conduiront à la mise en défaut de l'utilisation des fenêtres 2D et ce, quel que soit le type d'application.

De plus, si l'homologue d'un point d'une image se trouve sur la même ligne dans l'autre image et si l'information est suffisante, il ne semble pas nécessaire d'aller chercher de l'information sur les lignes voisines. Les travaux liés à la stéréovision linéaire, présentés précédemment, permettent de conforter cette idée. L'appariement semble être un problème 1D, puisqu'il s'agit de trouver l'homologue sur la droite épipolaire. A l'évidence, la notion d'« information suffisante » est en relation directe avec la taille et la géométrie de la fenêtre et la confiance accordée à un appariement.

Dans la suite de ce rapport, nous considérons (comme la plupart des auteurs) que les images stéréoscopiques utilisées sont parfaitement rectifiées. La légitimité de l'utilisation de fenêtres bidimensionnelles pour la mise en correspondance soulève de nombreuses interrogations que nous allons tenter de mettre en évidence grâce à la méthode présentée plus loin dans ce chapitre. La méthode d'évaluation utilisée est présentée dans la prochaine section.

2.4 Méthode d'évaluation

La multitude des travaux liés à la mise en correspondance stéréoscopique exige de pouvoir comparer la qualité de leurs résultats. Aussi, dans ce mémoire, nous proposons différents résultats qu'il convient d'évaluer et de comparer de la façon la plus rigoureuse possible.

L'évaluation proposée par Scharstein et Szeliski [9] sur le site de Middlebury[1] constitue, à notre connaissance, la comparaison de méthodes de mise en correspondance stéréoscopique la plus exhaustive qui existe. Elle a eu un écho considérable dans la communauté puisqu'elle a permis la comparaison et l'évaluation très précises de plusieurs dizaines d'algorithmes. Les critères d'évaluation utilisés ont ainsi été validés par l'ensemble de la communauté. Une synthèse de la classification de ces méthodes est disponible dans [10].

[1] www.middlebury.edu/stereo

2.4.1 Présentation

Dans le protocole d'évaluation de Middlebury, les couples d'images stéréoscopiques, sur lesquels chaque algorithme évalué est appliqué, sont synthétisés ou acquis dans des conditions réelles. Pour chaque couple d'images, une carte des disparités est donc calculée. La carte des disparités est représentée sous la forme d'une image dont le niveau de gris d'un pixel représente la valeur de disparité en ce pixel. Plus le niveau de gris est élevé et plus la valeur de disparité est importante. Les valeurs élevées de disparité caractérisent donc des objets proches du capteur stéréoscopique.

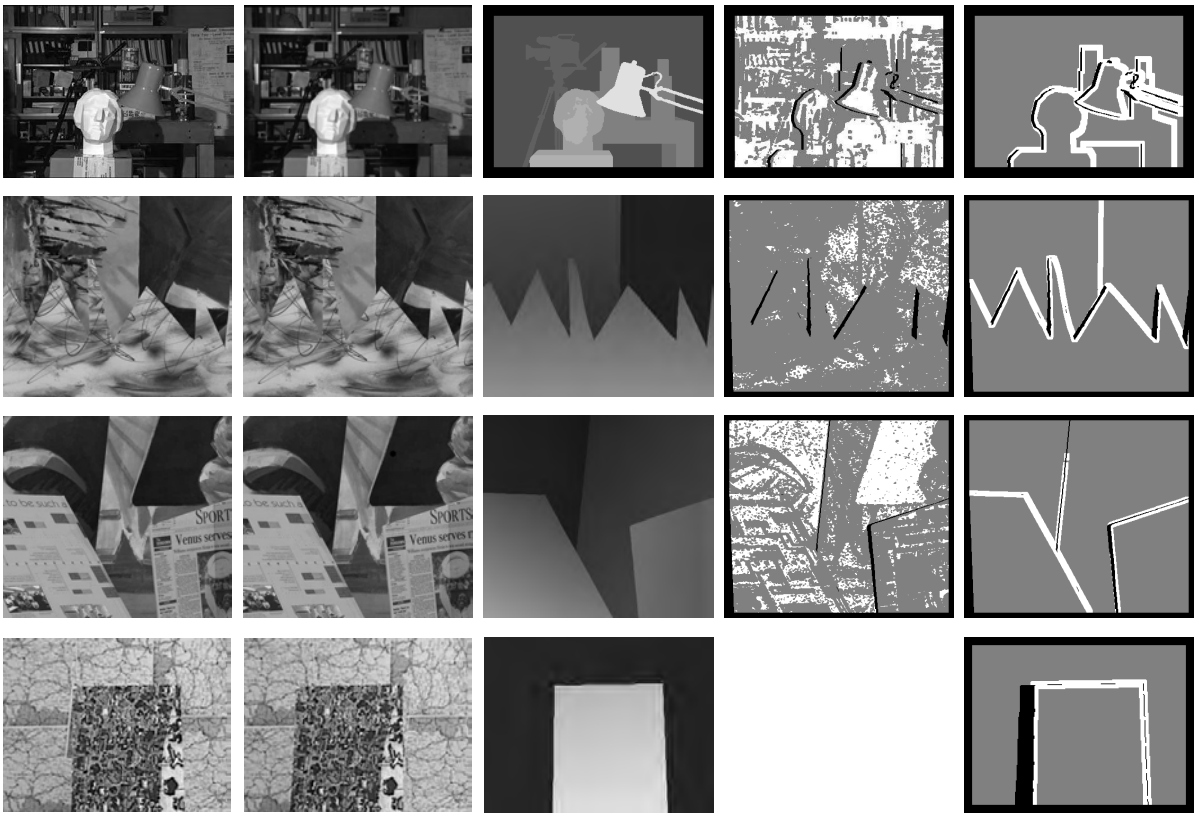
Pour chaque couple d'images stéréoscopiques, une vérité-terrain est fournie. Celle-ci représente une carte des disparités pour laquelle chaque valeur de disparité affectée est la valeur précise de disparité du pixel. Une vérité-terrain peut être calculée soit manuellement soit par utilisation de la vision active (système d'acquisition muni d'une source propre de lumière). Evaluer un algorithme de mise en correspondance revient à calculer la carte des disparités et à la comparer à la vérité-terrain. Dans notre étude, les couples d'images stéréoscopiques dépourvus de vérité-terrain ne sont pas considérés puisqu'il est impossible d'obtenir une étude quantitative précise des résultats de disparité obtenus.

En fonction de la nature de la zone où se trouve le pixel à apparier, les méthodes se comportent de différentes manières ; pour juger de ces différences de comportement, Scharstein et Szeliski classent les pixels selon trois zones :

- zones occultées – Pour lesquelles les pixels d'une image ne trouvent pas de correspondants dans l'autre image ;
- zones non texturées – Pour lesquelles la valeur de gradient horizontal, moyennée sur une fenêtre carrée (de taille 3×3), est inférieure à un seuil donné (ici égal à 4) ;
- zones des discontinuités – Pour lesquelles les pixels peuvent correspondre à des disparités très différentes (dont la différence absolue est, dans le cadre de cette évaluation, supérieure à deux pixels).

Le protocole d'évaluation présenté dans la suite de cette section correspond à la première version proposée par les auteurs. Celle-ci se base sur quatre couples d'images stéréoscopiques présentés à la figure 2.4 : respectivement de la première à la dernière ligne, les couples d'images stéréoscopiques *Head and Lamp*, *Sawtooth*, *Venus* et *Map* sont présentés.

Les images de vérité-terrain (figure 2.4(c)) fournissent la disparité à trouver en chaque pixel de l'image. Les autres cartes, associées aux images réelles, permettent une classification des pixels selon certains critères déjà présentés. Les figures 2.4(d) présentent à la fois les pixels appartenant à des zones texturées (en blanc) et à des zones non-texturées (en gris). Les figures 2.4(e) présentent quant à elles les pixels en zone de discontinuité (en blanc) et les pixels occultés (en noir). Cette classification de pixels en régions permet une comparaison détaillée des performances des différents algorithmes. On notera que le couple *Map* ne possède pas de carte des pixels en zones texturées et non-texturées.



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

Fig. 2.4 : (a) Image gauche , (b) Image droite, (c) Vérité-terrain, (d) Régions texturées (pixels blancs) et non-texturées (pixels gris), (e) Régions occultées (pixels noirs) et régions en zone de discontinuité (pixels blancs) pour les couples d'images (dans l'ordre) *Head and Lamp*, *Sawtooth*, *Venus* et *Map*.

Le critère d'évaluation consiste à calculer le nombre de disparités erronées, autrement dit le nombre de pixels pour lesquels la différence absolue pixel à pixel entre la carte des disparités et la vérité-terrain est supérieure à un seuil donné. Le taux d'erreur est ainsi calculé pour l'ensemble des pixels et pour chaque catégorie de pixels (occultés, en discontinuités et non-texturés). Dans la première version du protocole d'évaluation, le seuil est fixé à 1.

Dans la seconde version du protocole d'évaluation mise en ligne, les couples d'images *Sawtooth* et *Map* ont été retirés au profit des couples d'images *Cones* (figure 2.5(a)) et *Teddy* (figure 2.5(b)).

Les principales différences entre les deux versions de ce protocole d'évaluation sont les suivantes. La classification des pixels en zones texturées et non-texturées a été supprimée : les auteurs considèrent que ces catégories de pixels ne présentent plus de problèmes pour



(a) (b)

Fig. 2.5 : (a) Image gauche *Cones*, (b) Image gauche *Teddy*.

les algorithmes arrivant en tête de classement, autrement dit des algorithmes basés sur des méthodes globales. Aussi, pour les deux nouveaux couples d'images, les effets de bord ne sont plus pris en compte car selon les auteurs, seules les méthodes basées sur l'appariement de fenêtres ne peuvent pas y affecter de valeurs de disparité. Une des nouveautés de cette seconde version consiste en la modification du seuil d'erreur qui peut maintenant prendre les valeurs 0.5, 0.75, 1.0, 1.5 et 2.0. Ceci permet de juger plus finement de la précision des méthodes comparées.

Dans notre travail, la première version de ce protocole d'évaluation a été utilisée comme base de comparaison. En effet, la classification des pixels en fonction de la texture nous parait intéressante à conserver. Nos travaux sont fondés sur une approche locale, d'appariement de fenêtres, pour laquelle le manque de texture constitue un obstacle majeur pour l'efficacité de la mise en correspondance.

2.4.2 Les limites de la méthode d'évaluation

Même si la méthode d'évaluation présentée précédemment a été largement utilisée par la communauté, celle-ci comporte quelques limites sur lesquelles nous allons nous attarder ici.

En effet, on peut s'interroger sur la pertinence de l'affectation d'une disparité dans une zone non-texturée par les méthodes globales ou locales. Par définition, les zones non-texturées ne présentent pas suffisamment d'information pour obtenir un appariement efficace : certaines méthodes affectent néanmoins une valeur de disparité dans ces zones. Ceci constituent un défaut majeur puisqu'il n'y a, selon nous, aucune raison d'affecter une valeur de disparité si l'information locale ne le permet pas.

Aussi, l'ensemble des méthodes proposées dans cette comparaison affecte une valeur de disparité pour des pixels en zones d'occultation. Dans ce cas, il s'avère impossible de mettre en correspondance un pixel d'une image avec son homologue dans l'autre image si celui-ci n'existe pas. Ces méthodes sont en fait destinées à des applications de reconstruction 3D pour lesquelles il faut avoir un maximum de disparités affectées afin d'obtenir un rendu visuel « agréable » du modèle 3D de la scène observée.

La figure 2.6(b) présente la carte des disparités obtenues dans [11] : celle-ci est actuellement classée première selon l'évaluation de Middlebury pour un seuil d'erreur de un. Cependant, on observe figure 2.6(c) que pour ce seuil, un grand nombre de disparités erronées ne sont pas prises en compte (pixels encerclés) dans le calcul d'erreur. Pour une application de reconstruction 3D, cela n'est peut être pas crucial. Cependant, il en est autrement pour une application de « métrologie ». Dans notre approche, nous cherchons à faire de la mesure précise de disparité même si la carte finale des disparités n'est pas complètement dense.



(a) (b) (c)

Fig. 2.6 : (a) Vérité-terrain *Head and Lamp*, (b) Carte des disparités obtenues dans [11], (c) Mise en évidence de certaines erreurs d'appariement de (b) qui ne sont pas considérées comme telles par la méthode d'évaluation.

Dans la suite de ce chapitre, la pertinence de l'utilisation de fenêtre 2D dans le processus de mise en correspondance stéréoscopique va être remise en cause. Pour ce faire, une comparaison entre un algorithme classique, exploitant une fenêtre de similarité 2D et une fenêtre de similarité 1D, va être détaillée.

2.5 Comparaison 1D/2D

Cette partie vise à mettre en évidence des caractéristiques dues à l'utilisation de fenêtres de forme 2D et 1D pour la mise en correspondance stéréoscopique par corrélation. Nous présentons tout d'abord la méthode utilisée puis les résultats obtenus sur plusieurs couples d'images.

2.5.1 La méthode utilisée

L'objectif d'un algorithme de stéréovision dense par corrélation est de mettre en correspondance tous les pixels de l'image de manière à déterminer la disparité associée en chaque pixel et obtenir une carte dense des disparités.

Pour une fenêtre de forme 2D, l'indice de similarité $C_{2D}(x, y, s)$ est traditionnellement défini de la façon suivante :

$$C_{2D}(x, y, s) = \sum_{i=-w}^{i=w} \sum_{j=-w}^{j=w} C_{pixel}(x + i, y + j, s) \ . \tag{2.1}$$

Dans l'équation (2.1), $C_{pixel}(x, y, s)$ mesure la similarité entre le pixel de coordonnées (x, y) dans l'image de référence et le pixel décalé $(x + s, y)$ dans l'autre image. Dans le cas d'images rectifiées, la fenêtre 2D de dimension $(2w + 1)^2$ est centrée sur le pixel (x, y) dans la première image et $C_{2D}(x, y, s)$ est calculé pour plusieurs valeurs de décalage s de la fenêtre, le long de la même ligne de la deuxième image.

Dans le cas de mesures de similarité basées sur la différence d'intensité lumineuse, on peut décomposer (2.1) en une somme de coefficients $C_{1D}(x, y, s)$, chacun calculé sur des fenêtres de type ligne :

$$C_{2D}(x, y, s) = \sum_{j=-w}^{j=w} C_{1D}(x, y + j, s) \ , \tag{2.2}$$

avec

$$C_{1D}(x, y, s) = \sum_{i=-w}^{i=w} C_{pixel}(x + i, y, s) \ . \tag{2.3}$$

De manière à pouvoir comparer les résultats obtenus entre une mise en correspondance de fenêtres 2D et de fenêtres 1D, nous avons mis en place une méthode simple sans post ni pré-traitement. Nous calculons en chaque point les similarités C_{2D} et C_{1D} dans le cas particulier d'un algorithme fondé sur ces mesures de similarité SAD (pour « Sum of Absolute Differences ») et SSD (pour « Sum of Squared Differences »). Pour cette première comparaison, nous avons retenu ces deux mesures de similarité car ce sont les mesures les plus employées dans la littérature mais aussi parce qu'elles constituent les deux mesures les plus basiques.

Ainsi, le choix de la mesure nous permet de définir C_{pixel} comme suit :

Pour le SAD,

$$C_{pixel}(x, y, s) = |I_g(x, y) - I_d(x - s, y)| \ , \tag{2.4}$$

Pour le SSD,

$$C_{pixel}(x, y, s) = (I_g(x, y) - I_d(x - s, y))^2 \ , \tag{2.5}$$

où I_g et I_d sont respectivement les niveaux de gris des pixels des images gauche et droite. Chaque coefficient sera calculé pour les deux types de fenêtres, toutes deux centrées sur le pixel à apparier.

La similarité maximale entre le contenu de deux fenêtres de similarité sera caractérisée par le minimum de la courbe de similarité (de part la définition des mesures SAD et SSD). La disparité associée sera la valeur de décalage correspondant au minimum de la courbe de similarité et ceci quelle que soit sa forme : il s'agit ici de la méthode de recherche exhaustive WTA.

De manière formelle, dans le cas d'images rectifiées, on note $d(x, y)$ la disparité au pixel de coordonnées (x,y) qui minimise la fonction coût suivante :

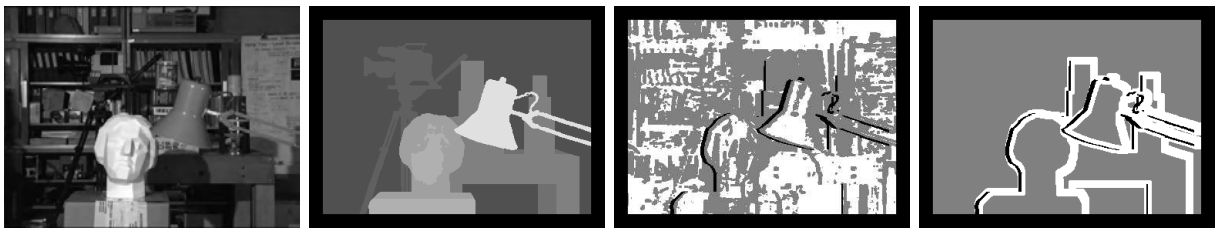
$$d(x, y) = \arg\min_{s} C_{xD}(x, y, s) \ , \tag{2.6}$$

où C_{xD} est la mesure de similarité C_{1D} ou C_{2D}.

Dans cette section, nous venons de présenter la méthode employée pour le calcul des cartes des disparités ; intéressons-nous maintenant aux résultats obtenus pour cette comparaison 1D/2D.

2.5.2 Résultats obtenus sur le couple d'images *Head and Lamp*

Pour cette première comparaison, nous avons utilisé un couple d'images de référence bien connu pour lequel de nombreux résultats sont disponibles : il s'agit du couple d'images *Head and Lamp* de l'université de *Tsukuba* présentées ci-après (figure 2.7). Ce couple d'images représente une scène réelle difficile à analyser puisqu'elle est composée de nombreux objets de tailles variées et placés à des profondeurs différentes.



(a) (b) (c) (d)

Fig. 2.7 : (a) Image gauche , (b) Vérité-terrain, (c) Régions texturées (pixels blancs) et non-texturées (pixels gris), (d) Régions occultées (pixels noirs) et régions en zone de discontinuité (pixels blancs) pour le couple d'images *Head and Lamp*.

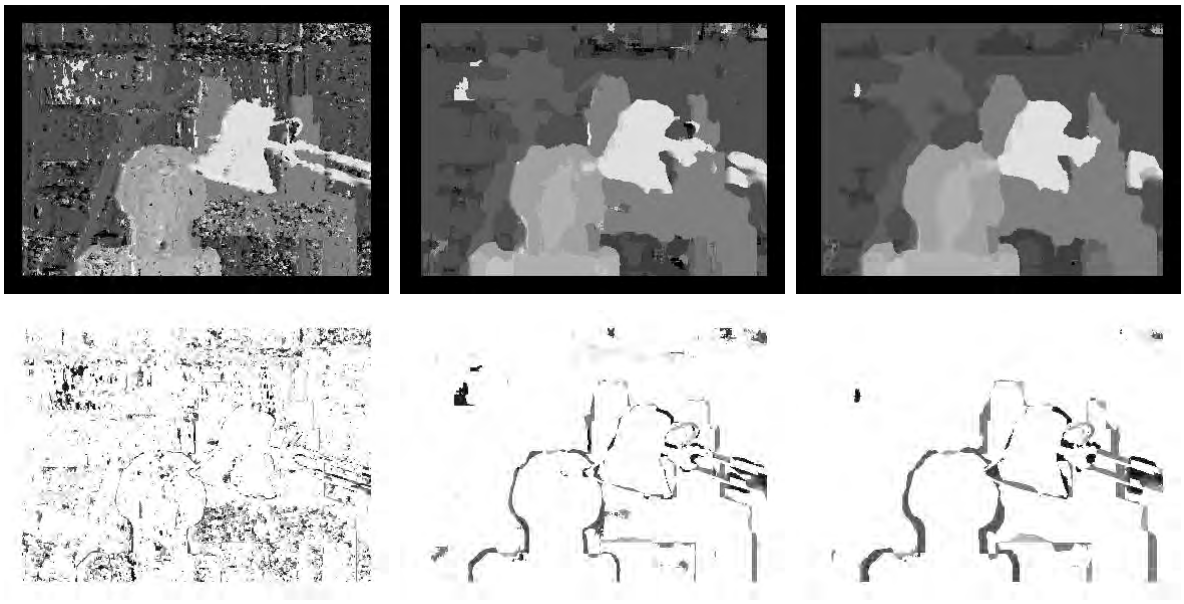
Les scores de similarité ont été calculés en faisant varier la taille de la fenêtre de 3×3 à 37×37 pour les voisinages bidimensionnels et de 3×1 à 37×1 pour les voisinages monodimensionnels, chaque fenêtre étant centrée sur le pixel à apparier.

Par la suite, les cartes des disparités obtenues avec les voisinages 1D et 2D sont comparées de manière qualitative et quantitative. L'évaluation qualitative est fondée sur l'impression ressentie lors d'une observation visuelle de la carte. L'évaluation quantitative est réalisée selon le critère proposé par Scharstein et Szeliski dans [9]. Ils calculent le pourcentage de pixels pour lesquels l'erreur absolue sur la disparité calculée est supérieure à un ($|\Delta\varepsilon| > 1.0$) par rapport à la valeur de disparité fournie par la vérité-terrain (les erreurs dues aux pixels occultés ne sont pas prises en compte).

2.5.2.1 Comparaison qualitative

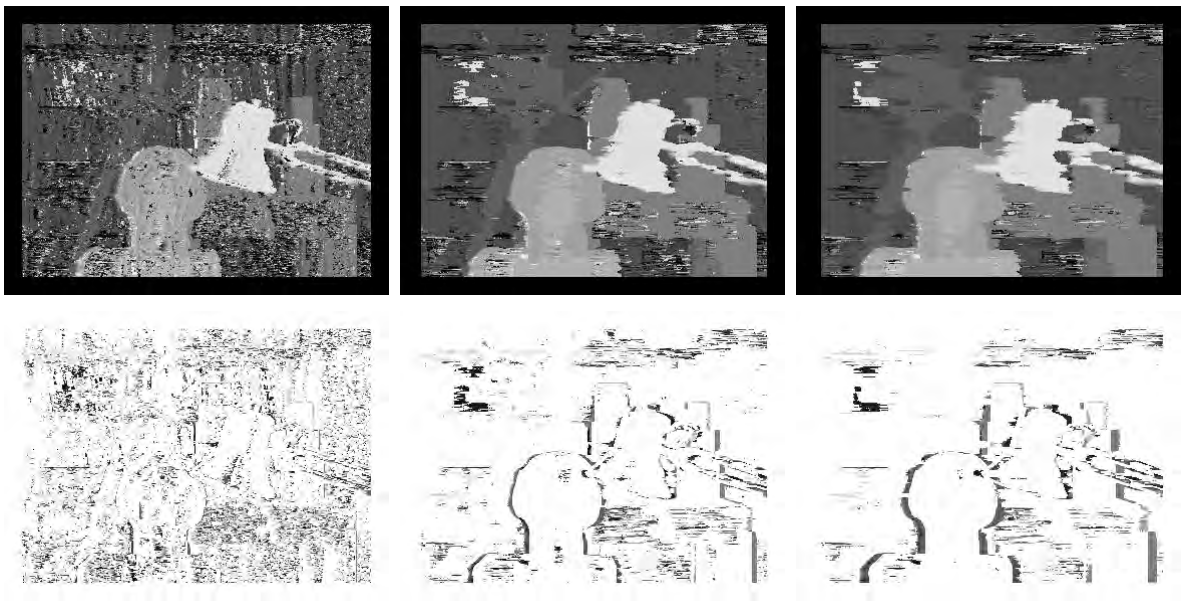
Les figures 2.8 et 2.9 présentent les cartes des disparités obtenues par la mesure SAD avec respectivement une fenêtre 2D et une fenêtre 1D. De même, les figures 2.10 et 2.11 présentent les cartes des disparités obtenues par la mesure SSD avec respectivement une fenêtre 2D et une fenêtre 1D. Les images de la seconde ligne de ces figures indiquent les pixels dont la différence absolue entre la disparité calculée et la vérité-terrain est supérieure à un. Dans ces images, plus l'erreur est importante, plus le niveau de gris du pixel est faible. Les résultats présentés sont caractéristiques de différentes tailles de fenêtre : petite (3 pixels de large), moyenne (11 pixels de large) et grande (19 pixels de large). La comparaison 1D et 2D des résultats est faite sur la base de fenêtres de même largeur plutôt que de fenêtres de même surface. Les résultats obtenus avec une fenêtre bidimensionnelle de taille 11×11 sont ainsi comparés à ceux obtenus avec une fenêtre monodimensionnelle de taille 11×1 plutôt que 121×1 : afin de pouvoir juger de la pertinence de l'utilisation (ou non) de la dimension verticale liée aux fenêtres 2D, il semble pertinent de comparer le comportement des fenêtres 1D et 2D sur la même base d'information horizontale, autrement dit pour une largeur identique. Aussi, on aurait pu envisager une étude du comportement de fenêtres proches de la forme 1D mais englobant quelques lignes de l'image comme par exemple une fenêtre de taille 11×3 ou 11×5. Cependant, cette première étude tend à montrer les avantages et inconvénients des fenêtres 2D et 1D ; utiliser une fenêtre de quelques lignes ne permettrait pas de mettre clairement ces caractéristiques en exergue.

Tout d'abord, on peut noter que visuellement les cartes des disparités obtenues à taille et forme de fenêtre identiques sont très semblables et ce, qu'il s'agisse de la mesure de similarité SAD ou SSD. Autrement dit, on observe que le comportement de ces deux mesures est assez similaire à fenêtre identique. Ensuite, en observant les figures 2.8, 2.9, 2.10 et 2.11, plusieurs constatations peuvent être faites. Pour des fenêtres de petites tailles, les erreurs d'appariement sont très nombreuses et réparties de manière uniforme sur toute la carte des disparités. Un même objet est parfois caractérisé par plusieurs valeurs de disparité. De larges fenêtres englobent plus d'information discriminante, ce qui semble améliorer les résultats de mise en correspondance dans les zones de l'image de même profondeur. D'un autre côté, les fenêtres de grande taille induisent des erreurs clairement



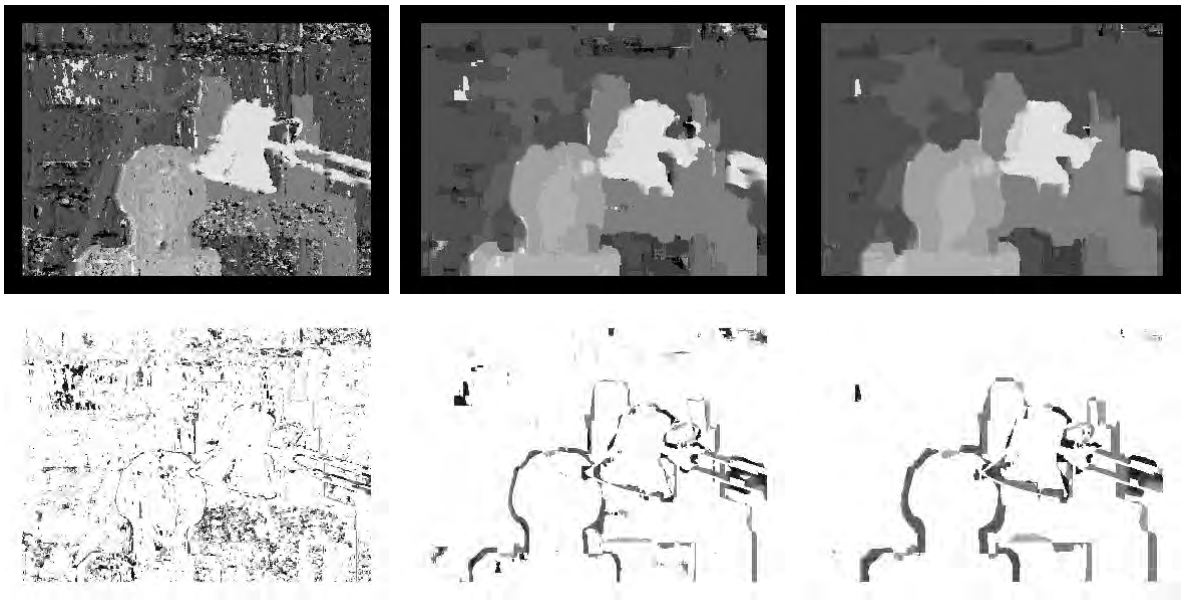
(a) (b) (c)

Fig. 2.8 : Comparaison qualitative des résultats obtenus par la mesure *SAD* avec des fenêtres de forme carrée de différentes tailles : (a) 3×3, (b) 11×11, (c) 19×19. La première ligne : cartes des disparités. La seconde ligne : cartes d'erreur pour le couple d'images *Head and Lamp*.



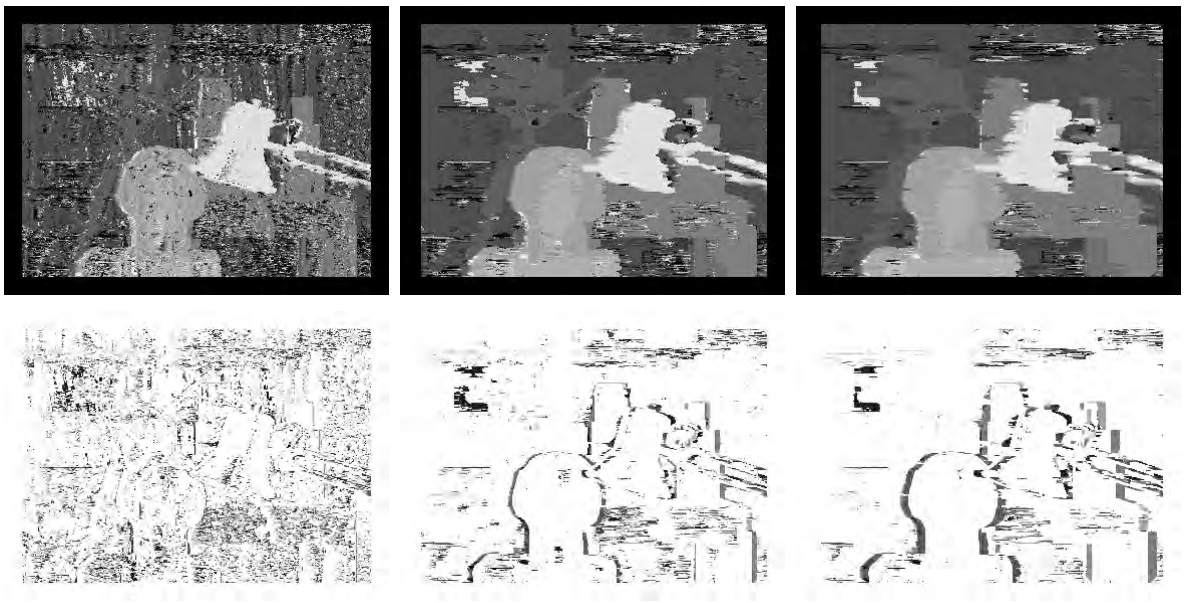
(a) (b) (c)

Fig. 2.9 : Comparaison qualitative des résultats obtenus par la mesure *SAD* avec des fenêtres de forme ligne de différentes tailles : (a) 3×1, (b) 11×1, (c) 19×1. La première ligne : cartes des disparités. La seconde ligne : cartes d'erreur pour le couple d'images *Head and Lamp*.



(a) (b) (c)

Fig. 2.10 : Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de forme carrée de différentes tailles : (a) 3×3, (b) 11×11, (c) 19×19. La première ligne : cartes des disparités. La seconde ligne : cartes d'erreur pour le couple d'images *Head and Lamp*.



(a) (b) (c)

Fig. 2.11 : Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de forme ligne de différentes tailles : (a) 3×1, (b) 11×1, (c) 19×1. La première ligne : cartes des disparités. La seconde ligne : cartes d'erreur pour le couple d'images *Head and Lamp*.

visibles en zone de discontinuité pour les différentes cartes des disparités. Quelle que soit la forme de la fenêtre, ces erreurs ne pourront pas être évitées si celles-ci couvrent des objets situés à des distances différentes. La disparité d'un pixel éloigné appartenant à la fenêtre et caractérisant un objet de profondeur différente, est affectée au pixel à apparier et engendre ainsi une erreur d'appariement : par abus de langage, ce phénomène est appelé « effet de moyennage » dans la suite de ce mémoire.

Dans les régions correspondant à un seul objet peu texturé, on peut vérifier que les fenêtres 2D fournissent de meilleurs résultats. Pour les images des figures 2.9 et 2.11, obtenues avec une fenêtre 1D, certaines régions présentent de nombreuses erreurs même pour de larges fenêtres. En fait, la taille de la fenêtre semble primordiale dans ces régions non-texturées et la fenêtre 2D semble tirer parti de sa dimension verticale en incluant plus d'information, nécessaire à un appariement efficace.

Pour les deux types de fenêtres, des erreurs apparaissent en zones de discontinuité. Le nombre d'erreurs augmente avec la taille des fenêtres. Cependant, l'effet de moyennage, qui introduit des erreurs en zones de discontinuité, semble plus marqué pour les cartes des disparités obtenues par la méthode 2D (figures 2.8 et 2.10) spécialement pour les fenêtres de plus grandes dimensions.

Le comportement des fenêtres 1D et 2D peut être analysé plus précisément en s'attardant sur des cas typiques tels que présentés figure 2.12. Ces quatre sous-images présentent soit une région non-texturée (figure 2.12(D)), soit des régions localisées en zone de discontinuité (figures 2.12(B) et 2.12(C)), soit une région présentant des motifs répétitifs (figure 2.12(A)).

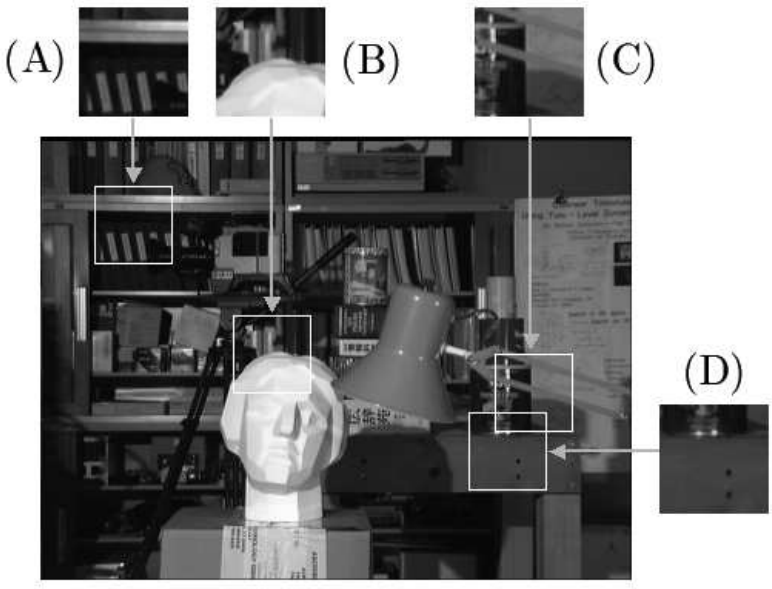


Fig. 2.12 : Quelques régions présentant des erreurs typiques.



Fig. 2.13 : Comparaison qualitative de résultats de disparité obtenus pour la sous-image (A) par la mesure *SSD* (Sous-image (A) - Vérité-terrain associée - Disparités obtenues avec une fenêtre carrée 11×11 - Disparités obtenues avec une fenêtre ligne 11×1).



Fig. 2.14 : Comparaison qualitative de résultats de disparité obtenus pour la sous-image (B) par la mesure *SSD* (Sous-image (B) - Vérité-terrain associée - Disparités obtenues avec une fenêtre carrée 11×11 - Disparités obtenues avec une fenêtre ligne 11×1).

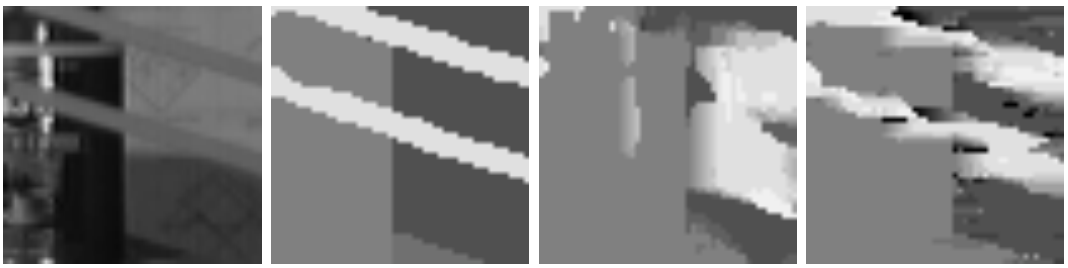


Fig. 2.15 : Comparaison qualitative de résultats de disparité obtenus pour la sous-image (C) par la mesure *SSD* (Sous-image (C) - Vérité-terrain associée - Disparités obtenues avec une fenêtre carrée 11×11 - Disparités obtenues avec une fenêtre ligne 11×1).



Fig. 2.16 : Comparaison qualitative de résultats de disparité obtenus pour la sous-image (D) par la mesure *SSD* (Sous-image (D) - Vérité-terrain associée - Disparités obtenues avec une fenêtre carrée 11×11 - Disparités obtenues avec une fenêtre ligne 11×1).

Les résultats de mise en correspondance pour ces différentes régions sont présentés figures 2.13 à 2.16 : il s'agit ici de parties de la carte des disparités obtenue avec la mesure de similarité SSD. Pour chacune de ces figures, la première image correspond à la zone étudiée, la seconde à la vérité-terrain associée et les troisième et quatrième aux disparités obtenues respectivement avec une fenêtre bidimensionnelle et monodimensionnelle. Nous comparons les résultats pour une largeur de fenêtre moyenne, soit 11×11 pour le 2D et 11×1 pour le 1D.

La figure 2.13 caractérise une zone présentant des motifs répétés dont la fréquence spatiale (en pixels) est inférieure à l'intervalle des disparités étudié. De ce fait, le processus de mise en correspondance devient très difficile et nécessite un maximum d'information discriminante. Ainsi, pour ce type de région, les fenêtres 2D fournissent de meilleurs résultats que ceux obtenus avec des fenêtres 1D et évitent de nombreuses ambiguïtés pour autant que tous les objets contenus dans la fenêtre sont caractérisés par la même valeur de disparité.

Avec une fenêtre 2D, les pixels caractérisant les livres sont facilement appariés grâce aux contours verticaux. L'étagère, située entre les deux séries de livres, présente peu d'information ; ainsi, par effet de moyennage, la disparité des livres est associée à cette étagère. Cependant, l'étagère est peut être située à une profondeur très différente de celle des livres et aucune information de l'image ne permet de mesurer cette profondeur. Dans ce cas, une information de confiance associée à la valeur de disparité pourrait trouver un grand intérêt.

Sur la figure 2.14, on note qu'avec des fenêtres bidimensionnelles, une partie du haut de la statue est assimilée à l'arrière-plan de l'image. En effet, ce type de fenêtres engendre des appariements erronés : la technique de mise en correspondance avec une fenêtre 2D suppose implicitement que tous les pixels inclus dans cette fenêtre aient la même disparité, ce qui n'est pas le cas pour les régions proches des discontinuités. Cependant, avec une fenêtre 1D, ces discontinuités sont limitées aux contours verticaux de l'image et les contours horizontaux sont mieux préservés. Une situation semblable apparaît dans des régions présentant des détails (dont la taille est inférieure à celle de la fenêtre) comme pour les bras de la lampe (illustré à la figure 2.15). Ceux-ci sont mieux préservés avec la fenêtre 1D. Avec une fenêtre bidimensionnelle, les petits objets sont assimilés par effet de moyennage à l'objet voisin dans l'image et disparaissent sur la carte des disparités.

La figure 2.16 montre que l'utilisation de fenêtres 2D entraîne peu de faux appariements en zone non-texturée puisque la carte des disparités obtenues est très semblable à la vérité-terrain. En zone non-texturée présentant des objets à la même profondeur, une fenêtre large est plus efficace et permet de tirer partie de plus d'information discriminante.

Le tableau suivant récapitule les observations induites par le comportement des fenêtres 1D et 2D et ce, pour les différentes zones de l'image.

Zones	Non-texturées	Texturées	En discontinuités
Fenêtre 2D de grande taille	++	−	−−
Fenêtre 2D de petite taille	−	+	+
Fenêtre 1D de grande taille	+	−	−
Fenêtre 1D de petite taille	−−	+	+

Tab. 2.1 : Synthèse des observations du comportement des fenêtres 1D et 2D en différentes zones de l'image.

De cette comparaison qualitative, on peut difficilement conclure de la supériorité d'une méthode sur l'autre. La méthode fondée sur l'utilisation de fenêtres bidimensionnelles est plus efficace dans des zones peu texturées. Cependant, les contours d'objets semblent mieux conservés avec l'utilisation de fenêtres 1D. La comparaison de ces deux méthodes est assez subjective puisqu'un observateur préférera probablement des cartes des disparités moins bruitées et plus homogènes comme celles obtenues avec la méthode 2D. Dans la suite de ce chapitre, nous allons conforter ces conclusions en calculant les taux de faux appariements dans ces différentes régions de l'image.

2.5.2.2 Comparaison quantitative

Scharstein et Szeliski [9] calculent en chaque pixel de l'image la différence en valeur absolue entre les résultats de disparité obtenus et la vérité-terrain. Chaque valeur absolue de cette différence, strictement supérieure à un, est considérée comme une erreur ($|\Delta\varepsilon| > 1.0$). De manière à comparer les méthodes 1D et 2D employées, les taux d'erreur ont été calculés pour les différentes catégories de pixels du couple *Head and Lamp*. Les trois catégories définies dans [9], pour lesquelles les pourcentages de disparités erronées ont été calculés, sont : les régions non-occultées ($B_{\overline{\mathcal{O}}}$), les régions non-texturées ($B_{\overline{\mathcal{T}}}$) et les régions proches des discontinuités ($B_{\mathcal{D}}$). A ces trois catégories, nous avons ajouté le calcul des taux d'erreur pour les régions texturées ($B_{\mathcal{T}}$) de manière à répondre à certaines interrogations. Dans toutes ces situations, les pixels occultés ne sont pas pris en compte.

Les résultats sont présentés figure 2.17 et caractérisent les erreurs d'appariement au sens de Scharstein et Szeliski pour les différentes catégories de pixels décrites ci-dessus. Chaque graphe représente une catégorie spécifique de pixels et montre les résultats obtenus en fonction de la dimension des fenêtres pour les deux types de méthodes 1D et 2D et pour les mesures de similarité SAD et SSD.

