



HAL
open science

Détection d'obstacles multi-capteurs supervisée par stéréovision

Mathias Perrollaz

► **To cite this version:**

Mathias Perrollaz. Détection d'obstacles multi-capteurs supervisée par stéréovision. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université Pierre et Marie Curie - Paris VI, 2008. Français. NNT: . tel-00656864

HAL Id: tel-00656864

<https://theses.hal.science/tel-00656864>

Submitted on 5 Jan 2012

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

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

**THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PIERRE ET MARIE CURIE**

spécialité :
Sciences Mécaniques, Acoustique et Électronique

présentée par

M. Mathias PERROLLAZ

pour obtenir le grade de

Docteur de l'Université Pierre et Marie Curie

**DÉTECTION D'OBSTACLES MULTI-CAPTEURS SUPERVISÉE
PAR STÉRÉOVISION**

Soutenue publiquement le 13 Juin 2008

devant le jury composé de :

Président : M. Milgram Professeur (UNIVERSITÉ PIERRE ET MARIE CURIE, Paris)

Rapporteurs : A. Bensrhair Professeur (INSA, Rouen)
F. Nashashibi Chercheur - HDR (ÉCOLE DES MINES, Paris)

Examineurs : D. Aubert Directeur de thèse - Directeur de recherche (INRETS, Paris)
R. Labayrade Encadrant - Enseignant chercheur (ENTPE, Lyon)

Laboratoire sur les Interactions Véhicule - Infrastructure - Conducteur (INRETS - LCPC)

Mis en page avec la classe thloria.

Remerciements

Les travaux présentés dans ce mémoire ont été réalisés au Laboratoire sur les Interactions Véhicules-Infrastructure-Conducteurs (LIVIC, unité mixte de l'INRETS et du LCPC). Je tiens à remercier sincèrement Messieurs Jean-Marc Blossville et Jacques Ehrlich, Directeurs successifs du LIVIC, qui m'ont accueilli dans leur équipe pendant la durée de ma thèse.

Je remercie tout particulièrement Monsieur Didier Aubert, Directeur de recherche à l'INRETS et responsable de l'équipe perception du LIVIC, pour avoir été mon directeur de thèse, pour ses conseils et sa bonne humeur.

Je remercie également Monsieur Raphaël Labayrade, Enseignant chercheur à l'ENTPE, qui m'a encadré au cours de ma thèse et a su me communiquer son savoir et son enthousiasme.

C'est un honneur que m'ont fait Messieurs Abdelaziz Bensrhair, Professeur à l'INSA de Rouen, et Fawzi Nashashibi, Enseignant chercheur à l'École des Mines de Paris, d'avoir accepté de rapporter ce travail.

Ce fut en outre un plaisir pour moi que d'avoir Monsieur Maurice Milgram, Professeur à l'Université Paris 6, comme président de mon jury.

Je tiens à remercier également Monsieur Axel von Arnim, Ingénieur de recherche au LIVIC et initiateur du projet d'identification optique, avec qui j'ai partagé de bons moments, dans le travail comme en dehors.

Un grand merci à Monsieur Laurent Smadja qui m'a beaucoup appris sur la vision artificielle, sa théorie comme sa pratique, dans la joie et la bonne humeur.

Je tiens à remercier avec une attention toute particulière Messieurs Cyril Royère, Benoit Lugetti et Jean Marie Chevreau, qui m'ont appris les joies de la mise en place d'expériences et de démonstrations dans les véhicules.

De façon plus générale, j'adresse mes remerciements à toutes les personnes que j'ai été amené à rencontrer au LIVIC, qui rendent ce laboratoire agréable et de qui j'ai beaucoup appris : les secrétaires (Séverine Somma et Isnîé Rusani), les joueurs (Romain Gallen, Guillaume Saint Pierre, Aurelien Cord, Steeve Pechberty, Jean Sebastien Barreiro), les maîtres de la blague subtile (Benjamin Mourllion, Sebastien Glaser, Nicolas Hautière), ceux qui m'ont supporté comme voisin de bureau (Sio Song Ieng, Nicolas Soquet, Jeremy Vrignon, Antoine Fusee, Dominique Guyer), les thésards suivants (Alexandre, Nicoleta, Nabil, Yazid, Jamil, Hong-Tu) ou encore Thomas, Simon, Guilhem, Arnaud, George, Francis, Joëlle, Ali, Nicolas, Benoit, Mariana, Lydie, Jean-Christophe, Frank, Christian et Didier, les stagiaires successifs, et tous ceux que j'aurais pu oublier. . .

Enfin, je remercie sincèrement Sophie, mes parents, mon frère et de façon générale ma famille et mes amis, qui m'ont toujours soutenu.

Table des matières

Table des figures	ix
Liste des tableaux	xiii
Introduction Générale	1

Chapitre 1

Capteurs et méthodes pour la détection d'obstacles : état de l'art et positionnement

1.1	Capteurs et méthodes pour la détection d'obstacles	12
1.1.1	Les télémètres	12
1.1.2	La vision monoculaire	15
1.1.3	La stéréovision	17
1.1.4	Les autres capteurs de vision tridimensionnelle	20
1.1.5	Les systèmes coopératifs	20
1.1.6	Conclusion	23
1.2	Définition d'une méthodologie de détection	24
1.2.1	Les contraintes liées au contexte routier	24
1.2.2	La combinaison multi-capteurs	24
1.2.3	Méthodologie de combinaison proposée	25
1.2.4	Choix d'un référentiel commun aux différents capteurs	27
1.3	Conclusion	28

Chapitre 2

Construction de points de mesure par stéréovision

2.1	Présentation du capteur stéréoscopique	30
2.1.1	Modèle d'un capteur de vision	30
2.1.2	Modèle d'un capteur stéréoscopique	34
2.2	Construction d'une carte de disparité	37
2.2.1	Principe	37

2.2.2	Méthodes de calcul	37
2.2.3	Principales améliorations des méthodes surfaciques	40
2.3	Détection et exploitation des plans de la scène	42
2.3.1	Modélisation de la scène	42
2.3.2	L'approche u-v-disparité	43
2.3.3	La surface de la route	44
2.3.4	Les plans verticaux	46
2.4	Vue d'ensemble algorithmique	48
2.4.1	Récapitulatif des méthodes retenues	48
2.4.2	Récapitulatif des filtrages utilisés	50
2.4.3	Algorithme final de construction des points de mesure	51
2.4.4	Remarques qualitatives sur les résultats	52
2.4.5	Exploitation du résultat : les points de mesure "obstacles"	53
2.5	Conclusion	54

Chapitre 3

Création d'hypothèses de détection

3.1	Définition des zones d'intérêt	56
3.2	Création d'hypothèses de détection par stéréovision	58
3.2.1	Création des points de mesure obstacles	58
3.2.2	Segmentation de l'espace de disparité	58
3.2.3	Intérêt de la suppression des plans non fronto-parallèles	59
3.2.4	Influence de la discrétisation des histogrammes	60
3.2.5	Construction des zones d'intérêt	60
3.2.6	Respect de la contrainte temps-réel	61
3.2.7	Résultats	62
3.3	Création d'hypothèses de détection par télémétrie laser	64
3.3.1	Modèle d'un télémètre laser à balayage 2D ou lidar	64
3.3.2	Données fournies par le lidar	65
3.3.3	Filtrage des données brutes	66
3.3.4	Construction des cibles lidar	67
3.3.5	Suivi des détections lidar	68
3.3.6	Construction des zones d'intérêt	69
3.3.7	Résultats	69
3.4	Création d'hypothèses de détection par identification optique	70
3.4.1	Présentation du capteur	70
3.4.2	Algorithme de traitement du récepteur	72

3.4.3	Construction des zones d'intérêt	74
3.4.4	Résultats	77
3.5	Perspectives : Association des différents capteurs	78
3.5.1	Combinaison de deux capteurs	78
3.5.2	Combinaison de plus de deux capteurs	80
3.5.3	Zones d'intérêt résultantes	80
3.6	Conclusion	80

Chapitre 4

Validation et caractérisation des hypothèses de détection

4.1	Obtention des données nécessaires	84
4.1.1	Construction des points de mesure par stéréovision locale	84
4.1.2	Intérêt de l'approche locale	85
4.1.3	Portée, précision et temps de calcul : des contraintes antinomiques	85
4.1.4	L'approche subpixelique	86
4.1.5	Approche proposée : l'utilisation d'un zoom numérique local	87
4.2	Validation des hypothèses de détections par stéréovision	90
4.2.1	Principales sources d'erreurs	90
4.2.2	Critère <i>points obstacles</i>	91
4.2.3	Critère <i>alignement principal</i>	92
4.2.4	Critère <i>hauteur de la base de l'obstacle</i>	93
4.3	Mesure des caractéristiques statiques des détections	94
4.3.1	Mesure de la géométrie	94
4.3.2	Mesure de la position	94
4.3.3	Mesure de l'orientation	95
4.4	Estimation de caractéristiques dynamiques	96
4.4.1	Nécessité d'un suivi temporel	96
4.4.2	Principe de l'algorithme de suivi	96
4.4.3	Modèle d'évolution des objets détectés	97
4.4.4	Observation et association	97
4.4.5	Résultats	98
4.5	Résumé de l'Approche et Conclusion	99

Chapitre 5

Evaluation expérimentale et applications

5.1	Capteurs utilisés	102
5.1.1	Moyens expérimentaux du LIVIC	102

5.1.2	Capteurs stéréoscopiques	103
5.1.3	Télémètre laser	104
5.1.4	Capteur d'identification optique	104
5.1.5	Calibrage et configuration des différents capteurs	104
5.2	Evaluation des performances de détection en situation routière	104
5.2.1	Performances de détection par stéréovision	105
5.2.2	Performances de détection par télémétrie laser	112
5.2.3	Performances de détection par identification optique	114
5.3	Evaluation statique des performances	115
5.3.1	Evaluation comparée de la précision de création d'hypothèses	115
5.3.2	Evaluation de la caractérisation d'un piéton	118
5.3.3	Evaluation de l'intérêt du zoom numérique	120
5.4	Applications aux aides à la conduite	121
5.4.1	Détection d'obstacles par stéréovision multi-résolution	121
5.4.2	Détection par stéréovision et télémétrie laser : de la mitigation de collisions au "stop & go" et suivi automatique	123
5.4.3	Enrichissement de la détection par perception coopérative	126
5.5	Conclusion	127
Conclusion Générale		129
<hr/>		
Annexe A Des caméras au système stéréoscopique		133
A.1	Technologie du capteur vidéo	133
A.2	Les paramètres intrinsèques du capteur	134
A.3	La base stéréoscopique	135
Annexe B Influence de la taille de la fenêtre de corrélation		139
Annexe C Transformée de Hough pour la détection de plans dans l'espace de disparité		145
C.1	Transformée de Hough pour la détection de droites dans l'image	145
C.2	Adaptation à la détection d'un plan : transformée en niveaux de gris	146
Annexe D Calibrage et configuration des capteurs		149
Annexe E Dimensionnement du capteur d'identification optique		151

Annexe F Production scientifique au cours de la thèse	153
F.1 Publications	153
F.2 Brevet	153
F.3 Participation à des projets de recherche	154
F.4 Séminaires et Démonstrations	154
F.4.1 Rapport d'articles soumis à la publication	154
F.5 Vulgarisation scientifique	154
Bibliographie	155

Table des figures

1	<i>Nombre de morts par an sur les routes françaises.</i>	2
2	<i>Illustration du lien entre le temps à collision et le mode d'action.</i>	5
3	<i>Portée de détection nécessaire pour différentes applications de la détection d'obstacles frontaux.</i>	6
1.1	<i>Télémetre à ultrasons conçu pour l'aide au stationnement.</i>	13
1.2	<i>Radars conçus pour des applications automobiles.</i>	13
1.3	<i>Deux lidars utilisés pour la détection d'obstacle.</i>	14
1.4	<i>Deux caméras CMOS conçues pour une intégration dans des véhicules.</i>	15
1.5	<i>Deux types de systèmes stéréoscopiques.</i>	18
1.6	<i>Caméra 3D exploitant une mesure de temps de vol.</i>	20
1.7	<i>Problème d'indétermination lors de l'utilisation de télécommunications V2V.</i>	22
1.8	<i>Utilisation d'un système d'identification optique pour supprimer l'indétermination lors de l'utilisation de télécommunications V2V.</i>	23
1.9	<i>Domaines de fonctionnement des capteurs les plus fréquemment utilisés pour la détection d'obstacles.</i>	23
1.10	<i>Vue d'ensemble de la méthodologie de détection par fusion interdépendante : les différents capteurs émettent des hypothèses qui sont ensuite confirmées et caractérisées par le capteur-superviseur.</i>	26
1.11	<i>Les différents capteurs du système se réfèrent à un même système de coordonnées. Cela permet de les faire coopérer.</i>	27
1.12	<i>Mise en œuvre proposée de la stratégie de détection.</i>	28
2.1	<i>Modèle de caméra en trou d'épingle ou sténopé.</i>	31
2.2	<i>Positionnement d'une caméra dans le repère absolu.</i>	32
2.3	<i>Principe de la stéréovision et de la géométrie épipolaire.</i>	34
2.4	<i>En haut : paire d'image non rectifiée. Les lignes images ne sont pas correctement alignées. En bas : la même paire après rectification de la géométrie épipolaire.</i>	35
2.5	<i>Positionnement du capteur stéréoscopique dans un repère absolu.</i>	36
2.6	<i>Observation d'un détail de la route.</i>	42
2.7	<i>Paire stéréoscopique et images v-disparité et u-disparité correspondantes.</i>	44
2.8	<i>Mise en correspondance d'une fenêtre rectangulaire à droite avec une fenêtre inclinée à gauche pour s'adapter à la mise en correspondance des pixels de la route.</i>	45
2.9	<i>Fiabilisation de l'extraction du plan de la route dans l'image v-disparité.</i>	46
2.10	<i>Plans verticaux vus en u-disparité. Les piétons apparaissent comme fronto-parallèles, pas la barrière.</i>	47
2.11	<i>Filtrage des plans non fronto-parallèles. La barrière a bien été supprimée de l'image de disparité.</i>	48