Dans le cas global présentant les résultats des pixels non-occultés, on constate sur la figure 2.17(a) que la méthode utilisant des fenêtres 2D fournit globalement de meilleurs résultats que l'autre méthode. Cela s'explique par le fait qu'une petite fenêtre de type ligne ne contient pas autant d'information qu'une fenêtre carrée de même côté.

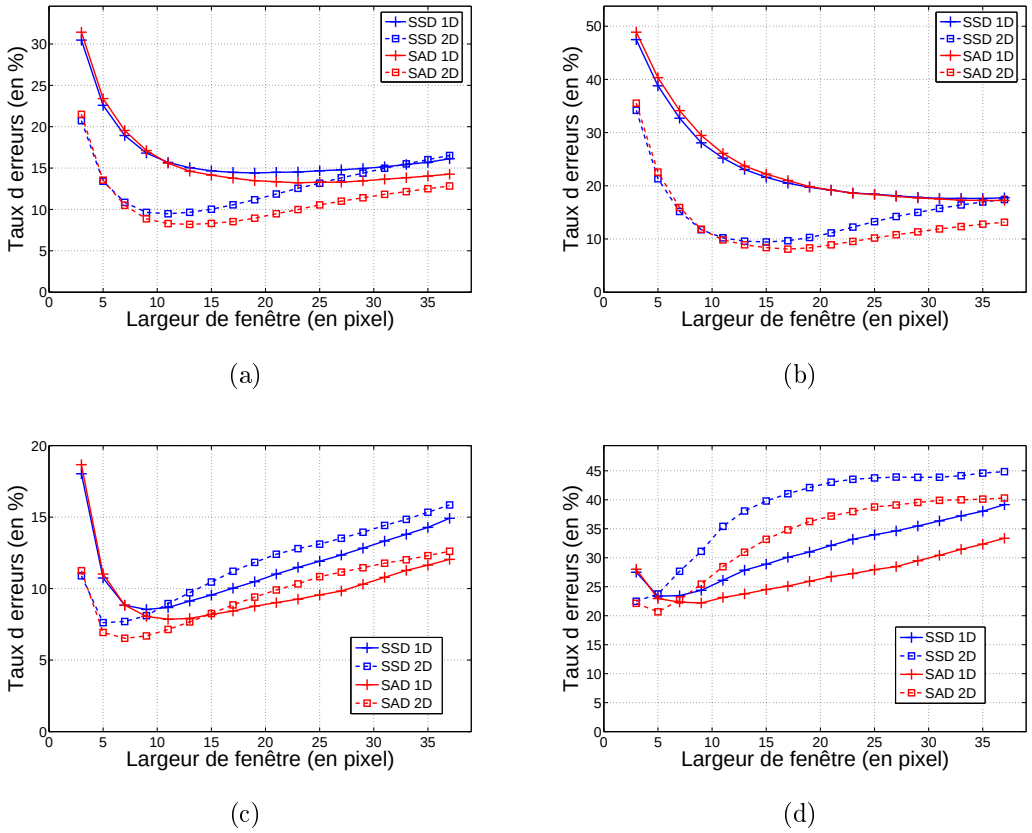


Fig. 2.17 : Comparaison quantitative des résultats de disparité obtenus avec les mesures SAD et SSD basées sur des fenêtres de formes carrée (2D) et ligne (1D) de différentes tailles dans les régions (a) non-occultées ($B_{\overline{\mathcal{O}}}$), (b) non-texturées ($B_{\overline{\mathcal{T}}}$), (c) texturées ($B_{\mathcal{T}}$), et (d) en discontinuités ($B_{\mathcal{D}}$) pour le couple d'images *Head and Lamp*.

Cependant, pour les deux mesures SAD et SSD, l'écart entre les deux taux d'erreur, 1D et 2D, diminue avec l'augmentation de la taille des fenêtres jusqu'à devenir nulle pour une largeur de fenêtre d'environ 33 pixels (figure 2.17(a)) pour la mesure SSD. Pour chacune de ces courbes, on peut définir une taille de fenêtre optimale engendrant un minimum d'erreurs. Pour ce couple d'images *Head and Lamp*, la largeur de fenêtre présentant le minimum d'erreurs est directement liée à la taille moyenne des objets constituant la scène observée. Passé ces valeurs optimales, on note que le taux d'erreur augmente plus rapidement pour la méthode se fondant sur les fenêtres 2D : l'effet néfaste de moyennage est de plus en plus marqué. Pour les fenêtres de petite taille, on constate que les deux mesures de similarité présentent des résultats équivalents. Néanmoins, l'augmentation de la taille des fenêtres rend le SSD moins efficace que le SAD.

En analysant la figure 2.17(b), on vérifie bien que les résultats de mise en correspondance sont meilleurs par l'utilisation de fenêtres 2D dans les régions non-texturées. Ceci confirme les observations établies lors de la comparaison qualitative notamment pour la

sous-image de la figure 2.16. Le taux minimum d'erreur est obtenu pour des largeurs respectives de 15 et 17 pixels pour une fenêtre 2D avec les mesures SSD et SAD et augmente au-delà de ces valeurs. D'autre part, les courbes relatives aux résultats de la méthode 1D sont monotones décroissantes. Plus la fenêtre 1D est large, moins elle génère d'erreurs en zones non ou peu-texturées et ceci, quelle que soit la mesure de similarité utilisée. Ainsi, pour améliorer les résultats obtenus avec la méthode 1D, il suffit d'augmenter la largeur de la fenêtre de similarité.

La figure 2.17(c) représente les taux d'erreur pour les zones texturées. On peut noter que la différence entre les taux d'erreur pour les deux méthodes est négligeable dans ce cas. Les deux courbes présentent un minimum pour de petites tailles de fenêtres alors que pour des fenêtres de grandes et moyennes tailles, les résultats de mise en correspondance obtenus avec la méthode 1D sont meilleurs pour chacune des mesures de similarité. Pour ce type de pixels, la mesure SAD prend de nouveau l'avantage par rapport au SSD. De manière générale, en zones texturées, on constate que les algorithmes utilisant des fenêtres d'appariement de type ligne fournissent des meilleurs résultats quand l'information locale contenue dans l'image est suffisante. Dans ce cas, on peut dire que l'information présente sur la droite épipolaire est suffisante pour effectuer un bon appariement.

Finalement, la figure 2.17(d) présente les taux d'erreur calculés pour des régions proches des discontinuités. On constate clairement que les fenêtres monodimensionnelles donnent de meilleurs résultats et ceci quelle que soit leur taille. Dans ce type de régions, les taux d'erreur sont assez élevés : ceci est dû à l'effet de moyennage observé sur les images des figures 2.14 et 2.15. Quand la fenêtre inclut des pixels ayant des disparités différentes, la méthode 2D affecte, en général, aux pixels appartenant aux petits objets la disparité caractérisant les plus grands. La mesure SAD présente une fois de plus une carte des disparités plus précise que celle obtenue avec la mesure SSD.

La méthode d'évaluation utilisée, présente selon nous quelques faiblesses. Sur la figure 2.10(c), on constate que la disparité du fond de l'image est affectée au pied de la caméra. Dans ce cas, selon le protocole d'évaluation, les disparités ne sont pas considérées comme erronées. Ainsi, toute erreur de disparité d'un pixel n'est pas considérée. Pour notre approche qui a pour finalité la précision du calcul de disparité, cela est inacceptable.

Le couple d'images *Head and Lamp* est assez difficile à analyser puisque la scène est constituée de nombreux objets situés à des profondeurs différentes. Dans une telle situation, les fenêtres 2D montrent leur faiblesse liée à l'effet de moyennage notamment pour des fenêtres 2D contenant des objets de profondeurs différentes. La suite de l'évaluation 1D/2D s'effectuera sur des images de nature différente de manière à mettre en évidence d'autres comportements des méthodes 1D/2D.

2.5.3 Résultats obtenus sur le couple d'images *Venus*

Comme pour le couple d'images *Head and Lamp*, *Venus* fait partie des couples d'images proposés dans l'évaluation de Scharstein et Szeliski [9]. Cette image est constituée de plusieurs plans faiblement inclinés dont certains sont quasiment verticaux. La comparaison se fait ici exclusivement de manière quantitative avec les deux méthodes 1D et 2D en se basant sur les deux mesures de similarité SAD et SSD.

La figure 2.18 présente les taux d'erreur (avec $|\Delta\varepsilon| > 1.0$) pour les différentes catégories de pixels (non-occultées ($B_{\overline{\mathcal{O}}}$), non-texturées ($B_{\overline{\mathcal{T}}}$), près des discontinuités ($B_{\mathcal{D}}$) et texturées ($B_{\mathcal{T}}$)). Les scores de similarité ont été calculés en faisant varier la taille de la fenêtre de 3×3 à 21×21 pour les fenêtres 2D et de 3×1 à 21×1 pour les fenêtres 1D.

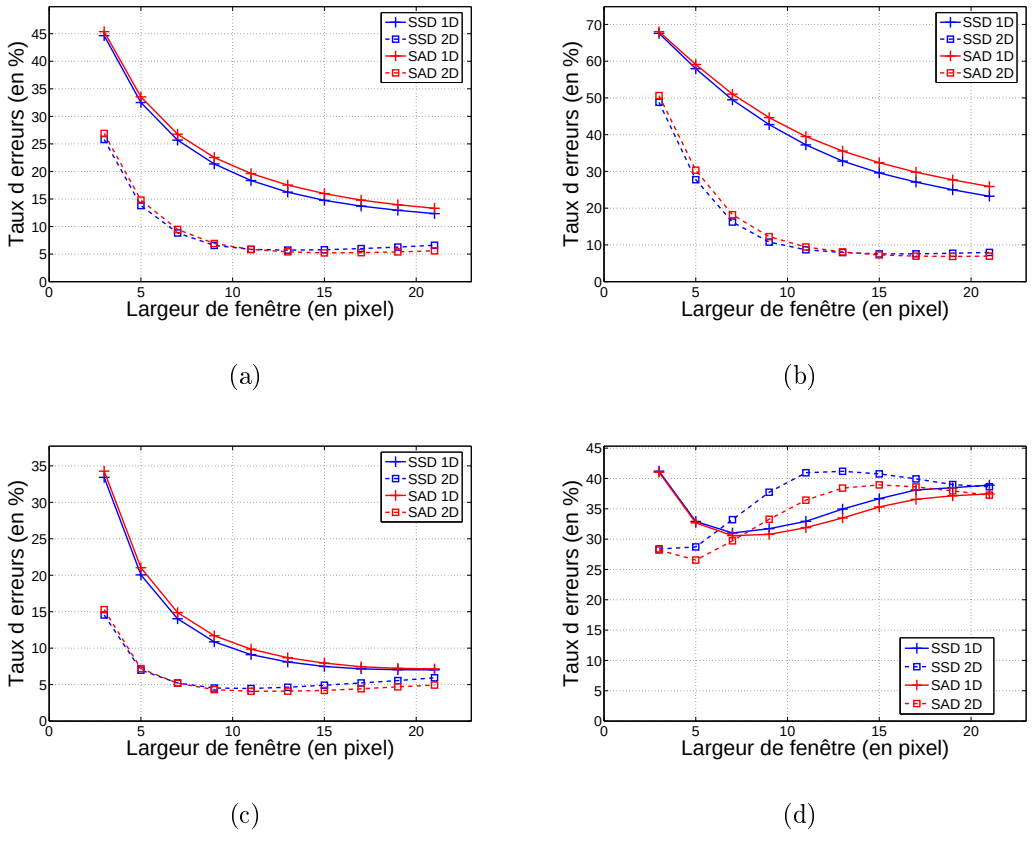


Fig. 2.18 : Comparaison quantitative des résultats de disparité obtenus avec les mesures SAD et SSD basées sur des fenêtre de formes carrée et ligne de différentes tailles dans les régions (a) non-occultées ($B_{\overline{\mathcal{O}}}$), (b) non-texturées ($B_{\overline{\mathcal{T}}}$), (c) texturées ($B_{\mathcal{T}}$), et (d) en discontinuités ($B_{\mathcal{D}}$).

Globalement, on constate sur les figures 2.18(a), 2.18(b) et 2.18(c) que la méthode utilisant des fenêtres 2D est plus précise quelle que soit la mesure de similarité utilisée, en dehors des zones de discontinuité.

Les plans étant très faiblement inclinés, les variations de profondeur autour du pixel à apparier sont moindres. L'hypothèse, stipulant que tous les pixels de la fenêtre doivent être caractérisés par la même valeur de disparité, est « quasiment » vérifiée, ce qui engendre peu d'erreurs d'appariement. La méthode 2D permet ainsi d'aller chercher l'information assez loin du pixel à apparier sans engendrer d'erreurs de disparité puisque les plans sont très peu inclinés, ce qui n'est pas en la faveur des fenêtres 1D.

Pour ces trois cas de figures, les résultats obtenus, à fenêtre identique, avec les deux mesures de similarité sont très semblables.

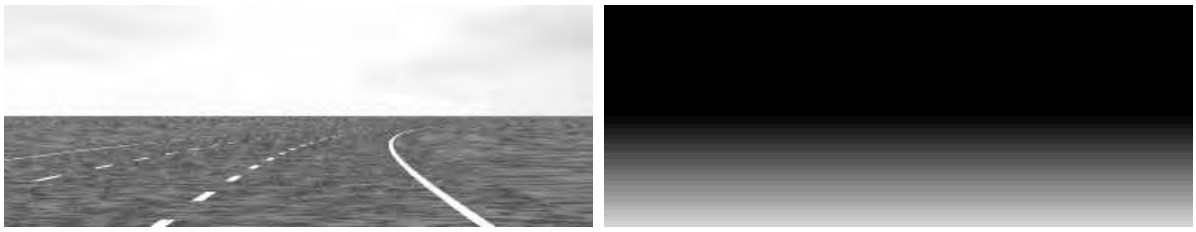
En zone de discontinuité (figure 2.18(d)), la méthode 1D génère moins d'erreurs d'appariement notamment pour les fenêtres de largeur moyenne comprise entre 7 et 19. La mesure SSD semble une fois de plus fournir des résultats moins précis ; ceci est encore plus marqué avec l'utilisation de fenêtres 2D.

Pour le cas précis du couple d'images *Venus*, on observe que les résultats obtenus pour les deux mesures de similarité ne diffèrent que très peu. Pour ce type d'images, la mise en correspondance sur la base de fenêtres 2D fournit de bien meilleurs résultats d'appariement qu'avec des fenêtres 1D. Ceci est directement lié à la composition de l'image qui est caractérisée par des plans quasi verticaux (et donc de disparité quasi constante) ; en effet, pour le plan de l'image *Venus* ayant l'inclinaison verticale la plus importante, on calcule suivant une colonne de ce plan, une évolution d'environ 0.04 disparité par ligne. Sur une fenêtre 2D de 21 pixels de hauteur, cela correspond à une variation inférieure à une disparité entre les lignes extrêmes de la fenêtre. Or, on comprend bien qu'en présence d'un plan véritablement incliné, l'effet de moyennage lié à la forme des fenêtres 2D devient un inconvénient certain. De ce fait, l'évaluation de Middlebury ne permet pas de mettre en évidence certains inconvénients de l'utilisation de fenêtres 2D puisque les plans sont très peu inclinés et les erreurs de disparité doivent être supérieures à un ($|\Delta\varepsilon| > 1.0$). Cependant, peu d'images de la littérature permettent de s'affranchir réellement de cette limite. Pour y remédier, nous avons généré un couple d'images de synthèse sur lequel une comparaison 1D/2D identique a été opérée.

2.5.4 Résultats obtenus sur une chaussée

Cette image de synthèse représentant une route a été générée par le logiciel POV-Ray^{TM} pour un système stéréoscopique dont les caractéristiques sont les suivantes :

- entraxe de 300 mm ;
- distance focale de 40 mm ;
- largeur du capteur de 8.46 mm ;
- résolution horizontale du capteur de 384 pixels.



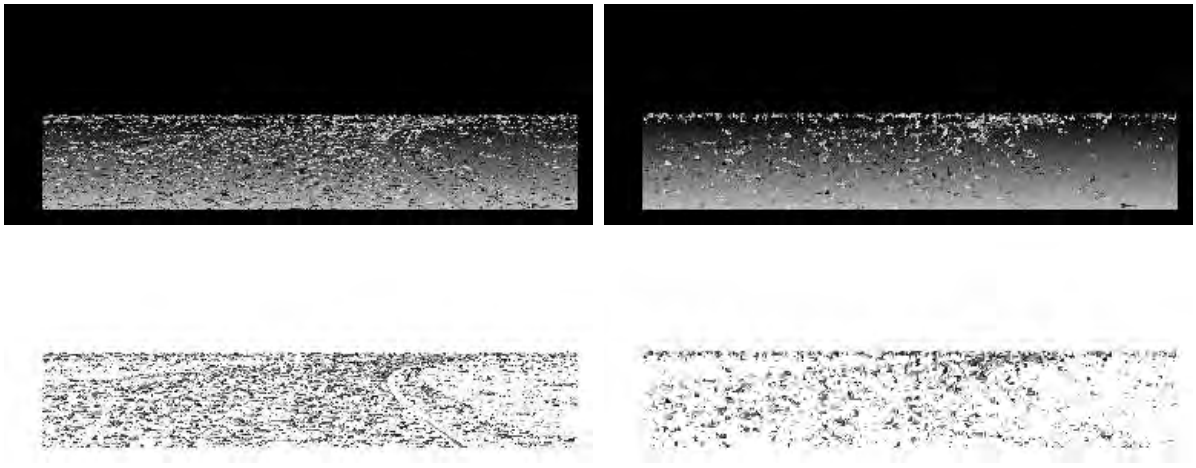
(a) (b)

Fig. 2.19 : (a) Image de synthèse gauche, (b) Vérité-terrain.

Dans cette configuration, l'inclinaison du plan de la route est d'environ 0.29 disparité par ligne, autrement dit une évolution d'environ 6 pixels entre les lignes extrêmes d'une fenêtre 2D de 21 pixels de hauteur. La figure 2.19 présente l'image gauche générée et la vérité-terrain associée.

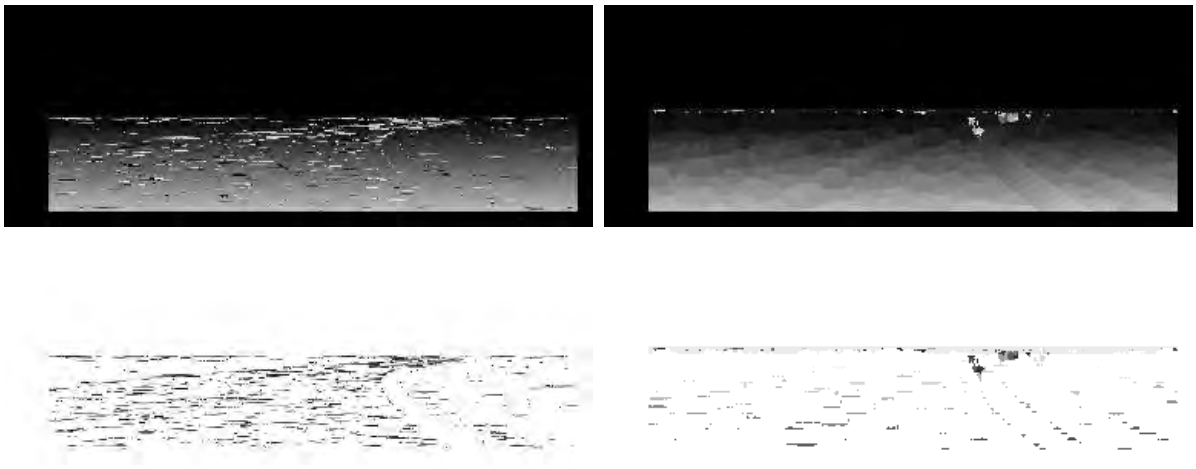
2.5.4.1 Comparaison qualitative

Comme pour le couple d'images *Head and Lamp*, les figures 2.20, 2.21 et 2.22 présentent les cartes des disparités obtenues par l'algorithme SSD avec respectivement une fenêtre 2D et une fenêtre 1D. La seconde ligne de ces figures montre les pixels dont la disparité calculée est erronée (dont la différence en valeur absolue avec la vérité-terrain est supérieure à un) : plus l'erreur est importante, plus le niveau de gris du pixel est faible.



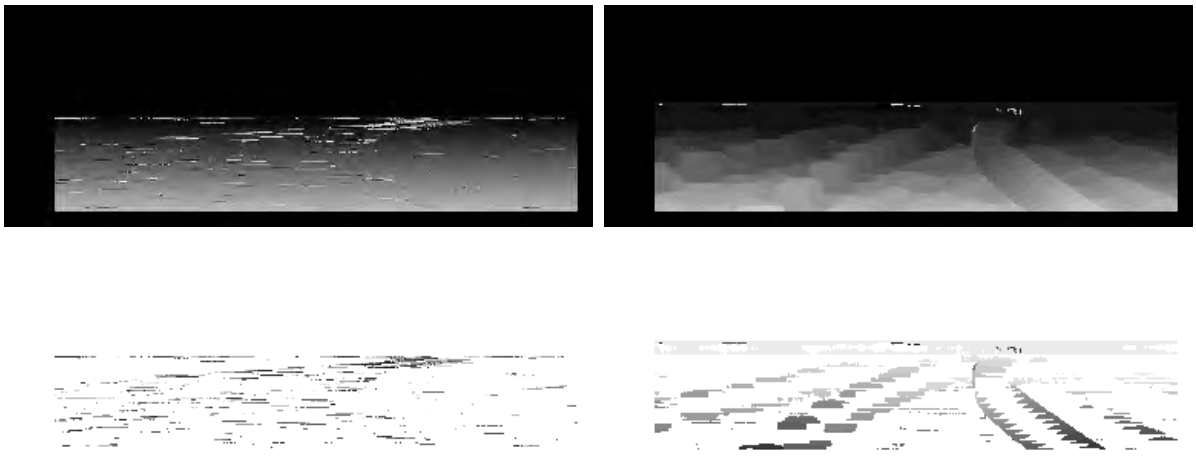
(a) (b)

Fig. 2.20 : Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de formes différentes de taille : (a) 3×1, (b) 3×3 ; la première ligne présente les cartes des disparités et la seconde les cartes d'erreur.



(a) (b)

Fig. 2.21 : Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de formes différentes de taille : (a) 11×1, (b) 11×11 ; la première ligne présente les cartes des disparités et la seconde les cartes d'erreur.



(a) (b)

Fig. 2.22 : Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de formes différentes de taille : (a) 19×1, (b) 19×19 ; la première ligne présente les cartes des disparités et la seconde les cartes d'erreur.

On constate visuellement qu'une fenêtre 2D de petite taille fournit de meilleurs résultats que les fenêtres 1D. Cependant, pour les fenêtres de taille plus importante, on constate que les fenêtres 2D engendrent des effets de moyennage et affectent la même valeur de disparité sur une importante zone de l'image ce qui génère des zones d'erreur importantes. Dans cette situation, la fenêtre 1D contient suffisamment d'information pour effectuer de bons appariements et affecter la bonne valeur de disparité.

2.5.4.2 Comparaison quantitative

La figure 2.23 présente les taux d'erreur calculés sur l'image complète ; la zone texturée représente toute l'image et aucune discontinuité n'est présente dans cette scène. Les scores de similarité ont été calculés en faisant varier la taille de la fenêtre de 3×3 à 21×21 pour les fenêtres 2D et de 3×1 à 21×1 pour les fenêtres 1D. Les différentes courbes nous montrent que la mesure de similarité SSD présente des résultats moins bons que ceux obtenus avec la mesure SAD.

Les taux d'erreur induits par l'utilisation de fenêtres de similarité 1D décroissent avec l'augmentation de la largeur de fenêtre. A l'inverse, l'augmentation de la taille de la fenêtre 2D entraîne une croissance importante des taux d'erreur.

Ainsi, pour les petites tailles de fenêtres, la fenêtre 1D ne contient pas assez d'information pour engendrer un appariement efficace. Pour les plus grandes tailles de fenêtres, la fenêtre 2D entraîne de nombreuses erreurs d'appariement car elle affecte la valeur de disparité d'une ligne donnée à ses voisines. Cet effet néfaste du 2D est moins marqué pour les fenêtres de petite taille : la tolérance des fenêtres 2D, par rapport au non-respect de l'hypothèse de constance de la disparité et aux erreurs induites, est directement liée à la taille de fenêtre utilisée et à l'inclinaison du plan dans la scène.

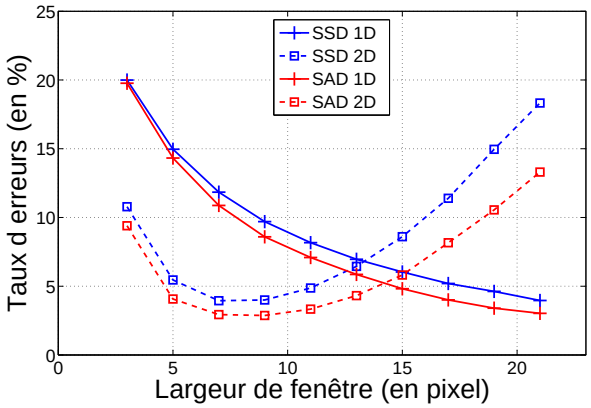


Fig. 2.23 : Comparaison quantitative des résultats de disparité obtenus avec les mesures SAD et SSD basées sur des fenêtres de formes carrée (2D) et ligne (1D) de différentes tailles.

2.6 Conclusion

Dans ce chapitre, des travaux basés sur un stéréoscope linéaire ont été présentés. Ces travaux utilisent uniquement l'information située sur la ligne pour mener à bien le processus de mise en correspondance. Les résultats obtenus sont de bonne qualité au prix d'une calibration méticuleuse.

Des questionnements quant à la pertinence de l'utilisation de fenêtres 2D pour la mise en correspondance ont été soulevés puisque l'information située sur la ligne courante semble suffisante pour effectuer des appariements efficaces. Ces résultats de mise en correspondance obtenus en stéréovision linéaire servent de caution et de point de départ à nos travaux.

Une comparaison brute, sur la base de deux mesures basiques de similarité, a permis de mettre en évidence l'impact des fenêtres 2D et 1D sur la qualité de la mise en correspondance stéréoscopique et d'appréhender ainsi leur comportement en diverses situations (zone de discontinuité, zones texturées et non-texturées).

Les différents résultats présentés sont obtenus à partir d'images de référence réelles ou de synthèse. Les images de synthèse présentent une succession de plans quasi-verticaux placés à différentes distances alors que les images réelles présentent des scènes caractérisant de petits espaces dépourvus de réelle perspective.

Dans ces situations, il semble que l'utilisation du 2D engendre des résultats discutables notamment en zones de discontinuité. En effet, le taux de mauvais appariements dus à la fenêtre 2D est, dans ce cas, supérieur à celui obtenu par l'utilisation de fenêtres 1D.

Aussi, pour les zones texturées, la mise en correspondance sur la base de fenêtres 1D est suffisante pour obtenir un appariement efficace. Il semble évident que si l'information est suffisante sur la ligne, une fenêtre 1D permet un appariement efficace. Dans ce cas précis, l'utilisation d'un voisinage 2D augmente le risque du non-respect de l'hypothèse de constance de la disparité, notamment dans le cas d'images à forte perspective.

De manière générale, les résultats obtenus avec une fenêtre 2D sont meilleurs. Cela est directement lié à la méthode d'évaluation de Middlebury qui ne considère pas comme erronée toute disparité dont la différence en valeur absolue avec la vérité-terrain est inférieure ou égale à un pixel. Pour une application, liée à la « métrologie », cela n'est pas acceptable.

Aussi, avec le couple *Sawtooth* composé de plans très légèrement inclinés, le protocole ne permet pas de montrer les limites du 2D dans de telles situations. A l'intérieur de ces plans inclinés, la fenêtre 2D va chercher de l'information très loin et permet un appariement et ce, même dans les zones non texturées. Cette affectation systématique de valeurs de disparité, même quand l'information locale ne le permet pas, peut être discutée.

Dans ce chapitre, nous avons montré que l'utilisation de fenêtres 2D n'est pas vraiment fondée puisque dans le cas d'images rectifiées, l'information se situe sur la ligne : la mise en correspondance est fondamentalement un problème 1D. Cependant, la mise en correspondance par corrélation sur la base de fenêtres 1D n'a pas été explorée à notre connaissance ; une méthode bien conçue pourrait mener à n'utiliser que l'information 1D et fournir des résultats précis et denses dans la plupart des situations. De plus, une information de confiance liée à chaque mesure de profondeur permettrait de pondérer

cette affectation de disparité en fonction du contenu informationnel de la fenêtre.

Aussi, on comprend aisément que pour traiter le cas d'une scène contenant des plans très inclinés voire horizontaux, on ne pourra pas utiliser des fenêtres bidimensionnelles de similarité pour l'algorithme de mise en correspondance, au risque de rencontrer de sérieux problèmes de précision. Dans cette situation, une fenêtre 2D est alors composée de pixels dont les disparités sont différentes, entraînant des erreurs lors de la phase d'appariement.

Dans le chapitre suivant, nous présentons une approche générique originale basée sur la mise en correspondance de fenêtres 1D. L'objectif est d'obtenir une carte des disparités aussi dense et sûre que possible.

Références

[1] J.L. BRUYELLE : *Conception et réalisation d'un dispositif de prise de vue stéréoscopique linéaire - Application à la détection d'obstacles à l'avant de véhicules routiers.* Thèse de doctorat, Université des Sciences et Technologies de Lille, 1994. 42, 46

[2] J.C. BURIE : *Mise en correspondance d'images stéréoscopiques linéaires - Application à la détection d'obstacles à l'avant de véhicules routiers.* Thèse de doctorat, Université des Sciences et Technologies de Lille, 1995. 43

[3] Y. RUICHEK : *Stéréovision linéaire par réseau de neurones de Hopfield.* Thèse de doctorat, Université des Sciences et Technologies de Lille, 1997. 13, 44

[4] Y. RUICHEK, H. ISSA et J.G. POSTAIRE : Genetic approach for obstacle detection using linear stereo vision. *IEEE International Vehicles Symposium*, pages 261–266, Dearborn, MI, USA, octobre 2000. 30, 31, 44

[5] D. DOOZE : *Conception et réalisation d'un stéréoscope bi-modal à portée variable - Application à la détection d'obstacles à l'avant de véhicules guidés automatisés.* Thèse de doctorat, Université des Sciences et Technologies de Lille, 2001. 45

[6] V. LEMONDE : *Stéréovision embarquée sur véhicule : De l'auto-calibrage à la détection d'obstacles.* Thèse de doctorat, Institut National des Sciences Appliquées de Toulouse, 2005. 2, 3, 33, 46, 110, 115

[7] S. NEDEVSCHI, C.VANCEA, T. MARITA et T. GRAF : On-line calibration method for stereovision systems used in vehicle applications. *IEEE International Transportation Systems Conference*, pages 957–962, Toronto, Canada, septembre 2006. 46

[8] A. FUSIELLO, E. TRUCCO et V. ROBERTO : A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000. 15, 46, 143

[9] D. SCHARSTEIN et R. SZELISKI : A taxomomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1): 7–42, avril 2002. **www.middlebury.edu/stereo**. 8, 35, 46, 47, 54, 60, 63, 72, 77, 86, 88, 89, 90, 91, 92

[10] S. CHAMBON : *Mise en correspondance stéréoscopique d'images couleur en présence d'occultations.* Thèse de doctorat, Université Paul Sabatier - Toulouse III, 2005. 30, 47

[11] Q. YÁNG, L. WANG, R. YANG, H. STEWÉNIUS et D. NISTÉR : Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2347–2354, New York, NY, USA, juin 2006. 35, 51

Chapitre 3

L'approche 1D proposée

Sommaire

3.1 Le principe de l'approche 1D 72
3.1.1 Calcul des scores de similarité 73
3.1.2 Modélisation de la confiance 75
3.1.3 La phase de décision 76
3.1.4 Validation directe/inverse 77
3.1.5 Raffinement des résultats 77
3.2 Etude de cas 77
3.2.1 Mise en correspondance 1D multi-configurations basée sur la mesure de dissimilarité *SSD* 77
3.2.2 Modélisation de la confiance par logique floue 79
3.2.3 Décision 83
3.3 Résultats 86
3.3.1 Evaluation de Middlebury 86
3.3.2 Méthodes comparables 87
3.3.3 Comparaison des résultats obtenus par les sept méthodes . . . 88
3.3.4 Evaluation sur un plan incliné 94
3.4 Conclusion 100
Références 102

LE CHAPITRE précédent a permis la mise en exergue du comportement des algorithmes de stéréovision dense utilisant deux types de fenêtres, 1D et 2D, lors du processus de mise en correspondance stéréoscopique par corrélation. L'utilisation systématique de fenêtres bidimensionnelles dans l'ensemble des travaux de stéréovision par corrélation relatés dans la littérature ne nous convainc pas. En effet, dans le cas d'images rectifiées, si l'information située sur la ligne du pixel à apparier est suffisante, il semble inutile d'aller chercher de l'information supplémentaire sur les lignes voisines : comme nous l'avons déjà écrit, la mise en correspondance stéréoscopique par corrélation est fondamentalement un problème 1D.

L'utilisation de fenêtres de similarité bidimensionnelles induit un certain nombre d'erreurs d'appariement qui, pour leur grande majorité, sont dues à l'effet de moyennage lié à cette forme de fenêtre. Pour une fenêtre 1D, le risque d'englober plusieurs valeurs de disparité est limité à la seule dimension horizontale.

De plus, l'affectation systématique de disparités dans des zones de l'image dépourvues d'information ne semble pas pertinente ; associer une valeur de confiance à chaque valeur de disparité semble être une voie plus intéressante.

Forts de ce constat, nous proposons dans ce chapitre une approche générique 1D originale capable de fournir une carte quasi-dense des disparités associée à sa carte des confiances. Dans un premier temps, l'approche générique 1D est explicitée de manière à en dégager les grandes idées. Dans un deuxième temps, une étude de cas permettra le choix de différentes techniques afin de remplir le rôle de chaque bloc fonctionnel de notre approche 1D (le choix de ces techniques sera aussi conditionné par leur potentiel d'implantation).

Puisque notre approche 1D a pour ambition d'être générique, nous l'évaluons avec la méthode académique très complète proposée par Scharstein et Szeliski [1] sur le site de Middlebury[1]. Ainsi, les résultats fournis par notre approche sont détaillés et comparés à ceux obtenus par des méthodes similaires décrites dans l'évaluation de Middlebury. Des résultats complémentaires seront présentés sur plusieurs couples d'images issues de l'observation d'un plan incliné.

3.1 Le principe de l'approche 1D

Comme explicité précédemment, la légitimité de l'utilisation systématique de fenêtres 2D lors de la phase de mise en correspondance peut être discutée. Avec notre méthode, nous voulons montrer que, dans la plupart des situations, une fenêtre de similarité 1D est suffisante pour établir un appariement efficace si nous nous plaçons dans le cas d'un couple d'images rectifiées.

L'approche 1D permettra d'associer une valeur de confiance à chaque valeur de disparité affectée : cette valeur de confiance quantifiera la véracité de la valeur de disparité affectée. Les zones de l'image peu texturées pourront trouver une valeur de disparité qui sera pondérée en fonction de l'information locale autour du pixel à apparier : si l'information est pauvre, la confiance associée sera faible.

Enfin, la méthode 1D exploitera uniquement l'information lumineuse portée par les pixels. Aucune information de plus haut-niveau, telle l'extraction de contours ou une quelconque étape de segmentation, ne sera prise en compte dans la phase d'appariement. Notre méthode 1D se situe au plus bas-niveau possible et se base sur l'étude des courbes de similarité.

[1] www.middlebury.edu/stereo

Dans sa globalité, l'approche 1D se décompose en cinq étapes distinctes (comme présenté figure 3.1) que nous allons décrire dans la suite de ce chapitre.

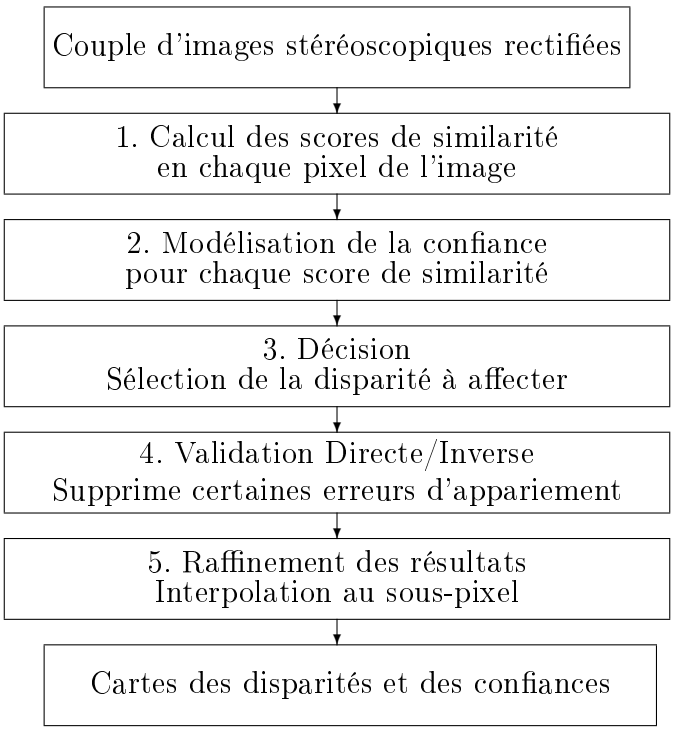


Fig. 3.1 : Descriptif de l'approche 1D.

3.1.1 Calcul des scores de similarité

Les performances des méthodes de stéréovision par corrélation sont particulièrement sensibles au choix des fenêtres de calcul. Trop petites, elles ne contiennent pas suffisamment d'information. Trop grandes, elles peuvent contenir des éléments appartenant à des objets différents situés à des distances différentes. Comme nous l'avons décrit dans le premier chapitre de ce mémoire, certaines méthodes proposent de trouver, en chaque pixel des images gauche et droite, la taille et la géométrie optimales du voisinage afin de lever les ambiguïtés qui apparaissent lorsque le voisinage est de taille fixe.

Les algorithmes classiques de stéréovision dense basés sur la mise en correspondance de fenêtres font l'hypothèse que tous les pixels contenus dans une fenêtre sont caractérisés par la même valeur de disparité. Ainsi, déterminer la fenêtre de similarité optimale est une étape cruciale pour l'obtention d'un appariement correct et le faire sans connaissance a priori de la scène observée est souvent impossible. Déterminer les paramètres de cette fenêtre (taille et position) revient à prendre une décision en début d'algorithme. Nous proposons par le biais de notre méthode 1D de décaler cette étape de décision en fin de

processus. De cette manière, nous calculons les scores de similarité pour plusieurs tailles de fenêtres 1D : cela permet d'obtenir de bons appariements avec une petite fenêtre si l'information locale le permet ou avec une fenêtre de largeur élevée si l'information lumineuse locale est insuffisante. Ce choix est fait sous l'hypothèse que la disparité est constante dans la fenêtre. De plus, plusieurs positions du pixel à apparier dans la fenêtre de similarité sont utilisées. Ceci permet de positionner la fenêtre dans l'image de telle sorte que la contrainte de constante de la disparité, notamment dans des zones de discontinuité, soit respectée.

Soit $w = 2n + 1$ la largeur d'une fenêtre 1D Ω_n, pour laquelle n est un entier compris entre $[1, n_{max}]$. Cette fenêtre correspond à un voisinage du pixel à apparier (x, y). Soit p la position du pixel à apparier dans la fenêtre Ω_n, avec $p \in [-n..n]$. Une fenêtre identique est placée dans l'autre image, au pixel (x, y), et décalée d'une valeur entière s telle que $s \in [s_{min}, s_{max}]$.

La position du pixel à apparier dans la fenêtre de similarité a une importance cruciale. La figure 3.2 présente schématiquement trois fenêtres 1D pour lesquelles le pixel à apparier est situé en différentes positions. Sur cet exemple, on constate que l'hypothèse de départ, stipulant que chaque pixel contenu dans la fenêtre doit avoir la même profondeur, est validée pour une seule position du pixel à apparier parmi les trois positions présentées (ici la première position). Dans ce cas, la courbe de similarité obtenue présente un extremum très marqué pour la disparité de la lampe. Pour les deux autres courbes, les disparités de l'arrière plan perturbent les courbes de similarité : l'extremum pour la disparité de la lampe est moins marqué, sa position pouvant même être modifiée.

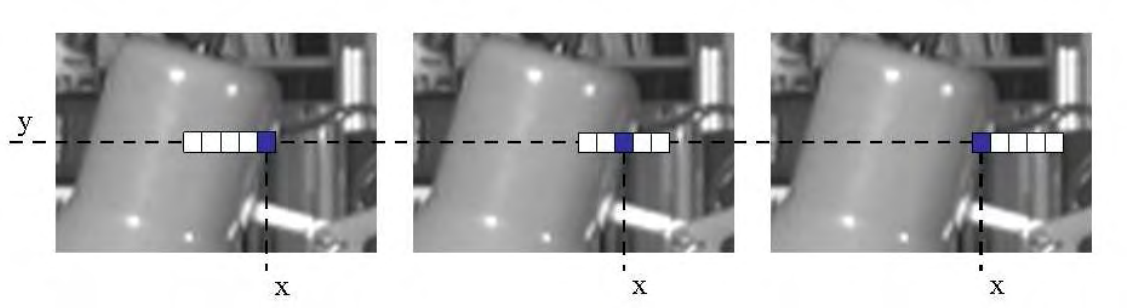


FIG. 3.2 : Pixel à apparier placé en différentes positions dans une fenêtre 1D.

Un score de similarité $C(x, y, p, w, s)$ est donc calculé pour une fenêtre 1D de largeur w ($w \in [3, 2n_{max} + 1]$) décalée de s et pour laquelle la position du pixel (x, y) à apparier est p. L'ensemble des scores de similarité forme ce que nous appelons le volume de similarité présenté figure 3.3 : chaque cube de ce volume correspond ainsi à un score de similarité $C(x, y, p, w, s)$. Chaque ligne de cubes dans ce volume de similarité suivant la direction s, représente une courbe de similarité pour une taille donnée de la fenêtre et une position donnée du pixel à apparier dans cette fenêtre. La mesure de similarité pourra être choisie parmi les nombreuses mesures disponibles dans la littérature.

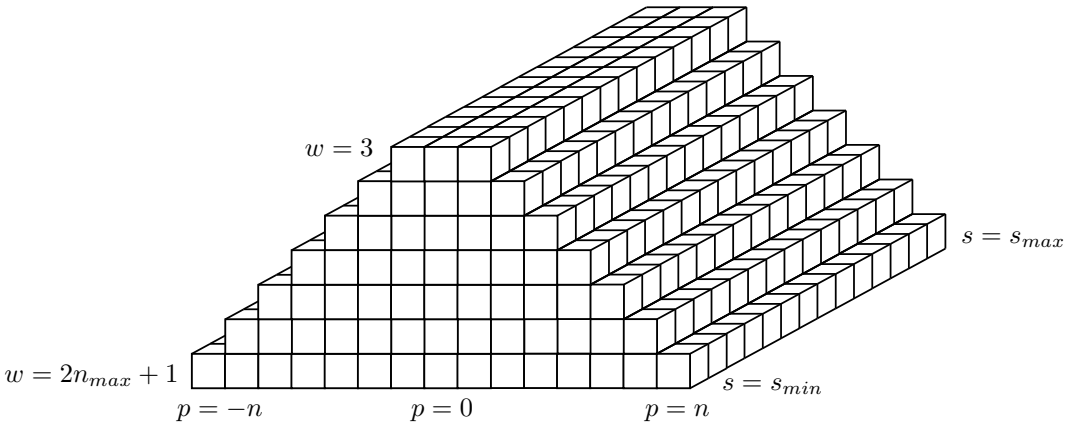


FIG. 3.3 : Volume de similarité, pour un pixel donné (x, y), défini par l'ensemble des scores de similarité toutes configurations confondues de la fenêtre.

Ce volume de similarité permet la mise à disposition, de façon exhaustive, des mesures de similarité pour chaque configuration de fenêtre 1D pour un pixel donné (x, y). L'étape suivante va consister à modéliser une valeur de confiance pour chaque score de similarité calculé.

3.1.2 Modélisation de la confiance

Cette étape de notre approche 1D a pour finalité l'association d'une valeur de confiance à chaque score de similarité, c'est-à-dire à chaque cube du volume de similarité. On obtient ainsi un volume de confiance équivalent à celui des mesures de similarité (figure 3.3).

Selon la configuration de la fenêtre, les courbes de similarité peuvent avoir des caractéristiques très variées et ce, pour un même pixel de l'image. La figure 3.4 présente, pour le pixel de coordonnées $(301, 51)$ dans l'image gauche *Head and Lamp*, plusieurs courbes de dissimilarité calculées selon différentes configurations de la fenêtre (différentes tailles et différentes positions du pixel à apparier). La valeur minimale de chaque courbe marque la même valeur de décalage ($s = 6$). Cependant, ces courbes ont des formes très différentes. Utilisée seule, la disparité retenue peut ne pas être associée à la même valeur de confiance. Pour les figures 3.4(c) et 3.4(e), la vallée étroite caractérise un appariement certain et semble marquer la valeur de disparité sans aucune ambiguïté. Par contre, les courbes des figures 3.4(a), 3.4(b) et 3.4(d), sont caractérisées par une ou plusieurs vallées peu marquées. Dans ce cas, la configuration de fenêtre utilisée pour l'appariement (même si la disparité retenue semble exacte) ne permet pas de lui affecter une confiance élevée.

Des auteurs ont déjà proposé l'association d'une valeur de confiance lors de l'étape de mise en correspondance [2] [3] [4]. Plusieurs méthodes de mesure de critères relatifs à la confiance ont déjà été décrites dans la littérature [5].

Dans [2], l'auteur analyse la différence entre les deux premiers minima de la courbe de

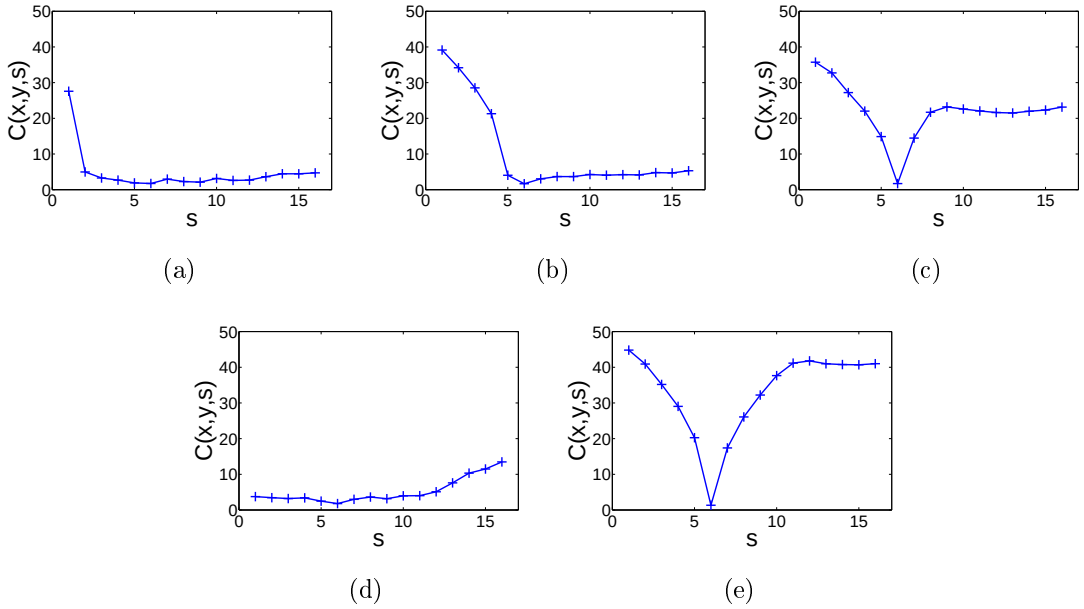


FIG. 3.4 : Courbes de dissimilarité pour différentes configurations de la fenêtre : (a) $w = 7$ et $p = 0$, (b) $w = 13$ et $p = 0$, (c) $w = 19$ et $p = 0$, (d) $w = 13$ et $p = -6$, (e) $w = 13$ et $p = 6$.

dissimilarité. Si la valeur est élevée, la configuration des fenêtres est validée et le coefficient de confiance élevé. Dans [3], l'appariement n'est validé que si le minimum de dissimilarité est suffisamment marqué. Si la confiance est faible, certains auteurs préfèrent ne pas affecter de valeur de disparité au pixel à apparier et calculer des cartes des disparités dites semi-denses ou quasi-denses [6] [7].

Cette étape de l'approche 1D est ainsi fondée sur l'extraction de critères pouvant quantifier la confiance accordée à la mesure de similarité pour un décalage donné. Dans la suite de ce chapitre, nous préciserons les critères utilisés lors de l'étude de cas.

3.1.3 La phase de décision

Cette partie de l'approche 1D a pour but d'analyser le volume de confiance afin de déterminer une valeur précise de disparité pour le plus grand nombre de pixels de l'image : ceci permet l'établissement de la carte des disparités et de sa carte des confiances associées.

Les possibilités d'analyse de ce volume sont multiples et le volume de données à traiter est important. De ce fait, il est difficile d'opter pour l'une ou l'autre méthode. Cependant, cette étape constitue l'une des étapes les plus importantes de notre méthode. Dans la suite de ce chapitre, l'étude de cas présente une technique simple et efficace de traitement du volume des confiances.

3.1.4 Validation directe/inverse

La validation directe/inverse (ou contrainte de symétrie) est une méthode classique largement relatée dans la littérature qui permet la suppression de certaines erreurs d'appariement [8]. Cette étape a été détaillée dans le premier chapitre de ce mémoire.

Lorsque la validation directe/inverse n'est pas réalisée, les valeurs associées (disparité et confiance) en ce pixel (x, y) sont mises à zéro dans les cartes des disparités et des confiances (prenant comme image de référence l'image gauche).

3.1.5 Raffinement des résultats

De nombreux auteurs ajoutent une étape de raffinement des résultats au sous-pixel après avoir effectué un appariement au pixel près [3].

Dans l'évaluation de Middlebury [1], cette étape est nécessaire puisque les cartes de vérité-terrain associées sont à la précision sous-pixel (sauf pour le couple *Head and Lamp*).

Ici, l'étape de raffinement est basée sur l'interpolation par une courbe du second ordre (ou parabole) des scores de similarité obtenus pour les pixels voisins du décalage retenu (comme décrit dans [9]).

Dans la première partie de ce chapitre, les grandes étapes de notre approche ont été présentées. Des scores de similarité sur la base de fenêtres 1D multi-configuration sont calculés. A chaque score de similarité, un indice de confiance est affecté. Ensuite, une étape de décision détermine la valeur de disparité, si l'information locale autour du pixel à apparier le permet, à laquelle il associe une confiance. La section suivante détaille une étude de cas présentant les techniques retenues lors d'une première mise en œuvre de notre méthode.

3.2 Etude de cas

Les scores de similarité, les indices de confiance et les cartes finales des disparités et des confiances peuvent être établis en utilisant des théories, outils ou méthodes très variés. Dans la suite de ce chapitre, nous présentons une étude de cas pour laquelle les méthodes les plus basiques ont été sélectionnées pour chaque bloc fonctionnel (figure 3.1) de notre approche. L'objectif est de valider la méthode 1D dans sa globalité et non pas de valider la complexité de chaque bloc fonctionnel.

3.2.1 Mise en correspondance 1D multi-configurations basée sur la mesure de dissimilarité *SSD*

Comme nous l'avons souligné dans le premier chapitre, de nombreuses expressions du calcul de similarité sont disponibles dans la littérature. Ils fournissent des résultats

dont la qualité est inégale et dont la rapidité de traitement est différente. Ces expressions s'articulent pour la plupart autour de l'étude de l'intensité lumineuse locale.

Dans [10], Mayoral *et al.* évaluent quatre méthodes basées sur différentes combinaisons de l'intensité lumineuse locale : la différence pixel à pixel (« Absolute Differences »), l'agrégation centrée des différences pixel à pixel sur une région support composée de plusieurs fenêtres, l'application de filtres basés sur l'opérateur de Sobel et une mesure ordinale. Ils concluent que l'approche par filtres associée à la méthode de décision locale WTA fournit les meilleurs résultats et que la mesure ordinale fournit de bons résultats pour les méthodes globales.

Afin de montrer que notre approche 1D fournit de bons résultats quelle que soit la mesure de similarité choisie, nous avons sélectionné l'une des plus classiques, c'est-à-dire le méthode SSD. De plus, nous avons vu dans le chapitre précédent que la mesure SSD engendrait plus d'erreurs d'appariement que la mesure SAD dans la majorité des cas. Le challenge semblait encore plus intéressant.

Ainsi, la mesure de dissimilarité $C(x, y, p, w, s)$ est définie par l'équation suivante :

$$C(x, y, p, w, s) = \sum_{i=-p-n}^{i=-p+n} (I_g(x+i, y) - I_d(x+i-s, y))^2 \tag{3.1}$$

$$\text{avec} \quad \begin{cases} n \in [0, n_{max}] \\ w = 2n+1 \\ p \in [-n, n] \\ s \in [s_{min}, s_{max}] \end{cases} , \tag{3.2}$$

où I_g et I_d représentent respectivement les niveaux de gris des pixels dans les images gauche et droite. Dans la suite de ce mémoire, les paramètres x et y ont été ôtés des équations afin de simplifier les notations.

Pour des valeurs données de p et w, la courbe de dissimilarité est définie comme l'ensemble des valeurs $C(p, w, s)$ avec $s \in [0, s_{max}]$:

$$\{C(p, w, s)\}_{s \in [0, s_{max}]} \ . \tag{3.3}$$

Les scores de dissimilarité SSD sont calculés sur la base de fenêtres 1D dont la taille est comprise entre 3×1 et 21×1 (soit $n \in [1, 10]$ et $n_{max} = 10$). Nous avons opté pour des fenêtres de moyenne et petite taille de manière à prouver leur capacité à fournir un appariement efficace comme illustré pour les résultats de la figure 2.17. Ainsi, nous analysons les scores de dissimilarité pour n_{max} tailles de fenêtres et $(2n_{max} + 1)$ positions du pixel à apparier. Chaque score de dissimilarité est considéré comme une source d'information qui pourra être combinée à d'autres sources afin de déterminer la disparité pour le pixel courant.

3.2.2 Modélisation de la confiance par logique floue

Il s'agit ici d'affecter une valeur de confiance à chaque valeur de dissimilarité. Dans ce travail, aucune connaissance globale de l'image n'est requise : seule l'information bas niveau, portée par les scores de dissimilarité, sera utilisée.

Afin de quantifier la confiance en un appariement, des caractéristiques issues des courbes de similarité peuvent être extraites ([2] [3] [11]). Dans la littérature, les caractéristiques extraites sont utilisées individuellement afin de quantifier la véracité de l'appariement. Ici, pour chaque courbe de dissimilarité, nous proposons de combiner plusieurs caractéristiques extraites, ce qui devrait rendre la modélisation de la confiance plus robuste. Différentes théories formalisant l'incertain peuvent être utilisées comme celles liées à la croyance, la possibilité et la probabilité.

Afin de combiner plusieurs critères, les techniques de logique floue semblent bien adaptées. Ces techniques sont souvent utilisées dans des applications pour lesquelles on dispose d'une expertise humaine pour résoudre un problème donné ; par ailleurs, ces techniques permettent de formaliser l'usage de termes vagues et de les rendre utilisables par les ordinateurs [12].

Par notre approche, une valeur de confiance associée en chaque score de dissimilarité sera modélisée par l'utilisation d'un filtre flou. Une valeur de confiance $Conf_{LF}(p, w, s)$ est ainsi calculée pour chaque position p du pixel à apparier dans la fenêtre de taille w et pour chaque décalage s. Une unique valeur de confiance est associée à chaque décalage de manière à quantifier la vraisemblance que cette valeur de décalage soit la disparité à affecter au pixel courant.

Ce filtre flou se décompose en quatre étapes :
- extraction des caractéristiques de la courbe de dissimilarité ;
- fuzzyfication des caractéristiques extraites ;
- mécanisme d'inférence ;
- défuzzification de manière à obtenir la valeur de confiance $Conf_{LF}(p, w, s)$.

Dans notre approche, nous avons fait le choix d'extraire trois caractéristiques de chaque courbe de dissimilarité :

- la courbure Cur : ce critère local mesure la courbure de la courbe de dissimilarité en chaque décalage (hors décalages extrêmes). Plus la vallée de dissimilarité est étroite, plus la courbure est élevée et plus l'appariement est efficace.
  La mesure de la courbure $Cur(p, w, s)$ est définie comme :

$$Cur(p, w, s) = \quad -2 \cdot C(p, w, s) + C(p, w, s+1) + C(p, w, s-1) \; ; \qquad (3.4)$$

- le rang $R(p, w, s)$. Les scores de dissimilarité sont classés par ordre décroissant et le rang du score de dissimilarité est défini par sa position dans ce classement. Dans

le cas idéal, une seule vallée caractérise la courbe de dissimilarité et le décalage qui apparaît en première position dans ce classement est la valeur de disparité. Cependant, les courbes de dissimilarité sont souvent composées de plusieurs vallées et la première valeur de décalage dans ce classement n'est pas toujours la bonne disparité. Dans ce cas, la confiance diminue ;

– le nombre $N(p,w)$ de points d'inflexion caractérisant le nombre de vallées de dissimilarité (parties convexes de la courbe). Une courbe de dissimilarité avec une seule vallée présente moins d'ambiguïté qu'une autre composée de plusieurs vallées. Cette caractéristique globale $N(p,w)$ est définie par le nombre de points pour lesquels :

$$\frac{\partial^2 C(p,w,s)}{\partial s^2} = 0\,. \tag{3.5}$$

Ces trois critères extraits sont modélisés en utilisant des ensembles flous selon trois états (mauvais, moyen et bon) définis par des fonctions d'appartenance de forme classique (triangulaire ou trapézoïdale).

Par exemple,

– si $Cur(p,w,s)$ est élevée, l'état de la fonction d'appartenance associée est *bon* ;
– si $N(p,w)$ et $R(p,w,s)$ sont élevés, l'état de leur fonction d'appartenance respective est *mauvais*.

Les figures 3.5(a), 3.5(b) et 3.5(c) présentent respectivement les fonctions d'appartenance des trois critères $Cur(p,w,s)$, $N(p,w)$ et $R(p,w,s)$ selon ces trois états : mauvais, moyen et bon.

Le filtre flou est fondé sur le mécanisme d'inférence standard **SI-ALORS-SINON**. Chaque règle d'inférence calcule une confiance intermédiaire CF pour une valeur donnée de p, w et s. Les 27 (3^3) règles d'inférence ont été définies en analysant le comportement de ces trois caractéristiques Cur, R, et N sur des voisinages typiques tels qu'en zones texturées, non texturées, près des discontinuités et ce, pour la mesure de dissimilarité SSD. La comparaison détaillée, présentée au chapitre précédent, a joué un rôle prépondérant dans l'acquisition de l'expertise nécessaire à l'établissement de ces règles d'inférence. Chaque règle fournit une valeur de confiance floue CF qui peut prendre cinq états : nul, mauvais, moyen, bon, excellent.

Certaines règles, comme la suivante, sont évidentes :

SI **Cur(p,w,s)** est *bon*,
ET **R(p,w,s)** est *bon*,
ET **N(p,w)** est *bon*
ALORS **CF(p,w,s)** est *excellent*.

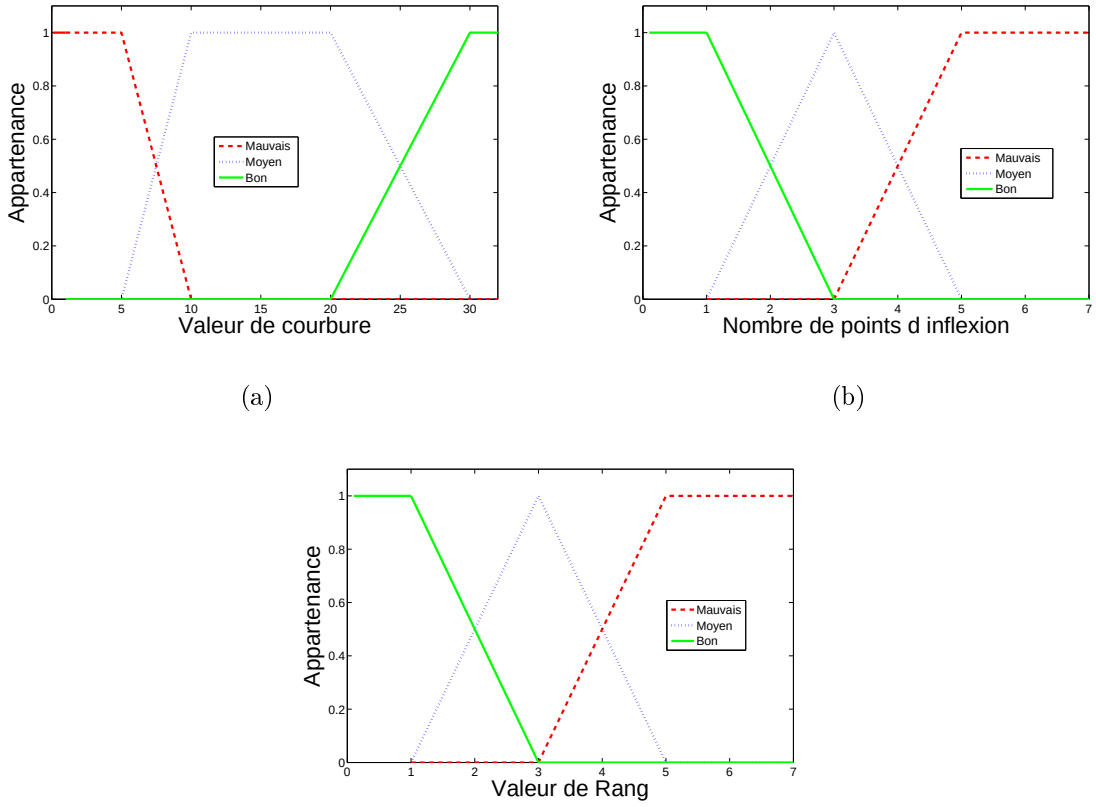


FIG. 3.5 : (a) Fonctions d'appartenance du critère Cur, (b) Fonctions d'appartenance du critère N et (c) Fonctions d'appartenance du critère R.

Cependant, pour de nombreuses règles, il est difficile de définir précisément les relations entre les variables floues.

Par exemple, pour deux valeurs de décalage, telles que $s = 1$ et $s = 2$, d'une même courbe de dissimilarité,

$$\text{si } (Cur(p, w, 1) = bon \ ET \ R(p, w, 1) = mauvais)$$
$$\text{et } (Cur(p, w, 2) = moyen \ ET \ R(p, w, 2) = bon),$$

il n'est pas aisé de définir laquelle de ces deux valeurs de décalage fournira la confiance la plus élevée ($N(p, w)$ étant identique). Dans ce cas, les confiances associées seront caractérisées comme *moyenne* : c'est l'étape de décision qui sera déterminante.

Les règles d'inférence, qui lient les entrées floues aux sorties floues du filtre, sont présentées par le tableau 3.1 : les entrées sont définies par les trois états mauvais, moyen et bon ; les cinq états de la sortie sont définis par nul, mauvais, moyen, bon et excellent.

	Entrées			Sortie
Règle	R	Cur	N	CF
1	Mauvais	Mauvais	Mauvais	Nul
2	Mauvais	Mauvais	Moyen	Nul
3	Mauvais	Mauvais	Bon	Nul
4	Mauvais	Moyen	Mauvais	Mauvais
5	Mauvais	Moyen	Moyen	Nul
6	Mauvais	Moyen	Bon	Mauvais
7	Mauvais	Bon	Mauvais	Mauvais
8	Mauvais	Bon	Moyen	Mauvais
9	Mauvais	Bon	Bon	Moyen
10	Moyen	Mauvais	Mauvais	Mauvais
11	Moyen	Mauvais	Moyen	Mauvais
12	Moyen	Mauvais	Bon	Mauvais
13	Moyen	Moyen	Mauvais	Mauvais
14	Moyen	Moyen	Moyen	Mauvais
15	Moyen	Moyen	Bon	Moyen
16	Moyen	Bon	Mauvais	Mauvais
17	Moyen	Bon	Moyen	Moyen
18	Moyen	Bon	Bon	Moyen
19	Bon	Mauvais	Mauvais	Mauvais
20	Bon	Mauvais	Moyen	Mauvais
21	Bon	Mauvais	Bon	Moyen
22	Bon	Moyen	Mauvais	Moyen
23	Bon	Moyen	Moyen	Moyen
24	Bon	Moyen	Bon	Bon
25	Bon	Bon	Mauvais	Moyen
26	Bon	Bon	Moyen	Bon
27	Bon	Bon	Bon	Excellent

Tab. 3.1 : Règles d'inférence du filtre flou.

Les règles d'inférences font appel aux opérateurs *ET, OU et NON*, qui s'appliquent aux variables floues. Dans le cas de la logique binaire, ces opérateurs sont définis de façon simple et univoque. Dans le cas de la logique floue, la définition de ces opérateurs n'est plus univoque et on utilise le plus souvent les relations simples telles que *minimum* ou *produit* pour l'opérateur *ET*, *maximum* ou *valeur moyenne* pour l'opérateur *OU*. Une généralisation des opérations de combinaison de type *minimum* et *maximum* est respectivement définie par les $t-normes$ et $t-conormes$. Les opérations *minimum* et *maximum* présentent l'avantage de la simplicité lors du calcul, par contre, elles privilégient l'une des deux variables. Les opérations de *produit* et *valeur moyenne* sont plus complexes à calculer mais elles produisent un résultat qui tient compte des valeurs des deux variables. Fidèles à notre volonté de choisir les outils/opérateurs les plus élémentaires, nous optons

pour le *minimum* et le *maximum* pour définir respectivement les opérateurs *ET* et *OU*.

Lors de l'étape de défuzzification, les 27 valeurs intermédiaires de confiance floue CF sont combinées en utilisant la méthode classique du centre de gravité de manière à déterminer une valeur numérique de confiance globale $Conf_{LF}(p, w, s)$ à affecter au score de dissimilarité $C(p, w, s)$.

En définitive, le filtre flou est donc basé sur la combinaison de trois caractéristiques issues de la courbe de dissimilarité par le biais de techniques floues très basiques. Les règles ont été définies de manière heuristique. Le problème de leur portabilité à un autre couple peut se poser. Nous verrons, notamment dans ce chapitre et le chapitre suivant, que ces règles sont robustes au changement de situations de test.

Dans la suite de ce mémoire, les résultats présentés sont tous basés sur ce même filtre flou (aucun paramètre n'est ajusté) de manière à évaluer l'efficacité de ce dernier quel que soit le type d'images traitées.

3.2.3 Décision

Dans cette section, nous définissons une méthode de décision capable d'affecter la disparité la plus vraisemblable pour la plupart des pixels. Cette méthode réduit le volume des confiances selon différentes coupes explicitées par la suite.

Tout d'abord, le volume est réduit selon les décalages s. Le décalage $s^*(p, w)$, correspondant au maximum de confiance pour chaque position p du pixel courant dans la fenêtre de taille w, est sélectionné :

$$s^*(p, w) = \arg[\max_s \{Conf_{LF}(p, w, s)\}] \ . \tag{3.6}$$

Le volume réduit suivant la relation (3.6) est présenté figure 3.6. Ici, le volume des données est défini pour n_{max} égal à 7 et s compris dans l'intervalle $[0, 13]$. De cette manière, le décalage de la fenêtre présentant la confiance maximale est retenu pour chaque courbe des similarités, celui-ci étant a priori la valeur de disparité pour la configuration de la fenêtre donnée. Dans un soucis de lisibilité, les valeurs de décalage $s^*(p, w)$ sont codées sur différentes couleurs. Dans l'exemple, on constate que le volume réduit propose deux valeurs potentielles de disparité (5 et 11) à affecter au pixel courant.

Ensuite, le volume initial est également réduit selon les positions p. Le nombre $N(w, s)$ d'occurrences, par valeur de décalage retenue $s^*(p, w)$, est calculé comme suit :

$$N(w, s) = Count_p[s = s^*(p, w)] \ . \tag{3.7}$$

Le volume réduit suivant la relation (3.7) est présenté figure 3.7. Seules les deux valeurs potentielles de disparité sont présentées ici ; les couleurs correspondent à la valeur de disparité et le nombre affiché dans chaque cube est le nombre d'occurrences $N(w, s)$ associé.

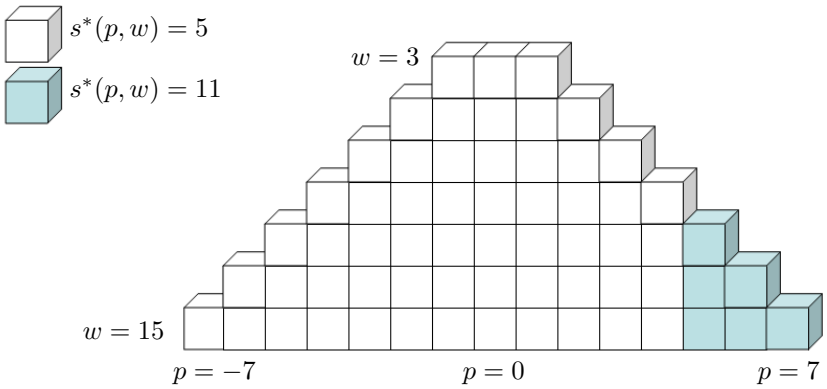


Fig. 3.6 : Volume réduit suivant la relation (3.6).

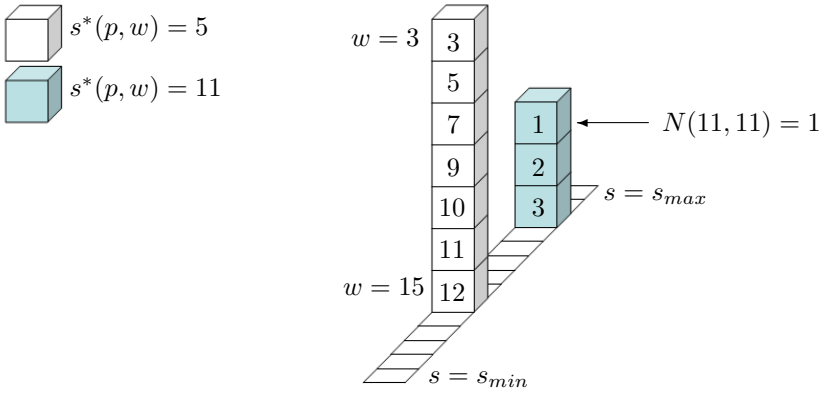


Fig. 3.7 : Volume réduit suivant la relation (3.7).

Par la suite, on peut déterminer la valeur de disparité $d^*(w)$ par le décalage $s^*(p,w)$ ayant le maximum d'occurrences, toutes positions du pixel à apparier confondues, par l'équation :

$$d^*(w) = \arg[\max_s \{N(w,s)\}] \; . \tag{3.8}$$

Le volume réduit suivant la relation (3.8) est présenté figure 3.8(a). On constate que la valeur de disparité $d^*(w) = 5$ propose l'occurrence maximale pour chaque largeur de fenêtre w. Seulement, il est possible que, malgré cela, toutes les positions du pixel à apparier dans la fenêtre ne proposent pas cette même valeur de décalage.

Ainsi, la proportion $T(w)$ du nombre d'occurrences $N(w,s)$ pour la largeur de fenêtre w et ce, pour chaque valeur de décalage $d^*(w)$, est établie de la manière suivante :

$$T(w) = \frac{N(w, d^*(w))}{w} \; . \tag{3.9}$$

Les données obtenues suivant la relation (3.9) sont présentées figure 3.8(b). On constate que la proportion n'est pas identique pour toutes les largeurs de fenêtres. Celle-ci diminue pour les largeurs de fenêtre les plus élevées ce qui indique que ces fenêtres prennent en

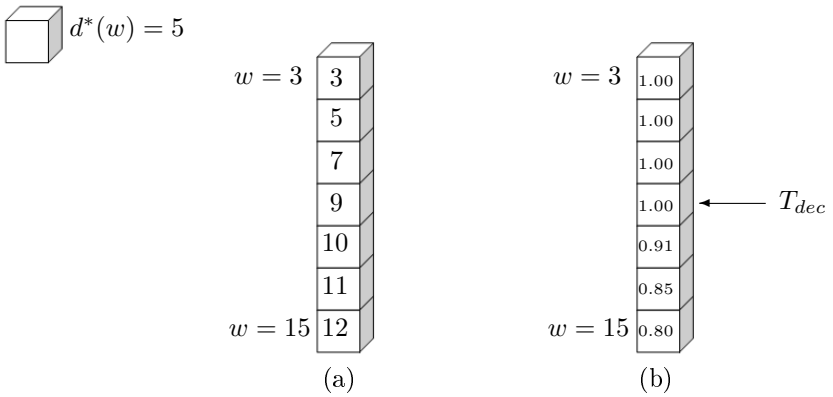


Fig. 3.8 : (a) Volume réduit suivant la relation (3.8) - (b) Données obtenues suivant la relation (3.9).

compte des pixels situés à des profondeurs différentes.

Au final, la valeur de disparité $d^*(w)$ est assignée au pixel (x, y) si la condition suivante est satisfaite :

$$T(w) \geq T_{dec} \ , \tag{3.10}$$

où T_{dec} définit un seuil. Lorsque T_{dec} vaut un, la condition présentée par la relation (3.10) signifie que la même valeur de disparité a été retenue pour toutes les positions p de la fenêtre de taille w. Si la condition n'est pas remplie, la disparité ne sera pas affectée pour cette largeur de fenêtre.

Les valeurs de disparité sont affectées en commençant par la fenêtre la plus grande et en diminuant la taille jusqu'à ce que la condition (3.10) soit remplie, cette taille de fenêtre étant notée $w_{T_{dec}}$.

Ce processus permet alors d'affecter les valeurs de disparité dans les zones de disparité homogènes puisque toutes les positions du pixel à apparier dans la fenêtre de similarité proposent la même valeur de disparité. En diminuant la taille de la fenêtre, on se rapproche des zones de discontinuité en affectant la bonne valeur de disparité puisqu'avec de petites fenêtres, seuls les pixels de l'objet sont pris en compte dans le processus d'appariement : ceci évite un grand nombre d'erreurs de décision.

Sur la figure 3.8(b), on constate que pour $T_{dec} = 1$, la condition n'est pas remplie pour $w = 15$, $w = 13$ et $w = 11$: la première largeur de fenêtre remplissant cette condition est la fenêtre de taille $w = 9$. Ainsi, la valeur de disparité 5 sera affectée au pixel (x, y), soit $d(x, y) = 5$.

Par la suite, la valeur de confiance finale $Conf(x, y)$ associée à la disparité affectée $d(x, y)$ est déterminée de telle sorte que $Conf(x, y) \in [0, 1]$ (1 étant la confiance maximale). Cette valeur est obtenue par la relation suivante :

$$Conf = \frac{1}{w}. \sum_{p} Conf_{LF}(p, w_{T_{dec}}, d^*(w_{T_{dec}})) \ . \tag{3.11}$$

Appliquer une étape de seuillage sur la carte des confiances permettra de réduire le nombre d'erreurs d'appariement, au détriment de la densité de la carte des disparités. Par la suite, le seuil pourra être judicieusement choisi par l'utilisateur en fonction de l'application visée.

3.3 Résultats

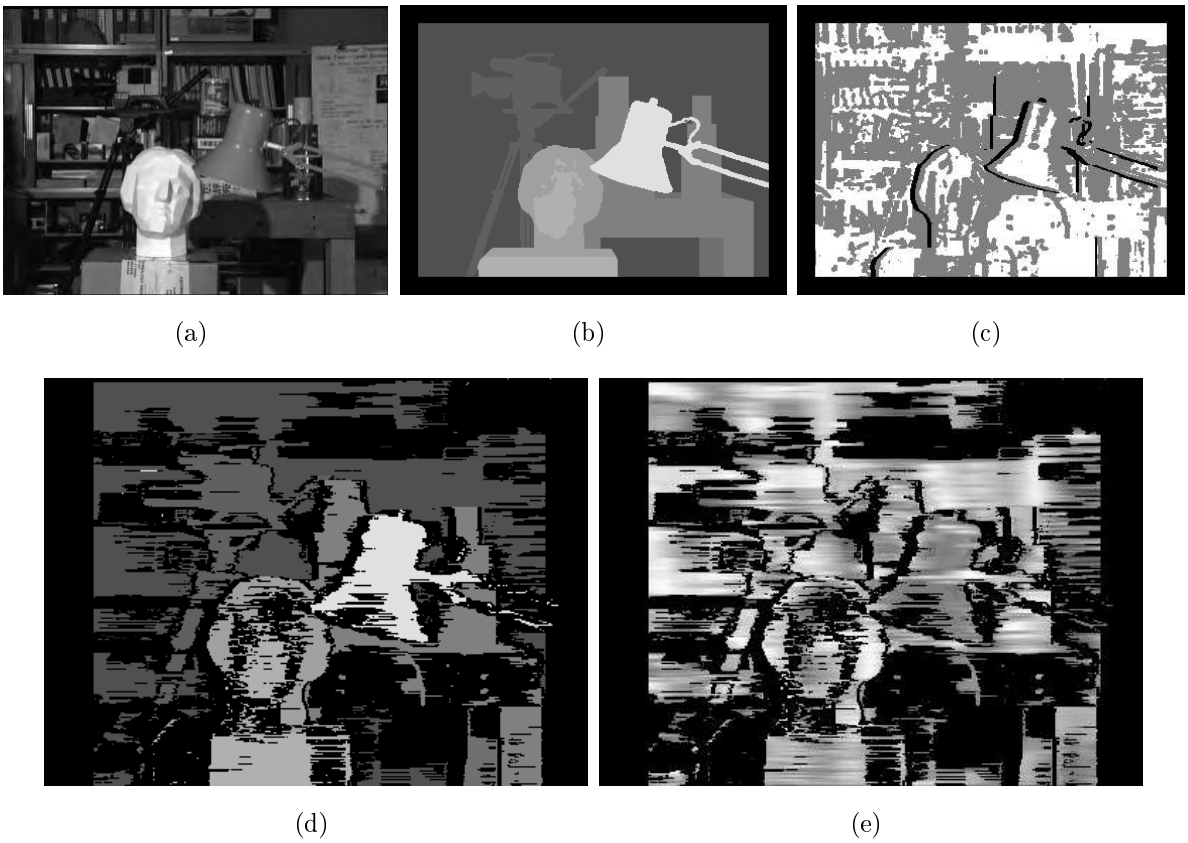
Dans cette partie, nous présentons tout d'abord les résultats obtenus par notre approche générique 1D dans le cadre de l'évaluation académique proposée dans [1]. Ces résultats sont comparés à ceux relatifs aux méthodes similaires que nous introduisons au préalable. Ensuite, nous présentons des résultats obtenus sur une scène composée d'un plan incliné de manière à montrer l'efficacité accrue de notre approche dans ce contexte.

3.3.1 Evaluation de Middlebury

L'évaluation proposée dans [1] a été largement utilisée dans un grand nombre de publications. Même si cette méthode d'évaluation comporte certaines limites (cf section 2.4.2) et qu'elle est surtout dédiée à l'évaluation de méthodes 2D, il semblait important de se plier à l'exercice et ainsi valider notre méthode générique dans un cadre académique. Ici, nous présentons les résultats associés à l'image *Head and Lamp* de l'université de *Tsukuba*. Nous avons obtenu ces résultats avec des tailles de fenêtres de dissimilarité variant entre 3×1 et 21×1. La figure 3.9 présente l'image gauche, la vérité-terrain, la carte des zones texturées et non-texturées ainsi que les cartes des disparités et des confiances obtenues sur le couple d'images *Head and Lamp*.

De manière qualitative, la carte des disparités que nous obtenons, semble assez proche de la vérité-terrain avec un assez grand nombre de pixels affectés. Globalement, on constate, sur la carte des disparités 3.9(d), que les pixels non affectés (pixels noirs) par notre méthode correspondent aux pixels situés en zones non texturées (pixels blancs) de la carte 3.9(c). Pour chaque pixel pour lequel une valeur de disparité est calculée, une valeur de confiance est associée ; la figure 3.9(e) montre la carte des confiances, les niveaux de gris élevés représentant les confiances les plus importantes.

Ces résultats, issus de la méthode 1D, vont être comparés de manière quantitative dans le suite de ce chapitre à d'autres méthodes largement relatées dans la littérature. Ces méthodes sont présentées dans la section suivante.



(a) (b) (c)

(d) (e)

FIG. 3.9 : (a) Image gauche *Head and Lamp*, (b) Vérité-terrain, (c) Zones texturées (pixel gris) et non texturées (pixels blancs), (d) Carte des disparités, (e) Carte des confiances.

3.3.2 Méthodes comparables

Nous avons comparé notre méthode à six méthodes « similaires », toutes basées sur la méthode locale de recherche exhaustive appelée « Winner-Take-All » (WTA). De telles méthodes sont utilisables pour nos systèmes de vision embarqués contrairement aux algorithmes reposant sur une approche globale. Ces dernières sont basées sur une recherche itérative nécessitant des temps de calcul trop importants, mais surtout dont le point de convergence associé n'est pas prévisible.

Le tableau 3.2 présente une synthèse des méthodes comparables retenues pour cette comparaison. En premier lieu, il précise la mesure de similarité utilisée (MES), le type de fenêtres ($SUPP$) et aussi la méthode d'optimisation (OPT) ; comme nous l'avons précisé précédemment, ces méthodes sont toutes basées sur la technique WTA. Ensuite, les différentes étapes des méthodes sont mises en évidence : le prétraitement (PRE), le raffinement des résultats (RAF), l'utilisation de l'information de couleur (COU) et la prise en compte des erreurs d'occultation (OCC).

REFERENCE	MES	SUPP	OPT	PRE	RAF	COU	OCC
Realtime [2]	SAD	2D mult.	WTA	-	Filtrage	-	oui
SSD+MF [1]	SSD	2D	WTA	-	Sous-pixel	oui	oui
Improved Coop. [13]	SAD,NCC	2D	WTA	Filtrage	Sous-pixel	oui	oui
Adapt. Weights [14]	SAD,SSD	2D adapt.	WTA	-	-	oui	-
Comp. Win. [15]	SAD,SSD	2D adapt.	WTA	-	-	-	-
MMHM colors [3]	SAD	2D	WTA	Filtrage	Sous-pixel	oui	oui

Tab. 3.2 : Tableau récapitulatif des méthodes comparables à notre approche 1D et relatées dans l'évaluation de Middlebury.

3.3.3 Comparaison des résultats obtenus par les sept méthodes

Pour chaque couple d'images de référence, les cartes des disparités calculées par notre méthode sont basées sur des fenêtres 1D de taille 3×1 à 21×1 avec $T_{dec} = 1$ (valeur par défaut). Les cartes des disparités calculées sont présentées figures 3.10(c), 3.11(c), 3.12(c) et 3.13(c). Les tableaux 3.3, 3.4, 3.5 et 3.6 présentent les résultats obtenus sur respectivement les couples d'images *Head and Lamp*, *Sawtooth*, *Venus* et *Map* pour différentes classes de pixels : les pixels non occultés $B_{\overline{\mathcal{O}}}$, les pixels situés en régions non-texturées $B_{\overline{\mathcal{T}}}$, en zones de discontinuité $B_{\mathcal{D}}$ et en régions texturées $B_{\mathcal{T}}$.

A ce jour, il n'existe aucune méthode de comparaison entre les cartes des disparités denses et les cartes des disparités quasi-denses. Nous avons donc mis en œuvre une méthode de comparaison simple (basée sur l'évaluation de Middlebury) que nous détaillons ici. Les taux d'erreur sont calculés en effectuant la différence pixel à pixel entre les cartes des disparités calculées et la vérité-terrain correspondante. La comparaison est réalisée sur l'ensemble des pixels affectés par notre méthode 1D, intersection commune des ensembles de pixels affectés par notre méthode quasi-dense et les méthodes denses utilisées. En effet, les autres méthodes n'associent pas de cartes des confiances à leurs cartes des disparités, aussi peut-on en déduire que chacune des disparités affectées par ces méthodes, est caractérisée par la même confiance. On peut donc sélectionner « à l'aveugle » une partie des pixels des cartes denses puisqu'ils sont a priori aussi bons les uns que les autres. Ensuite, si la valeur absolue de cette différence pixel à pixel est supérieure à un seuil donné $|\Delta\varepsilon|$, la disparité est considérée comme erronée : ici, la condition est fixée à $|\Delta\varepsilon| > 0.5$ soit le seuil minimum proposé dans l'évaluation de Middlebury. Ceci permettra de mettre en avant l'aspect métrologique de notre méthode qui privilégie la précision des résultats par rapport à la densité de la carte calculée.