2.12	<i>Trois plages de disparité correspondant à trois types d'hypothèses d'appariements : sous la route (bleu), route ou obstacle (jaune), obstacle uniquement.</i>	49
2.13	<i>Vue d'ensemble de l'algorithme de calcul des images de disparité obstacles et v-disparité route.</i>	51
2.14	<i>Paires d'images stéréoscopiques et cartes de disparité obstacles associées.</i>	52
2.15	<i>Représentation d'une carte de disparité comme un ensemble de points représentés dans R_{Δ} et dans R_a.</i>	53
3.1	<i>Définition des zones d'intérêt utilisées pour la suite.</i>	57
3.2	<i>Représentation d'une zone d'intérêt dans une paire d'images stéréoscopiques.</i>	57
3.3	<i>Algorithme de segmentation de l'espace de disparité : de l'image de disparité à la construction d'une classe obstacle.</i>	59
3.4	<i>a) La présence d'une barrière oblique forme un continuum de disparité très préjudiciable à la détection. b) Sa suppression facilite la segmentation.</i>	60
3.5	<i>Influence de la taille des bins sur le résultat de la segmentation.</i>	61
3.6	<i>Réduction d'un facteur 2 d'une paire d'images stéréoscopique par moyennage, et image de disparité obstacle résultante.</i>	62
3.7	<i>Hypothèses de détection fournies par la stéréovision.</i>	63
3.8	<i>Repères polaire et cartésien liés au télémètre laser.</i>	64
3.9	<i>Positionnement du télémètre dans un repère absolu.</i>	65
3.10	<i>Impacts d'une trame lidar représentés en vue de dessus dans le repère cartésien R_t (a) et projetés dans l'image de droite de la paire stéréoscopique (b).</i>	66
3.11	<i>Filtrage des impacts laser par l'utilisation d'un profil de route estimé par stéréovision.</i>	67
3.12	<i>Résultats de la détection par le lidar.</i>	68
3.13	<i>Construction d'une zone d'intérêt dans les images stéréoscopiques à partir d'une détection laser.</i>	69
3.14	<i>Zones d'intérêt générées par le lidar.</i>	70
3.15	<i>Prototype de lampe émettrice.</i>	70
3.16	<i>Protocole d'émission de la trame d'identification.</i>	71
3.17	<i>Image d'une scène routière acquise avec la caméra utilisée comme récepteur sans (a) et avec (b) le filtre passe bande 880 nm.</i>	71
3.18	<i>Algorithme de bas niveau de traitement des images infrarouge.</i>	72
3.19	<i>Suivi d'une lampe infrarouge.</i>	73
3.20	<i>Une piste correcte étiquetée 4 (cyan), et des pistes incorrectes (autres couleurs) justement invalidées par l'algorithme de décodage/filtrage.</i>	74
3.21	<i>Construction d'une zone d'intérêt dans les images stéréoscopiques à partir de l'identification et la localisation d'un émetteur infrarouge.</i>	74
3.22	<i>Identifiants localisés dans l'image de droite de la paire stéréoscopique en utilisant la hauteur connus de la lampe (a,c) et zones d'intérêt correspondantes (b,d).</i>	76
3.23	<i>Identifiants localisés dans l'image de droite de la paire stéréoscopique en utilisant l'identification en stéréovision (a,c) et zones d'intérêt correspondantes (b,d).</i>	77
3.24	<i>Résultats des zones détectées par les trois capteurs embarqués.</i>	78
3.25	<i>Les zones d'intérêt de la figure 3.24 ont été combinées (en jaune).</i>	80
3.26	<i>Détection multi-capteurs lors du suivi d'un véhicule.</i>	81
4.1	<i>Scène comportant un candidat pour la détection (a), et données 3D associées représentées sous la forme d'images de disparité (d), u-disparité (b), et v-disparité (c).</i>	85

4.2	<i>Principe du zoom numérique variable.</i>	88
4.3	<i>a) Dégradation d'un détail d'image stéréoscopique par l'application de deux homographies successives (une rectification et un zoom de facteur 2), et (b) comparaison au cas où ces homographies ont été combinées en une seule.</i>	89
4.4	<i>Exemples de fausses détections.</i>	90
4.5	<i>Valeur du seuil de confirmation en pixels en fonction de la surface exposée minimum et de la disparité des pixels (pour $b_s = 1$ m) dans le cas d'un obstacle fronto-parallèle.</i>	92
4.6	<i>Détection représentée par sa zone d'intérêt (a) et son volume réel fronto-parallèle (b).</i>	95
4.7	<i>Définition du point de contrôle utilisé pour la mesure de position.</i>	95
4.8	<i>Mesure de l'orientation d'une cible d'après ses points obstacles.</i>	96
4.9	<i>Principe de l'algorithme de suivi appliqué aux données stéréoscopiques.</i>	97
4.10	<i>Limites du point de contrôle utilisé pour modéliser la position d'un objet déformable.</i>	98
4.11	<i>Abscisse mesurée pour la position du piéton de la figure 4.10 qui traverse à vitesse constante. La déformation due à l'écartement et au resserrement des jambes laisse entrevoir des alternances de position fixe et de sauts, pouvant mesurer jusqu'à 20 cm.</i>	99
4.12	<i>Vue d'ensemble de l'algorithme de confirmation des détections.</i>	100
5.1	<i>Trois des véhicules expérimentaux du LIVIC. De gauche à droite : VIPER, CARLLA, LOLA.</i>	103
5.2	<i>Résultats obtenus par stéréovision pour la séquence "autoroute".</i>	109
5.3	<i>Résultats obtenus par stéréovision pour la séquence "ville".</i>	110
5.4	<i>Résultats obtenus par stéréovision pour la séquence "mixte".</i>	111
5.5	<i>Illustration des deux séquences de test utilisées pour la détection lidar.</i>	112
5.6	<i>Intérêt du suivi lors de la création d'hypothèses par le lidar.</i>	113
5.7	<i>Illustration de la séquence de test utilisée pour la détection par identification optique.</i>	114
5.8	<i>Lorsque la lampe émettrice se reflète sur une surface réfléchissante, ici la carrosserie d'un autre véhicule, le reflet peut être vu comme une autre source transmettant le même code.</i>	115
5.9	<i>Comparaison des mesures de distance obtenues par chacun des trois capteurs avec une référence donnée par un topomètre.</i>	116
5.10	<i>Positionnement latéral mesuré par les différents capteurs lors de l'éloignement d'un véhicule.</i>	117
5.11	<i>Parcours d'un piéton qui s'arrête sur différents points du site de calibrage.</i>	119
5.12	<i>Caractérisation d'un piéton se déplaçant sur des points connus du banc de calibrage.</i>	120
5.13	<i>Zone de filtrage latérale utilisée dans le projet LOVE.</i>	122
5.14	<i>Diagramme RTMaps de la partie perception de l'application de "stop & go" et mitigation de collisions.</i>	125
5.15	<i>Détection enrichie par communications V2V.</i>	127
5.16	<i>Panneau de signalisation équipé d'un émetteur infrarouge, et imagerie affichée sur l'interface conducteur lorsque ce panneau est identifié.</i>	128
A.1	<i>Les deux principales technologies de caméras.</i>	134