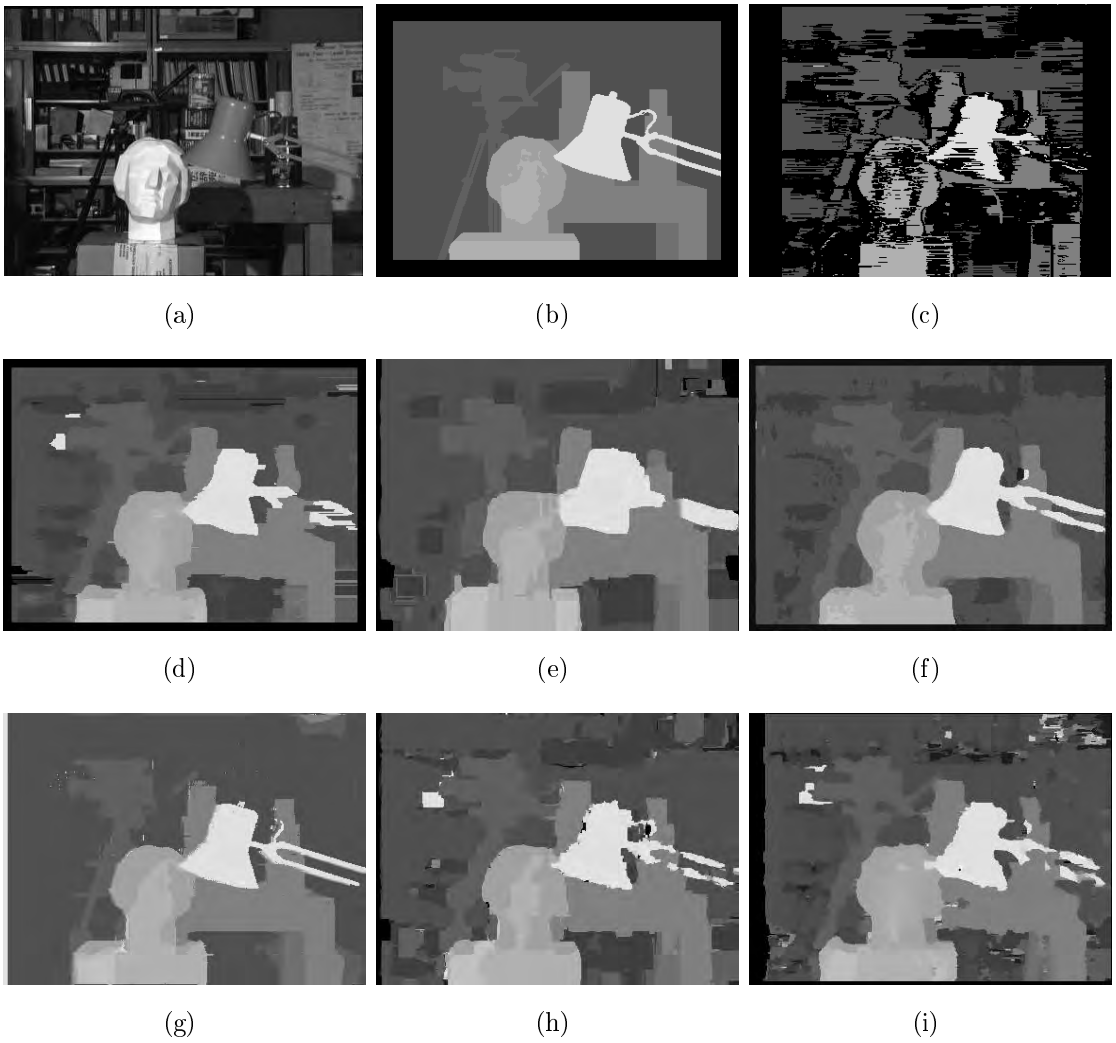
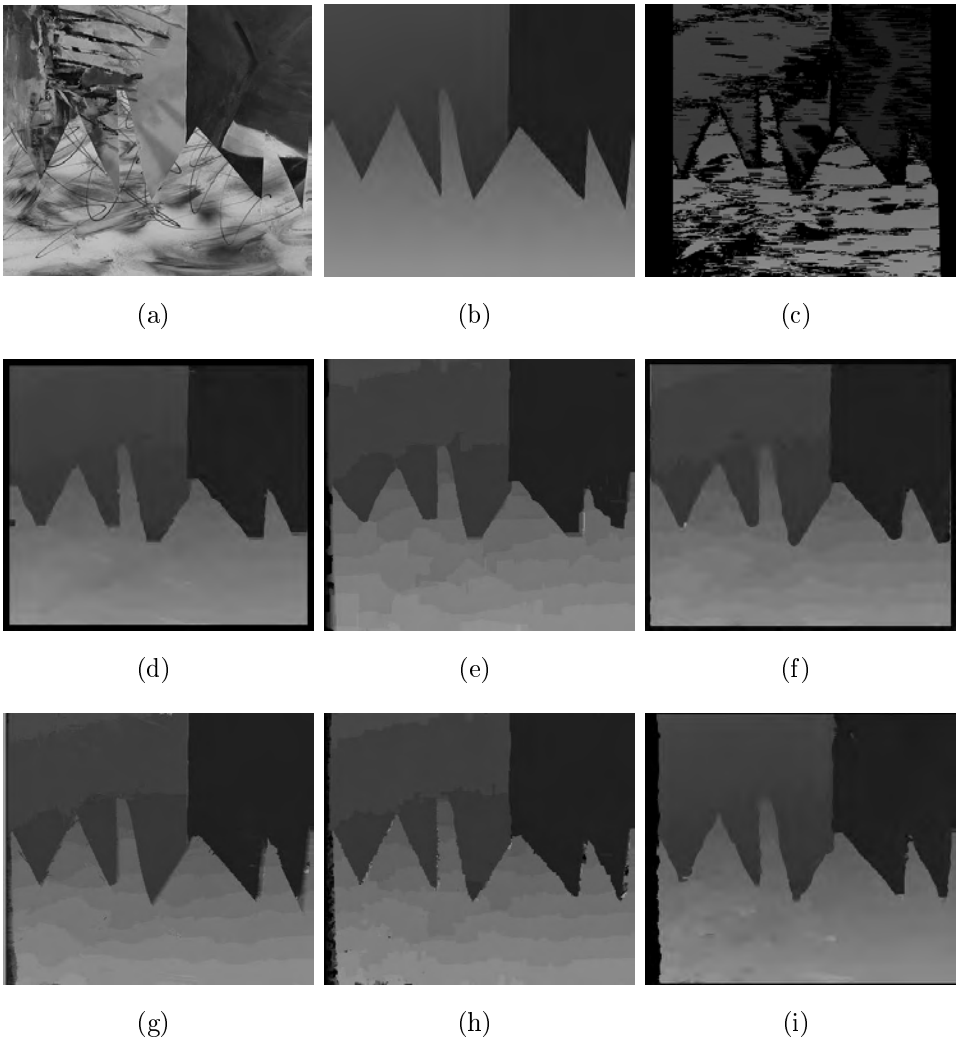


Fig. 3.10 : Image gauche *Head and Lamp* (a), Vérité-terrain (b), Notre méthode 1D (c), Realtime (d), $SSD + MF$ (e), Improved Coop. (f), Adapt. Weights (g), Comp. Win. (h), MMHM colors (i).

Méthodes	$B_{\overline{\mathcal{O}}}$	$B_{\overline{\mathcal{T}}}$	$B_{\mathcal{D}}$	$B_{\mathcal{T}}$
Realtime [2]	3.67^{6}	2.89^{6}	4.32^{4}	4.24^{5}
SSD+MF [1]	5.80^{7}	2.97^{7}	9.88^{7}	7.87^{7}
Improved Coop. [13]	3.10^{3}	2.35^{4}	3.75^{2}	3.65^{3}
Adapt. Weights [14]	2.47^{2}	0.80^{1}	4.39^{5}	3.69^{4}
Comp. Win. [15]	3.35^{5}	2.09^{3}	4.09^{3}	4.27^{6}
MMHM colors [3]	3.24^{4}	2.78^{5}	5.46^{6}	3.59^{2}
Notre méthode 1D	2.19^{1}	1.24^{2}	3.03^{1}	2.88^{1}
Densité	45.15	23.11	46.45	62.94

Tab. 3.3 : Résultats obtenus à partir de l'évaluation de Middlebury sur le couple d'images *Head and Lamp* (avec $|\Delta\varepsilon| > 0.5$).



(a) (b) (c)

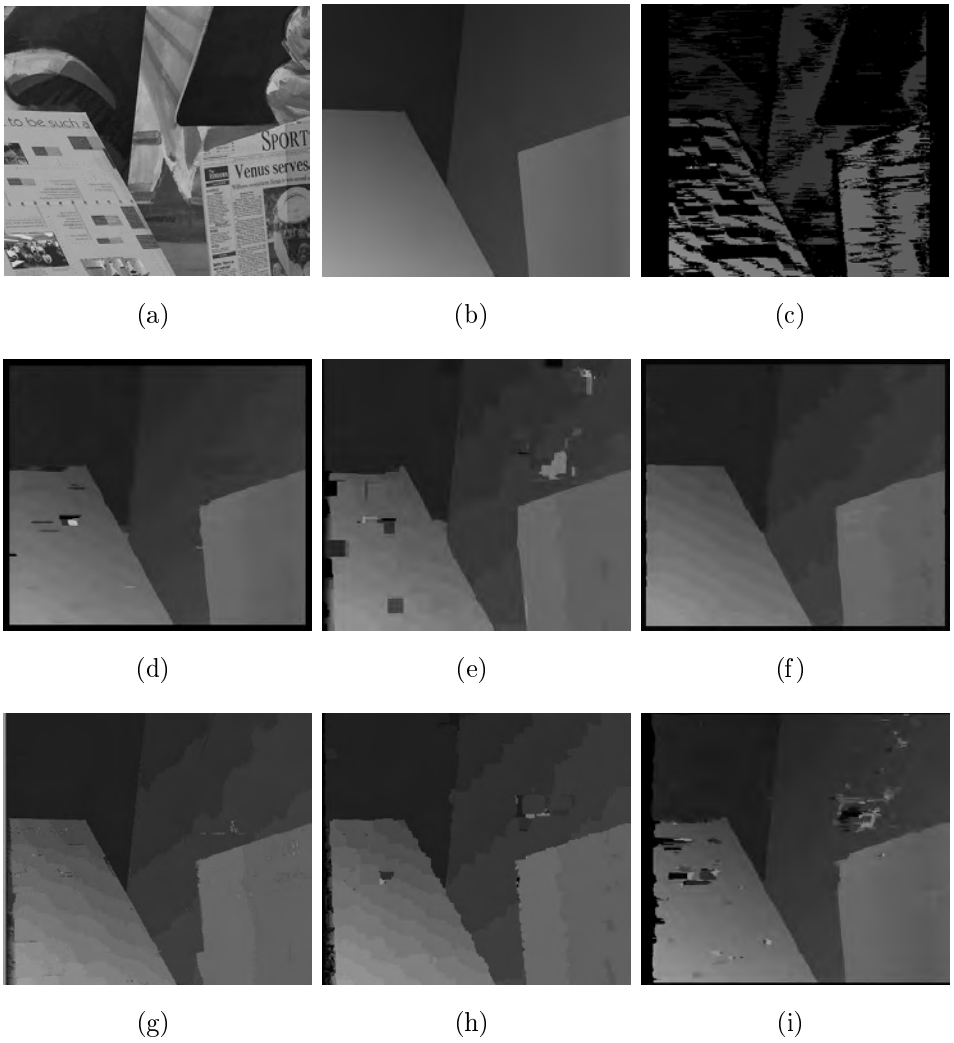
(d) (e) (f)

(g) (h) (i)

Fig. 3.11 : Image gauche *Sawtooth* (a), Vérité-terrain (b), Notre méthode 1D (c), Realtime (d), $SSD + MF$ (e), Improved Coop. (f), Adapt. Weights (g), Comp. Win. (h), MMHM colors (i).

Méthodes	$B_{\overline{\mathcal{O}}}$	$B_{\overline{\mathcal{T}}}$	$B_{\mathcal{D}}$	$B_{\mathcal{T}}$
Realtime [2]	2.95^{1}	0.16^{1}	4.02^{1}	3.24^{1}
SSD+MF [1]	4.66^{7}	0.93^{7}	8.13^{7}	5.04^{7}
Improved Coop. [13]	4.61^{6}	0.50^{4}	6.45^{6}	5.03^{6}
Adapt. Weights [14]	4.20^{5}	0.31^{2}	5.87^{5}	4.60^{5}
Comp. Win. [15]	3.86^{4}	0.44^{3}	4.98^{4}	4.21^{4}
MMHM colors [3]	3.84^{3}	0.56^{5}	4.59^{2}	4.17^{3}
Notre méthode 1D	3.51^{2}	0.60^{6}	4.87^{3}	3.81^{2}
Densité	52.95	20.44	43.17	57.52

Tab. 3.4 : Résultats obtenus à partir de l'évaluation de Middlebury sur le couple d'images *Sawtooth* (avec $|\Delta\varepsilon| > 0.5$).



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

FIG. 3.12 : Image gauche *Venus* (a), Vérité-terrain (b), Notre méthode 1D (c), Realtime (d), $SSD + MF$ (e), Improved Coop. (f), Adapt. Weights (g), Comp. Win. (h), MMHM colors (i).

Méthodes	$B_{\overline{\mathcal{O}}}$	$B_{\overline{\mathcal{T}}}$	$B_{\mathcal{D}}$	$B_{\mathcal{T}}$
Realtime [2]	0.32^{1}	0.09^{1}	1.30^{1}	0.43^{1}
SSD+MF [1]	1.59^{6}	0.36^{6}	5.70^{7}	2.19^{6}
Improved Coop. [13]	1.43^{5}	0.40^{7}	2.11^{3}	1.93^{5}
Adapt. Weights [14]	1.80^{7}	0.25^{3}	3.23^{6}	2.56^{7}
Comp. Win. [15]	0.75^{3}	0.26^{4}	2.76^{5}	0.99^{3}
MMHM colors [3]	0.37^{2}	0.13^{2}	2.26^{4}	0.48^{2}
Notre méthode 1D	0.93^{4}	0.29^{5}	1.91^{2}	1.24^{4}
Densité	37.79	11.71	40.17	51.22

TAB. 3.5 : Résultats obtenus à partir de l'évaluation de Middlebury sur le couple d'images *Venus* (avec $|\Delta\varepsilon| > 0.5$).

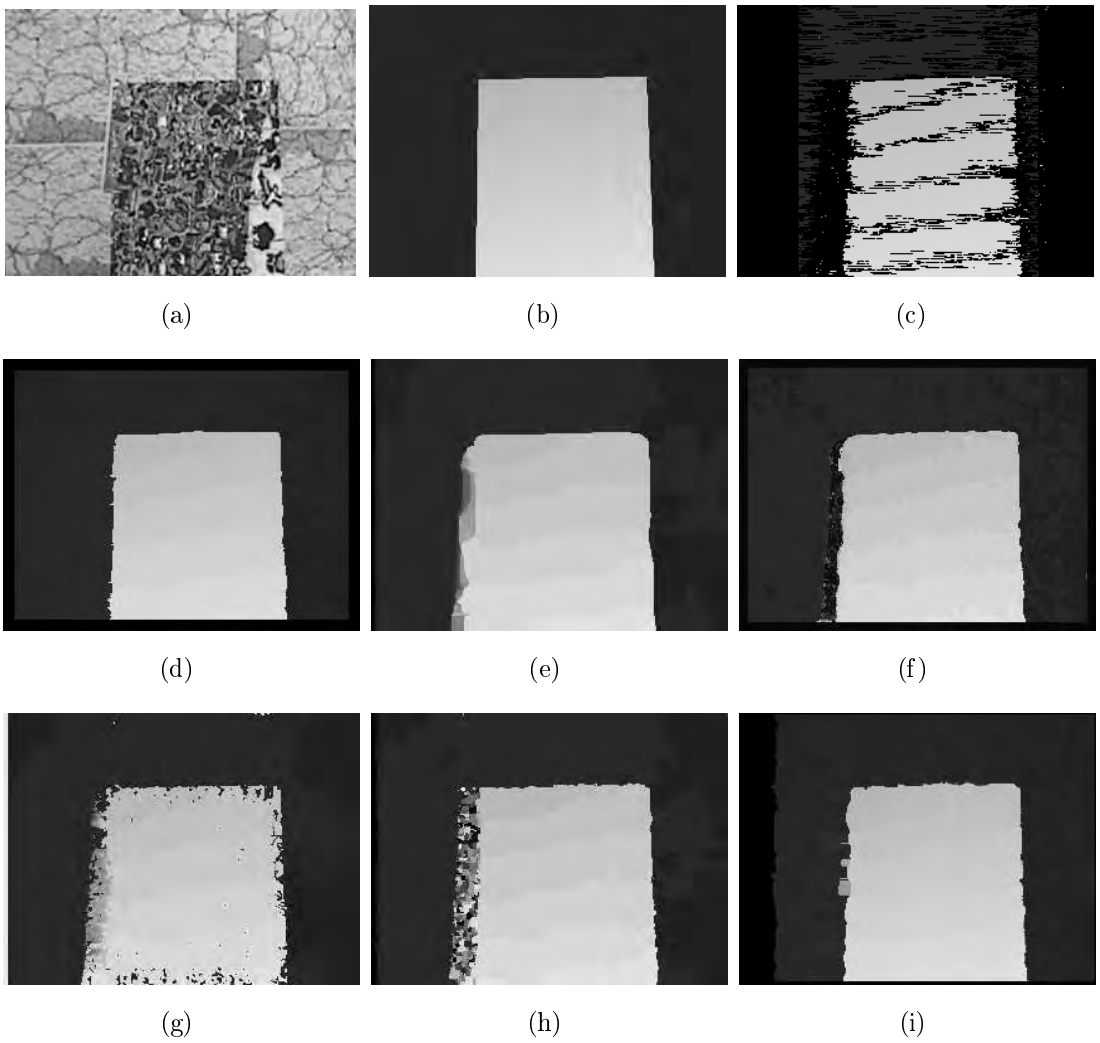


Fig. 3.13 : Image gauche *Map* (a), Vérité-terrain (b), Notre méthode 1D (c), Realtime (d), $SSD + MF$ (e), Improved Coop. (f), Adapt. Weights (g), Comp. Win. (h), MMHM colors (i).

Méthodes	$B_{\overline{\mathcal{O}}}$	$B_{\mathcal{D}}$
Realtime [2]	0.13^{1}	1.53^{2}
SSD+MF [1]	3.91^{6}	8.66^{6}
Improved Coop. [13]	3.56^{4}	3.07^{4}
Adapt. Weights [14]	4.43^{7}	9.12^{7}
Comp. Win. [15]	3.86^{5}	4.89^{5}
MMHM colors [3]	0.14^{2}	1.45^{1}
Notre méthode 1D	1.02^{3}	2.84^{3}
Densité	47.53	45.80

Tab. 3.6 : Résultats obtenus à partir de l'évaluation de Middlebury sur le couple d'images *Map* (avec $|\Delta\varepsilon| > 0.5$).

Comme le montrent les différents tableaux, les cartes des disparités calculées sont assez denses puisque la plupart des pixels sont affectés. On constate que notre méthode présente de bons résultats au regard des autres méthodes notamment pour les pixels en zones de discontinuité et texturées, autrement dit, lorsqu'il y a de l'information locale dans les images. L'exposant lié à chaque taux précise le classement de la méthode correspondante parmi les sept méthodes comparées.

Notre méthode 1D se positionne souvent en tête de classement. Les résultats sont tout particulièrement bons pour l'image *Head and Lamp*. Dans cette scène d'intérieur, les principaux objets sont caractérisés par une valeur de disparité constante mais ils sont situés à des profondeurs différentes ; de cette façon, sur la carte de vérité-terrain, ces objets sont représentés par des plans verticaux de disparité (la disparité étant constante). Ainsi, si la fenêtre est disposée sur deux objets différents, celle-ci contient deux valeurs de disparité réellement différentes et conduit à une erreur d'appariement d'au moins un pixel puisque cette vérité-terrain est au pixel près : dans l'évaluation de Middlebury, cela engendre des taux d'erreur élevés.

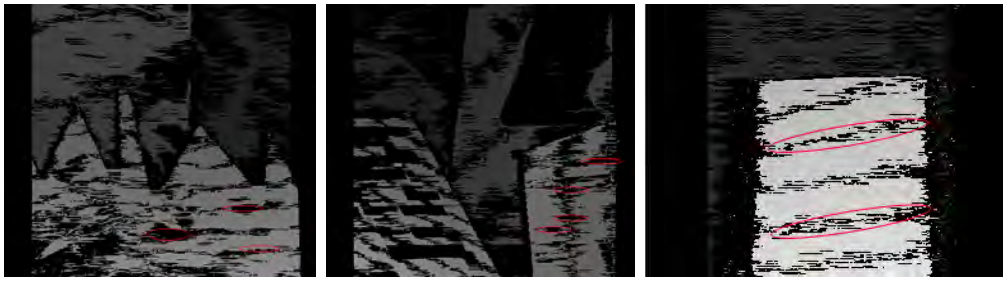
En définitive, ces résultats montrent bien que le choix d'utiliser une combinaison de plusieurs configurations de fenêtres 1D est pertinent, même si la méthode d'évaluation utilisée n'est, a posteriori, pas favorable à la mise en valeur de méthodes utilisant des fenêtres 1D.

Aussi, on peut remarquer l'originalité de notre approche puisque selon les zones de l'image, la densité diffère. Par exemple, elle diminue considérablement en zones non texturées, autrement dit dans les zones où l'information locale est moins riche.

La densité des cartes des disparités peut être augmentée en faisant varier le seuil de décision T_{dec} (ici $T_{dec} = 1$) au détriment de la précision. Seulement, comme chaque disparité est associée à une valeur de confiance, l'utilisateur pourra toujours revenir à une précision plus importante en filtrant les disparités de confiance faible.

La figure 3.14 permet de mettre en évidence certains comportements de notre méthode 1D. Dans un premier temps, on constate pour les couples d'images *Sawtooth*, *Venus* et *Map* (constitués de plans légèrement inclinés) que certaines zones continues (entourées en rouge) ne sont pas affectées par une valeur de disparité. Cependant ces zones devraient être affectées puisqu'elles sont considérées comme texturées. Ce phénomène sera détaillé dans la fin de ce mémoire puisqu'il constitue une perspective intéressante n'ayant pas encore fait l'objet d'un travail approfondi.

Avec notre approche 1D, le lecteur peut néanmoins se demander si l'étape de validation directe/inverse n'est pas en grande partie la source de ces bons résultats. Cette étape de suppression des erreurs d'appariement est utilisée par une grande majorité des méthodes de mise en correspondance par corrélation. Cependant, pour nos cartes des disparités quasi-denses, elle comporte un inconvénient certain. En effet, lorsque la validation



(a) (b) (c)

FIG. 3.14 : Cartes des disparités calculées avec notre méthode 1D sur les couples d'images (a) *Sawtooth*, (b) *Venus* et (c) *Map*.

directe/inverse est effectuée sur un couple composé d'un pixel affecté et d'un autre non affecté, la disparité est supprimée. Or, cette dernière peut être la bonne valeur de disparité à affecter au pixel. Le tableau 3.7 montre les résultats d'erreur calculés sur les cartes des disparités obtenues avant et après validation directe/inverse (VDI) et sur la carte des disparités supprimées par cette méthode de validation. On constate globalement que pour chaque catégorie de pixels, la grande majorité des pixels supprimés était caractérisée par la bonne valeur de disparité. Sur les 6.20 % de pixels supprimés seuls 20 % étaient erronés ce qui correspond à 1.28 % des pixels. Cet effet gênant de la validation directe/inverse pourrait être atténué en affectant la disparité non-nulle (de l'une des deux cartes des disparités) au pixel, tout en diminuant la confiance associée.

		$B_{\overline{\mathcal{O}}}$	$B_{\overline{\mathcal{T}}}$	$B_{\mathcal{D}}$	$B_{\mathcal{T}}$
Après VDI	Erreurs	2.19	1.24	3.03	2.88
	Densité	45.15	23.11	46.45	62.94
Avant VDI	Erreurs	3.47	2.35	4.55	4.28
	Densité	51.34	27.11	53.62	70.28
Pixels supprimés par la VDI	Erreurs	1.28	1.11	1.52	1.40
	Densité	6.20	4.00	7.17	7.34

TAB. 3.7 : Effets de la validation directe/inverse (avec $|\Delta\varepsilon| > 0.5$).

3.3.4 Evaluation sur un plan incliné

Nous avons vu précédemment que les images utilisées dans la procédure d'évaluation de Middlebury n'intègrent pas de régions pour lesquelles les valeurs de disparité varient rapidement. De ce fait, cette section présente une comparaison complémentaire sur un couple d'images stéréoscopiques d'une scène caractérisée par un plan incliné et une perspective marquée. La figure 3.15 présente le couple d'images et la vérité-terrain associée. Dans la suite, nous comparons les résultats obtenus par notre approche 1D et ceux obte-



(a) (b) (c)

Fig. 3.15 : (a) Image gauche et (b) image droite du couple d'images *plan incliné* et (c) la vérité-terrain associée.

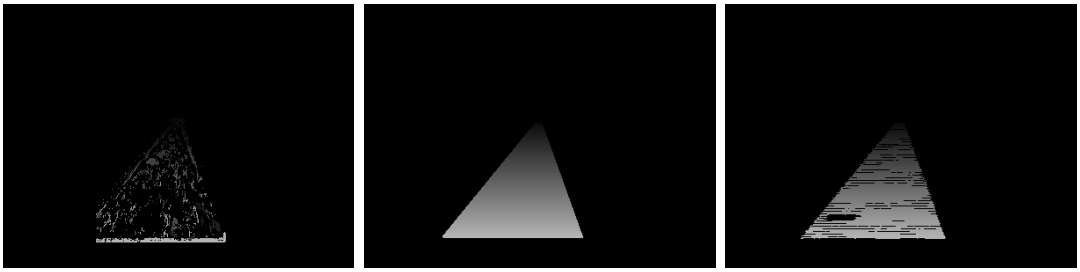
nus par la méthode 2D proposée par Hirschmüller qui présente les meilleurs classements dans la comparaison de la section précédente.

Hirschmüller combine les scores de similarité obtenus sur une fenêtre centrée sur le pixel à apparier et les scores de similarité obtenus sur plusieurs fenêtres support [2]. Pour notre comparaison, parmi toutes les configurations proposées par l'auteur, nous retenons celle basée sur une fenêtre centrale et quatre fenêtres « support ». Premièrement, les scores de dissimilarité de type SSD sont calculés sur la base de cette configuration. Le décalage présentant le score de dissimilarité le plus faible est affecté comme étant la disparité du pixel à apparier. Les cartes sont établies en prenant à tour de rôle l'image gauche puis l'image droite en référence. Ensuite, une étape de validation directe/inverse permet la suppression de certaines erreurs d'appariement.

Avec notre méthode, la densité de la carte des disparités peut être ajustée en faisant varier le seuil de décision T_{dec}. Cependant, nous avons conservé le seuil par défaut ($T_{dec} = 1$) puisque la méthode de Hirschmüller fournit des cartes peu denses.

La figure 3.16 présente les cartes des disparités obtenues par notre approche 1D et la méthode 2D de Hirschmüller. Nous avons utilisé la fenêtre optimale 2D de taille 7×9 pour la méthode 2D comme présenté dans [2]. On constate que les deux cartes ont des densités très différentes : les pixels noirs représentent les pixels pour lesquels aucune valeur de disparité n'a été affectée. La méthode 1D (cf figure 3.16(c)) estime plus efficacement les disparités caractérisant les plans inclinés puisque la carte des disparités est assez dense et présente visuellement peu d'erreurs.

Sur la carte 3.16(a), la méthode 2D affecte peu de valeurs de disparité. Ceci est dû au fait que cette méthode génère de nombreuses erreurs d'appariement qui sont différentes selon l'image qui est prise comme référence. De cette façon, la validation directe/inverse supprime un grand nombre d'ambiguïtés d'appariement et diminue ainsi la densité de la carte finale des disparités. Cependant, les quelques disparités affectées forment des zones de disparité constante alors que celles-ci doivent évoluer au fur et à mesure de l'inclinaison du plan des disparités.



(a) (b) (c)

Fig. 3.16 : (a) Carte des disparités calculée par la méthode de Hirschmüller avec une fenêtre de taille 7×9, (b) Vérité-terrain et (c) Carte des disparités calculée avec notre méthode 1D.

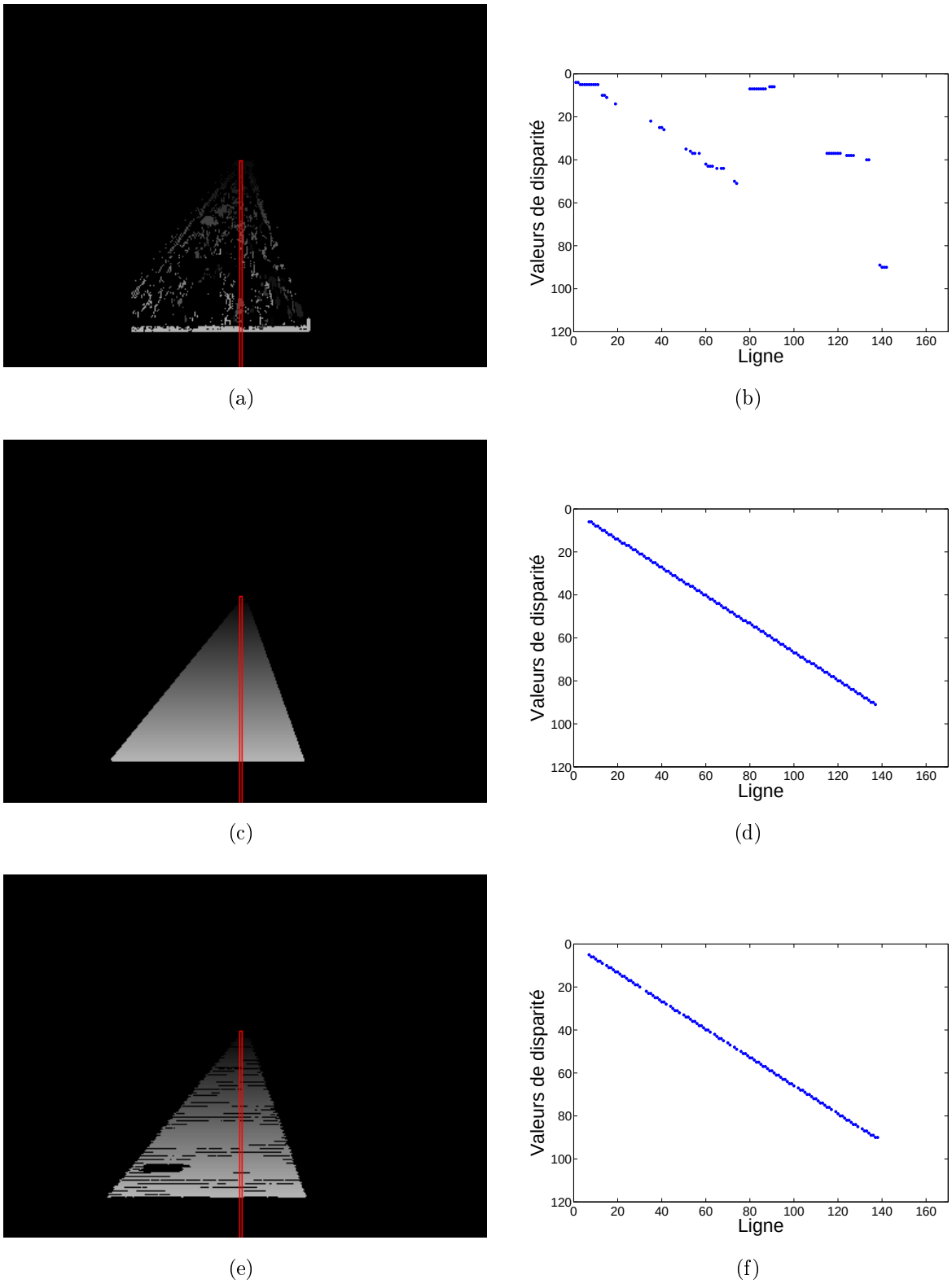
La figure 3.17 présente le profil des disparités le long de la même portion de colonne de l'image (encadrée en rouge) pour les trois cartes des disparités : méthode de Hirschmüller (Fig. 3.17(b)), vérité-terrain (Fig. 3.17(d)) et notre méthode 1D (Fig. 3.17(f)). La colonne présentant le plus grand nombre de pixels affectés pour la méthode 2D a été retenue.

Comme on peut le constater sur la figure 3.17(d), le profil des disparités de la colonne encadrée dans la vérité-terrain est composé d'un segment de droite oblique : ceci est typique des plans inclinés caractérisés par une rampe des disparités. En comparant l'ensemble des résultats, on peut noter que le profil associé à la méthode 1D est plus proche de celui de la vérité-terrain que dans le cas de la méthode 2D. Notre méthode fournit des disparités très semblables à celles de la vérité-terrain pour la colonne étudiée.

La méthode 2D donne des disparités complètement erronées pour la plupart des pixels de la colonne. Ceci est lié à la forme de la fenêtre 2D qui moyenne et affecte à l'aveugle les disparités de lignes voisines à la ligne courante. Pour la méthode 1D, les différences entre les valeurs de disparité erronées et celle de la vérité-terrain sont minimes.

Le tableau 3.3.4 présente les taux d'erreur d'appariement pour les deux méthodes 1D et 2D selon le protocole de l'évaluation de Middlebury. Les taux d'erreur sont calculés en utilisant la valeur absolue de la différence pixel à pixel $|\Delta\varepsilon|$ entre les cartes des disparités calculées et la vérité-terrain. Ici, comme les cartes des disparités sont déterminées au pixel près, les taux d'erreur d'appariement seront calculés pour $|\Delta\varepsilon|$ supérieur à un.

On peut constater que notre méthode 1D fournit des résultats bien meilleurs que la méthode 2D de Hirschmüller. Le taux d'erreur calculé sur le même nombre de pixels (i.e. les pixels appartenant au plan incliné dans la vérité-terrain), est bien plus élevé pour la méthode 2D. En effet, sur les 17.39 % de pixels affectés environ 70 % sont erronés ce qui correspond à 11.80 % des pixels sur la totalité des pixels appartenant au plan dans la vérité-terrain.



(a) (b)

(c) (d)

(e) (f)

Fig. 3.17 : Cartes des disparités et profil des disparités de la colonne encadrée en rouge pour respectivement la méthode de Hirschmüller (a)(b), la vérité-terrain (c)(d) et notre méthode 1D (e)(f).

Avec notre méthode 1D, 85.72 % des pixels sont affectés dont un taux quasi nul de disparités erronées qui correspond à 0.02 % des pixels sur la totalité des pixels appartenant au plan dans la vérité-terrain. Ces résultats sont éloquents ; avec ce couple d'images, l'effet de moyennage dû à la forme de fenêtre 2D induit de nombreuses erreurs d'appariement puisque la grande majorité de ces fenêtres 2D incluent des pixels correspondant à des profondeurs différentes.

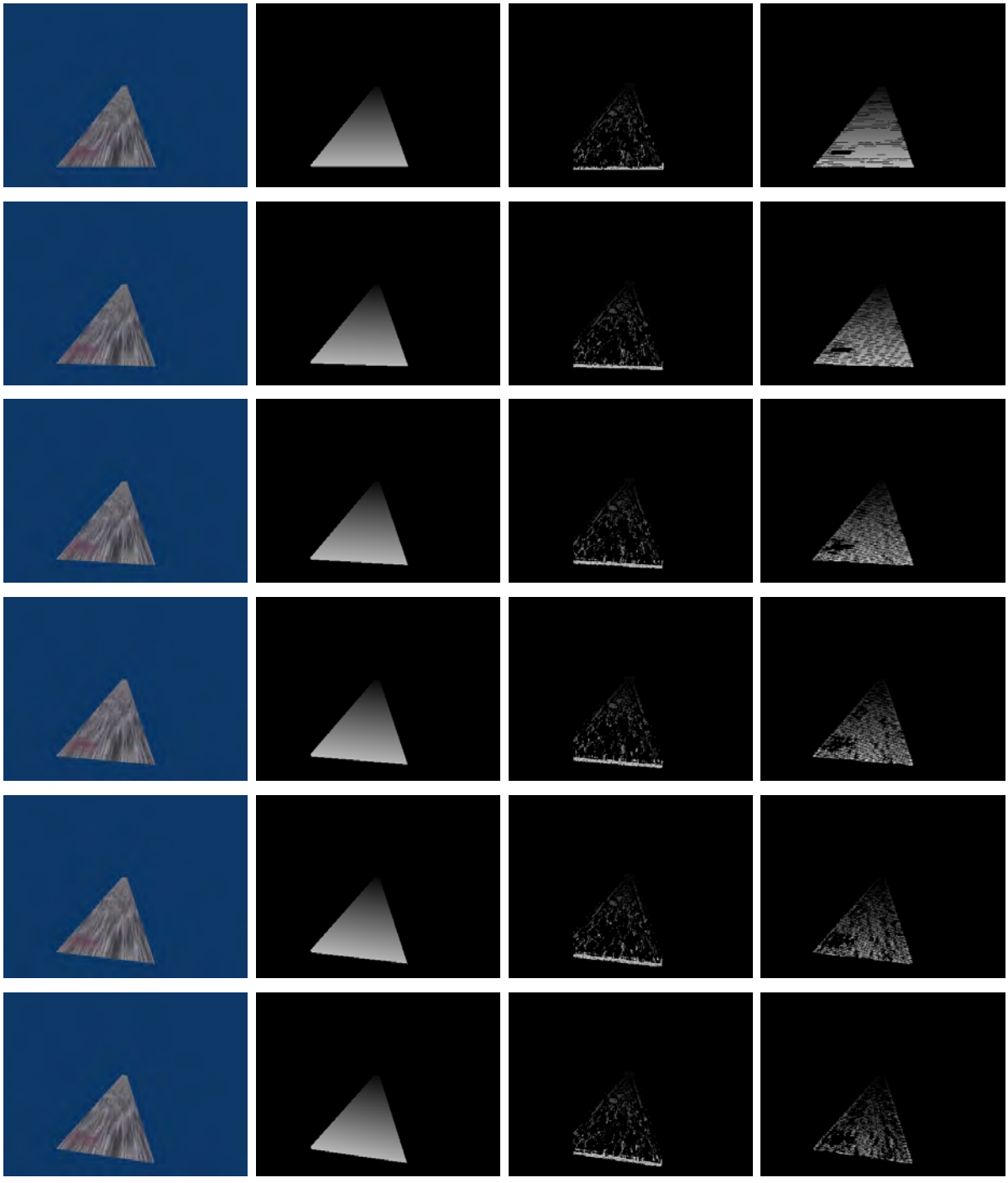
Méthode	ε_1 ($\|\Delta\varepsilon\| > 1$)	Densité
Hirschmüller (fenêtre de taille 9×7)	11.80 %	17.39 %
Notre méthode 1D	0.02 %	85.72 %

Tab. 3.8 : Comparaison des taux d'erreur pour la méthode de Hirschmüller et notre méthode 1D (avec $T_{dec} = 1$) pour le couple d'images *plan incliné*.

Pour conclure cette comparaison, différents couples de ce plan incliné ont été synthétisés en modifiant le dévers du plan d'un angle compris entre 0° et 10° par pas de 2°. L'image droite de ces couples est présenté figure 3.18 par la colonne (a). La vérité-terrain, la carte des disparités obtenue par la méthode 2D et celle calculée par notre méthode 1D sont présentées respectivement pour les colonnes (b), (c) et (d). On constate que les cartes des disparités générées par la méthode de Hirschmüller restent de densité équivalente quel que soit l'angle de dévers du plan : les erreurs restent toujours aussi nombreuses. Avec la méthode 1D, la densité diminue au fur et à mesure que le plan s'incline et le nombre de faux appariements engendrés augmente progressivement mais reste faible.

De la même manière, les taux d'erreur et la densité peuvent être calculés pour les cartes des disparités obtenues. Le taux d'erreur est toujours fixé à un et seuls les pixels du plan incliné de la vérité-terrain sont pris en compte. Les résultats sont présentés dans le tableau 3.3.4.

Pour la méthode de Hirschmüller, les taux d'erreur et la densité évoluent peu malgré l'angle croissant de dévers du plan. La densité reste assez faible puisque moins de 20% des pixels du plan sont affectés ; de plus, les erreurs engendrées sont nombreuses puisque plus de la moitié des disparités calculées sont fausses. Avec la méthode 1D, la densité décroit au fur et à mesure de l'augmentation du dévers tout en restant bien supérieure à celle obtenue avec la méthode 2D. Les taux d'erreur restent très faibles (inférieurs à 1%). Ces résultats montrent une particularité très intéressante de la méthode 1D qui module la densité de la carte en fonction de l'information locale tout en maintenant des taux d'erreur très faibles.



(a) (b) (c) (d)

Fig. 3.18 : (a) Image droite, (b) Vérité-terrain, (c) Carte des disparités calculée par la méthode de Hirschmüller avec une fenêtre de taille 7×9 et (d) Carte des disparités calculée avec notre méthode 1D (avec $T_{dec} = 1$).

Angle de dévers	Hirschmüller (Fenêtre de taille 9×7)		Notre méthode 1D ($T_{dec} = 1$)	
0°	11.80 %	17.39 %	0.02 %	85.72 %
2°	10.94 %	17.23 %	0.05 %	73.65 %
4°	10.15 %	16.90 %	0.12 %	57.87 %
6°	10.28 %	17.41 %	0.18 %	45.10 %
8°	10.13 %	16.91 %	0.20 %	33.49 %
10°	9.58 %	16.75 %	0.36 %	24.22 %

Tab. 3.9 : Comparaison des taux d'erreur avec $|\Delta\varepsilon| > 1$ pour la méthode de Hirschmüller et notre méthode 1D (avec $T_{dec} = 1$) pour plusieurs angles de dévers du plan incliné.

3.4 Conclusion

Dans ce chapitre, nous avons présenté une approche générique 1D originale. Celle-ci est fondée sur le calcul de scores de similarité pour un ensemble de fenêtres 1D de propriétés différentes (largeur de la fenêtre et position du pixel à apparier). Une valeur de confiance, quantifiant la vraisemblance que le décalage soit la bonne valeur de disparité pour le pixel, est modélisée pour chaque score de similarité associé. Enfin, une étape de décision finale permet d'établir une carte des disparités aussi dense et précise que possible. Une carte des confiances associées permet de pondérer l'affectation d'une valeur de disparité dans des zones peu sûres (dépourvues d'information). Dans cette étude, la confiance n'a pas été exploitée. Toutefois, elle ouvre des perspectives intéressantes notamment pour des étapes de post-traitement supplémentaires ou encore pour des applications multi-capteur pour lesquelles l'étape de fusion des données issues des différents capteurs sera prépondérante.

La méthode permet ainsi de juger de l'efficacité d'un appariement grâce à la mesure de confiance. Aussi, les fenêtres multi-configuration permettent de bénéficier des fenêtres plus grandes quand l'information locale est rare et des fenêtres de petites tailles en zone de discontinuité.

Les résultats obtenus par notre méthode générique ont été comparés à ceux issus de méthodes similaires en termes d'implantation (la méthode d'optimisation est basée sur la technique *WTA*). Nos résultats se classent parmi les meilleurs notamment pour l'image *Head and Lamp* qui est constituée d'objets situés à des profondeurs réellement différentes. Dans ce cas, la précision des résultats ne trompe pas. Les résultats obtenus montrent ainsi la pertinence de l'utilisation de fenêtres 1D dans le cadre de la mise en correspondance stéréoscopique par corrélation.

L'évaluation a d'abord été réalisée à partir de la méthode académique proposée par Scharstein et Szeliski qui proposent des couples d'images de compositions différentes et diverses classifications des pixels. Cette méthode d'évaluation, bien que conçue pour l'évaluation de méthodes globales et locales utilisant des fenêtres 2D, a permis de montrer les bons résultats obtenus par notre méthode 1D et ce, dans la plupart des cas.

Dans un deuxième temps, notre méthode 1D a été confrontée à la méthode 2D de Hirschmüller (qui a obtenu les meilleurs résultats lors de la première partie de l'évaluation) sur la base de couples d'images composées d'un plan incliné soumis à différents angles de dévers. Dans cette situation, les résultats obtenus par la méthode 1D sont incontestablement supérieurs à ceux de la méthode 2D de Hirschmüller en termes de densité et de taux d'erreur. De plus, on peut constater l'adaptabilité de la méthode 1D qui diminue la densité de la carte des disparités de manière à conserver un taux d'erreur faible et de ce fait une précision importante.

Cette propriété est fondamentale dans une application de la stéréovision passive à des fins de « métrologie ». En effet, il s'agit alors de garantir la précision de la mesure à défaut de garantir la densité de la carte. Les comparaisons présentées dans ce chapitre prouvent aussi la supériorité de notre approche dans ce contexte particulier.

L'approche 1D a été mise en œuvre sur la base de techniques très basiques de manière à vraiment juger de l'efficacité due à la propriété 1D de la fenêtre. Ceci laisse à penser que l'utilisation de techniques plus évoluées mènerait à des résultats encore plus probants.

L'étude de cas présentée précédemment pourra, par la suite, être implantée efficacement sur une architecture dédiée pour des applications embarquées. Dans le chapitre suivant, la méthode générique 1D mise en œuvre est employée et évaluée dans le cadre d'une application liée à la sécurité des transports.

Références

[1] D. Scharstein and R. Szeliski. A taxomomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1) :7–42, avril 2002. **www.middlebury.edu/stereo**. 8, 35, 46, 47, 54, 60, 63, 72, 77, 86, 88, 89, 90, 91, 92

[2] H. Hirschmüller. Improvements in real-time correlation-based stereo vision. In *IEEE Workshop on Stereo and Multi-Baseline Vision*, Kauai, HI, USA, décembre 2001. 75, 79, 88, 89, 90, 91, 92, 95

[3] K. Mühlmann, D. Maier, J. Hesser, and R. Männer. Calculating dense disparity maps from color stereo images, an efficient implementation. In *IEEE Workshop on Stereo and Multi-Baseline Vision*, pages 30–36, Kauai, HI, USA, juin 2001. 29, 34, 75, 76, 77, 79, 88, 89, 90, 91, 92

[4] A. Fusiello, V. Roberto, and E. Trucco. Symmetric stereo with multiple windowing. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(8) :1053–1066, 2000. 28, 75

[5] G. Egnal, M. Mintz, and P. Wildes. A stereo confidence metric using single view imagery with comparison to five alternative approaches. *Image and vision computing*, 22(12) :943–957, octobre 2004. 75

[6] R. Šára. Finding the largest unambiguous component of stereo matching. In *European Conference on Computer Vision*, page 919, Berlin, Germany, mai 2002. Springer. 76

[7] M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8) :1140–1146, 2002. 6, 76, 106, 114

[8] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. Rapport de recherche RR-1369, Institut National de Recherche en Informatique et en Automatique, INRIA, janvier 1991. 34, 77

[9] C. Sun. Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques. *International Journal of Computer Vision*, 47(1-3) :99–117, 2002. 33, 34, 77

[10] R. Mayoral, G. Lera, and M.J. Pérez-Ilzarbe. Evaluation of correspondance errors for stereo. *Image and Vision Computing*, (24) :1288–1300, 2006. 78

[11] G. Egnal, M. Mintz, and P. Wildes. A stereo confidence metric using single view imagery. In *International Conference on Vision Interface*, pages 162–169, Calgary, Canada, mai 2002. 79

[12] L.A. Zadeh. Fuzzy sets. In *Information and Control*, volume 8, pages 338–353, juin 1965. 79

[13] H. Mayer. Analysis of means to improve cooperative disparity estimation. In *ISPRS Conference on Photogrammetric Image Analysis*, pages 25–31, septembre 2003. 88, 89, 90, 91, 92

[14] K.J. Yoon and I.S. Kweon. Locally adaptive support-weight approach for visual correspondence search. In *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 924–931, San Diego, CA, USA, juin 2005. 88, 89, 90, 91, 92

[15] O. Veksler. Stereo matching by compact windows via minimum ratio cycle. In *International Conference on Computer Vision*, pages 540–547, Vancouver, Canada, juillet 2001. 88, 89, 90, 91, 92

Chapitre 4

La détection d'obstacles routiers

Sommaire

4.1 Un état de l'art 105
4.2 Calcul de cartes des disparités quasi-denses et précises 107
4.3 Projet RaViOLi 109
4.3.1 Présentation générale 110
4.3.2 Stéréoscope mono-caméra orientable 110
4.3.3 Détection et extraction des obstacles 113
4.3.4 Comportement de la méthode 1D en situations critiques 127
4.3.5 Quelques images réelles 136
4.4 Implantation matérielle 139
4.5 Conclusion 143
Références 144

LE DOMAINE des transports dits « intelligents » a connu une activité de recherche très intense ces deux dernières décennies. De nombreuses thématiques scientifiques liées à ce domaine ont été explorées dans le cadre de projets comme CARSENSE, ROADSENSE, STATUE, ARCOS, etc. Le fil conducteur de ces projets est l'utilisation simultanée de différents capteurs embarqués dans le but de bénéficier de leur complémentarité et/ou de leur redondance. Par exemple, dans [1], les auteurs utilisent la fusion de données issues d'un RADAR et d'un système de vision : le RADAR permet la localisation des zones d'intérêt dans l'image ; le système de vision permet la validation des données RADAR et l'augmentation de leur précision.

La précision des mesures issues de ces capteurs est un aspect fondamental pour notre application de détection d'obstacles. Dans ce chapitre, notre méthode générique 1D va être appliquée sur des images en situation routière, l'objectif étant la détection précise d'obstacles à longue distance en milieu autoroutier.

Dans ce début de chapitre, les particularités des scènes routières sont introduites. Ensuite, nous présentons des résultats quasi-denses et précis à partir d'images utilisées dans

la littérature. Le projet RaViOLi, dans lequel s'intègre ce travail de thèse, est présenté. Des images de synthèse, générées pour cette application particulière, sont utilisées et ce, dans le but de montrer que les cartes des disparités calculées par notre méthode 1D, sont plus précises que celles obtenues par une méthode 2D classique dans un contexte « transport » : ceci permet une détection plus précise des obstacles potentiels présents dans la scène observée avec une méthode simple. Le comportement de la méthode 1D face au manque de texture et aux imperfections de calibration est étudié. Enfin, le potentiel d'implantation de la méthode 1D sur une architecture parallèle est introduit.

4.1 Un état de l'art

De nombreux travaux ont été menés dans le cadre du développement des aides à la conduite et tout particulièrement de la fonction de détection et de suivi des obstacles. Deux types d'approches ont été étudiés : l'approche mono-caméra et l'approche multi-caméra. Cette section présente un bref état de l'art des différentes approches.

Les approches mono-caméra procèdent à la reconstruction 3D de la scène à partir de la mise en correspondance de certaines primitives extraites dans plusieurs images consécutives. Ces primitives peuvent être des régions texturées, des points de contour ou des zones en mouvement. La perte de l'information de profondeur est la première conséquence de la projection d'une scène 3D sur un capteur 2D. Cette conséquence fait du problème de reconstruction un problème mal posé qui se manifeste par l'apparition d'un facteur d'échelle qu'il est difficile de déterminer. Afin d'y parvenir, ces méthodes ont systématiquement recours à des connaissances a priori sur l'environnement ou sur la typologie et la géométrie des obstacles à détecter.

Dans [2], Broggi *et al.* présentent une méthode de détection de véhicule dans un contexte mono-caméra. Partant des propriétés de symétrie d'un véhicule, la solution propose une analyse des images de contours afin de détecter des zones présentant une forte symétrie verticale. Les auteurs adoptent une technique multi-résolution afin de détecter tous les obstacles présents dans les images tout en limitant les temps de calcul. La prise en compte des contours horizontaux situés entre deux contours verticaux permet alors de déterminer une boîte englobante autour de chaque véhicule.

Les approches multi-caméra exploitent la segmentation d'une carte des disparités calculées à partir de l'analyse des images fournies par un stéréoscope calibré. Les cartes des disparités sont généralement éparses afin de garantir une cadence raisonnable des traitements et ainsi respecter la contrainte temps-réel requise pour un organe de sécurité.

La phase de segmentation consiste tout d'abord à délimiter la surface correspondant à la chaussée. Dans une seconde phase, on procède à la détection de tous les objets situés au dessus du plan de la route ainsi obtenu. Dans [3], l'auteur propose un moyen robuste

et rapide de détection des obstacles et du plan de la route par l'analyse combinée des histogrammes appelés *V-disparité* et *H-disparité*. Depuis, cette méthode est utilisée par de nombreux auteurs et a été portée dans le cadre d'applications de détection de piétons.

Afin de localiser au mieux le véhicule porteur sur la route, certains auteurs proposent de détecter les voies de circulation. Dans [4], un détecteur de lignes blanches est présenté. Celui-ci utilise plusieurs instances d'un modèle générique de route déjà présenté dans [5], chacune détectant un type particulier de marquage au sol. Les résultats fournis par chaque instance sont utilisés afin d'obtenir une description plus complète de la configuration de la route et de déterminer la voie du véhicule porteur. Aussi, certains auteurs proposent de localiser toutes les voies et de limiter la recherche d'obstacles à ces zones d'intérêt [6]. Cette orientation permet de limiter le nombre de fausses détections dues aux obstacles localisés en bordure de route tels que les panneaux, les arbres et les glissières de sécurité.

Afin d'améliorer la robustesse et la précision de la détection des obstacles, Toulminet et al. proposent dans [7], une méthode hybride fondée à la fois sur une technique mono et multi-caméra. La méthode opère en deux étapes. Dans la première étape, une carte éparse des disparités est calculée à partir de la mise en correspondance des points de contour extraits dans chaque image du couple stéréoscopique. La segmentation de la carte des disparités se répartit en deux classes : une classe « route » et une classe « obstacle ». Les paramètres du plan de la route sont calculés grâce à la méthode décrite dans [3]. Les obstacles sont alors détectés à partir de certains segments 3D extraits de la carte des disparités suivi par une analyse des symétries telle que décrite dans [2].

Toutes ces méthodes montrent de premiers résultats intéressants mais nécessitent encore une amélioration des performances en matière de précision de la localisation et de robustesse de la détection. Ces améliorations sont directement liées au calcul des paramètres 3D du plan de la route. Le plan de la route est représenté par une rampe dans la carte des disparités. Comme nous l'avons montré dans la section 3.3.4 sur des images de synthèse, les méthodes 2D ne fournissent pas une carte suffisamment précise afin de résoudre le problème de segmentation posé. Ce problème s'avère être encore plus délicat en présence d'un roulis du véhicule et d'un dévers de la route (cf tableau 3.3.4).

Une première solution est d'adapter les solutions précédemment décrites à des cartes denses des disparités. Labayrade et al. [8] proposent une première méthode à partir de la méthode *quasi-dense* de Lhuillier décrite dans [9].

Dans la suite de ce chapitre, nous proposons d'appliquer notre méthode générique 1D décrite dans le chapitre précédent au domaine des transports. La section suivante présente une comparaison de la qualité des cartes des disparités obtenues par la méthode 2D de Hirschmüller [10] et par notre méthode sur une séquence d'images de synthèse de type route. Puis, nous consacrons une section à la description des résultats obtenus spécifiquement dans le cadre du projet RaViOLi. Nous clôturons le chapitre par une conclusion.

4.2 Calcul de cartes des disparités quasi-denses et précises

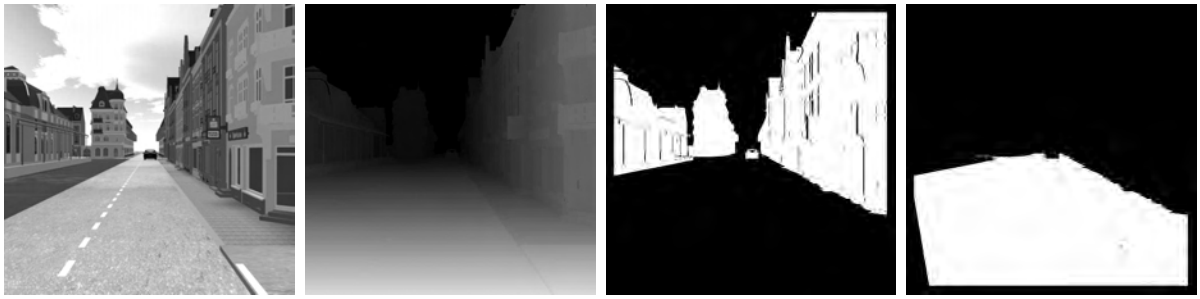
Notre méthode 1D décrite dans la section précédente, fournit une carte des disparités dont la densité dépend du niveau de confiance et de la précision souhaités. Plus la confiance seuil, fixée par l'utilisateur, est élevée, moins la carte des disparités est dense ; elle est cependant plus précise. Nous montrons ici que lorsque la vérité-terrain présente une rampe des disparités, notre méthode s'avère plus efficace qu'une méthode 2D. La méthode 2D retenue est la méthode de Hirschmüller [10] qui a obtenu les meilleurs résultats lors de l'évaluation académique de la section 3.3.3.

Il est important de noter que les cartes des disparités sont calculées pour des valeurs entières de disparité : aucune étape de mise en correspondance ni de raffinement au sous-pixel n'est utilisée. Par ailleurs, les paramètres de réglage de notre algorithme (i.e. les fonctions d'appartenance) sont identiques à ceux utilisés lors de la phase d'évaluation de Middlebury présentée au chapitre précédent. Nous soulignons ainsi la bonne adaptabilité de la méthode 1D.

La comparaison est réalisée sur la séquence d'images de synthèse proposée par Wannes van der Mark *et al.* dans [11]. Cette séquence illustre une scène routière en milieu urbain. Elle est accompagnée d'une vérité-terrain et d'une classification des pixels en 5 classes : pixels occultés, pixels de chaussée, pixels en arrière-plan, pixels au premier plan et pixels de ciel. Dans ce travail, les auteurs comparent sept méthodes différentes de mise en correspondance stéréoscopique pour la détection d'obstacles en milieu routier. La quasi totalité des méthodes utilisées exploite la mesure de similarité SAD calculée sur un ensemble de fenêtres 2D. L'étape de décision applique la stratégie du vote majoritaire WTA à toutes les mesures de similarité ainsi calculées. Les auteurs ont aussi introduit, dans le but de comparer ces méthodes, une séquence d'images synthétiques présentant une scène routière en milieu urbain.

La figure 4.1 présente l'image qui nous a servi pour cette première comparaison (image 95 de la séquence présentée dans [11]). Pour cette comparaison, la vérité-terrain (4.1(b)), les classifications de pixels d'arrière-plan (4.1(c)) et de chaussée (4.1(d)) sont utilisées.

La carte des disparités de la figure 4.2(a) présente les résultats obtenus par la méthode 2D de Hirschmüller pour la fenêtre optimale de dimensions 7×9. Nous rappelons que l'indice de dissimilarité calculé en chaque pixel est le résultat de la combinaison des indices de dissimilarité calculés sur plusieurs fenêtres support. Finalement, la disparité retenue est le décalage qui minimise la dissimilarité totale. Dans cette comparaison, nous utilisons la fenêtre centrée sur le pixel considéré et quatre fenêtres support. La figure 4.2(b) représente la carte des disparités obtenue grâce à notre méthode 1D. La densité de la carte peut être ajustée par le paramètre T_{dec}. Ici, ce paramètre est fixé à 0.8 afin d'obtenir une densité équivalente à celle obtenue par la méthode 2D.



(a) (b) (c) (d)

Fig. 4.1 : (a) Image gauche, (b) Vérité-terrain, (c) Pixels de l'arrière-plan, (d) Pixels de la chaussée du couple 95 de la séquence présentée dans [11].



(a) (b)

Fig. 4.2 : Carte des disparités calculée par (a) la méthode de Hirschmüller avec la fenêtre de taille 7×9, (b) notre méthode 1D avec $T_{dec} = 0.8$ pour le couple 95 de la séquence présentée dans [11].

Les taux d'erreur d'appariement ont été calculés pour les deux méthodes, pour les pixels appartenant à la chaussée et ceux appartenant à l'arrière-plan de l'image. Seuls les pixels affectés, communs aux deux cartes des disparités, ont été pris en compte lors de la comparaison. Sur les cartes des disparités, les pixels non affectés ont un niveau de gris nul. Comme le montre le tableau 4.1, à densité égale, les disparités calculées pour les pixels de chaussée sont bien plus précises avec notre méthode 1D qu'avec la méthode 2D. Cette différence est particulièrement marquée lorsqu'on souhaite une précision sur la disparité inférieure à 0.5 (la vérité-terrain étant au sous-pixel) comme introduit dans l'évaluation de Middlebury.

Comme le montre le tableau 4.2, pour les pixels appartenant à l'arrière-plan de l'image, les taux d'erreur générés par les deux méthodes sont équivalents. Ceci s'explique par le fait que les bâtiments sont des plans verticaux dans la scène et qu'ils sont représentés

par de larges régions 2D dont la disparité est constante dans la carte des disparités. Les résultats ne sont donc pas affectés lorsque des voisinages 2D sont utilisés.

Méthode	$\|\Delta\varepsilon\| > 1$	$\|\Delta\varepsilon\| > 0.5$	Densité
Hirschmüller (7×9)	0.43 %	6.43 %	43.39 %
Méthode 1D ($T_{dec} = 0.8$)	0.28 %	2.17 %	43.39 %

Tab. 4.1 : Comparaison des taux d'erreur d'appariement pour les pixels caractérisant la chaussée.

Méthode	$\|\Delta\varepsilon\| > 1$	$\|\Delta\varepsilon\| > 0.5$	Densité
Hirschmüller (7×9)	0.99 %	3.92 %	32.93 %
Méthode 1D ($T_{dec} = 0.8$)	0.90 %	3.77 %	32.93 %

Tab. 4.2 : Comparaison des taux d'erreur d'appariement pour les pixels caractérisant l'arrière-plan.

Des résultats similaires ont été obtenus sur l'ensemble des images de la séquence permettant ainsi de conforter les résultats précédents.

4.3 Projet RaViOLi

Dans cette section, nous présentons les résultats obtenus dans le cadre du projet RaViOLi. Le projet s'inscrit dans le cadre du développement d'un système de détection d'obstacles situés à longue distance et à l'avant d'un véhicule sur autoroute. Le système est composé d'un capteur catadioptrique grâce auquel il est possible d'acquérir un couple d'images stéréoscopiques d'une zone de la route située à longue distance.

Cette section du chapitre se décompose en 3 parties. La première constitue une présentation générale du projet. Dans la seconde partie, nous décrivons le capteur catadioptrique développé dans le cadre du projet. La dernière section est consacrée à la présentation de l'algorithme de détection d'obstacles développé à partir de notre méthode de mise en correspondance stéréoscopique. Nous étudions le comportement de la méthode en présence d'images peu texturées ou de couples d'images décalibrées.

Le prototype actuel du système de vision ne permet pas encore l'acquisition d'images à bord d'un véhicule. Nous validons donc la méthode sur des couples d'images générés à partir du logiciel de création d'images de synthèse POV-Ray[1]. Ce logiciel permet aussi la simulation de scènes complexes et la génération de cartes de vérité-terrain indispensables à l'évaluation objective des méthodes.

[1] www.povray.org

4.3.1 Présentation générale

RaViOLi est un projet débuté en 2003 et financé par la région Nord-Pas de Calais et le FEDER[1] dans le cadre du contrat de plan état-région 2000-2006.

Ce projet de recherche a comme objectif l'amélioration de la sécurité de la conduite grâce à la perception multi-capteur de l'environnement du véhicule. Le contexte d'application du projet RaViOLi est celui de la détection d'obstacles à l'avant de véhicules sur autoroute. Plus la vitesse est élevée, plus la détection des obstacles doit se faire tôt, autrement dit à de longues distances. Le système de perception est composé des capteurs hétérogènes suivants : un RADAR, un LIDAR et un système de vision catadioptrique. Tous sont des capteurs logiques fournissant chacun une description du monde dont la fusion permet d'améliorer la robustesse de la détection et la précision de la localisation de chaque obstacle.

Le projet est divisé en trois volets. Le premier, lié à la vision par ordinateur, a pour objectif l'augmentation de la portée d'un système de prise de vue stéréoscopique couleur afin de permettre la visualisation de l'environnement jusqu'à une distance de 120 mètres (pour mémoire, le projet ARCOS s'arrêtait à des distances de 45 mètres [12]). Dans le deuxième volet, l'objectif est de développer un capteur RADAR à corrélation permettant de déterminer la distance et la position angulaire de tous les obstacles situés sur la chaussée et dont il est possible de diriger le faisceau vers une zone d'intérêt. Le dernier volet vise à développer une interface unique d'acquisition et de traitement des données de bas niveau, issues des multiples capteurs installés sur le véhicule.

L'originalité du projet repose sur celle des capteurs développés et tout particulièrement sur la possibilité pour le RADAR et la caméra d'orienter leur champ de perception. En effet, il semble assez pertinent de pouvoir modifier dynamiquement le champ de perception de chaque capteur en fonction de la configuration de la route et de la zone d'intérêt à surveiller. Par exemple à longue distance et à l'approche d'un virage, le champ de perception des capteurs est orienté vers le bas de côté de la route. Il est donc nécessaire de corriger ce biais.

Ce travail de thèse s'inscrit dans le premier volet du projet dont l'objectif est la mise en œuvre d'un capteur stéréoscopique mono-caméra orientable que nous décrivons dans la section suivante.

4.3.2 Stéréoscope mono-caméra orientable

Dans la littérature, le capteur stéréoscopique est le plus souvent utilisé dans le cadre de l'étude des systèmes de détection d'obstacles. Comme nous l'avons précisé dans les chapitres 1 et 2, l'étape de mise en correspondance nécessite une phase de calibrage et

[1] Fonds Européen de DEveloppement Régional

de rectification et requiert l'acquisition simultanée des deux images, deux pré-requis dont dépend la précision de la carte des disparités obtenue.

Certains auteurs ont proposé une solution en développant des systèmes « catadioptriques » utilisant une seule caméra et un jeu de miroirs. Dans [13] et [14], les auteurs utilisent un biprisme central pour la projection des deux sous-images gauche et droite sur le capteur de la caméra. Ce capteur possède plusieurs avantages au regard d'autres stéréoscopes classiques. Tout d'abord, le problème de synchronisation entre caméras est éliminé puisque le système ne comporte qu'une seule caméra. Par ailleurs, le fait d'utiliser une seule caméra diminue les différences d'intensité lumineuse dans les deux sous-images. Finalement, les problèmes liés à l'utilisation de deux objectifs, dont les distances focales sont généralement différentes, sont également éliminés.

La figure 4.3 présente une prise de vue du prototype du stéréoscope mono-caméra orientable développé dans le projet RaViOLi. Celui-ci est inspiré des travaux de [14]. Il est composé d'une caméra, d'un jeu de miroirs dont un miroir central de type biprisme motorisé [15].

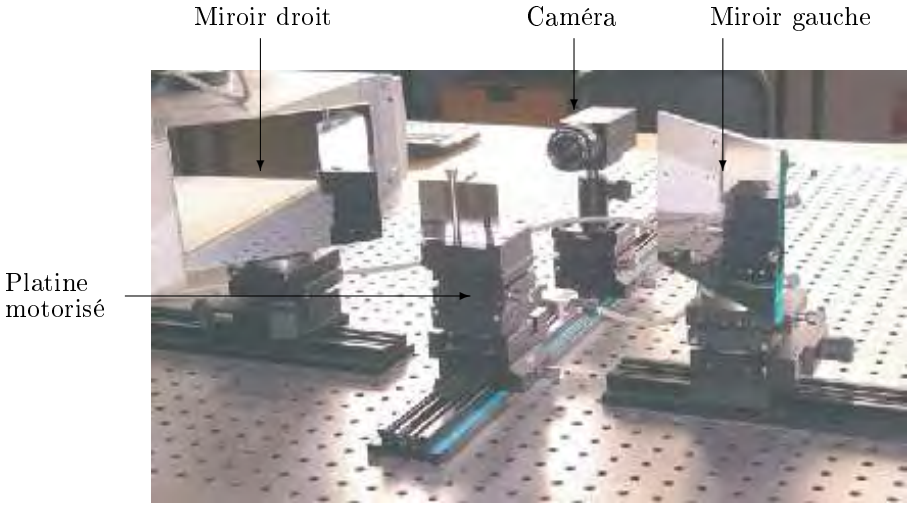


Fig. 4.3 : Prototype du stéréoscope mono-caméra à axes orientables.

Grâce à la motorisation du miroir central, le système est capable de suivre une zone d'intérêt dans la scène observée. Cette opération serait bien moins aisée s'il fallait orienter deux caméras simultanément. Les figures 4.4 et 4.5 illustrent ces propos. Sur la figure 4.4, le système installé sur le véhicule est orienté selon une direction fixe avec un objectif à courte distance focale. Ce système possède un champ de vision très large ; ainsi, un obstacle situé à longue distance sera caractérisé par très peu de pixels dans l'image (résolution faible) ce qui rendra sa détection plus complexe. Ce système fixe ne permet pas le suivi de la zone d'intérêt de la scène ; en phase de virage sur autoroute (cf figure 4.4(b)), la scène observée dans la zone de perception ne sera pas toujours des plus pertinentes.

A l'instar de la vision humaine, le système doit être capable, pour pallier ces inconvénients, de modifier la zone de perception de manière à toujours visualiser la zone d'intérêt de l'image. Par ailleurs, associé à l'utilisation d'un objectif à longue distance focale, le système permet d'obtenir une résolution élevée pour des obstacles situés à longue distance. La figure 4.5(a) montre qu'avec une distance focale élevée, la zone de perception est restreinte à l'obstacle et son environnement proche, permettant ainsi une augmentation significative de la résolution de ce véhicule dans l'image. Cependant en phase de virage (cf figure 4.5(b)), le véhicule n'est plus dans le champ de vision du système stéréoscopique ; cette lacune est comblée par l'orientation de la zone de perception du système suivant la zone d'intérêt comme illustré figure 4.5(c).

L'un des défauts majeurs concerne la réappropriation des procédures classiques de calibrage pour ce capteur atypique. Ce travail fait l'objet d'un sujet de thèse mené au laboratoire LAGIS.

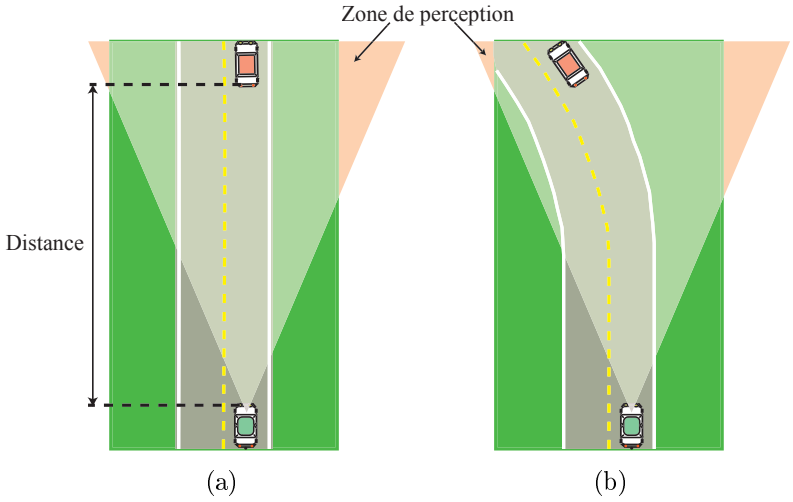


Fig. 4.4 : Vision fixe à faible distance focale.

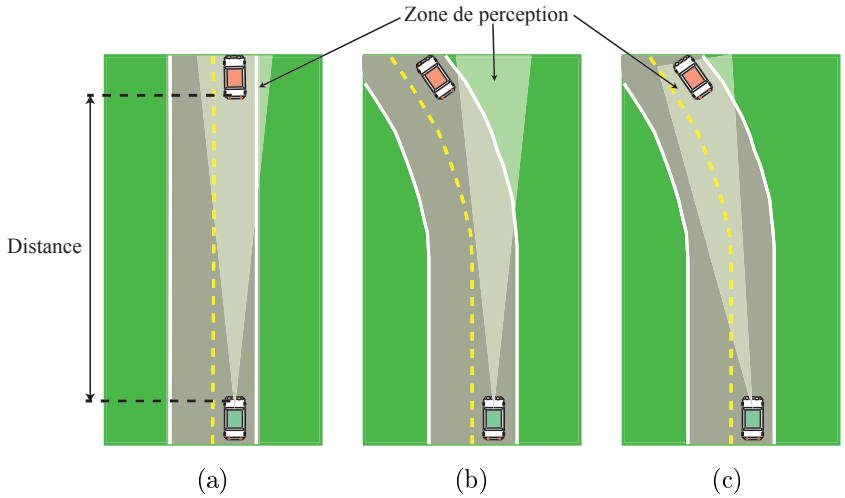


Fig. 4.5 : Vision orientable à longue distance focale.

4.3.3 Détection et extraction des obstacles

Détecter et localiser les obstacles situés sur la route par stéréovision nécessitent dans un premier temps de calculer une carte des disparités puis de procéder à sa segmentation afin d'isoler et de retrouver la position de chaque objet sur la route. De nombreuses méthodes ont été proposées dans la littérature. Dans cette section, nous commençons par présenter la stratégie des méthodes les plus classiques. La grande majorité de ces méthodes proposent d'utiliser des cartes éparses des disparités principalement pour des raisons de temps de calcul. Or, il est assez clair qu'utiliser une carte dense des disparités est incontournable afin d'augmenter la qualité des résultats (meilleure discrimination et détectabilité des obstacles et meilleure localisation). Dans la partie suivante, nous comparons la qualité des cartes des disparités calculées à partir d'une méthode 2D et de notre méthode 1D. Nous terminons par la description de l'étape de segmentation de la route et des véhicules que nous avons associée à notre méthode de mise en correspondance stéréoscopique 1D.

4.3.3.1 Etat de l'art

La détection d'obstacles, par des techniques de stéréovision, exploite une segmentation des cartes des disparités qui se décompose en deux étapes : l'estimation du plan de la route puis la détection des zones situées au dessus de ce plan, considérées alors comme des obstacles. La première distingue les pixels appartenant à la route de ceux appartenant aux obstacles. La deuxième regroupe les pixels de manière à détecter et séparer tous les obstacles situés à l'avant du véhicule.

Certaines méthodes considèrent que la chaussée et un véhicule peuvent être modélisés respectivement par un plan horizontal et un plan vertical dans le monde réel. Ainsi, un pixel appartiendra à un obstacle s'il se situe au dessus du plan horizontal modélisant la chaussée. La chaussée constitue un environnement relativement homogène, sur lequel plusieurs hypothèses peuvent être faites pour réduire la complexité des traitements. La plus classique de ces hypothèses, sur laquelle de nombreux travaux sont fondés, est de considérer que la chaussée est plane. Dans [16], une méthode en deux étapes est proposée : la première consiste à déterminer les paramètres du plan de la route en utilisant la méthode des moindres carrés. La deuxième étape vérifie si le pixel se situe au dessus de ce plan.

Certains travaux, comme l'approche de Labayrade *et al.* [17], analysent l'histogramme des valeurs des disparités d'une carte éparse. Cette méthode est fondée sur l'hypothèse qu'en chaque ligne pour laquelle la route est visible, la valeur dominante de disparité sera celle caractérisant la route. En chaque ligne, une « V-disparité » est calculée : il s'agit de l'histogramme des valeurs des disparités de la ligne.

La figure 4.6(b) présente la V-disparité établie à partir de la carte éparse des disparités associées de la figure 4.6(a). Dans cette illustration, la V-disparité a été binarisée de manière à en apprécier au mieux ses caractéristiques. Chaque ligne extraite, représente ainsi le profil longitudinal de la route. Cette dernière, perçue comme un plan horizontal dans la scène réelle, est représentée par un segment de droite oblique caractéristique des rampes des disparités. Les obstacles, associés à des valeurs de disparité quasi-identiques, sont représentés par des segments de droite quasi-verticaux dans la V-disparité. L'ensemble des points du véhicule est caractérisé par la même profondeur. De cette manière, le plan de la route est estimé par extraction dans l'image V-disparité des surfaces globales (extraction des droites par transformée de Hough). La V-disparité permet ainsi l'extraction des obstacles, du profil de route et du point de fuite.

Cette méthode, rapide et efficace, a été largement utilisée dans la littérature. Seulement, le fait d'utiliser des données éparses comporte quelques inconvénients. En effet, avec cette méthode [17], il est difficile de déterminer les dimensions d'un obstacle. Les contours sont essentiellement verticaux ce qui complique la localisation des limites horizontales de l'obstacle. De plus, devant le faible nombre de disparités utilisées, l'extraction du profil de route peut être altérée par les erreurs potentielles d'appariement et ce, même si l'utilisation des contours limite cet effet. La transformée de Hough sera dans ce cas plus approximative.

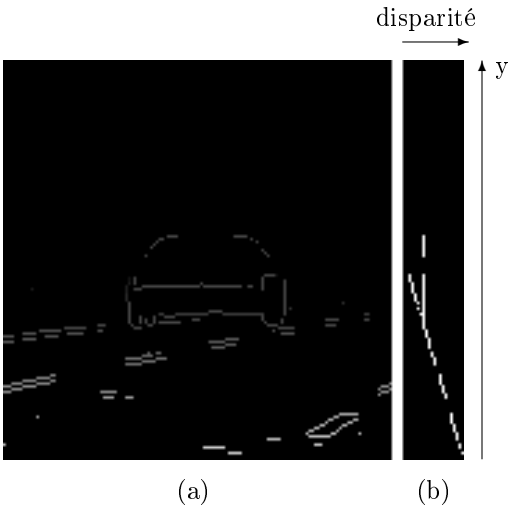


Fig. 4.6 : Illustration de la V-disparité.

Pour pallier ces problèmes, ces auteurs ont densifié leur carte des disparités par une méthode de croissance de germes [8] [9]. La première étape calcule toujours une carte éparse des disparités sur la base des contours extraits de chacune des deux images de manière à déterminer les paramètres du plan de la route. Ensuite, la méthode de croissance de germes remplit si possible le plan prédéfini : si le germe candidat correspond au plan de la route, celui-ci est accepté, dans le cas contraire il est rejeté. Dans cette approche,

les contours horizontaux des obstacles sont plus efficacement localisés puisque la carte des disparités est plus dense. Cependant, la propagation des disparités est fondée sur la carte éparse qui reste sensible aux erreurs de mise en correspondance : une erreur faite sur l'extraction des coordonnées du plan de la chaussée aurait des conséquences gênantes sur les traitements postérieurs.

Toutes ces méthodes reposent sur l'hypothèse d'un monde plan. Cependant, la route ne peut être considérée comme un plan horizontal en chaque instant. En effet, le caractère dynamique du véhicule, associé aux caractéristiques de la route, invalide cette hypothèse de travail. Dans [18] et [19], une méthode de détection d'obstacles a été mise au point dans le cadre d'une navigation autonome d'un véhicule de chantier. L'environnement est très difficile et aucune hypothèse ne peut être utilisée : la route n'est pas plane, la pente varie souvent, aucun marquage au sol n'est disponible. Ainsi, la présence de dévers sur la route et/ou de roulis du véhicule porteur rend assez difficile l'extraction du profil de la route. Cependant, la précision de la détection des obstacles va dépendre de la qualité d'extraction du profil de la route.

Certains auteurs ont travaillé sur des méthodes d'extraction du profil de route et de détection des véhicules dans le cas de devers et/ou roulis. Ces effets de devers de la route et/ou de roulis du véhicule porteur ne sont pas observables indépendamment l'un de l'autre ; le terme de devers sera, par abus de langage, utilisé génériquement [12].

Dans [12], l'auteur crée le profil longitudinal de la route dans l'espace 3D de la V-disparité et ce, ligne par ligne ; ceci permet de corriger ligne par ligne l'effet de dévers. Par interpolation linéaire des points 3D extraits de la ligne courante, l'auteur calcule l'angle d'inclinaison dû au dévers. Ensuite, le profil longitudinal de la ligne courante en dévers est corrigé, ce qui revient à étudier le profil longitudinal de cette même ligne sans le dévers. Dans ce cas, l'extraction des points d'intérêt de la ligne est plus précise, ce qui constitue un avantage certain par rapport à la V-disparité. L'exactitude de l'estimation du plan de la chaussée va permettre une localisation plus précise des obstacles en présence. Cependant, cette méthode n'est utilisable que si le devers sur la ligne étudiée est constant.

4.3.3.2 Comparaison des cartes des disparités obtenues par une méthode 2D et notre méthode 1D sur un couple d'images de route

Grâce au logiciel POV-Ray, nous avons synthétisé un couple d'images stéréoscopiques présentant une situation de conduite sur une autoroute à trois voies. Trois véhicules évoluent dans cette scène ; un quatrième (non visible) est le véhicule porteur du système stéréoscopique. Une carte de vérité-terrain est associée à chaque couple de manière à établir une comparaison quantitative. Les figures 4.8(a) et 4.9(a) représentent deux images gauches de la séquence l'une comportant un véhicule et l'autre pas.

Dans cette section, nous comparons les taux d'erreur de la carte des disparités obtenue à partir de notre méthode 1D et celle fournie par l'algorithme classique de Hirschmüller du fait que celui-ci a obtenu les meilleurs classements sur l'évaluation académique du chapitre précédent. Comme dans le chapitre précédent, la fenêtre 2D utilisée dans Hirschmüller est de dimension 7×9. De même, la méthode 1D utilise des fenêtres de dimensions 3×1 à 21×1 (c'est-à-dire pour $w \in \{1..10\}$ et $w_{max} = 10$). Pour ces couples d'images, T_{dec} est fixé à 0.9 afin d'obtenir des cartes des disparités de densité équivalente pour les deux méthodes comparées.

Pour cette application, la précision sur la mesure de disparité est fondamentale puisqu'elle affecte directement le calcul de distance du stéréoscope à l'objet. Les images sont synthétisées pour un système stéréoscopique réaliste dont les caractéristiques sont les suivantes :

- entraxe de 300 mm ;
- distance focale de 40 mm ;
- largeur du capteur de 8.46 mm ;
- résolution horizontale du capteur de 384 pixels.

Ces images sont synthétisées pour une résolution de capteur assez faible (i.e 384 pixels). En effet, si la méthode donne de bons résultats pour une résolution moindre, cette même méthode fournira des résultats bien meilleurs pour des résolutions plus importantes puisque la quantité d'information présente dans la scène observée sera ainsi augmentée. Ceci sera particulièrement vrai pour les objets situés à très grande distance.

La courbe de la figure 4.7 présente le calcul de distance en fonction de la valeur de disparité ; la distance est inversement proportionnelle à la valeur de disparité comme détaillé par l'équation (1.5). Par exemple, pour ce capteur, à une distance d'environ soixante mètres, une erreur de 1 pixel sur la disparité engendre une erreur de dix mètres sur la distance.

Les figures 4.8(a), 4.8(b), 4.8(c) et 4.8(d) présentent respectivement l'image gauche d'une autoroute, la vérité-terrain de la carte des disparités, les cartes des disparités calculées par notre méthode 1D et la méthode 2D. De la même manière, les figures 4.9(a), 4.9(b), 4.9(c) et 4.9(d) représentent respectivement l'image gauche d'une autoroute occupée par plusieurs véhicules, la vérité-terrain de la carte des disparités, les cartes des disparités calculées par notre méthode 1D et la méthode 2D. Sur les cartes des disparités, les pixels dont le niveau de gris est élevé marquent des disparités élevées correspondant à des objets proches ; les pixels dont le niveau de gris est nul ne sont pas affectés.

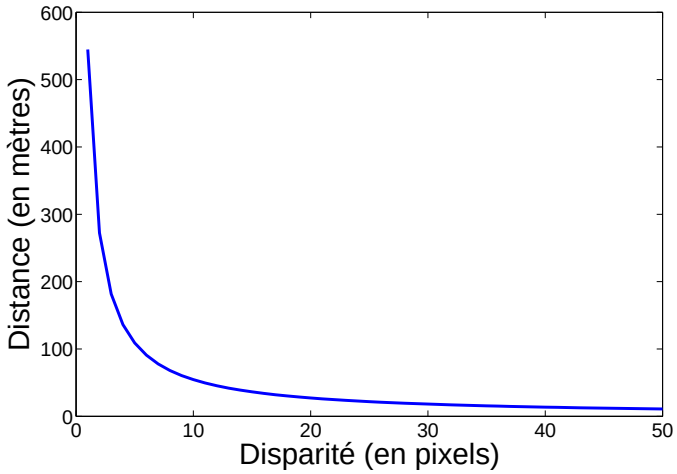


FIG. 4.7 : Distance en fonction de la valeur de disparité.

Comme le montrent les cartes des disparités des figures 4.8(c) et 4.9(c), notre méthode 1D fournit de bons résultats présentant peu d'erreurs et ce, pour une densité élevée. A contrario, les cartes des disparités générées par la méthode 2D présentent de nombreuses erreurs d'appariement sur la chaussée mais aussi aux contours des obstacles. Visuellement, on constate sur les différentes cartes d'erreur, des figures 4.8 et 4.9, que notre méthode 1D fournit de meilleurs résultats qu'il s'agisse des pixels appartenant à la chaussée ou de ceux appartenant aux différents obstacles en comparaison avec les résultats obtenus par la méthode 2D. Les résultats obtenus sur les 250 images composant la séquence synthétisée avec le logiciel POV-Ray permettent de conforter ces conclusions.

De manière quantitative, les taux d'erreur d'appariement ont été calculés pour des seuils d'erreur différents (1 et 0) et sont présentés pour les deux couples d'images dans les tableaux 4.3 et 4.4. Ici, on ne peut pas calculer de taux d'erreur au sous-pixel puisque les cartes des disparités à disposition (y compris la vérité-terrain) sont en valeurs entières de disparité. De cette façon, il semble important d'évaluer de façon précise ($|\Delta\varepsilon| > 0$) toute erreur d'appariement commise lors de la phase de mise en correspondance : une erreur d'un pixel sur la disparité à longue distance peut conduire à une imprécision de plusieurs mètres sur le calcul de distance.