A.2	<i>a) Une image de droite d'une paire stéréoscopique acquise avec des caméras CMOS. c) La carte de disparité résultante. Le bruit caractéristique des images CMOS est très visible sur l'image originale et perturbe l'étape d'appariements. b) et d) : le même couple d'images lorsqu'un filtre gaussien a été appliqué préalablement. . .</i>	135
A.3	<i>Influence de la base stéréoscopique sur la représentation v-disparité : a) base de 65 cm b) base de 103 cm.</i>	137
B.1	<i>Paire d'image utilisée pour illustrer l'influence de la taille de la fenêtre de corrélation sur la carte de disparité.</i>	140
B.2	<i>Influence des dimensions de la fenêtre de corrélation sur une image de disparité. .</i>	141
B.3	<i>Carte de disparité obtenue avec une fenêtre de corrélation inclinable de taille 15*15 pixels.</i>	141
B.4	<i>Images issues de la première version du simulateur de capteurs SiVIC, utilisées pour comparer les performances de différentes tailles de fenêtres de corrélation. .</i>	142
B.5	<i>Evaluation du taux de bons appariements (en %) sur une scène synthétique en fonction de la taille de la fenêtre : à gauche avec une fenêtre classique, à droite avec une fenêtre inclinable.</i>	143
B.6	<i>Temps de calcul d'une image de disparité (en ms) en fonction de la taille de la fenêtre de corrélation utilisée.</i>	143
C.1	<i>Paramétrisation d'une droite dans l'espace de Hough.</i>	145
C.2	<i>Les deux approches de la transformée de Hough appliquées à une image v-disparité bruitée (a) pour trouver le profil global de la route.</i>	147
C.3	<i>Limites de la transformée de Hough en niveaux de gris.</i>	148
D.1	<i>Banc de calibrage du site de Satory utilisé pour réaliser la configuration des différents capteurs.</i>	149
E.1	<i>Image d'une scène routière acquise avec la caméra utilisée comme récepteur sans (a) et avec le filtre passe bande 880 nm (b). La même scène acquise en résolution réduite à une fréquence de 514 images par seconde (c et d).</i>	152

Liste des tableaux

5.1	<i>Caractéristiques des différentes caméras utilisées dans cette thèse.</i>	103
5.2	<i>Configuration du capteur stéréoscopique pour les différents véhicules ayant servi à l'évaluation.</i>	104
5.3	<i>Performances de la détection par stéréovision pour la séquence "autoroute". Le taux de détection est meilleur sur les voies latérales car il correspond à des dépassements de véhicules, et donc à une portée de détection plus faible.</i>	107
5.4	<i>Performances de la détection par stéréovision pour la séquence "ville".</i>	108
5.5	<i>Performances de la détection par stéréovision pour la séquence "mixte".</i>	108
5.6	<i>Nombre de fausses détections pour la séquence "parking", avec le seul lidar, et après validation par stéréovision.</i>	113
5.7	<i>Fausse détections lors d'un scénario de roulage sur autoroute.</i>	114
5.8	<i>Portée d'identification.</i>	116
5.9	<i>Influence du zoom numérique sur la portée de confirmation du lidar.</i>	121
5.10	<i>Temps de calculs obtenus par l'application de détection d'obstacles utilisant 3 résolutions.</i>	123
A.1	<i>Caractéristiques des différentes caméras utilisées dans cette thèse.</i>	136

Introduction Générale

Au cours du vingtième siècle, la voiture est progressivement devenue un moyen de transport indispensable au quotidien. Sa généralisation a été telle que le parc automobile français compte désormais plus de 26 millions de véhicules. Mais celle-ci a entraîné un véritable problème de santé publique. En effet, parallèlement à l'accroissement du nombre de véhicules en circulation, le nombre de décès sur les routes de France a augmenté jusqu'à atteindre un maximum de 17000 morts en un an, en 1973 (figure 1). Cette année là, les limitations de vitesse et l'obligation du port de la ceinture de sécurité ont alors été introduites, ce qui a permis d'inverser la tendance et d'initier une baisse de la mortalité sur les routes.

Ces deux mesures illustrent les principaux outils de lutte contre les dangers de la route. En effet, si les différentes évolutions de la législation relative au code de la route prennent une part importante dans la baisse de cette mortalité, une autre partie de l'explication peut être trouvée dans la sécurisation croissante des véhicules mis sur le marché. Les constructeurs automobiles suivent cette voie depuis longtemps, en concevant des systèmes de sécurité passive. Ces derniers permettent de sécuriser les passagers d'un véhicule ou les autres usagers de la route, sans pour autant agir sur la conduite. Les exemples les plus emblématiques de tels systèmes sont la ceinture de sécurité ou les airbags¹. Néanmoins, de nombreux travaux ont également été entrepris pour optimiser les déformations de l'habitacle et les pare-chocs, afin de diminuer la gravité d'une éventuelle collision.

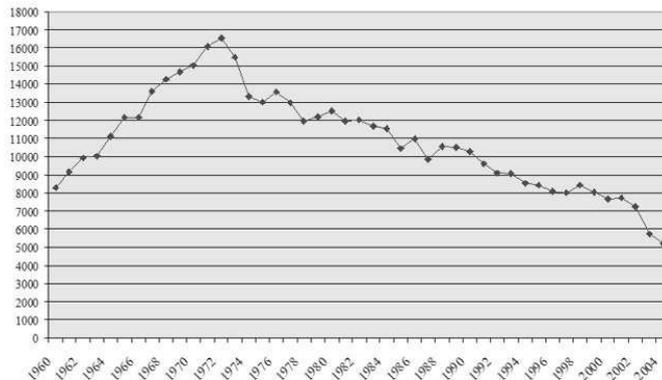


FIG. 1 – *Nombre de morts par an sur les routes françaises.*

Désormais, ces gisements de progrès étant partiellement épuisés, les nouvelles avancées en matière de sécurité doivent être recherchées parmi les systèmes de sécurité active. Ceux-ci interviennent directement sur la conduite, pour aider le conducteur ou pour diminuer la gravité d'un éventuel accident. Dans ce domaine, l'aide au freinage ABS² ou le programme électronique de stabilité ESP³ sont de plus en plus répandus et ont démontré leur efficacité en matière de sécurité. Néanmoins, de tels systèmes agissent comme une réponse immédiate à un problème survenu lors de la conduite mais ne permettent pas d'anticiper les situations à risque. La réponse à cette faiblesse peut alors être trouvée dans l'émergence d'ADAS⁴, systèmes d'aide à la conduite dits "intelligents", capables d'interagir avec le conducteur sans nécessairement s'y substituer. De tels dispositifs apportent potentiellement aux usagers de la route un gain important en confort et en

¹Coussins gonflables de sécurité

²AntiBlockierSystem

³Electronic Stability Program

⁴Advanced Driver Assistance Systems

sécurité. Dans un cas idéal, certaines études jugent même que 50% des accidents pourraient être évités grâce à eux [col05].

Pour leur fonctionnement, les ADAS exploitent un système de perception permettant d'accéder à des informations sur l'environnement du véhicule. En conséquence, de nombreuses recherches portent actuellement sur des aspects perceptifs comme la détection automatique de la signalisation horizontale ou verticale, l'exploitation de données cartographiques, la détection d'obstacles, l'estimation de la visibilité... La majorité des accidents corporels étant dus à une collision⁵, la détection d'obstacles s'avère particulièrement prometteuse pour ses applications, car elle permet d'anticiper ces collisions. Nous allons maintenant présenter certaines de ces applications, mais auparavant, attachons nous à préciser la notion de détection d'obstacles.

La détection d'obstacles

Définition

Au sens large, la détection d'obstacles est une tâche perceptive, dont le but est la détection et la caractérisation des objets présents dans une scène, pour déterminer les trajectoires possibles. Tous les objets de la scène n'étant pas des obstacles, comment alors peut-on définir un obstacle ? Le dictionnaire de l'académie française donne la définition suivante :

"Ce qui arrête ou ralentit le passage, ce qui empêche d'avancer."

Dans le contexte automobile auquel nous nous restreignons dans cette thèse, la notion d'obstacle est donc liée à l'évolution même du véhicule. Un objet devient obstacle lorsque, à un instant donné, l'intersection de sa trajectoire avec celle du véhicule est non nulle. Cependant, au delà d'une certaine durée, la probabilité de cette intersection n'est plus estimable et perd son sens. Compte tenu du temps de réaction d'un conducteur et des vitesses mise en jeu sur les routes, la durée maximale au delà de laquelle on peut considérer que l'évolution du système véhicule-infrastructure-conducteur n'est plus prédictible sur la base des observations passées et présentes est de l'ordre de quelques secondes. Nous proposons donc d'affiner notre notion d'obstacle en le définissant comme un :

Objet de taille significative dont la trajectoire coupe l'une des trajectoires possibles de notre véhicule dans les secondes à venir.

Un objet présent sur la route ne peut donc être considéré comme un obstacle que s'il est effectivement susceptible d'entraver la trajectoire du véhicule, c'est-à-dire si la collision du véhicule contre celui-ci entraîne des dégâts physiques. De fait, des dimensions minimums sont nécessaires pour qualifier un objet d'obstacle. Le matériau constituant l'objet est également important, mais sa caractérisation par un système de détection peut s'avérer particulièrement complexe, et se situe hors du champ de cette thèse.

La trajectoire future estimée du véhicule permet de définir des limites latérales de la zone de risque. Dans un contexte autoroutier, la voie de circulation constitue intrinsèquement un encadrement de la trajectoire du véhicule. Dans un cadre plus complexe, en milieu urbain

⁵<http://www.securiteroutiere.equipement.gouv.fr/infos-ref/observatoire/accidentologie/>

par exemple, il est préférable de revenir à la notion cinématique de trajectoire, en prédisant la position future du véhicule en fonction de sa trajectoire passée et de sa dynamique [LRA05].

L’horizon temporel est également une notion essentielle pour la classification d’un objet comme obstacle. L’indicateur le plus pertinent pour cela est le temps à collision (TTC⁶) qui est très largement utilisé pour la détection d’obstacles routiers. Il s’agit, en combinant les estimations de distance et de vitesse relatives d’un obstacle par rapport à notre véhicule, d’estimer la durée restant avant collision, si aucun événement ne perturbe le système entre temps. Cette notion de temps à collision présente l’intérêt d’être facilement comparable au temps de réaction du conducteur ou des systèmes d’aide. Le TTC intègre donc une notion de risque en qualifiant l’urgence d’une situation.

Au delà de cette définition assez générique, il est notable que le positionnement dans un contexte routier apporte une spécificité au problème de la détection d’obstacles. D’une part, les scènes dans lesquelles évolue notre véhicule peuvent généralement être considérées comme très structurées. En effet, la signalisation horizontale comme verticale, les bords de routes ou les connaissances sur la forme des piétons et des autres véhicules sont autant d’éléments qui participent à la simplification du problème en fournissant des *a priori* forts. En contrepartie, la présence d’humains dans le système Véhicule-Infrastructure-Conducteur et les conséquences potentiellement importantes de défaillances des dispositifs de détection imposent une fiabilité exemplaire de ceux-ci. Cette fiabilité est d’ailleurs une composante essentielle de l’acceptabilité de tels dispositifs par le grand public. En particulier, les faux positifs (détection d’un obstacle alors que celui-ci n’existe pas) sont difficilement acceptables, et il est généralement admis qu’un système de détection d’obstacles est suffisamment fiable s’il produit moins d’une fausse alarme pour un million de kilomètres parcourus.

Applications

Les applications de la détection d’obstacles au domaine automobile sont nombreuses. L’application la plus intrusive est l’automatisation complète de la conduite, idée qui a donné naissance à des scénarii de route automatisée. Cependant, dès lors que le véhicule équipé est plongé dans un réseau routier dans lequel sont impliqués des conducteurs humains, les aléas du comportement de ceux-ci sont quasiment antinomiques avec une automatisation complète. Pour cette raison, les applications visent plutôt des fonctions d’aide à la conduite susceptibles d’améliorer activement la sécurité ou le confort des usagers de la route, sans pour autant décharger le conducteur de sa fonction et de sa responsabilité [GTAN02]. Les recherches s’orientent donc vers des modes de coopération d’autant plus intrusifs que le délai d’intervention est court. Cela est illustré par le positionnement de différents modes d’action sur la ligne temporelle de la figure 2.

Deux secondes d’écart au véhicule précédent correspondant à la recommandation du code de la route français, il est justifié de prévenir le conducteur lorsque celui-ci se rapproche d’un objet en dessous de cette limite. Si malgré tout le rapprochement continue sans que le conducteur n’ait réagi, il est envisageable que le système intervienne pour corriger cette déficience. Il va alors faire ralentir le véhicule pour accompagner une éventuelle manœuvre de freinage ou d’évitement, ou véritablement déclencher un freinage d’urgence, moins d’une seconde avant l’impact. Ce type de

⁶Time To Collision



FIG. 2 – Illustration du lien entre le temps à collision et le mode d'action.

Source : PREVENT-APALACT⁷.

manœuvre est appelé mitigation de collision car, à défaut d'empêcher un choc, elle permet d'en réduire considérablement la gravité.

Les dernières millisecondes précédant l'impact peuvent être mises à profit pour préparer les différents systèmes de sécurité passive disponibles, comme des pré-tensionneurs de ceinture de sécurité ou des airbags. L'extérieur du véhicule peut être également équipé de dispositifs de sécurité active, pour protéger les usagers vulnérables de la route (piétons, cyclistes, ...) lors d'un accident, en utilisant par exemple un airbag externe pour piétons ou un capot actif pouvant amortir le choc. Ces applications intervenant dans les derniers instants sont regroupées sous l'appellation pré-crash.

La sécurité des usagers de la route peut également tirer bénéfice d'une détection d'obstacles par un véhicule à l'arrêt. C'est le cas pour les poids lourds, depuis lesquels le chauffeur ne peut pas voir les obstacles situés dans les premiers mètres devant la calandre, en raison d'un angle mort important. Un système de détection placé dans cet angle mort et avertissant le conducteur ou empêchant le camion de démarrer si un piéton traverse représente donc une avancée importante en terme de sécurité [BBM⁺06].

Mises à part ces applications sécuritaires de la détection d'obstacles, d'autres visent à améliorer le confort de conduite. C'est le cas de l'ACC⁸, qui permet d'alléger la tâche du conducteur en assurant la régulation d'une distance de sécurité sur voie rapide, par une action sur les commandes d'accélération et de décélération [Jau02]. En cas de défaillance du système, le conducteur dispose encore du temps nécessaire pour réagir et éviter le danger.

Dans le même esprit, le "stop & go" assure un suivi automatique du véhicule précédent, mais à plus basse vitesse. Celui-ci est particulièrement adapté à la conduite dans des situations de congestion du trafic et se conçoit comme une véritable application de confort.

Au vue de toutes ces applications, embarquer un système de détection d'obstacles dans un véhicule présente un intérêt réel. Cependant, si l'on tient compte du contexte dans lequel se place chacune de ces applications (ville, autoroute, route nationale) et des vitesses associées, il apparaît que les portées de détection nécessaires sont très variées, comme cela est représenté sur la figure 3. Nous verrons au chapitre 1 que cela induit l'usage de plusieurs types de capteurs. Un enjeu de recherche sur la détection d'obstacles, pour donner une véritable attractivité à son déploiement à grande échelle dans des véhicules de série, consiste donc à maximiser le nombre d'applications possibles, et par conséquent la plage de fonctionnement, tout en conservant un nombre de capteurs minimal.