On peut noter qu'à densité équivalente, notre méthode 1D fournit de bien meilleurs résultats que la méthode 2D pour laquelle les taux d'erreur sont plusieurs fois supérieurs à ceux obtenus avec notre méthode, et ce quel que soit le seuil d'erreur choisi.

Dans la suite de ce chapitre, nous proposons une méthode assez simple d'extraction du profil de route et de détection d'obstacles basée sur nos cartes des disparités quasi-denses. La densité de cette carte, associée à sa précision, permet de se passer des techniques par histogramme et ainsi appréhender plus finement les caractéristiques de la scène observée.

4.3.3.3 La détection des véhicules

La méthode proposée dans cette section est fondée sur l'analyse du profil des colonnes de la carte des disparités. L'objectif est de mettre en évidence qu'une méthode simple d'extraction permet d'obtenir de bons résultats sur une carte des disparités dense et précise. Cette méthode permet d'extraire les caractéristiques de la chaussée dans de nombreuses situations. Ici, l'extraction du profil de la route et la détection des obstacles exploitent nos cartes des disparités dont nous avons fixé le ratio entre la densité et la précision. Nous montrons que grâce à cette carte quasi-dense, il n'est pas nécessaire d'utiliser une méthode robuste comme le cumul d'histogramme de la V-disparité puisque les pixels pour lesquels une disparité a été calculée y sont en nombre suffisant.

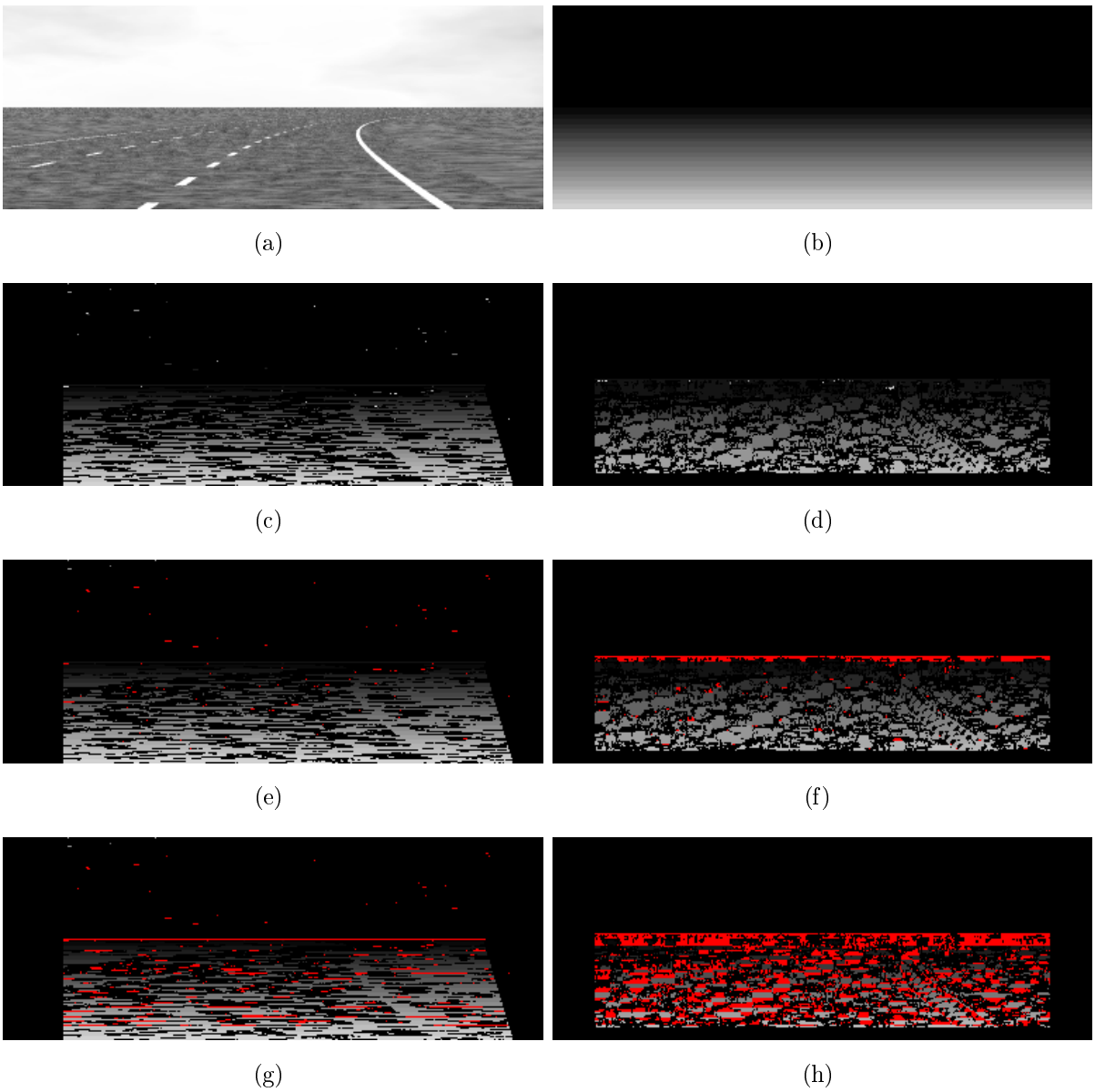
Cette méthode repose sur l'hypothèse que chaque obstacle est représenté par un plan vertical dont chaque point est situé à la même distance du stéréoscope. Nous appliquons la méthode sur les cartes des disparités obtenues par notre méthode 1D et par la méthode de Hirschmüller. Les résultats associés sont présentés par les figures 4.12(b) et 4.13(b) sur lesquelles nous pouvons visualiser les boîtes englobantes reportées sur l'image gauche de référence. La figure 4.11(b) représente les boîtes englobantes générées à partir de la vérité-terrain et servira de base pour cette première comparaison.

Le principe repose sur une analyse par colonne de la carte des disparités. La figure 4.10(c) présente le profil de la colonne surlignée dans la figure 4.10(b). Celui-ci est composé de deux segments de droite. Le premier est un segment de droite oblique qui représente la rampe des disparités, caractéristique de la chaussée. Le deuxième segment est un segment vertical qui caractérise l'obstacle dans l'image 4.10(b). En isolant ce segment vertical, on extrait les pixels de la colonne courante appartenant à l'obstacle potentiel.

Dans cette phase de détection des obstacles, on considère P_{total} pixels consécutifs d'une colonne de la carte des disparités. Si le nombre des pixels affectés P_{aff} est suffisamment important et si chacun de ces pixels est associé à la même valeur de disparité alors ces pixels consécutifs sont définis comme appartenant à un obstacle. Le profil dense des valeurs de disparité nous permet de délimiter facilement et précisément les contours de l'obstacle pour la colonne utilisée. En effectuant la même opération pour l'ensemble des colonnes de l'image, on peut localiser précisément les contours de l'obstacle comme illustré figure 4.10(d). Dans cette figure, les pixels considérés comme appartenant à la chaussée sont mis à zéro de manière à visualiser plus facilement les pixels supposés appartenir à l'image d'un plan vertical.

Méthode	$\|\Delta\varepsilon\| > 1$	$\|\Delta\varepsilon\| > 0$	Densité
Méthode 1D avec $T_{dec} = 0.9$	0.39 %	4.94 %	62.30 %
Hirschmüller (fenêtre de taille 7×9)	2.93 %	18.83 %	62.12 %

Tab. 4.3 : Comparaison des taux d'erreur calculés sur les cartes des disparités des figures 4.8(c) et 4.8(d).

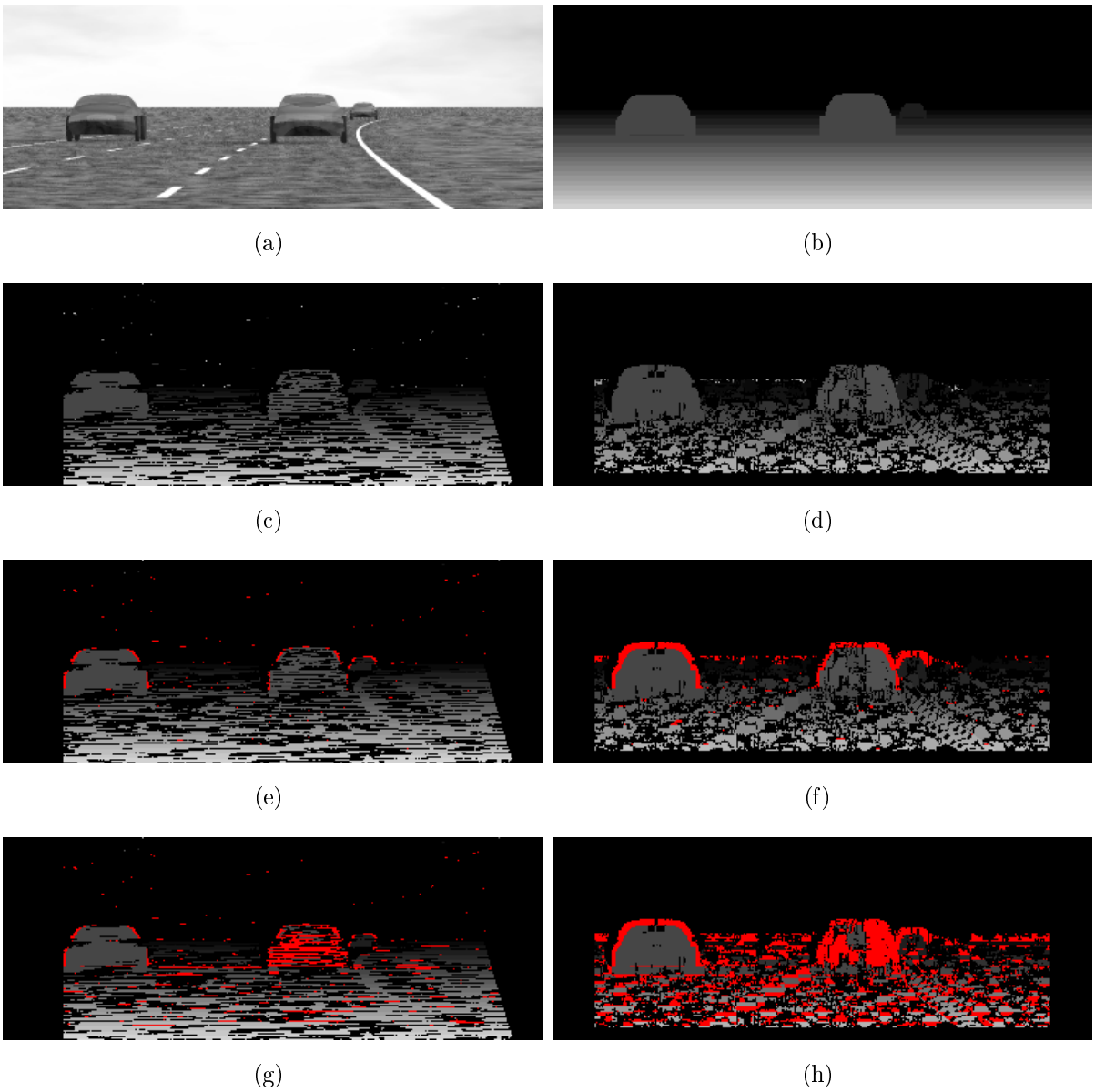


(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4.8 : (a) Image gauche d'autoroute sans véhicule, (b) Vérité-terrain, (c) Carte des disparités obtenue par notre méthode 1D, (d) Carte des disparités obtenue par la méthode de Hirschmüller - Les figures (e) et (g) présentent les erreurs générées par la méthode 1D et (f) et (h) pour la méthode 2D pour des seuils d'erreur respectifs de 1 et 0.

Méthode	$\|\Delta\varepsilon\| > 1$	$\|\Delta\varepsilon\| > 0$	Densité
Méthode 1D avec $T_{dec} = 0.9$	0.88 %	5.60 %	65.03 %
Hirschmüller (fenêtre de taille 7×9)	3.34 %	18.42 %	66.53 %

Tab. 4.4 : Comparaison des taux d'erreur calculés sur les cartes des disparités des figures 4.9(c) et 4.9(d).



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

Fig. 4.9 : (a) Image gauche d'autoroute avec véhicules, (b) Vérité-terrain, (c) Carte des disparités obtenue par notre méthode 1D, (d) Carte des disparités obtenue par la méthode de Hirschmüller - Les figures (e) et (g) présentent les erreurs générées par la méthode 1D et (f) et (h) pour la méthode 2D pour des seuils d'erreur respectifs de 1 et 0.

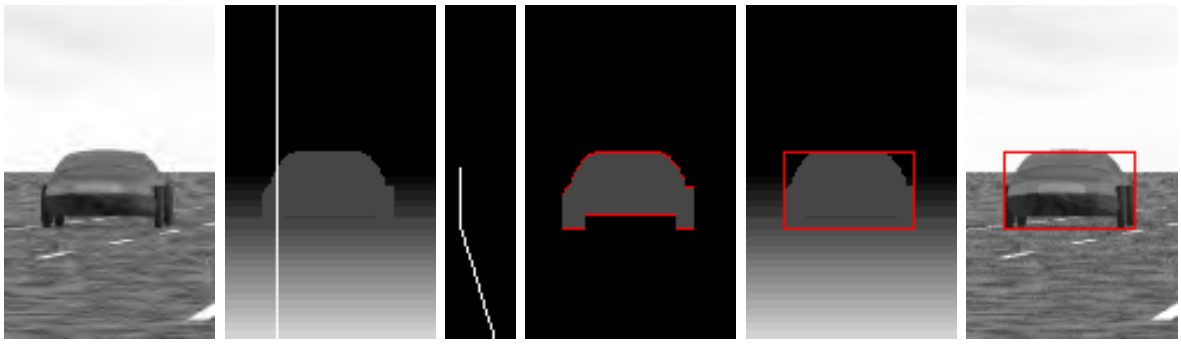
Les boîtes englobantes sont générées autour de ces points de contour grâce à une analyse colonne par colonne puis ligne par ligne. L'analyse colonne par colonne permet de déterminer les limites verticales des objets tandis que l'analyse ligne par ligne fournit les limites horizontales. Lors de la première étape, on considère que les pixels de deux colonnes consécutives doivent être associés à une même valeur de disparité. Une colonne contenant plusieurs pixels associés à une disparité différente de la colonne précédente constitue alors la fin de la boîte englobante courante et le début d'une autre. Il est ainsi possible de discriminer deux objets proches situés à des distances différentes. Lors de la seconde étape, la hauteur de l'obstacle fait l'hypothèse que, d'une ligne à l'autre, la disparité ne varie pas. Seuls les pixels localisés entre les limites verticales des obstacles précédemment établies sont utilisés. Les figures 4.10(e) et 4.10(f) présentent les boîtes englobantes générées respectivement sur la carte des disparités et sur l'image gauche de référence.

Il faut noter que si la densité de la carte des disparités est localement insuffisante, plusieurs boîtes englobantes de taille réduite peuvent être associées à un même obstacle. Dans ce cas, si chaque boîte englobante est caractérisée par la même disparité, une étape de fusion est lancée afin de n'en former qu'une seule.

En définitive, si les valeurs des disparités affectées aux véhicules et à la chaussée sont précises, la détection des obstacles par boîtes englobantes n'en sera que plus précise.

On peut constater que la méthode d'extraction des obstacles, fondée sur les cartes des disparités générées par notre méthode 1D, fournit de très bons résultats puisque les obstacles sont clairement détectés. Un objet situé à une grande distance est également correctement détecté et efficacement discriminé par rapport à l'obstacle voisin (dans cette image, le véhicule le plus éloigné se situe à une distance d'environ 180 mètres). Cette performance est atteinte grâce, à la fois, à une densité plus importante et à une précision garantie de la carte des disparités.

La présence de pixels non affectés ne gêne en aucun cas la bonne détection des obstacles dans la carte des disparités à condition que celle-ci soit suffisamment dense. Par la figure 4.13(b), on peut constater que la méthode d'extraction des obstacles sur la base de la carte des disparités établie par la méthode de Hirschmüller ne permet pas de calculer des boîtes englobantes précises. En effet, elles apparaissent autour de pixels caractérisant la route et apportent ainsi de fausses détections. Ceci est toujours lié à la forme des fenêtres 2D qui inclut des pixels de différentes profondeurs, moyenne les valeurs de disparité et crée des effets « paliers » sur la rampe des disparités caractérisant la chaussée.

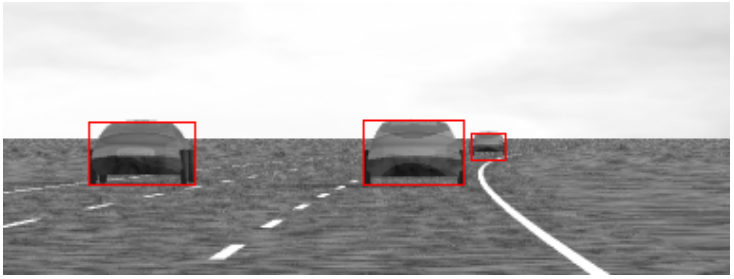


(a) (b) (c) (d) (e) (f)

Fig. 4.10 : (a) Image gauche, (b) Colonne utilisée, (c) Profil des disparités, (d) Détection des contours de l'obstacle, (e) Boîtes englobantes associées à la carte des disparités, (f) Boîtes englobantes associées à l'image gauche.



(a) Vérité-terrain

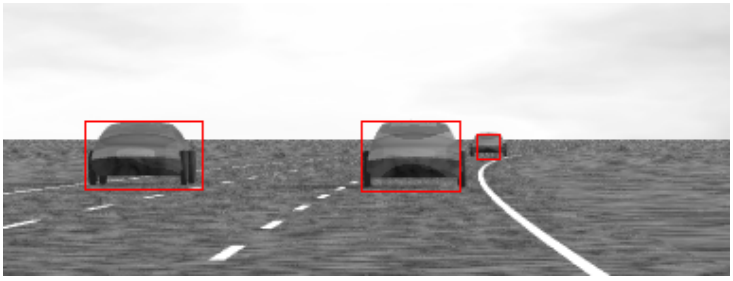


(b) Boîtes englobantes obtenues sur la carte des disparités (a)

Fig. 4.11 : Résultats obtenus sur la vérité-terrain du couple d'images 10.



(a) Carte des disparités obtenues avec notre méthode 1D pour $T_{dec} = 0.9$

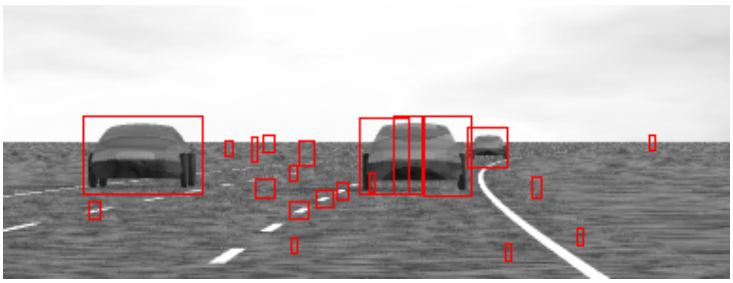


(b) Boîtes englobantes obtenues sur la carte des disparités (a)

Fig. 4.12 : Résultats obtenus sur le couple d'images 10 avec la méthode 1D.



(a) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9)



(b) Boîtes englobantes obtenues sur la carte des disparités (a)

Fig. 4.13 : Résultats obtenus sur le couple d'images 10 avec la méthode 2D de Hirschmüller.

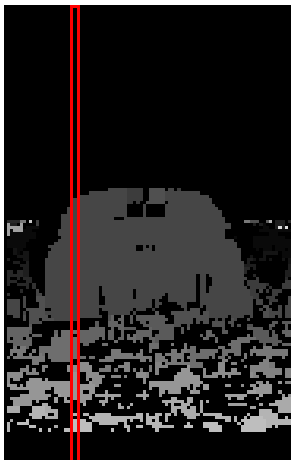
Afin d'illustrer ce propos, les profils des disparités pour une colonne donnée de la vérité-terrain (figure 4.14(c)), de la carte 1D des disparités (figure 4.14(e)) et de la carte 2D des disparités (figure 4.14(a)), ont été extraits sur respectivement les courbes des figures 4.14(d), 4.14(f) et 4.14(b). On constate que pour le profil des disparités calculé avec la méthode 2D, la rampe des disparités clairement visible sur la figure 4.14(d), est réduite à quelques paliers de disparité sur le profil de la figure 4.14(a).

De même, la méthode 2D amène de nombreuses erreurs au niveau des discontinuités (c'est-à-dire les contours des objets) dans la carte des disparités. Ceci engendre une localisation imprécise de la boîte englobante, laquelle peut être surdimensionnée ou subdivisée en plusieurs boîtes de disparité différente. Avec notre méthode, les erreurs dues aux discontinuités sont beaucoup moins nombreuses et les boîtes englobantes sont générées au plus près des contours des obstacles. De cette manière, les boîtes englobantes ne contiennent jamais plus d'un seul véhicule et l'étape de suivi des obstacles est beaucoup plus efficace.

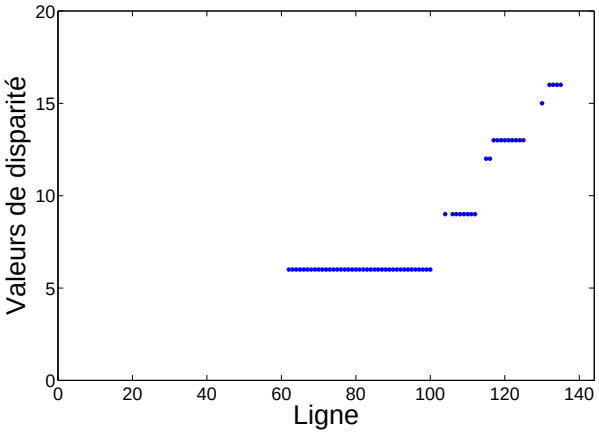
Ainsi, le compromis densité/précision semble être la clef pour détecter de façon optimale les différents obstacles. La figure 4.15 présente les cartes des disparités obtenues pour différentes valeurs de T_{dec} telles que $T_{dec} \in [0.6..1.0]$; les boîtes englobantes sont générées respectivement sur ces cartes des disparités et présentées figures 4.15. Pour les seuils les plus élevés, on constate que la carte est déjà suffisamment dense pour obtenir des boîtes englobantes précises avec notre méthode simple d'extraction des obstacles. Pour les différents seuils T_{dec}, le compromis densité/précision est suffisamment satisfaisant pour générer ces boîtes englobantes précises. On constate cependant que, lorsque la densité augmente, les boîtes englobantes sont légèrement moins précises puisque que le nombre d'erreurs d'appariement augmente avec la densité.

Aussi, traiter le profil des disparités colonne par colonne, permet de résoudre le problème d'extraction des obstacles dans le cas d'une route présentant un devers non constant. En effet, en cas de dévers (et/ou roulis), une ligne de l'image représentant la chaussée n'est plus caractérisée par une valeur constante de disparité. Ainsi, les techniques cumulatives ou autres techniques utilisant les disparités ligne par ligne perdront en précision alors que notre méthode colonne par colonne ne sera pas affectée.

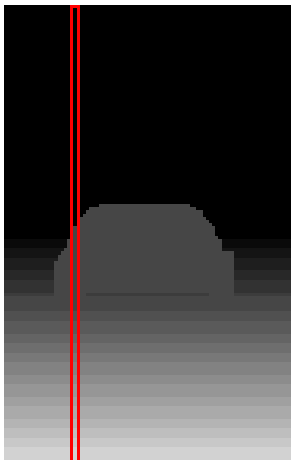
Comme dans [20], une phase de suivi des boîtes englobantes (i.e des obstacles potentiels) basée sur les techniques de filtrage de Kalman pourrait être implémentée ; cela permettrait de lever des ambiguïtés en gérant le phénomène d'apparition et/ou disparition des boîtes englobantes d'une image à l'autre. De même, certaines boîtes englobantes pourraient être éliminées en prenant en compte la taille de la boîte par rapport à la profondeur des points englobés. Les petites boîtes englobantes, caractérisées par une faible profondeur, pourraient ainsi être supprimées.



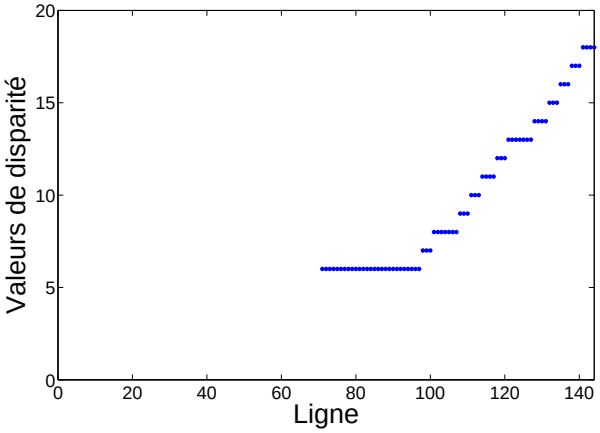
(a)



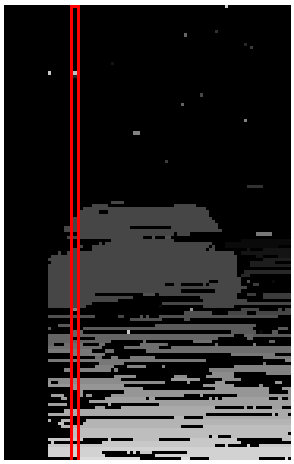
(b)



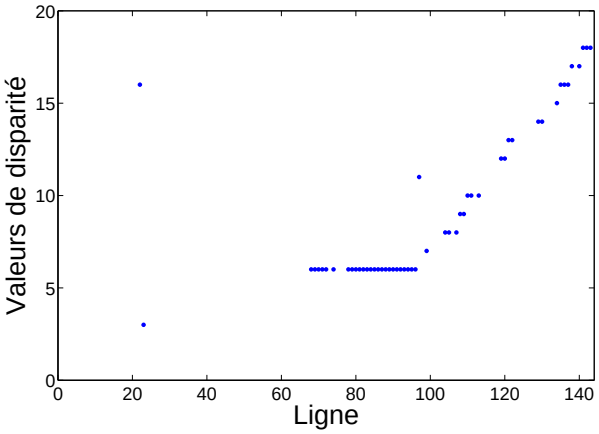
(c)



(d)

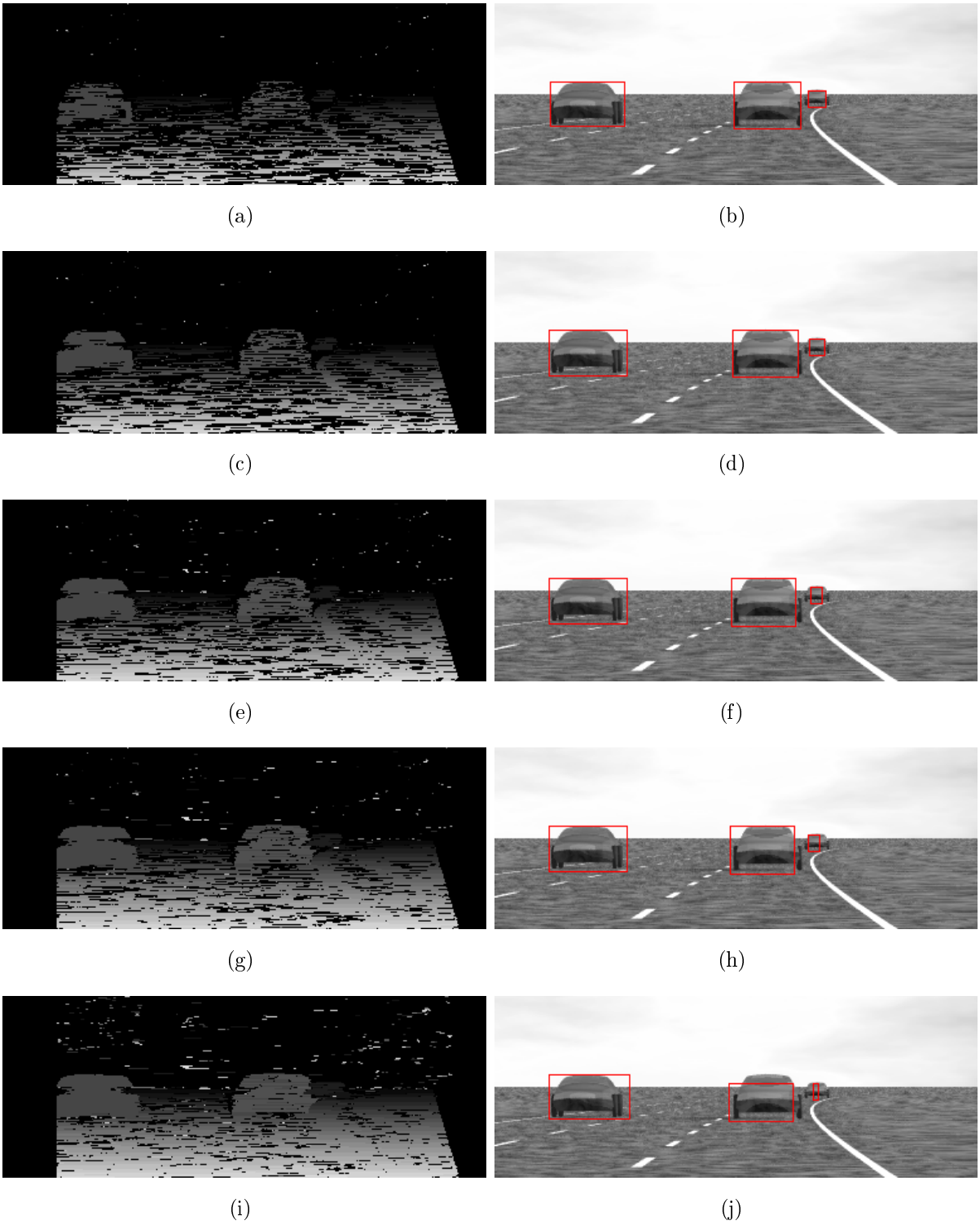


(e)



(f)

Fig. 4.14 : Cartes des disparités et profil des disparités de la colonne encadrée en rouge pour respectivement la méthode de Hirschmüller (a)(b), la vérité-terrain (c)(d) et notre méthode 1D (e)(f).



(a) (b)

(c) (d)

(e) (f)

(g) (h)

(i) (j)

Fig. 4.15 : Cartes des disparités obtenues par notre méthode 1D et images gauche pour lesquelles les obstacles ont été détectés pour respectivement T_{dec} égal à : 1.0 (a)(b), 0.9 (c)(d), 0.8 (e)(f), 0.7 (g)(h), 0.6 (i)(j).

4.3.4 Comportement de la méthode 1D en situations critiques

Dans cette section, nous étudions le comportement de notre méthode 1D dans le cas d'images non texturées et de couples d'images décalibrées. Ces deux problèmes sont très souvent rencontrés dans le contexte des applications visées.

La première étude porte sur le comportement de la méthode 1D face à un manque de texture pour les zones de l'image caractérisant la chaussée. Notre méthode 1D permet de calculer la profondeur en un pixel si et seulement si l'information locale le permet, autrement dit, si le voisinage de ce pixel est assez texturé. Cependant, dans le contexte routier, on sait qu'il est difficile d'obtenir de la texture sur une image de chaussée, notamment avec les faibles résolutions des caméras composant les stéréoscopes embarqués actuels.

Les figures 4.16 et 4.17 présentent respectivement les résultats obtenus sur un couple d'images texturées et le même couple d'images pour lequel la chaussée n'est pas texturée cette fois-ci. Les images sont obtenues avec le logiciel POV-Ray.

Les résultats de la figure 4.16 présentent les cartes des disparités obtenues avec notre méthode 1D (avec $T_{dec} = 0.9$) et avec la méthode de Hirschmüller (avec des fenêtres de taille 7×9) pour un couple d'images stéréoscopiques texturées ; à chacune de ces cartes des disparités est appliquée la méthode d'extraction des obstacles par boîtes englobantes.

Les résultats obtenus sont similaires à ceux présentés précédemment. Les obstacles y sont précisément détectés dans la carte des disparités obtenue avec la méthode 1D (même pour le véhicule situé à longue distance). La méthode 2D permet, elle aussi, une détection des obstacles mais engendre quelques fausses détections dues à l'effet de moyennage et des boîtes englobantes moins précises qu'avec la méthode 1D.

Les résultats de la figure 4.17 présentent les cartes des disparités obtenues avec notre méthode 1D (avec $T_{dec} = 0.9$) et avec la méthode de Hirschmüller (avec des fenêtres de taille 7×9) pour un couple d'images stéréoscopiques pour lequel la chaussée n'est pas texturée ; à chacune de ces cartes des disparités est appliquée la méthode d'extraction des obstacles par boîtes englobantes.

Les résultats obtenus sont assez éloquents. Sur ce couple d'images stéréoscopiques pour lequel la chaussée est dépourvue de texture, seuls les pixels de chaussée proches de contours et ceux caractérisant les obstacles sont affectés par une valeur de disparité. Néanmoins, les obstacles sont correctement détectés avec la carte des disparités obtenue avec la méthode 1D (même pour le véhicule situé à longue distance) puisque celle-ci est suffisamment dense et précise notamment pour les pixels appartenant aux différents véhicules. Comme précédemment, la méthode 2D permet la détection des obstacles mais engendre quelques fausses détections dues à l'effet de moyennage et des boîtes englobantes moins précises qu'avec la méthode 1D.

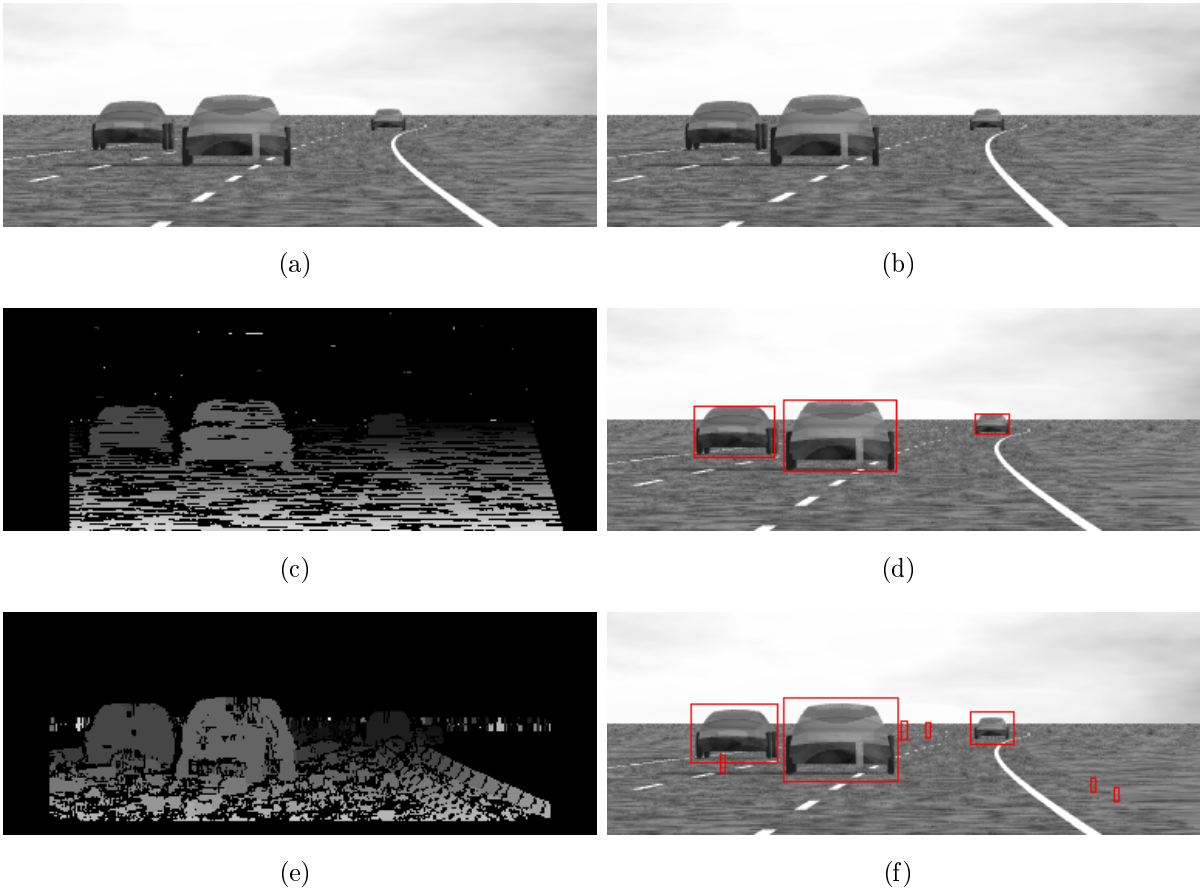
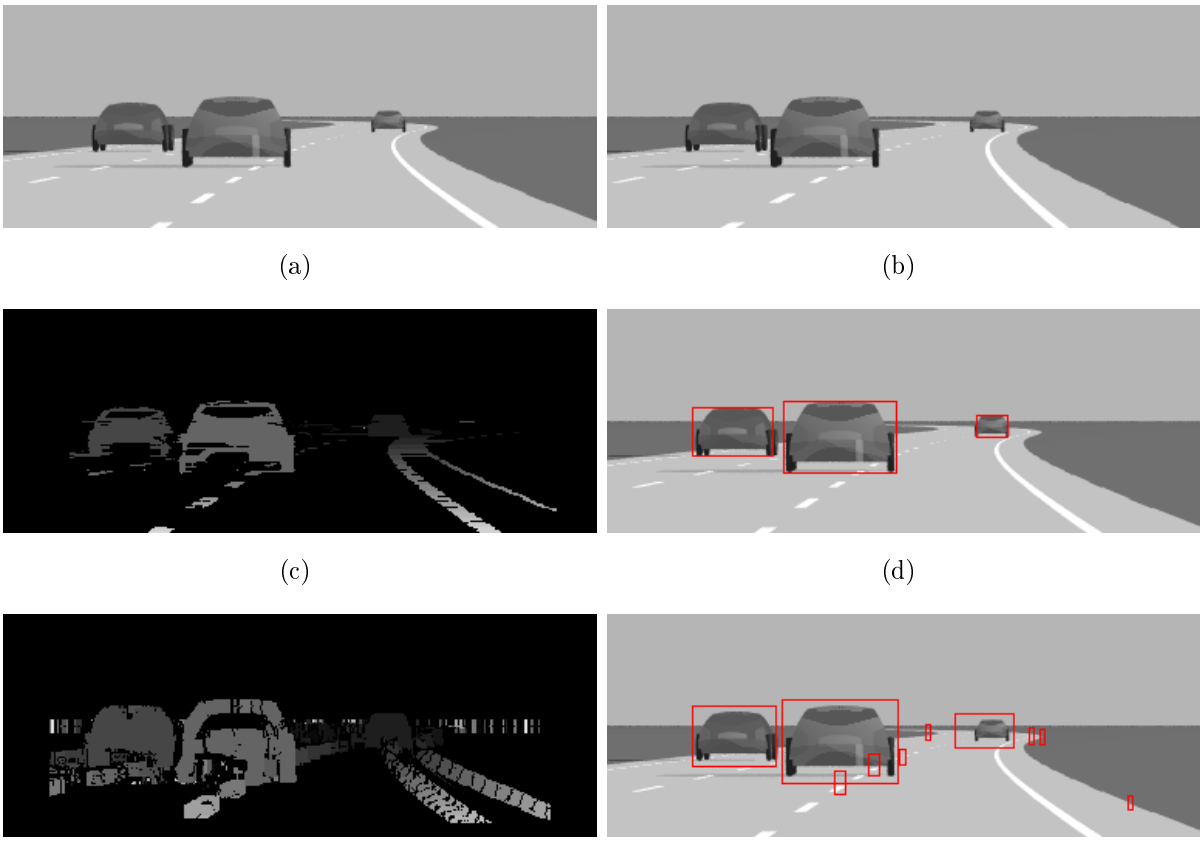


Fig. 4.16 : (a) Image gauche, (b) Image droite, (c) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (f) Boîtes englobantes obtenues sur la carte des disparités (e) - pour un couple d'images stéréoscopiques texturées.

Sur la carte des disparités calculée par la méthode 2D, on observe clairement des zones de disparité constante et ce, même sur les rampes des disparités. Ceci montre que, dans la carte des disparités, si les véhicules sont représentés par un ensemble de pixels suffisamment dense, la méthode d'extraction va permettre leur détection et ce, même si la chaussée est peu texturée.

Dans la suite de cette section, le comportement de la méthode 1D face à des défauts de calibration du stéréoscope est analysé. Comme précédemment, plusieurs couples d'images stéréoscopiques sont générés avec le logiciel POV-Ray pour différents défauts de calibration d'un stéréoscope classique. Afin d'appréhender les répercutions de ces défauts sur la densité et la précision des cartes des disparités obtenues, nous choisissons de travailler avec des images texturées.



(a) (b)

(c) (d)

(e) (f)

FIG. 4.17 : (a) Image gauche, (b) Image droite, (c) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (f) Boîtes englobantes obtenues sur la carte des disparités (e) - pour un couple d'images stéréoscopiques pour lequel la chaussée n'est pas texturée.

Chaque défaut de calibration est directement lié à la position de la caméra droite du stéréoscope. Les trois configurations suivantes de la caméra droite (par rapport à la caméra gauche) sont analysées, indépendamment l'une de l'autre :

- caméra droite sur-élevée de 5 centimètres (figure 4.18) ;
- caméra droite inclinée vers la chaussée d'un angle de 0.05 degré (figure 4.19) ;
- caméra droite orientée vers l'extérieur d'un angle de 0.5 degré (figure 4.20).

Chacune des figures, citées précédemment, présentent les images gauche et droite du couple stéréoscopique, la carte des disparités obtenue avec la méthode 2D retenue, les cartes des disparités calculées avec notre méthode 1D pour des valeurs de T_{dec} de 0.9 et 0.7 ; pour ces trois cartes des disparités sont établies les boîtes englobantes associées.