⁸Automatic Cruise Control

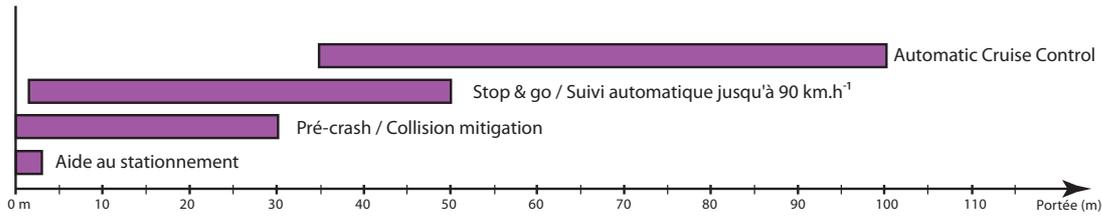


FIG. 3 – Portée de détection nécessaire pour différentes applications de la détection d’obstacles frontaux.

Objectifs et contexte de cette thèse

L’ambition de cette thèse est la proposition puis la mise en œuvre d’une méthodologie originale et efficace de détection d’obstacles exploitant plusieurs capteurs. Celle-ci doit respecter trois contraintes fondamentales : une grande robustesse de détection, une plage de portées de fonctionnement élevée et une approche réaliste en terme de ressources de calcul nécessaires.

La nécessité d’une plage de portées élevée a orienté le choix des capteurs vers des systèmes différents et complémentaires dans leur fonctionnement, comme dans leurs portées respectives. Pour cela, l’étude de l’état de l’art proposée au chapitre 1 et la disponibilité technologique ont orienté notre choix vers : un télémètre laser à balayage, rapide et précis, un capteur stéréoscopique, fournissant une grande quantité d’informations et de possibilités de traitement, et un système d’identification optique de véhicules, permettant une détection à grande portée et une collaboration avec des télécommunications inter-véhiculaires.

Cette thèse a donné lieu à une participation et des développements dans plusieurs projets de recherche.

Le projet D.O.30⁹ vise à concevoir un système de détection d’obstacles par stéréovision, destiné à des applications de pré-crash et de mitigation de collisions. Le LIVIC est en charge de la réalisation de l’algorithme de détection, qui doit répondre à des spécifications strictes en termes de précision et de fiabilité. Cet algorithme sera ensuite évalué par les constructeurs automobiles français Renault et PSA, puis porté par le CEA¹⁰ sur une architecture matérielle dédiée.

Le projet LOVE¹¹ rassemble 12 partenaires et traite quant à lui de la détection et de la reconnaissance d’usagers vulnérables de la route, par une approche multi-capteurs. La contribution de cette thèse au projet LOVE concerne la conception d’un algorithme de détection d’obstacles par stéréovision.

Le projet européen TRACKSS¹² enfin, traite de capteurs coopératifs. Il vise d’une part à concevoir de tels capteurs, et d’autre part à définir un modèle d’échange de données pour la coopération d’un grand nombre de capteurs, que ceux-ci soient situés dans un ou plusieurs véhicules, ou encore sur l’infrastructure routière. Dans le cadre de ce projet, le LIVIC propose un capteur d’identification optique et sa coopération avec les capteurs proposés par deux autres partenaires, Bosch et l’université de Newcastle.

La participation à ces trois projets a contribué à orienter cette thèse vers des méthodes

⁹Détection d’Obstacles à 30 mètres

¹⁰Commisariat à l’Energie Atomique

¹¹Logiciel d’Observation de Vulnérables

¹²Technologies for Road Advanced Cooperation Knowledge Sharing Sensors

prioritairement efficaces et réalistes. De fait, l'effort a particulièrement été porté sur la proposition et l'utilisation de techniques génériques, peu coûteuses en temps de calcul et robuste au bruit.

Contributions proposées dans cette thèse

Les contributions proposées dans cette thèse se situent à plusieurs niveaux. Elles concernent autant des aspects théoriques et techniques liés à la définition d'une méthodologie de détection, que des aspects expérimentaux et pragmatiques relatifs à sa mise en œuvre.

Tout d'abord, un apport méthodologique est proposé sous la forme d'une stratégie de coopération entre plusieurs capteurs. Celle-ci suppose la création d'hypothèses de détection par différents capteurs et leur confirmation *a posteriori* par un capteur qui joue un rôle de superviseur, en l'occurrence la stéréovision.

Nous avons proposé pour la création d'hypothèses de détection d'exploiter trois capteurs complémentaires : un capteur stéréoscopique, un télémètre laser à balayage et un capteur d'identification optique. Nous proposons donc pour chacun d'entre eux une méthode permettant de l'exploiter dans ce cadre. Les résultats fournis par ces méthodes sont alors des volumes d'intérêt, localisés dans l'espace de disparité lié aux images stéréoscopiques.

Compte tenu des choix qui ont été faits, la stéréovision joue un rôle central dans la méthodologie proposée, puisqu'elle intervient à deux niveaux distincts : la création d'hypothèses de détection et la validation de celles-ci. De nombreux efforts ont donc été portés sur son utilisation. En premier lieu, nous nous sommes intéressés à la construction de points de mesure dans l'espace de disparité à partir des images stéréoscopiques. Pour cela, nous nous sommes attachés à choisir, parmi le grand nombre de méthodes existantes, celles qui permettent d'obtenir des données bien adaptées au contexte routier structuré. Par ailleurs, nous avons montré que l'exploitation des caractéristiques des plans les plus représentés dans les scènes routières, à savoir les plans verticaux et horizontaux, permet d'obtenir des données de meilleure qualité, et ainsi de fiabiliser l'estimation de la surface de la route et la détection des obstacles.

C'est par la segmentation de l'espace de disparité contenant les points de mesure qu'est réalisée la création d'hypothèses de détection par stéréovision. Nous proposons en ce sens un algorithme de segmentation rapide, exploitant efficacement l'approche u-v-disparité proposée par Labayrade *et al.* [Lab03]. Cet algorithme débute par une observation de la scène dans sa globalité puis resserre son observation autour des différents objets détectés. L'utilisation d'histogrammes de cumul le rend robuste au bruit causé par des erreurs d'appariements, si celles-ci sont décorrélées.

L'étude des données stéréoscopiques de façon plus locale, c'est à dire limitée à des zones d'intérêt correspondant chacune potentiellement à un objet unique, présente également un intérêt fort. Cela permet de traiter des images de résolution importante pour un temps de calcul raisonnable, et surtout d'observer certaines caractéristiques des détections, comme l'orientation, qui ne seraient pas mesurables en considérant l'espace dans sa totalité. Nous proposons donc d'exploiter cette approche de la stéréovision pour valider et caractériser des hypothèses de détection issues de la stéréovision ou d'autres capteurs. Pour cela, nous proposons notamment trois critères de validation des hypothèses de détection.

Nous verrons également que le traitement par zones d'intérêt permet d'adapter la précision voulue à chacune des détections, pour optimiser l'inévitable compromis entre le temps de calcul et la précision. Nous proposons donc de réaliser cela par un ré-échantillonnage local des images stéréoscopiques.

Enfin, la composante expérimentale de cette thèse regroupe une évaluation des méthodes proposées et sa mise en œuvre fonctionnelle, à travers des applications implantées et testées dans les véhicules expérimentaux du LIVIC.

Organisation du mémoire

Ce rapport s'articule autour de 5 chapitres.

Le chapitre 1 pose le cadre général de travail dans lequel nous avons décidé de nous placer. Il propose dans un premier temps un état de l'art des capteurs les plus couramment utilisés pour la détection d'obstacles et des méthodes proposées par la communauté scientifique pour les utiliser. Fort de la connaissance de ce contexte, nous décrivons ensuite la méthodologie de détection originale que nous avons envisagée et mise en œuvre.

Le chapitre 2 est consacré à la construction de points de mesure dans l'espace de disparité lié aux images stéréoscopiques. Ces points de mesure sont obtenus à partir d'une carte de disparité éparsée de la scène. Nous allons donc présenter un rapide état de l'art des méthodes de construction de cartes de disparité, puis la méthode que nous avons retenue. Enfin, nous proposerons également d'exploiter les plans détectés dans la scène par l'approche u-v-disparité pour améliorer le résultat obtenu.

Le chapitre 3 aborde le principe de création d'hypothèses de détection, sous la forme de volumes d'intérêt dans l'espace de disparité lié aux images stéréoscopiques. Ce procédé est présenté pour les trois méthodes de détection que nous avons retenues : la stéréovision, la télémétrie laser et l'identification optique.

Le chapitre 4 est l'occasion de proposer l'utilisation de la stéréovision de façon locale dans les zones d'intérêt correspondant aux hypothèses de détection. En particulier, nous proposons trois critères de validation de celles-ci, puis leur caractérisation par des mesures de position, de géométrie et de vitesse.

Le chapitre 5 fournit une évaluation des performances des différents éléments du système de détection proposé. Celle-ci est réalisée à travers des séquences statiques, acquises sur le banc de calibrage du site de Satory, et des séquences issues de situations réelles de conduite, dans des contextes variés.

Après cette évaluation, nous proposons trois applications de notre stratégie de détections et leur implantation sur des véhicules instrumentés : une architecture de détection d'obstacles par stéréovision exploitant trois résolutions pour optimiser le temps de calcul nécessaire, un système de mitigation de collisions et de "stop & go" fondé sur la coopération entre stéréovision et télémétrie laser, et un système de perception enrichie exploitant les télécommunications.

Enfin, nous concluons ce rapport sur les perspectives que présente la méthode proposée, en

termes d'évolutions possibles d'une part et d'applications encore envisageables d'autre part.

CAPTEURS ET MÉTHODES POUR LA DÉTECTION D'OBSTACLES : ÉTAT DE L'ART ET POSITIONNEMENT

Sommaire

1.1	Capteurs et méthodes pour la détection d'obstacles	12
1.1.1	Les télémètres	12
1.1.2	La vision monoculaire	15
1.1.3	La stéréovision	17
1.1.4	Les autres capteurs de vision tridimensionnelle	20
1.1.5	Les systèmes coopératifs	20
1.1.6	Conclusion	23
1.2	Définition d'une méthodologie de détection	24
1.2.1	Les contraintes liées au contexte routier	24
1.2.2	La combinaison multi-capteurs	24
1.2.3	Méthodologie de combinaison proposée	25
1.2.4	Choix d'un référentiel commun aux différents capteurs	27
1.3	Conclusion	28

UN système perceptif est constitué de capteurs pour récupérer des informations brutes sur l'environnement dans lequel il est placé et d'une ou plusieurs unités de traitement dont la tâche est l'analyse de ces informations. Cette structure se retrouve aussi bien dans les systèmes naturels qu'artificiels. La détection d'obstacles embarquée dans un véhicule ne déroge donc pas à cette règle.

Les capteurs mis en œuvre à cette fin sont variés et peuvent être actifs (radar, laser,...) ou passifs (vision artificielle). L'analyse, quant à elle, est réalisée par des algorithmes de traitement, exécutés par un matériel informatique. Nous allons au cours de ce chapitre présenter les différents capteurs, ainsi que les algorithmes de traitements subséquents qui ont été proposés dans la littérature scientifique.

Nous aborderons dans un premier temps les télémètres, capteurs actifs capables de mesurer des distances.

Nous nous intéresserons ensuite à la vision artificielle, qui suscite de grands efforts de recherche pour la détection d'obstacles routiers. De nombreux algorithmes ont été proposés en ce sens, dont nous allons présenter un aperçu.

Par ailleurs, lorsque plusieurs caméras sont utilisées dans un même système de vision, il est possible de mettre en œuvre des algorithmes de stéréovision, qui fournissent une information tri-dimensionnelle. Leur intérêt étant largement reconnu par la communauté de recherche sur les véhicules intelligents, nous y consacreront également une partie de ce chapitre, d'autant que nous choisirons d'utiliser cet outil par la suite.

La détection coopérative est une piste qui commence à être explorée pour rendre la perception plus performante. Nous présenterons donc ce concept, ainsi que les premiers essais qui ont été faits en ce sens.

Forts de la connaissance de ces différentes techniques, nous nous intéresserons aux approches multi-capteurs, qui permettent de tendre vers une fiabilité suffisante au développement de systèmes actifs d'aide à la conduite fortement intrusifs, puis nous clorons ce chapitre en proposant la méthodologie de détection que nous avons développée et mise en œuvre au cours de cette thèse.

1.1 Capteurs et méthodes pour la détection d'obstacles

1.1.1 Les télémètres

La première catégorie de capteurs utilisés pour la détection d'obstacles regroupe les télémètres. Ceux-ci sont des capteurs actifs émettant une onde qui se réfléchit à la surface des objets situés dans la direction de propagation. La comparaison entre l'onde incidente avec son écho permet d'estimer un temps de vol, un déphasage ou un éventuel glissement de fréquence, qui fournissent ensuite la distance séparant le capteur de la surface formant l'écho. Trois types de télémètres sont généralement utilisés pour des applications automobiles, la différence fondamentale résidant dans la nature de l'onde émise et les traitements associés.

Le télémètre à ultrasons

Il est aisé et très peu coûteux de construire des télémètres exploitant des ondes acoustiques. La fréquence de celles-ci ne pouvant pas être située dans le spectre audible, ce sont des ultrasons qui sont utilisés. Ceux-ci présentent également les avantages de pouvoir être émis par une source de petite taille et d'être réfléchis par des objets à taille humaine.

La portée de ces capteurs est très faible (2 m) et leur champ d'application est donc restreint, se limitant principalement à l'aide au stationnement.



FIG. 1.1 – *Télémètre à ultrasons conçu pour l'aide au stationnement.*

Le RADAR

La découverte à la fin du XIX^{ème} siècle par H. Hertz du phénomène de réflexion des ondes électromagnétiques sur un matériau conducteur a donné naissance quelques années plus tard au radar¹³. Le principe est l'émission d'une onde électromagnétique et la réception de l'écho formé par cette onde sur une surface conductrice pour estimer la distance à cette surface. Par ailleurs, si l'objet réfléchissant est animé d'une certaine vitesse relativement au radar, l'onde réfléchie aura une fréquence sensiblement différente de celle de l'onde incidente (en supposant que celle-ci soit monochromatique). Cet effet, dit Doppler, permet d'obtenir immédiatement une mesure de vitesse relative entre le radar et l'objet réfléchissant, en estimant le glissement de fréquence.