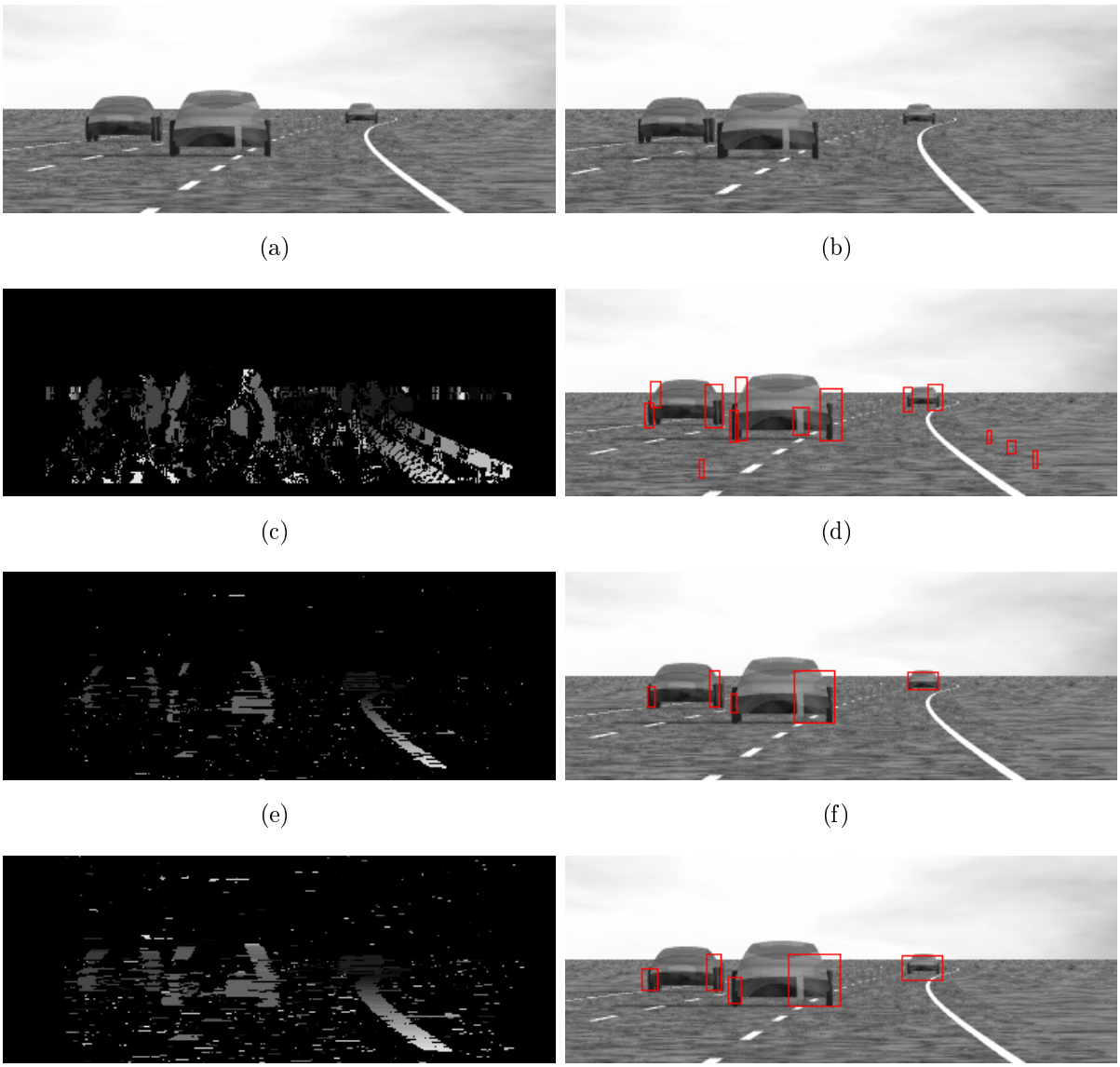


Fig. 4.18 : (a) Image gauche, (b) Image droite (caméra droite sur-élevée de 5 cm), (c) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (f) Boîtes englobantes obtenues sur la carte des disparités (e), (g) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (h) Boîtes englobantes obtenues sur la carte des disparités (g) - pour un couple d'images stéréoscopiques texturées.

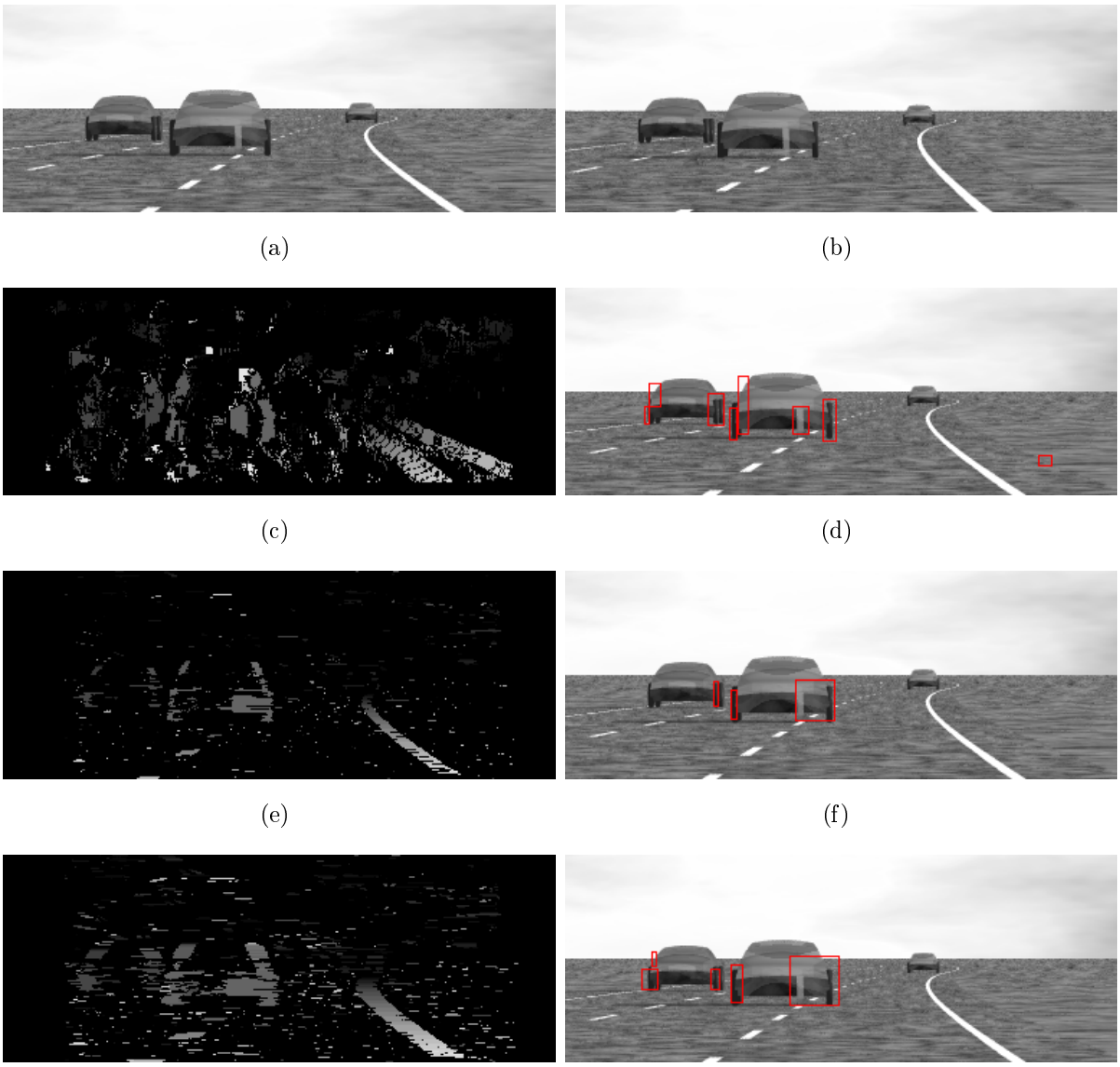


Fig. 4.19 : (a) Image gauche, (b) Image droite (caméra droite inclinée vers la chaussée d'un angle de 0.05 degré), (c) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (f) Boîtes englobantes obtenues sur la carte des disparités (e), (g) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (h) Boîtes englobantes obtenues sur la carte des disparités (g) - pour un couple d'images stéréoscopiques texturées.

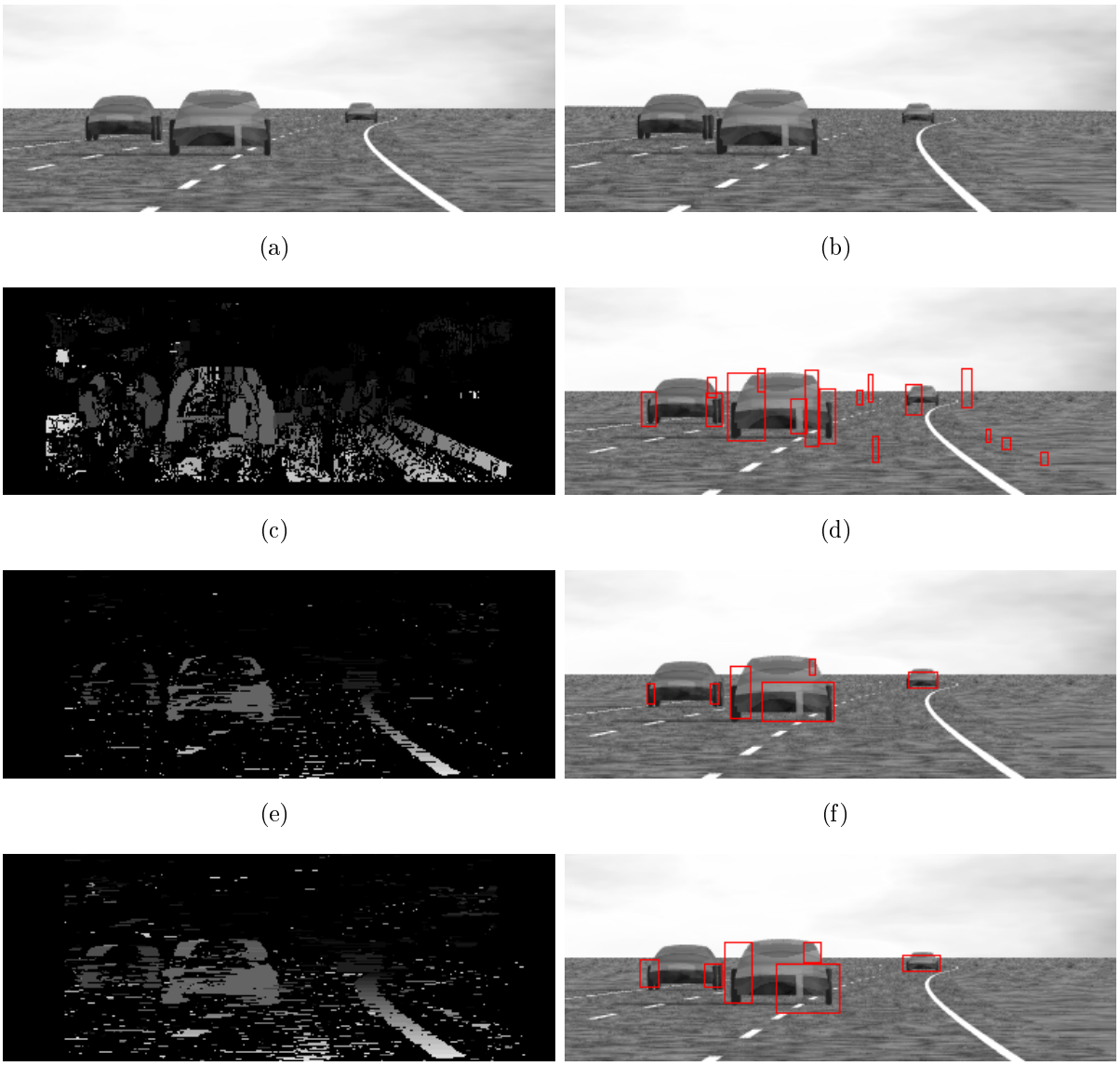


Fig. 4.20 : (a) Image gauche, (b) Image droite (caméra droite orientée vers l'extérieur d'un angle de 0.5 degré), (c) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (f) Boîtes englobantes obtenues sur la carte des disparités (e), (g) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (h) Boîtes englobantes obtenues sur la carte des disparités (g) - pour un couple d'images stéréoscopiques texturées.

Dans chacune de ces configurations, une imperfection quant à la calibration est engendrée. Les droites épipolaires n'étant plus confondues, un pixel situé sur une ligne dans l'image gauche peut trouver son homologue dans l'image droite plusieurs lignes au dessus ou au dessous. Dans les trois cas synthétisés ici, l'imperfection de calibration pour deux pixels homologues est comprise entre 1 et 3 lignes.

Lorsque le stéréoscope est décalibré, les cartes des disparités obtenues sont moins denses et moins précises et ce, quelle que soit la méthode de mise en correspondance utilisée. En effet, puisque l'information nécessaire à l'appariement ne se trouve plus sur les mêmes lignes dans les images gauche et droite, la mise en correspondance est moins efficace. De manière générale, les cartes obtenues avec la méthode 2D semblent plus denses puisque la fenêtre 2D va pouvoir aller chercher l'information sur les lignes voisines. Cependant, on constate que ces mêmes cartes des disparités sont caractérisées par de nombreuses zones d'erreur qui engendrent de fausses détections en termes de boîtes englobantes. On remarque qu'un véhicule est représenté par plusieurs boîtes englobantes très proches qui n'ont pas été fusionnées car elles sont caractérisées par des disparités différentes ou parce que le véhicule est moins dense que celui obtenu dans le cas calibré. Par ailleurs, le véhicule lointain reste toujours non détecté.

La méthode 1D génère des cartes des disparités moins denses mais celles-ci semblent plus précises notamment avec une valeur de T_{dec} élevée (égale à 0.9). La diminution de la valeur T_{dec} permet de rendre la carte des disparités plus dense mais quelques erreurs de mise en correspondance apparaissent. Les erreurs engendrées sont très ponctuelles et pourraient être éliminées facilement (notamment en filtrant les disparités ayant les confiances les plus faibles). Comme dans le cas 2D, les véhicules proches sont représentés par plusieurs boîtes englobantes pour les mêmes raisons qui ont été évoquées précédemment. Par ailleurs, dans les figures 4.18 et 4.20, il est important de noter que le véhicule situé à longue distance est précisément détecté. Ceci s'explique par le fait que les effets de ces configurations de décalibration s'estompent pour les distances les plus élevées. Lorsque la caméra droite est inclinée vers la chaussée (figure 4.19), l'erreur de calibration reste constante quelle que soit la distance et ne permet pas la détection du véhicule le plus éloigné.

Dans chacun de ces cas, les boîtes englobantes générées sont moins précises que celles de la figure 4.16 calculées pour un couple d'images rectifiées ; ceci est directement lié au manque de densité et de précision des cartes des disparités calculées.

Dans ce travail, nous avons montré que l'utilisation de fenêtres 1D pour la mise en correspondance d'images rectifiées se révélait efficace. Lorsque les images ne sont pas parfaitement rectifiées, la méthode 1D présente des limites mais les résultats sont comparables avec ceux obtenus avec la méthode 2D. Toutefois, dans certaines configurations, la méthode 1D est plus robuste pour des véhicules lointains.

Afin de compenser le décalage de quelques lignes de deux points homologues, nous avons modifié la méthode 1D en l'appliquant à des fenêtres composées de plusieurs lignes consécutives. Le nombre de lignes est cependant faible de manière à ne pas trop s'éloigner de la forme 1D. Ce choix permet de conserver les avantages de notre méthode 1D tout en augmentant sa robustesse face aux imperfections de calibration.

Il est important de noter qu'en utilisant des fenêtres de plusieurs lignes, nous changeons le cadre de notre approche et que nous ne sommes plus dans un contexte de « métrologie » ; dans ce cas, l'hypothèse de constance de la disparité dans la fenêtre de similarité peut être invalidée.

Les figures 4.21, 4.22 et 4.23 présentent respectivement les résultats obtenus pour les positions suivantes de la caméra droite : sur-élevée de 5 centimètres, inclinée vers la chaussée d'un angle de 0.05 degré ou orientée vers l'extérieur d'un angle de 0.5 degré ; les cartes des disparités sont calculées avec notre méthode 1D (pour $T_{dec} = 0.7$) en utilisant des fenêtres de 5 pixels de hauteur. A chacune des cartes des disparités calculées est associée sa carte des boîtes englobantes.

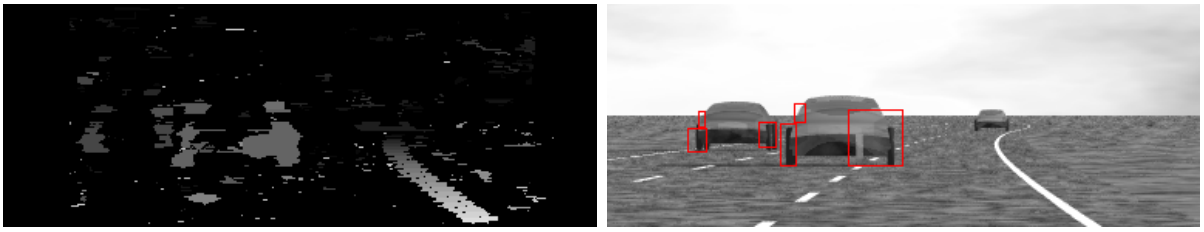
On constate clairement, sur ces figures, que les cartes des disparités calculées avec des fenêtres composées de 5 lignes présentent moins d'erreurs ponctuelles ; seulement, cela entraîne déjà l'effet de moyennage lié à l'utilisation de fenêtres 2D. Les deux véhicules situés à plus longues distances ne sont plus détectés sur la figure 4.21(b) (fenêtres de 5 pixels de hauteur) alors qu'ils l'étaient sur la figure 4.18(f) (fenêtres 1D) : ceci est une conséquence directe de l'extension de la fenêtre 1D à plusieurs lignes.

Pour des images qui ne sont pas parfaitement rectifiées, on peut conclure que les méthodes de mise en correspondance testées ont leurs limites. En effet, qu'ils s'agissent de la méthode 2D ou de notre méthode 1D (même étendue à des fenêtres de quelques lignes), les résultats de disparité sont moins denses et moins précis. Les résultats obtenus avec notre méthode 1D semblent meilleurs au regard des boîtes englobantes générées (moins de fausses détections).



(a) (b)

Fig. 4.21 : (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et une hauteur de fenêtre de 5 pixels, (b) Boîtes englobantes obtenues sur la carte des disparités (a) (ici, la caméra droite est sur-élevée de 5 cm).



(a) (b)

FIG. 4.22 : (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et une hauteur de fenêtre de 5 pixels, (b) Boîtes englobantes obtenues sur la carte des disparités (a) (ici, la caméra droite est inclinée vers la chaussée d'un angle de 0.1 degré).



(a) (b)

FIG. 4.23 : (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et une hauteur de fenêtre de 5 pixels, (b) Boîtes englobantes obtenues sur la carte des disparités (a) (ici, la caméra droite est orientée vers l'extérieur d'un angle de 0.5 degré).

On constate que l'extension de la méthode 1D à des fenêtres composées de plusieurs lignes dégrade les résultats puisque les boîtes englobantes associées sont moins précises.

Ainsi, l'extension des fenêtres 1D à des fenêtres composées de plusieurs lignes ne semble pas fournir des résultats très probants dans ce contexte. Même si les cartes des disparités obtenues sur des fenêtres de 5 pixels de hauteur semblent visuellement moins erronées, l'effet de moyennage induit une perte de précision quant aux valeurs de profondeur et par conséquent, aux boîtes englobantes générées.

D'après les résultats obtenus, nous pouvons dire que l'utilisation de fenêtres 1D pour la mise en correspondance stéréoscopique est efficace puisqu'elle fournit un compromis densité/précision très intéressant. Dans cette section, nous avons vu que le manque de texture sur la chaussée n'était pas un frein à la détection efficace des obstacles si ceux-ci sont suffisamment denses en termes de disparité : l'essentiel étant de ne pas affecter de valeur de disparité si l'information locale ne le permet pas et de conserver ainsi une précision élevée.

Toutefois, les limitations des capteurs stéréoscopiques actuels soulèvent quelques interrogations quant à l'efficacité de notre méthode 1D pour des couples d'images réelles.

4.3.5 Quelques images réelles

Dans cette section, différents couples d'images réelles caractérisant différentes situations sont utilisés. Nous nous intéressons ici aux résultats de mise en correspondance sur ces couples d'images réelles. Les couples d'images réelles acquises en environnement non contrôlé sont souvent caractérisés par de faibles contrastes et présentent des imperfections de calibration plus ou moins fortes. Les résultats présentés ci-après ont pour objectif de montrer dans quelles proportions les performances, obtenues par notre méthode sur des images de synthèse (cf section précédente), sont dégradées. Nous comparons les performances à celles obtenues par la méthode 2D dans ces mêmes environnements transport.

4.3.5.1 Environnement « transport guidé »

Le couple d'images de la figure 4.24 représente une scène de transport guidé ; le stéréoscope est ici à l'intérieur d'une voiture du « VAL[1] » de la société Transpole de Lille. Ces scènes typiques dans un environnement « transport guidé » permettent la comparaison des résultats de disparité obtenus avec notre méthode 1D (pour différentes valeurs de T_{dec}) à ceux obtenus avec la méthode 2D de Hirschmüller. Nous comparons aussi la méthode 1D initiale à la méthode 1D étendue à plusieurs lignes.

Sur les cartes des disparités obtenues avec notre méthode 1D (images 4.24(a) et 4.24(c)), on constate que les cartes sont assez denses et contiennent visuellement peu d'erreurs de mise en correspondance pour des fenêtres de hauteur respective 1 pixel et 3 pixels. Le fait d'utiliser des images réelles ne nous permet pas d'obtenir des résultats d'erreur précis puisqu'aucune carte de vérité-terrain n'est disponible ici.

Etendre la fenêtre 1D à quelques lignes (ici 3 lignes) permet d'augmenter la densité de la carte des disparités tout en conservant une précision supérieure à celle de la carte des disparités calculée avec la méthode 2D (figure 4.24(e)). En effet, la rampe des disparités obtenues semble plus précise et on observe des disparités très élevées (i.e. de niveau de gris élevé) au centre de la rame de tramway ; or, il est évident que ces valeurs sont erronées.

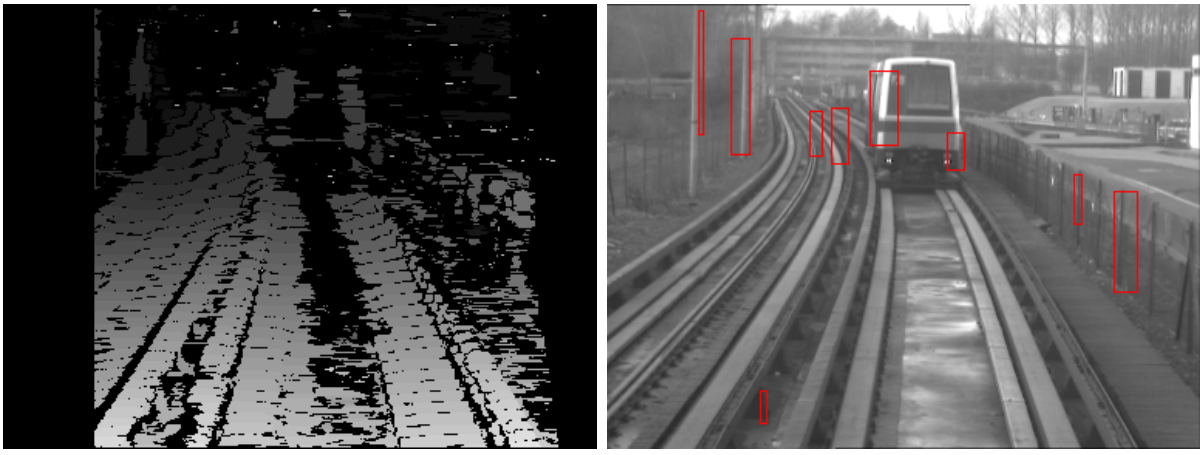
Les boîtes englobantes générées sur les cartes des disparités montrent que la méthode 2D souffre encore des effets de moyennage en générant des fausses alarmes. Notre méthode permet de détecter l'obstacle tout en évitant ces fausses alarmes. Dans ce cas, il est préférable d'avoir des valeurs de disparité précises même si celles-ci sont peu nombreuses.

On constate que les disparités sont calculées principalement dans les zones de l'image situées à proximité des contours bien marqués du VAL ; le centre de cette obstacle ne présente pas beaucoup d'information de texture et de ce fait, il semble difficile d'y affecter une valeur de disparité et ce, même en utilisant la méthode 2D de Hirschmüller.

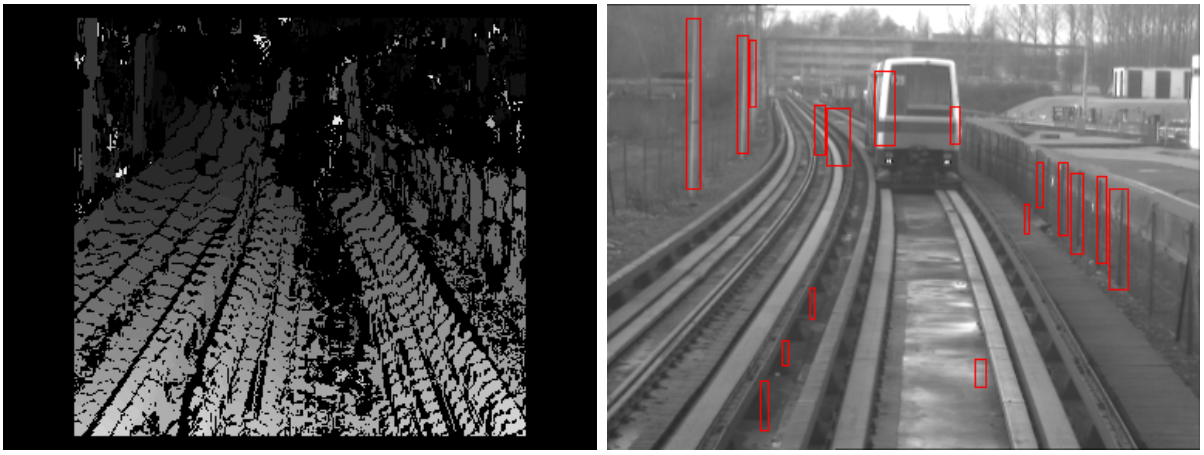
[1] Véhicule Automatique Léger



(a) (b)



(c) (d)



(e) (f)

FIG. 4.24 : (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (b) Boîtes englobantes obtenues sur la carte des disparités (a), (c) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et des fenêtres de 3 pixels de hauteur, (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (f) Boîtes englobantes obtenues sur la carte des disparités (e).



(a) (b)

FIG. 4.25 : (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et $n \in [1..20]$, (b) Boîtes englobantes obtenues sur la carte des disparités (a).

Afin de pallier ce problème, nous présentons sur la figure 4.25 les résultats obtenus avec des fenêtres de tailles supérieures (i.e. $n \in [1..20]$ sachant que jusqu'ici les résultats ont été obtenus pour $n \in [1..10]$). On remarque visuellement que la carte des disparités obtenue est très proche de celles calculées avec la méthode 1D de la figure 4.24. Plus particulièrement, on constate que la rame de tramway est plus dense en termes de disparité ce qui permet sa détection (illustrée par la boîte englobante générée). En effet, en augmentant la taille de fenêtre, la méthode 1D va chercher de l'information plus loin sur la ligne tout en évitant les erreurs de mise en correspondance dues aux discontinuités.

Les premiers résultats de mise en correspondance sur ces images réelles montre clairement la viabilité de notre méthode 1D dans ce cadre applicatif de « transports guidés ».

4.3.5.2 Environnement « transport routier »

Dans un deuxième temps, nous avons appliqué notre méthode 1D sur des images présentant de faibles contrastes mises à disposition par le LIVIC[1]. Ces images sont acquises dans un contexte de « transport routier » ; les résultats, illustrés figure 4.26, sont obtenus pour un intervalle des disparités limité (i.e. $s \in [0..85]$) de manière à ne détecter que l'environnement extérieur et non l'intérieur du véhicule.

On constate que notre méthode 1D fournit peu de valeurs de disparité pour les pixels de chaussée ; ceci est directement lié au manque de texture pour ces zones de l'image. Cependant, la méthode 1D arrive à calculer des valeurs de disparité pour les pixels appartenant aux différents véhicules.

[1]Laboratoire sur les Interactions Véhicules-Infrastructure-Conducteurs

En étendant la taille de la fenêtre 1D à quelques lignes (ici 3 lignes), la densité de la carte des disparités est augmentée tout en conservant une précision supérieure à celle de la carte des disparités calculée avec la méthode 2D (figure 4.26(e)). Cette dernière est caractérisée par des zones de l'image pour lesquelles les disparités calculées sont erronées.

La méthode d'extraction permet de mettre en évidence les compromis densité/précision des différentes cartes des disparités obtenues. Avec des fenêtres 1D composées d'une seule ligne, certains obstacles sont détectés ; toutefois la faible densité de la carte ne nous permet pas de détecter tous les obstacles. Avec des fenêtres de 3 pixels de hauteur, la carte des disparités (plus dense) permet de détecter l'ensemble des véhicules présents dans l'image et ce, sans fausse détection. Il est possible de définir une boîte englobante pour chaque véhicule ; mais la faible densité des cartes des disparités fait que toutes ces boîtes n'englobent pas parfaitement les véhicules.

La figure 4.26(f) montre clairement que la carte des disparités calculée par la méthode 2D est assez imprécise ; les obstacles ne sont pas tous détectés (le véhicule de la voie centrale n'est pas détecté) et davantage de fausses détections sont générées.

De plus, on peut remarquer sur les cartes des disparités obtenues que notre méthode 1D n'affecte pas de valeurs aux pixels situés à des profondeurs n'appartenant pas à l'intervalle des disparités traitées ; la méthode 2D affecte des valeurs de disparité aux zones de l'image les plus proches (comme l'intérieur du véhicule) alors que l'intervalle des disparités utilisé ne le permet pas.

Les résultats obtenus montre que la méthode 1D obtient des performances dégradées lorsque les contrastes sont faibles. Toutefois, les performances affichées ne sont pas moins bonnes que celles obtenues par la méthode 2D de Hirschmüller. Lorsque l'information locale est suffisante, notre méthode 1D présente de meilleurs résultats que la méthode 2D.

4.4 Implantation matérielle

Dans ce travail, l'approche 1D est fondée sur des analyses de bas-niveau. Les avantages sont que (1) l'algorithme est régulier, grâce à la forme des fenêtres de similarité et la structure des filtres flous, et que (2) aucun paramètre ne doit être adapté en fonction des images traitées. Implantée de manière traditionnelle sur des calculateurs de type séquentiel, cette méthode est incompatible avec la contrainte temps-réel d'une application. Cependant, on peut pallier cette difficulté en l'implantant sur une architecture dédiée de manière à atteindre des performances requises (cadence vidéo).



(a) (b)



(c) (d)



(e) (f)

Fig. 4.26 : (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (b) Boîtes englobantes obtenues sur la carte des disparités (a), (c) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et des fenêtres de 3 pixels de hauteur, (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (f) Boîtes englobantes obtenues sur la carte des disparités (e).

Lors d'un précédent projet de recherche régional, les laboratoires LAGIS et LEOST ont développé une carte électronique reconfigurable au format PCI appelée « STREAM[1] ». Elle est composée d'un composant logique programmable (de type FPGA de marque ALTERA). Une implantation matérielle de notre méthode 1D est donc envisagée sur cette architecture ce qui nécessitera une décomposition des traitements en opérations élémentaires et indépendantes. Cette idée d'implantation sur l'architecture *STREAM* a guidé les différents choix lors de l'élaboration de la méthode 1D. Même si l'implantation matérielle n'est pas inclus dans ce travail de thèse, nous allons introduire dans cette section les possibilités d'implantation de notre méthode.

Sur le *STREAM*, le traitement des informations est de type « flot de données » : l'image est balayée ligne par ligne puis pixel par pixel rendant ainsi disponible, à chaque instant t, un unique pixel [22]. Pour traiter le voisinage d'un pixel, il est nécessaire d'insérer des éléments de retard dans le flot de données. Avec ces éléments de retard, le calcul peut être effectué du fait que toutes les valeurs des pixels du voisinage deviennent disponibles au même instant. Le nombre d'opérations ainsi que la quantité de ressources matérielles nécessaires à l'implantation sont alors considérablement diminués par l'utilisation de fenêtres 1D, comparativement à l'implantation de méthodes 2D qui nécessitent des retards plus importants.

L'utilisation de traitements récursifs permet une diminution considérable du nombre d'opérations. Ces techniques ont déjà été introduites dans la littérature [23] et peuvent être transposées à notre problème. Ainsi, l'image est traitée ligne par ligne indépendamment l'une de l'autre. Sur chaque ligne, la fenêtre de largeur w est utilisée afin de mesurer la similarité pour des décalages s tels que $s \in [s_{min}..s_{max}]$ et pour la position p. Ces calculs exploitant des fenêtres 1D peuvent être implantés très efficacement de manière récursive.

La relation suivante formalise ce calcul :

$$C(x, y, i, w, s) = C(x - i, y, 0, w, s) \quad \text{avec} \quad i \in [-n..n]. \tag{4.1}$$

Les calculs de similarité pour la position centrale 0 des fenêtres de largeur w_n sont ainsi effectués. Cependant, les calculs de similarité pour des positions différentes du pixel à apparier dans la fenêtre ne seront effectués qu'une seule fois puisqu'il suffira d'utiliser la mesure d'une fenêtre de largeur identique w_n centrée en la position 0 pour un pixel voisin en tirant parti de la relation 4.1. Autrement dit, la mesure de similarité $C(x, y, 2, 7, 2)$ équivaut à la mesure $C(x - 2, y, 0, 7, 2)$ tout comme $C(x, y, -2, 15, 2)$ équivaut à $C(x + 2, y, 0, 15, 2)$.

[1] Système Temps-Réel d'Extraction et d'Analyse du Mouvement

Ainsi, pour chaque pixel (x, y), le volume des similarités et, a posteriori, le volume des confiances pourront être établis en réutilisant les valeurs calculées pour les pixels voisins. Cette astuce d'implantation fera décroître de manière très significative le nombre d'opérations à effectuer pour chaque pixel de l'image.

Concernant la suite des traitements, notons qu'un seul filtre identique, exploitant la logique floue, est appliqué pour toute la phase de modélisation de la confiance. Le filtre flou ainsi que la méthode de décision décrite dans ce manuscrit pourront être implantés dans la carte $STREAM$, puisqu'ils sont réguliers et n'intègrent que des opérations arithmétiques élémentaires.

Les opérations permettant le calcul des trois critères, comme le calcul de la valeur de courbure, le classement des scores de similarité ou encore le calcul du nombre de points d'inflexion, sont facilement implantables. Le calcul de la valeur de courbure se résume en une multiplication et deux additions. Le classement par valeur croissante des scores de similarité a déjà été implanté dans [22]. Le calcul du nombre de points d'inflexion se décompose en deux étapes qui sont (1) mise à un d'une variable d'un tableau si les scores de similarité voisins sont supérieurs à celle du décalage courant, (2) compter le nombre de valeur égale à un dans le tableau.

Ensuite, la fuzzyfication de ces critères pourra être effectuée par l'intermédiaire d'une « LUT[1] » (de complexité nulle) stockée dans un composant mémoire de l'architecture.

De la même manière, le mécanisme d'inférence est caractérisée par une LUT directement liée au tableau 3.1 (du chapitre 3) présentant les règles d'inférence du filtre flou.

Enfin l'étape de défuzzyfication, permettant d'obtenir la valeur finale de confiance, pourra être implantée par le calcul standard du centre de gravité qui peut être décomposé en opérations arithmétiques élémentaires. Ainsi, chacune de ces étapes pourra être efficacement implantée dans l'architecture $STREAM$.

La partie décision peut être perçue comme moins régulière et de ce fait, moins implantable sur une architecture dédiée. Toutefois, on peut déterminer pour chaque couple « volume des disparités/volume des confiances », les volumes réduits de la figure 3.8 ; ensuite, il suffira de retenir les valeurs finales de la disparité et de la confiance pour le pixel, en fonction de la valeur $T(w)$ fixée. Aussi, l'étape de validation directe/inverse a déjà été implantée dans l'architecture $STREAM$ par Pérez dans [22].

Dans une architecture parallèle telle le $STREAM$, les différents modulent travaillent en même temps mais sur des données différentes. De cette manière, le volume de données traitées est largement supérieur à celui que pourrait traiter une architecture séquentielle. Un résultat de calcul est donné à chaque top d'horloge et l'obtention des cartes des disparités et des confiances est réalisée avec un retard pur donné.

[1] Look-Up Table

L'étude sur l'implantation matérielle de la méthode 1D sur l'architecture *STREAM* sera approfondie au cours de travaux postérieurs à ce travail de thèse. De ce fait, cette section constitue une introduction à l'implantation matérielle de notre méthode 1D.

4.5 Conclusion

Dans ce chapitre, nous avons introduit le contexte lié aux transports routiers. Ce sujet de thèse est intégré dans le projet RaViOLi et plus précisément dans le volet vision de ce dernier. Notre méthode générique 1D élaborée peut être appliquée à un capteur spécifique catadioptrique pour une utilisation de détection d'obstacles à longue distance sur autoroutes. Les mesures de disparité présentent des cartes quasi-denses et précises (par comparaison avec celles obtenues par une méthode 2D classique) ; ces caractéristiques permettent l'utilisation d'une méthode simple et efficace d'extraction du profil de route et de détection d'obstacles. Celle-ci est fondée sur l'estimation du profil de route colonne par colonne puisque la densité élevée des cartes des disparités calculées ne nécessite pas de techniques cumulatives comme présentées au préalable. Les boîtes englobantes générées autour des obstacles de la scène sont plus précises avec les cartes des disparités générées par notre méthode 1D qu'avec celles calculées par la méthode 2D. Le présent système est capable de détecter efficacement des obstacles situés à une distance d'environ 180 mètres. L'objectif de la méthode d'extraction est de mettre en évidence l'intérêt d'utiliser des cartes des disparités denses et précises ; cette méthode d'extraction assez simple pourra, par la suite, être largement améliorée. Ainsi, nous avons montré la supériorité en termes de précision de notre méthode générique 1D appliquée à la détection d'obstacles à longue distance.

Le comportement de la méthode 1D face à des défauts de calibration ou encore au manque de texture a permis de montrer les limites de la méthode. Cependant, dans ces mêmes situations, la méthode 2D de Hirschmüller montre encore plus de faiblesses. Comme nous l'avons indiqué au début du deuxième chapitre, des auteurs ont mis en place une technique de rectification très légère présentée dans [21]. L'étape de rectification pourrait être appliquée à ces trois couples obtenus avec un stéréoscope décalibré de manière à améliorer les résultats de mise en correspondance.

Des premiers résultats sur des images réelles de type « transport » ont montré la viabilité de notre méthode 1D dans ce contexte embarqué.

Enfin, nous avons montré le potentiel d'implantation de notre méthode sur une architecture parallèle.

Références

[1] G. ALESSANDRETTI, A. BROGGI et P. CERRI : Vehicle and guard rail detection using radar and vision data fusion. *IEEE Transactions on Intelligent Transportation Systems*, 8(1):95–105, mars 2007. 104

[2] A. BROGGI, P. CERRI et P. C. ANTONELLO : Multi-resolution vehicle detection using artificial vision. *Proceedings of the IEEE Intelligent Vehicles Symposium 2004*, pages 310–314, Parma, Italy, juin 2004. 105, 106

[3] R. LABAYRADE : *Détection générique, robuste et rapide d'obstacles routiers par stéréovision embarquée.* Thèse de doctorat, Université Paris VI, 2004. 3, 105, 106

[4] R. LABAYRADE, J. DOURET et D. AUBERT : A multi-model lane detector that handles road singularities. *IEEE International Transportation Systems Conference*, pages 1143–1148, Toronto, Canada, septembre 2006. 106

[5] R. AUFRERE, R. CHAPUIS et F. CHAUSSE : A model driven approach for real time road recognition. *International Journal of Machine Vision and Applications*, pages 95–107, décembre 2001. 106

[6] R. CHAPUIS, F. MARMOITON, R. AUFRERE, F. COLLANGE et J. DERUTIN : Road detection and vehicles tracking by vision for an on-board acc system in the velac vehicle. *Proceedings of the Third International Conference on Information Fusion*, Paris, France, juillet 2000. 106

[7] G. TOULMINET, M. BERTOZZI, S. MOUSSET, A. BENSRHAIR et A. BROGGI : Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis. *IEEE Transactions on Image Processing*, 15(8):2364–2375, août 2006. 106

[8] N. HAUTIÈRE, R. LABAYRADE, M. PERROLLAZ et D. AUBERT : Road scene analysis by stereovision : a robust and quasi-dense approach. *Proceedings of the IEEE International Conference control, automation, robotics and vision*, pages 1453–1458, Singapore, décembre 2006. 106, 114

[9] M. LHUILLIER et L. QUAN : Match propagation for image-based modeling and rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8): 1140–1146, 2002. 6, 76, 106, 114

[10] H. HIRSCHMÜLLER : Improvements in real-time correlation-based stereo vision. *IEEE Workshop on Stereo and Multi-Baseline Vision*, Kauai, HI, USA, décembre 2001. 106, 107

[11] W. van der MARK et D.M. GAVRILA : Real-time dense stereo for intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):38–50, mars 2006. 9, 107, 108

[12] V. LEMONDE : *Stéréovision embarquée sur véhicule : De l'auto-calibrage à la détection d'obstacles*. Thèse de doctorat, Institut National des Sciences Appliquées de Toulouse, 2005. 2, 3, 33, 46, 110, 115

[13] D. H. LEE et I. KWEON : A novel stereo camera system by a biprism. *IEEE Transactions on Robotics and Automation*, 16(5):528–541, 2000. 111

[14] H. MATHIEU et F. DEVERNAY : Systèmes de miroirs pour la stéréoscopie. Rapport technique RR-0172, Institut National de Recherche en Informatique et en Automatique, INRIA, juin 1995. 12, 111

[15] L. DUVIEUBOURG, S. AMBELLOUIS et F. CABESTAING : Single-camera stereovision setup with orientable optical axes. *International Conference on Computer Vision and Graphics*, Warsaw, Poland, septembre 2004. 111

[16] S. SE et M. BRADY : Ground plane estimation, error analysis and applications. *Robotics and Autonomous Systems*, 39:59–71, mai 2002. 113

[17] R. LABAYRADE, D. AUBERT et J.P. TAREL : Real time obstacle detection in stereovision on non flat road geometry through V-disparity representation. *Procs. IEEE Intelligent Vehicles Symposium 2002*, pages 646–651, Versailles, France, juin 2002. 113, 114

[18] A. BROGGI, C. CARAFFI, P. PORTA et P. ZANI : The single frame stereo vision system for reliable obstacle detection used during the 2005 darpa grand challenge on terramax. *IEEE International Transportation Systems Conference*, pages 745–752, Toronto, Canada, septembre 2006. 115

[19] A. BROGGI, C. CARAFFI, R. FEDRIGA et P. GRISLERI : Obstacle detection with stereo vision for off-road vehicle navigation. *Proceedings of the IEEE International Workshop on Machine Vision for Intelligent Vehicles*, San Diego, USA, juin 2005. 115

[20] R. AUFRERE, F. MARMOITON, R. CHAPUIS, F. COLLANGE et J.P DÉRUTIN : Détection de route et suivi de véhicules par vision pour l'acc. *Traitement du Signal*, 17(3):233–248, 2000. 124

[21] A. FUSIELLO, E. TRUCCO et V. ROBERTO : A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000. 15, 46, 143

[22] M. PÉREZ : *Stéréovision dense par traitement adaptatif temps réel : Algorithmes et implantation*. Thèse de doctorat, Université des Sciences et Technologies de Lille, France, février 2005. 18, 35, 141, 142

[23] O. FAUGERAS, B. HOTZ, H. MATHIEU, T. VIÉVILLE, Z. ZHANG, P. FUA, E. THÉRON, L. MOLL, G. BERRY, J. VUILLEMIN, P. BERTIN et C. PROY : Real time correlation-based stereo : Algorithm, implementation and applications. Rapport de recherche RR-2013, Institut National de Recherche en Informatique et en Automatique, INRIA, août 1993. 8, 141

Conclusion générale

CE TRAVAIL de thèse présente une nouvelle méthode générique de mise en correspondance stéréoscopique par corrélation. La méthode proposée permet d'obtenir une carte des disparités à partir de l'analyse des courbes de similarité calculées sur un ensemble de fenêtres monodimensionnelles. La carte des disparités est associée à une carte des confiances qui permet de mesurer la validité de chaque disparité calculée. Précision et densité sont deux paramètres liés sur lesquels l'utilisateur peut jouer en fonction de l'application visée. Toute cette étude a d'abord été menée dans un cadre académique bien connu dans le domaine de la stéréovision de manière à montrer la généricité de la méthode ; cette dernière a ensuite été appliquée à la détection d'obstacles à longue distance à l'avant d'un véhicule routier. Le mémoire est composé de 4 chapitres.

Dans le chapitre 1, nous avons présenté un état de l'art sur la stéréovision passive et détaillé toutes les étapes de la chaîne allant de l'acquisition des couples d'images par un capteur stéréoscopique au calcul de profondeur. Une attention particulière a été accordée aux techniques locales de mise en correspondance de pixels qui possèdent des propriétés d'implantation matérielle intéressantes. Pour ces techniques, l'utilisation de fenêtres rectangulaires 2D est systématique et fait l'hypothèse que la disparité est constante dans le voisinage considéré. Or, cette hypothèse n'est pas vérifiée dans la plupart des cas et donc exploiter une fenêtre 2D n'est pas toujours légitime. Par ailleurs, ces méthodes déterminent, sans prendre aucune précaution, une disparité en chaque pixel y compris ceux localisés dans des zones de l'image non ou faiblement texturées ou encore dans les zones d'occultation.

Dans le chapitre 2, nous avons fait état des principaux travaux antérieurs sur la mise en correspondance réalisée à partir d'un capteur stéréoscopique linéaire. Ces méthodes n'exploitent que l'information projetée sur la ligne de deux capteurs CCD linéaires sous la contrainte d'un calibrage fort. Nous avons décrit les résultats de la comparaison de deux algorithmes classiques de mise en correspondance utilisant les mesures de similarité SAD et SSD appliquées sur des fenêtres 1D et 2D. Cette étude a montré que l'efficacité de la méthode utilisant des fenêtres 2D par rapport à celle exploitant des fenêtres 1D n'est pas accrue. En utilisant une fenêtre de similarité 2D, l'algorithme recueille un maximum d'informations pour la phase d'appariement. Ceci est particulièrement intéressant dans

des zones de l'image peu ou pas texturées ou présentant des motifs répétés, pour lesquelles les pixels sont caractérisés par la même valeur de disparité. Cependant, nous avons montré que les fenêtres de similarité 1D fournissent de meilleurs résultats, ou des résultats au moins équivalents, à ceux obtenus avec les fenêtres 2D pour des régions texturées de l'image ou proches des discontinuités.

Dans le troisième chapitre, nous avons présenté une nouvelle approche générique exploitant un ensemble de fenêtres 1D de mise en correspondance. Cette méthode intègre l'information contenue par plusieurs configurations de fenêtres 1D pour lesquelles la taille de la fenêtre et la position du pixel à apparier varient. Pour chaque décalage des fenêtres 1D, un score de similarité est calculé donnant lieu à un volume de similarité en chaque pixel à apparier. A chaque score de similarité est associé un indice de confiance quantifiant la vraisemblance que la valeur de décalage associée soit la valeur de disparité à affecter au pixel. Ces indices de confiance sont calculés en combinant plusieurs critères extraits des courbes de similarité. Un volume de confiance dont la géométrie est similaire à celle du volume de similarité est ainsi disponible. Une ultime étape de décision permet la sélection de la valeur de disparité à associer au pixel, à partir de l'analyse du volume des confiances. Nous avons validé cette méthode à partir d'une mesure de similarité SSD et en modélisant la confiance et l'étape de décision dans le cadre particulier de la logique floue.

Les résultats des disparités obtenus par notre méthode ont ensuite été comparés à ceux obtenus par six méthodes similaires exploitant des fenêtres 2D. Selon la méthode d'évaluation académique de Middlebury, la méthode 1D présente de bons résultats notamment pour les pixels en zones de discontinuité ou en zones texturées. Une étude complémentaire a présenté les résultats obtenus sur des couples d'images composées d'un plan fortement incliné. Notre méthode 1D a été comparée à la méthode 2D ayant obtenu le meilleur classement dans la première évaluation. Les résultats prouvent qu'une méthode locale exploitant des fenêtres 2D ne peut obtenir de bons résultats dans cette situation. A contrario, notre méthode 1D obtient des résultats très convainquants.

Dans le dernier chapitre, nous décrivons un travail dans lequel nous avons utilisé notre méthode générique 1D de mise en correspondance dans le cadre applicatif de la détection d'obstacles à l'avant d'un véhicule routier. Dans ce contexte, nous avons obtenu des cartes des disparités bien plus précises avec notre approche 1D qu'avec la méthode 2D précédemment utilisée. Une méthode assez simple d'extraction des obstacles sur la chaussée a permis de montrer que nos cartes quasi-denses et précises permettent une détection et une localisation efficaces des obstacles et ce, même s'ils sont situés à longue distance. Cette même méthode, appliquée sur les cartes des disparités obtenues avec la méthode 2D, engendre de nombreuses fausses détections et une localisation imprécise des

obstacles. Le comportement de la méthode 1D face à des défauts de calibration ou encore au manque de texture a permis de montrer les limites de la méthode. Les premiers résultats sur des images réelles de type « transport » ont montré la viabilité de notre méthode 1D dans ce contexte embarqué. Enfin, nous avons montré le potentiel d'implantation de notre méthode sur une architecture dédiée.

Plusieurs perspectives de travail sont présentées dans la suite de ce mémoire.

Perspectives

LE TRAVAIL décrit dans ce mémoire propose une méthode générique de mise en correspondance stéréoscopique exploitant des fenêtres 1D. Pour chaque couple d'images stéréoscopiques, cette technique calcule une carte des disparités quasi-dense et précise ainsi que sa carte des confiances associées. Cette méthode fait l'hypothèse que la confiance associée à chaque disparité est d'autant plus faible que le voisinage du pixel considéré est peu texturé. Par ailleurs, lorsque le pixel appartient à une zone de discontinuité, certaines ambiguïtés apparaissent et la méthode n'y affecte pas de disparité. Cette dernière partie du mémoire présente quelques axes de recherche qui visent à améliorer les performances de notre méthode en se focalisant tout particulièrement sur la prise en compte des zones des pixels non affectés.

P.1 Pixels non affectés en zones texturées

Aucune valeur de disparité n'est affectée en certains pixels pourtant localisés dans des zones texturées de l'image. L'ambiguïté qui apparaît se manifeste par une structure particulière du volume réduit des décalages défini en ce pixel. Nous proposons de l'analyser afin de lever cette ambiguïté. Deux phénomènes peuvent alors apparaître.

P.1.1 Décision au sous-pixel

Le premier phénomène est illustré par la figure P.1. Il est représentatif des situations pour lesquelles la valeur de disparité évolue régulièrement et de manière oblique le long d'une surface (effet de dévers). Nous l'avons vu apparaître dans les chapitres précédents lors de notre étude des couples d'images de plans inclinés (cf figure 3.14) pour lesquels la valeur de disparité évolue en fonction de l'angle de dévers du plan le long des lignes des images. Dans la figure P.1, le volume réduit est composé de deux valeurs voisines de décalage (ici $s^*(p,w) = 5$ et $s^*(p,w) = 6$). Cette configuration ne permet pas, selon la méthode de décision actuelle, de définir la valeur de disparité du pixel puisque la condition spécifiée par l'équation (3.10) n'est jamais remplie (pour $T_{dec} = 1$). Deux solutions sont alors possibles pour pallier cette difficulté :

1. diminuer le seuil T_{dec} afin d'affecter l'une ou l'autre des deux valeurs voisines de disparité. L'étape de raffinement des résultats (par interpolation de la courbe des similarités) déterminerait ensuite une valeur de disparité au sous-pixel ;

2. calculer la disparité au sous-pixel directement à partir des décalages candidats et de leurs confiances associées.

Dans ce deuxième cas, la valeur de disparité au sous-pixel pourrait être déterminée comme la moyenne des deux valeurs potentielles de disparité, pondérées par les valeurs de confiance associées, calculée de la manière suivante :

$$d(x,y) = \frac{\sum_{p,w} s^*(p,w).Conf_{LF}(p,w,s^*(p,w))}{\sum_{p,w} Conf_{LF}(p,w,s^*(p,w))} \,. \tag{P.2}$$

Par ailleurs, cette solution pourrait être étendue à l'ensemble des pixels des images afin d'obtenir une carte des disparités avec une précision sub-pixellique. Les résultats obtenus seraient comparés à ceux obtenus avec les méthodes classiques d'interpolation.

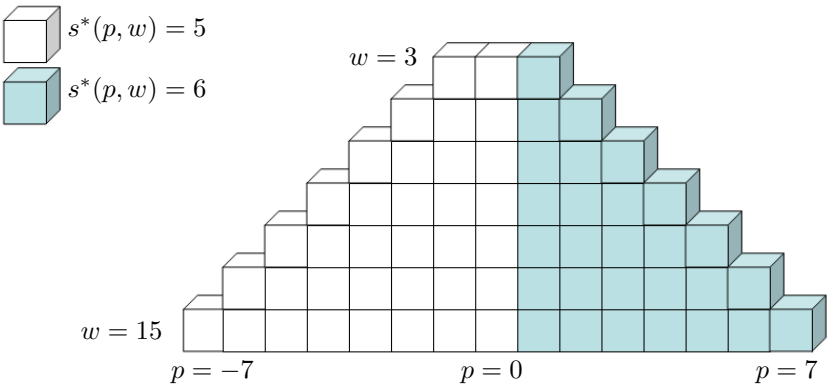
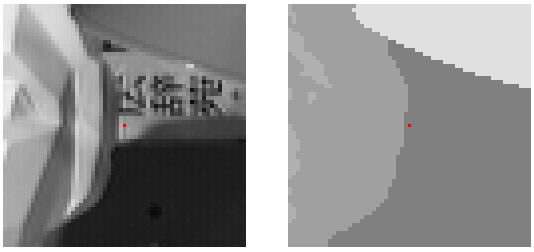


Fig. P.1 : Volume réduit (après la relation (3.6)) des décalages candidats, pour un pixel donné (x, y), défini par l'ensemble des valeurs de décalage ayant la confiance maximale, toutes configurations confondues de la fenêtre de similarité.

P.1.2 Ambiguïté d'une discontinuité

Le second phénomène est illustré par la figure P.2. Sur cette figure, un pixel proche d'une zone de discontinuité dans le couple d'images *Head and Lamp* est affiché en rouge au centre des images 4.2(a) et 4.2(b), représentant respectivement une partie de l'image gauche et une partie de la vérité-terrain associée. Le volume des confiances généré par notre méthode 1D est présenté figure P.3. Deux valeurs voisines de décalage (ici $s^*(p,w) = 11$ et $s^*(p,w) = 9$ caractérisant respectivement la disparité de la statue et celle de l'arrière-plan de l'image) composent le volume réduit. Selon la méthode de décision actuelle, cette configuration ne permet pas une décision de la valeur de disparité à affecter au pixel puisque la condition spécifiée par l'équation (3.10) n'est jamais remplie (pour $T_{dec} = 1$).



(a) (b)

Fig. P.2 : Pixel en zone de discontinuité (a) dans l'image gauche et (b) dans la vérité-terrain.

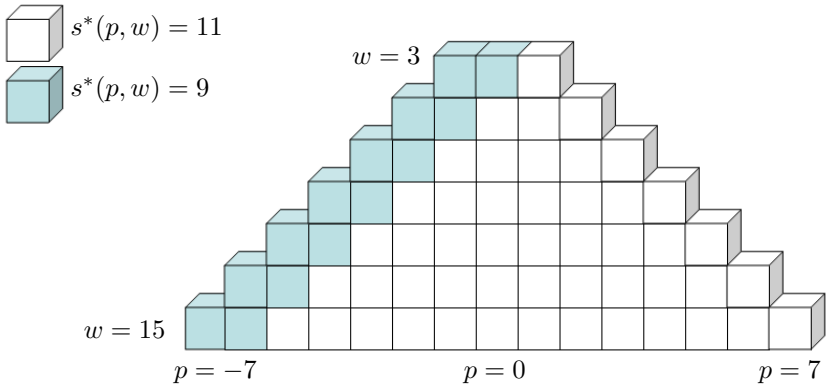


Fig. P.3 : Volume réduit (après la relation (3.6)) des décalages candidats, pour un pixel donné (x, y), défini par l'ensemble des valeurs de décalage ayant la confiance maximale, toutes configurations confondues de la fenêtre de corrélation.

Sur la figure P.3, on se rend compte que, dès lors que le contour « statue/fond » est pris en compte dans la fenêtre de similarité, pour une position donnée du pixel à apparier, la disparité appartenant à la statue (i.e. $s^*(p, w) = 11$) est affectée au pixel appartenant à l'arrière-plan (i.e. $s^*(p, w) = 9$), créant ainsi une erreur de mise en correspondance. A contrario, la valeur réelle de disparité de l'arrière-plan est correctement définie pour les positions du pixel dans la fenêtre localisée sur la partie gauche du volume réduit. En fait, seules ces positions définissent des fenêtres composées uniquement de pixels appartenant à l'arrière-plan, fenêtres pour lesquelles l'hypothèse de constance de la disparité est respectée. Pour cet ensemble de fenêtres, la mise en correspondance est alors correcte.

Là encore, nous pourrions envisager deux solutions pour limiter les imprécisions liées à la prise en compte des fenêtres les plus larges (fenêtres qui incluent des pixels dont la disparité est différente) lors de la phase de décision. Il pourrait s'agir :

1. d'affecter la disparité appartenant à la position centrale ($p = 0$) pour la fenêtre de largeur minimale $w = 3$. Dans l'exemple de la figure P.3, la disparité $s^*(p, w) = 9$ serait affectée au pixel (x, y) ;

2. de parcourir, pour chaque pixel non encore affecté, le volume réduit des décalages candidats, cette fois-ci dans le sens croissant de la taille des voisinages et pour une valeur inférieure du seuil T_{dec}.

P.2 Utilisation de l'information couleur

La méthode proposée a été conçue pour des couples d'images en niveau de gris. De manière à obtenir des cartes des disparités plus denses et lever certaines ambiguïtés de mise en correspondance, l'utilisation de l'information couleur semble être intéressante. Les résultats obtenus lors d'un premier travail sur l'intégration de l'information couleur dans notre méthode 1D ont été présentés au colloque *IEEE Conference on Signal-Image Technology and Internet-based Systems 2007*. Dans ce papier, notre méthode a été appliquée sur chaque canal du couple d'images stéréoscopiques *Head and Lamp* dans l'espace couleur initial (i.e. l'espace RGB). De cette manière, une carte des disparités et une carte des confiances ont été calculées pour chacune des composantes rouge, vert et bleu. Une étape simple de combinaison des résultats, obtenus indépendamment sur chaque composante, a été proposée. Les résultats obtenus montrent l'intérêt de l'utilisation de l'information couleur puisque la carte finale des disparités est légèrement plus dense et plus précise que celle obtenue à partir des images en niveaux de gris.

Ces résultats sont prometteurs puisque l'image et l'espace couleur utilisés pour cette comparaison ne sont pas les plus discriminants. En effet, l'image *Head and Lamp* est très peu colorée. Les résultats seraient plus probants sur des images pour lesquelles l'information couleur est omniprésente. D'autre part, dans l'espace couleur RGB, les trois composantes sont très corrélées. L'utilisation d'espaces couleur à axes indépendants (comme l'espace $I_1I_2I_3$) ou luminance-chrominance (comme l'espace AC_1C_2) apporterait plus d'information discriminante nécessaire pour lever certaines ambiguïtés de mise en correspondance. Cependant, le choix de l'espace à utiliser reste un problème important.

Une idée serait de calculer en parallèle les cartes des disparités et des confiances pour chacune des composantes de plusieurs espaces couleur. Ainsi, pour chaque pixel de l'image, un ensemble de valeurs potentielles de la disparité permettrait de choisir la disparité la plus vraisemblable puisqu'à chacune d'entre-elles est associé un indice de confiance quantifiant la vraisemblance que cette valeur de disparité soit la bonne.

L'explosion du volume de données est importante sur une architecture séquentielle. Cependant, cette approche pourrait être efficacement implantée sur une architecture parallèle puisque les traitements seront identiques sur les trois canaux de l'image couleur.

P.3 Augmentation de la densité de la carte

Les deux premières sections de cette partie avaient pour objectif l'affectation d'une valeur de disparité en tout pixel situé dans des zones présentant certaines ambiguïtés dues aux discontinuités. Il est ici question de compléter, grâce à une étape de post-traitement, les zones non affectées par la méthode 1D et qui ne le seraient toujours pas après l'application de l'une des méthodes proposées en perspectives.

Il s'agit donc de calculer une disparité en des pixels inclus dans des régions peu ou non texturées. Tout au long de ce mémoire, nous avons défendu l'idée qu'aucune disparité ne doit être calculée en un pixel d'une zone peu ou non texturée et il n'est pas question ici d'y déroger. Toutefois, il nous parait important de proposer une méthode qui permettrait à un utilisateur qui le souhaiterait de calculer une carte dense à partir de la carte quasi-dense que notre première méthode fournirait. Il faut souligner que cette extension de la méthode sous-entend que l'on change le cadre de travail : il ne s'agirait plus de « métrologie », mais de reconstruction de scène par le biais d'une carte dense.

Il est aisé de déterminer si un pixel appartient ou non à une zone non texturée à partir d'un ensemble de courbes de similarité, comme le montre la figure P.4 illustrant une courbe de similarité obtenue avec la mesure SSD pour un pixel (x, y) situé dans une zone très peu texturée. Pour chacun de ces pixels, nous pourrions alors proposer de retenir plusieurs couples de valeurs « décalage/confiance » grâce à l'analyse du volume des similarités. Ensuite, ces décalages candidats seraient combinés avec les disparités calculées aux pixels voisins lors d'une étape itérative de post-traitement. Cette combinaison pourrait tenir compte des valeurs de confiance associées à chaque décalage et chaque disparité intervenant dans le calcul. Ce traitement itératif de plus haut niveau, du fait qu'il n'est pas régulier, devrait être réalisé par une architecture séquentielle de calcul, limitant de ce fait ses performances en temps-réel.

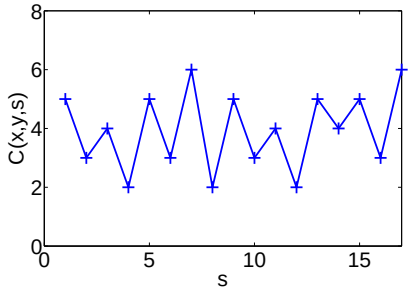


Fig. P.4 : Courbe de similarité obtenue avec la mesure SSD pour un pixel (x, y) situé dans une zone très peu texturée.

Liste des publications

Communications internationales avec actes et comité de lecture

◇ A colour correlation-based stereo matching using 1D windows. S. Lefebvre, S. Ambellouis et F. Cabestaing. Dans *IEEE Conference on Signal-Image Technology and Internet-based Systems*, Shanghai, Chine. Décembre 2007.

◇ Obstacle detection using a single camera stereo sensor. L Duvieubourg, S. Ambellouis, S. Lefebvre et F. Cabestaing. Dans *IEEE Conference on Signal-Image Technology and Internet-based Systems*, Shanghai, Chine. Décembre 2007.

◇ Obstacles detection on a road by dense stereovision with 1D correlation windows and fuzzy filtering. S. Lefebvre, S. Ambellouis et F. Cabestaing. Dans *IEEE Intelligent Transportation Systems Conference*, pp. 739–744, Toronto, Canada. Septembre 2006.

Communication nationale avec actes et comité de lecture

◇ Approche monodimensionnelle de la mise en correspondance stéréoscopique par corrélation. S. Lefebvre, S. Ambellouis, et F. Cabestaing. Dans XXI^{eme} *colloque GRETSI*, pp. 189–192, Troyes, France. Septembre 2007.

Séminaires internes avec actes et comité de lecture

◇ Mise en correspondance stéréoscopique de fenêtres monodimensionnelles par logique floue - Application à la détection d'obstacles routiers. S. Lefebvre, S. Ambellouis, et F. Cabestaing. Dans *Journée des doctorants de l'INRETS*, pp. 89–101, Villeneuve d'Ascq, France. Avril 2006.

◇ Légitimité de l'utilisation d'un voisinage 2D pour la corrélation en stéréovision dense. S. Lefebvre, S. Ambellouis, et F. Cabestaing. Dans *Journée des doctorants de l'INRETS*, pp. 63–71, Villeneuve d'Ascq, France. Février 2005.

Rapports techniques internes

◇ Rapport d'activité 2006, S. Lefebvre. *INRETS/RA-06-715-FR*. Octobre 2006.

◇ Rapport d'activité 2005, S. Lefebvre. *INRETS/RA-05-720-FR*. Octobre 2005.

Publications liées à la thématique GNSS (travaux antérieurs)

Communications internationales avec actes et comité de lecture

◇ Satellite availability in a railway mountainous environment : can we use satellite positioning for safety applications ? J. Marais, A. Flancquart et S. Lefebvre. Dans *Proceedings of the* 7^{th} *World Congress on Railway Research*, Montréal, Canada. Juin 2006.

◇ Satellite propagation path model along a railway track for GNSS applications. J. Marais, S. Lefebvre et M. Berbineau. Dans *IEEE Vehicular Technology Conference*, Los Angeles, CA, USA. Septembre 2004.

Table des figures

I.1 Illustration de l'hypothèse 1 : (a)(c) vérifiée et (b)(d) non vérifiée avec respectivement l'image et l'histogramme associé. 6
I.2 Illustration de l'hypothèse 2 non vérifiée : (a) pour *Head and Lamp* et (b) pour une image de route. 7

1.1 Le stéréoscope en configuration quelconque. 13
1.2 Illustration des épipoles. 14
1.3 Configuration parallèle d'un stéréoscope binoculaire. 15
1.4 Triangulation géométrique dans le plan $(O_r, (X_r, Z_r))$. 16
1.5 Grille de mise en correspondance basée sur la contrainte épipolaire. . . . 19
1.6 Application de la contrainte d'unicité, l'image gauche étant l'image de référence. 20
1.7 Application de la contrainte d'ordre. 21
1.8 Application de la contrainte de position. 21
1.9 Exemple d'une courbe de dissimilarité. 24
1.10 Illustration des fenêtres utilisées par Fusiello *et al.* dans [28]. 28
1.11 Illustration des fenêtres utilisées par Hirschmüller dans [29] (le pixel noir représente la position du pixel à apparier dans la fenêtre). 29

2.1 Principe d'une caméra linéaire. 42
2.2 Plan de visée du système. 43
2.3 Application des contraintes épipolaire et de signe du gradient afin d'éliminer des couples candidats à l'appariement. 44
2.4 (a) Image gauche , (b) Image droite, (c) Vérité-terrain, (d) Régions texturées (pixels blancs) et non-texturées (pixels gris), (e) Régions occultées (pixels noirs) et régions en zone de discontinuité (pixels blancs) pour les couples d'images (dans l'ordre) *Head and Lamp*, *Sawtooth*, *Venus* et *Map*. 49
2.5 (a) Image gauche *Cones*, (b) Image gauche *Teddy*. 50
2.6 (a) Vérité-terrain *Head and Lamp*, (b) Carte des disparités obtenues dans [11], (c) Mise en évidence de certaines erreurs d'appariement de (b) qui ne sont pas considérées comme telles par la méthode d'évaluation. 51

2.7 (a) Image gauche , (b) Vérité-terrain, (c) Régions texturées (pixels blancs) et non-texturées (pixels gris), (d) Régions occultées (pixels noirs) et régions en zone de discontinuité (pixels blancs) pour le couple d'images *Head and Lamp*. 53
2.8 Comparaison qualitative des résultats obtenus par la mesure SAD avec des fenêtres de forme carrée de différentes tailles : (a) 3×3, (b) 11×11, (c) 19×19. La première ligne : cartes des disparités. La seconde ligne : cartes d'erreur pour le couple d'images *Head and Lamp*. 55
2.9 Comparaison qualitative des résultats obtenus par la mesure SAD avec des fenêtres de forme ligne de différentes tailles : (a) 3×1, (b) 11×1, (c) 19×1. La première ligne : cartes des disparités. La seconde ligne : cartes d'erreur pour le couple d'images *Head and Lamp*. 55
2.10 Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de forme carrée de différentes tailles : (a) 3×3, (b) 11×11, (c) 19×19. La première ligne : cartes des disparités. La seconde ligne : cartes d'erreur pour le couple d'images *Head and Lamp*. 56
2.11 Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de forme ligne de différentes tailles : (a) 3×1, (b) 11×1, (c) 19×1. La première ligne : cartes des disparités. La seconde ligne : cartes d'erreur pour le couple d'images *Head and Lamp*. 56
2.12 Quelques régions présentant des erreurs typiques. 57
2.13 Comparaison qualitative de résultats de disparité obtenus pour la sous-image (A) par la mesure SSD (Sous-image (A) - Vérité-terrain associée - Disparités obtenues avec une fenêtre carrée 11×11 - Disparités obtenues avec une fenêtre ligne 11×1). 58
2.14 Comparaison qualitative de résultats de disparité obtenus pour la sous-image (B) par la mesure SSD (Sous-image (B) - Vérité-terrain associée - Disparités obtenues avec une fenêtre carrée 11×11 - Disparités obtenues avec une fenêtre ligne 11×1). 58
2.15 Comparaison qualitative de résultats de disparité obtenus pour la sous-image (C) par la mesure SSD (Sous-image (C) - Vérité-terrain associée - Disparités obtenues avec une fenêtre carrée 11×11 - Disparités obtenues avec une fenêtre ligne 11×1). 58
2.16 Comparaison qualitative de résultats de disparité obtenus pour la sous-image (D) par la mesure SSD (Sous-image (D) - Vérité-terrain associée - Disparités obtenues avec une fenêtre carrée 11×11 - Disparités obtenues avec une fenêtre ligne 11×1). 58

2.17 Comparaison quantitative des résultats de disparité obtenus avec les mesures SAD et SSD basées sur des fenêtres de formes carrée (2D) et ligne (1D) de différentes tailles dans les régions (a) non-occultées ($B_{\overline{\mathcal{O}}}$), (b) non-texturées ($B_{\overline{\mathcal{T}}}$), (c) texturées ($B_{\mathcal{T}}$), et (d) en discontinuités ($B_{\mathcal{D}}$) pour le couple d'images *Head and Lamp*. 61
2.18 Comparaison quantitative des résultats de disparité obtenus avec les mesures SAD et SSD basées sur des fenêtre de formes carrée et ligne de différentes tailles dans les régions (a) non-occultées ($B_{\overline{\mathcal{O}}}$), (b) non-texturées ($B_{\overline{\mathcal{T}}}$), (c) texturées ($B_{\mathcal{T}}$), et (d) en discontinuités ($B_{\mathcal{D}}$). 63
2.19 (a) Image de synthèse gauche, (b) Vérité-terrain. 65
2.20 Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de formes différentes de taille : (a) 3×1, (b) 3×3 ; la première ligne présente les cartes des disparités et la seconde les cartes d'erreur. . . 65
2.21 Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de formes différentes de taille : (a) 11×1, (b) 11×11 ; la première ligne présente les cartes des disparités et la seconde les cartes d'erreur. 66
2.22 Comparaison qualitative des résultats obtenus par la mesure SSD avec des fenêtres de formes différentes de taille : (a) 19×1, (b) 19×19 ; la première ligne présente les cartes des disparités et la seconde les cartes d'erreur. 66
2.23 Comparaison quantitative des résultats de disparité obtenus avec les mesures SAD et SSD basées sur des fenêtres de formes carrée (2D) et ligne (1D) de différentes tailles. 67

3.1 Descriptif de l'approche 1D. 73
3.2 Pixel à apparier placé en différentes positions dans une fenêtre 1D. 74
3.3 Volume de similarité, pour un pixel donné (x, y), défini par l'ensemble des scores de similarité toutes configurations confondues de la fenêtre. 75
3.4 Courbes de dissimilarité pour différentes configurations de la fenêtre : (a) $w = 7$ et $p = 0$, (b) $w = 13$ et $p = 0$, (c) $w = 19$ et $p = 0$, (d) $w = 13$ et $p = -6$, (e) $w = 13$ et $p = 6$. 76
3.5 (a) Fonctions d'appartenance du critère Cur, (b) Fonctions d'appartenance du critère N et (c) Fonctions d'appartenance du critère R. 81
3.6 Volume réduit suivant la relation (3.6). 84
3.7 Volume réduit suivant la relation (3.7). 84
3.8 (a) Volume réduit suivant la relation (3.8) - (b) Données obtenues suivant la relation (3.9). 85

3.9 (a) Image gauche *Head and Lamp*, (b) Vérité-terrain, (c) Zones texturées (pixel gris) et non texturées (pixels blancs), (d) Carte des disparités, (e) Carte des confiances. 87
3.10 Image gauche *Head and Lamp* (a), Vérité-terrain (b), Notre méthode 1D (c), Realtime (d), $SSD + MF$ (e), Improved Coop. (f), Adapt. Weights (g), Comp. Win. (h), MMHM colors (i). 89
3.11 Image gauche *Sawtooth* (a), Vérité-terrain (b), Notre méthode 1D (c), Realtime (d), $SSD + MF$ (e), Improved Coop. (f), Adapt. Weights (g), Comp. Win. (h), MMHM colors (i). 90
3.12 Image gauche *Venus* (a), Vérité-terrain (b), Notre méthode 1D (c), Realtime (d), $SSD+MF$ (e), Improved Coop. (f), Adapt. Weights (g), Comp. Win. (h), MMHM colors (i). 91
3.13 Image gauche *Map* (a), Vérité-terrain (b), Notre méthode 1D (c), Realtime (d), $SSD+MF$ (e), Improved Coop. (f), Adapt. Weights (g), Comp. Win. (h), MMHM colors (i). 92
3.14 Cartes des disparités calculées avec notre méthode 1D sur les couples d'images (a) *Sawtooth*, (b) *Venus* et (c) *Map*. 94
3.15 (a) Image gauche et (b) image droite du couple d'images *plan incliné* et (c) la vérité-terrain associée. 95
3.16 (a) Carte des disparités calculée par la méthode de Hirschmüller avec une fenêtre de taille 7×9, (b) Vérité-terrain et (c) Carte des disparités calculée avec notre méthode 1D. 96
3.17 Cartes des disparités et profil des disparités de la colonne encadrée en rouge pour respectivement la méthode de Hirschmüller (a)(b), la vérité-terrain (c)(d) et notre méthode 1D (e)(f). 97
3.18 (a) Image droite, (b) Vérité-terrain, (c) Carte des disparités calculée par la méthode de Hirschmüller avec une fenêtre de taille 7×9 et (d) Carte des disparités calculée avec notre méthode 1D (avec $T_{dec} = 1$). 99

4.1 (a) Image gauche, (b) Vérité-terrain, (c) Pixels de l'arrière-plan, (d) Pixels de la chaussée du couple 95 de la séquence présentée dans [11]. 108
4.2 Carte des disparités calculée par (a) la méthode de Hirschmüller avec la fenêtre de taille 7×9, (b) notre méthode 1D avec $T_{dec} = 0.8$ pour le couple 95 de la séquence présentée dans [11]. 108
4.3 Prototype du stéréoscope mono-caméra à axes orientables. 111
4.4 Vision fixe à faible distance focale. 112
4.5 Vision orientable à longue distance focale. 112
4.6 Illustration de la V-disparité. 114
4.7 Distance en fonction de la valeur de disparité. 117

4.8 (a) Image gauche d'autoroute sans véhicule, (b) Vérité-terrain, (c) Carte des disparités obtenue par notre méthode 1D, (d) Carte des disparités obtenue par la méthode de Hirschmüller - Les figures (e) et (g) présentent les erreurs générées par la méthode 1D et (f) et (h) pour la méthode 2D pour des seuils d'erreur respectifs de 1 et 0. 119
4.9 (a) Image gauche d'autoroute avec véhicules, (b) Vérité-terrain, (c) Carte des disparités obtenue par notre méthode 1D, (d) Carte des disparités obtenue par la méthode de Hirschmüller - Les figures (e) et (g) présentent les erreurs générées par la méthode 1D et (f) et (h) pour la méthode 2D pour des seuils d'erreur respectifs de 1 et 0. 120
4.10 (a) Image gauche, (b) Colonne utilisée, (c) Profil des disparités, (d) Détection des contours de l'obstacle, (e) Boîtes englobantes associées à la carte des disparités, (f) Boîtes englobantes associées à l'image gauche. . . 122
4.11 Résultats obtenus sur la vérité-terrain du couple d'images 10. 122
4.12 Résultats obtenus sur le couple d'images 10 avec la méthode 1D. 123
4.13 Résultats obtenus sur le couple d'images 10 avec la méthode 2D de Hirschmüller. 123
4.14 Cartes des disparités et profil des disparités de la colonne encadrée en rouge pour respectivement la méthode de Hirschmüller (a)(b), la vérité-terrain (c)(d) et notre méthode 1D (e)(f). 125
4.15 Cartes des disparités obtenues par notre méthode 1D et images gauche pour lesquelles les obstacles ont été détectés pour respectivement T_{dec} égal à : 1.0 (a)(b), 0.9 (c)(d), 0.8 (e)(f), 0.7 (g)(h), 0.6 (i)(j). 126
4.16 (a) Image gauche, (b) Image droite, (c) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (f) Boîtes englobantes obtenues sur la carte des disparités (e) - pour un couple d'images stéréoscopiques texturées. 128
4.17 (a) Image gauche, (b) Image droite, (c) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (f) Boîtes englobantes obtenues sur la carte des disparités (e) - pour un couple d'images stéréoscopiques pour lequel la chaussée n'est pas texturée. 129

4.18 (a) Image gauche, (b) Image droite (caméra droite sur-élevée de 5 cm), (c) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (f) Boîtes englobantes obtenues sur la carte des disparités (e), (g) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (h) Boîtes englobantes obtenues sur la carte des disparités (g) - pour un couple d'images stéréoscopiques texturées. 130

4.19 (a) Image gauche, (b) Image droite (caméra droite inclinée vers la chaussée d'un angle de 0.05 degré), (c) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (f) Boîtes englobantes obtenues sur la carte des disparités (e), (g) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (h) Boîtes englobantes obtenues sur la carte des disparités (g) - pour un couple d'images stéréoscopiques texturées. . . 131

4.20 (a) Image gauche, (b) Image droite (caméra droite orientée vers l'extérieur d'un angle de 0.5 degré), (c) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.9$, (f) Boîtes englobantes obtenues sur la carte des disparités (e), (g) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (h) Boîtes englobantes obtenues sur la carte des disparités (g) - pour un couple d'images stéréoscopiques texturées. 132

4.21 (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et une hauteur de fenêtre de 5 pixels, (b) Boîtes englobantes obtenues sur la carte des disparités (a) (ici, la caméra droite est sur-élevée de 5 cm). . 134

4.22 (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et une hauteur de fenêtre de 5 pixels, (b) Boîtes englobantes obtenues sur la carte des disparités (a) (ici, la caméra droite est inclinée vers la chaussée d'un angle de 0.1 degré). 135

4.23 (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et une hauteur de fenêtre de 5 pixels, (b) Boîtes englobantes obtenues sur la carte des disparités (a) (ici, la caméra droite est orientée vers l'extérieur d'un angle de 0.5 degré). 135

4.24 (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (b) Boîtes englobantes obtenues sur la carte des disparités (a), (c) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et des fenêtres de 3 pixels de hauteur, (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (f) Boîtes englobantes obtenues sur la carte des disparités (e). 137
4.25 (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et $n \in [1..20]$, (b) Boîtes englobantes obtenues sur la carte des disparités (a). 138
4.26 (a) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$, (b) Boîtes englobantes obtenues sur la carte des disparités (a), (c) Carte des disparités obtenues avec notre méthode 1D avec $T_{dec} = 0.7$ et des fenêtres de 3 pixels de hauteur, (d) Boîtes englobantes obtenues sur la carte des disparités (c), (e) Carte des disparités obtenues avec la méthode de Hirschmüller (fenêtre de taille 7×9), (f) Boîtes englobantes obtenues sur la carte des disparités (e). 140

P.1 Volume réduit (après la relation (3.6)) des décalages candidats, pour un pixel donné (x, y), défini par l'ensemble des valeurs de décalage ayant la confiance maximale, toutes configurations confondues de la fenêtre de similarité. 150
P.2 Pixel en zone de discontinuité (a) dans l'image gauche et (b) dans la vérité-terrain. 151
P.3 Volume réduit (après la relation (3.6)) des décalages candidats, pour un pixel donné (x, y), défini par l'ensemble des valeurs de décalage ayant la confiance maximale, toutes configurations confondues de la fenêtre de corrélation. 151
P.4 Courbe de similarité obtenue avec la mesure SSD pour un pixel (x, y) situé dans une zone très peu texturée. 153

Liste des tableaux

1.1 Illustration de la Transformée *Rank*. 26
1.2 Illustration de la Transformée *Census*. 27

2.1 Synthèse des observations du comportement des fenêtres 1D et 2D en différentes zones de l'image. 60

3.1 Règles d'inférence du filtre flou. 82
3.2 Tableau récapitulatif des méthodes comparables à notre approche 1D et relatées dans l'évaluation de Middlebury. 88
3.3 Résultats obtenus à partir de l'évaluation de Middlebury sur le couple d'images *Head and Lamp* (avec $|\Delta\varepsilon| > 0.5$). 89
3.4 Résultats obtenus à partir de l'évaluation de Middlebury sur le couple d'images *Sawtooth* (avec $|\Delta\varepsilon| > 0.5$). 90
3.5 Résultats obtenus à partir de l'évaluation de Middlebury sur le couple d'images *Venus* (avec $|\Delta\varepsilon| > 0.5$). 91
3.6 Résultats obtenus à partir de l'évaluation de Middlebury sur le couple d'images *Map* (avec $|\Delta\varepsilon| > 0.5$). 92
3.7 Effets de la validation directe/inverse (avec $|\Delta\varepsilon| > 0.5$). 94
3.8 Comparaison des taux d'erreur pour la méthode de Hirschmüller et notre méthode 1D (avec $T_{dec} = 1$) pour le couple d'images *plan incliné*. 98
3.9 Comparaison des taux d'erreur avec $|\Delta\varepsilon| > 1$ pour la méthode de Hirschmüller et notre méthode 1D (avec $T_{dec} = 1$) pour plusieurs angles de dévers du plan incliné. 100

4.1 Comparaison des taux d'erreur d'appariement pour les pixels caractérisant la chaussée. 109
4.2 Comparaison des taux d'erreur d'appariement pour les pixels caractérisant l'arrière-plan. 109
4.3 Comparaison des taux d'erreur calculés sur les cartes des disparités des figures 4.8(c) et 4.8(d). 119
4.4 Comparaison des taux d'erreur calculés sur les cartes des disparités des figures 4.9(c) et 4.9(d). 120

Résumé/Abstract

Approche monodimensionnelle de la mise en correspondance stéréoscopique par corrélation

Application à la détection d'obstacles routiers

En stéréovision, une carte dense des disparités peut être déterminée grâce à des méthodes locales exploitant la mise en correspondance des contenus de voisinages bidimensionnels. Toutefois, dans le contexte d'une application de métrologie 3D, il n'est pas nécessaire que la carte soit dense, mais elle doit être très précise. Par ailleurs, en présence d'un fort effet de perspective, d'importantes variations locales de la disparité apparaissent et les méthodes locales exploitant des fenêtres 2D ne garantissent plus une précision suffisante. Dans ce travail, nous proposons une méthode générique de mise en correspondance qui exploite des caractéristiques extraites d'un ensemble de courbes de similarité calculées sur des fenêtres 1D. Nous établissons une notion de confiance reliant directement la précision à la densité de la carte des disparités obtenue, entre lesquelles il est alors possible de fixer un compromis. La méthode générique proposée est évaluée en comparant ses résultats à ceux obtenus par six autres méthodes locales similaires. Nous montrons ainsi son efficacité, tout particulièrement quand la scène contient un plan fortement incliné. Nous testons finalement notre méthode dans le cadre d'une application de détection d'obstacles à l'avant d'un véhicule routier. Nous montrons qu'en segmentant notre carte des disparités par un procédé assez simple, la détection des obstacles est possible et surtout plus précise qu'avec des cartes des disparités calculées sur des voisinages 2D. Le comportement de la méthode 1D face à différents défauts de texture ou de calibration est étudié à la fois sur des images de synthèse et des images réelles.

Mots clefs : Stéréovision passive, mise en correspondance, 1D, carte des disparités, quasi-dense, confiance, transport, détection d'obstacles.

A 1D approach to correlation-based stereo matching

Application to obstacle detection on a road

In stereovision, a dense disparity map can be computed thanks to local methods that match 2D image neighbourhoods. In the context of 3D metrology, obtaining a dense disparity map is not mandatory, but high accuracy is required. Moreover, because of important local variations of disparity values due to high perspective conditions, we show that 2D methods do not reach a sufficient level of accuracy. In this work, we propose a new matching method based on the analysis of several characteristics extracted from a set of similarity curves computed using 1D neighbourhoods. We compute confidence values directly linked to accuracy and density properties of the disparity map. This confidence values are used to tune a trade-off between density and accuracy of the disparity maps. The generic proposed method is evaluated by comparing its results with those computed with six similar local methods. In this way, we show its efficiency, particularly with scenes having a strong perspective. Finally, we validate our 1D method in an obstacle detection application. We show that a quite basic segmentation process applied to our disparity maps yields better results than with disparity maps computed using 2D neighbourhoods. The behaviour of the 1D method with regard to untextured areas or uncalibration defaults is studied for synthetic and real images.

Keywords : Passive stereovision, matching process, 1D, disparity map, quasi-dense, confidence, transport, obstacle detection.