FIG. 1.2 – *Radar conçu pour des applications automobiles.*

Une portée et une précision importantes, couplées à un fonctionnement assuré dans la plupart des conditions météo (brouillard, pluie et neige) sont les avantages indéniables du radar dans le contexte routier. Différents équipementiers automobiles se sont donc déjà penchés sur cette technologie et commercialisent des modèles de radar à grande échelle.

Cependant, le radar présente un certain nombre de limitations. Tout d'abord, l'angle de vue du capteur est faible (environ 10 degrés pour un radar longue portée) et ne permet donc pas de percevoir des objets situés autour de ce cône d'ouverture. La résolution angulaire étant par ailleurs très limitée, elle ne permet pas nécessairement de distinguer deux objets situés côte à côte.

D'autre part, l'étude de données radar laisse apparaître de nombreux échos parasites dans la chaussée ou sur les bords de route. Pour éviter que cela ne provoque un grand nombre de fausses

¹³RAdio Detection And Ranging

détections, les objets immobiles sont généralement filtrés en comparant la mesure de leur vitesse à la vitesse du véhicule équipé. Un véhicule à l'arrêt au milieu de la voie ne sera donc pas détecté. Enfin, seules quelques longueurs d'onde permettent un fonctionnement relativement peu sensible aux perturbations atmosphériques. Les technologies existantes exploitent donc une bande de fréquence centrée sur 77 GHz pour des radars longue portée (150 m) à faible angle de vue (10 degrés) et sur 24 GHz pour des radars à plus courte portée (30m) mais bénéficiant d'un grand angle de vue (70 degrés)¹⁴. Si ces technologies sont efficaces pour la détection de véhicules partiellement métalliques, elles sont peu efficaces pour détecter des piétons, ce type de cible créant peu d'écho.

Les radars sont généralement envisagés pour des applications de type ACC sur autoroute en raison de leur portée [Jau02]. L'utilisation d'un couple de radars courte portée dans des systèmes de mitigation de collision, "stop & go" ou pré-crash, est également envisagée [SMW03]. Mais le filtrage des objets immobiles, actuellement utilisé, tend à les exclure de ce type d'applications.

Le LIDAR

Fondé sur l'utilisation d'un laser, le lidar¹⁵ (ou télémètre laser) exploite la réflexion d'un rayon lumineux pour mesurer une distance. Le principe général de fonctionnement de ce capteur est décrit dans [Han99]. Il exploite généralement le temps de vol de l'onde pour estimer une distance. Il est possible d'utiliser par ailleurs une mesure de réflectance pour caractériser les surfaces de réflexion et ainsi distinguer des objets dont la surface est différente.



FIG. 1.3 – Deux lidars utilisés pour la détection d'obstacle.

a) Un télémètre à balayage 2D générant un plan de coupe. b) Un télémètre à balayage 2D générant 4 plans.

Le lidar fournit une mesure très précise de distance sur une grande plage de portée. Il permet de plus de détecter tout type d'objet (voiture, piéton, carton . . .) pourvu que ceux-ci renvoient une partie de la lumière incidente. Il est de fait adapté à un grand nombre d'applications.

Les technologies actuelles ne sont cependant pas exemptes de défauts. Un télémètre laser simple est constitué d'un unique rayon et non d'un cône d'émission comme le radar. Il ne peut donc détecter que dans cette unique direction. Pour corriger ce défaut, la plupart des systèmes utilisent un télémètre à balayage. Celui-ci exploite un miroir rotatif, commandé par un moteur électrique, qui permet d'orienter le rayon selon un grand nombre de directions, toutes coplanaires. Ainsi, un

¹⁴la commission européenne a autorisé en 2005 l'utilisation de la bande de fréquences 24 GHz pour ce type de radars

¹⁵Light Detection And Ranging

large angle de vue peut être exploré, avec une grande précision angulaire. La détection est alors réalisée en agglomérant les impacts proches en objets [MBN04][LRGA05]. Cependant, ce type de laser est très sensible au tangage du véhicule instrumenté, qui peut amener le plan de balayage à couper la surface de la route, percevant celle-ci comme un obstacle. Ce défaut est parfois corrigé par l'utilisation d'un laser multi-plans (généralement 4) qui permet de reconstruire une vue 3D partielle de la scène. Si ce type de capteur est de plus en plus étudié [BCPZ06b], il reste néanmoins une solution à long terme pour le contexte automobile, tant l'utilisation de parties mécaniques mobiles à l'intérieur le rendent fragile et onéreux.

Le lidar étant finalement un capteur précis, générique vis à vis du type d'obstacles et disposant d'une grande plage de fonctionnement, il s'adapte à un grand nombre d'applications, notamment des fonctions de mitigation de collision, pré-crash ou "stop & go".

1.1.2 La vision monoculaire

Parallèlement à l'utilisation des capteurs actifs que nous venons de présenter, grand nombre d'études portent sur la vision artificielle. L'utilisation d'une caméra permet, pour un coût très faible, d'accéder à une grande quantité d'informations sur un large champ de vue. Cependant, comme une caméra ne fournit pas immédiatement de mesure de distance, contrairement aux télémètres, elle doit faire appel à des algorithmes de traitement d'images plus ou moins complexes pour réaliser la détection.

Nous allons présenter rapidement les algorithmes les plus couramment utilisés pour la détection d'obstacles dans le contexte routier. La plupart d'entre eux visent à trouver dans l'image un modèle d'objet connu, généralement un véhicule ou un piéton. Dans ce cas, une première étape consiste à extraire des primitives pouvant appartenir à des obstacles, pour ensuite vérifier que celles-ci correspondent effectivement au modèle d'obstacle recherché. Néanmoins, nous verrons qu'une détection générique est également possible en monovision, notamment par l'exploitation de la composante temporelle.



FIG. 1.4 – Deux caméras CMOS conçues pour une intégration dans des véhicules.

Les primitives à rechercher peuvent être de types très variés et nous allons en présenter un aperçu non exhaustif.

Les contours image des objets sont souvent utilisés, car faciles à détecter et très présents sur le type d'objets recherchés. En particulier, les véhicules contiennent des contours horizontaux et verticaux marqués.

La symétrie est également présentée comme très caractéristique des véhicules. La recherche de maxima dans une carte de symétrie permet de trouver le positionnement latéral et la largeur des régions à plus forte symétrie, fournissant des rectangles englobant les détections [BBFN00].

La texture et les ombres sont également des primitives exploitées pour la détection.

L'approche présentée dans [MGS⁺07] propose quant à elle de combiner des mesures d'orientation, de contours et de mouvement pour construire une carte de salience, qui mesure la propension de chaque pixel à retenir l'attention. Ce type d'algorithme demande cependant à être paramétré très finement, en fonction de l'application.

Lorsque la caméra utilisée permet d'acquérir des images en couleur, d'autres primitives sont envisageables. Par exemple, la recherche de zones de forte intensité dans le plan rouge permet de trouver les feux arrières d'un véhicule [BN98]. La couleur permet également de détecter la surface de la route. Ainsi, dans [BHD97], les auteurs proposent d'utiliser un système d'apprentissage pour construire un modèle de couleur de la surface de la route, utilisé par la suite pour classer les pixels image selon leur appartenance à cette surface.

Les longueurs d'ondes au delà du visible peuvent également se révéler pertinentes. L'utilisation d'une caméra sensible au rayonnement infrarouge lointain est une solution bien adaptée à la détection des piétons, car elle permet de distinguer efficacement les sources de chaleur, et donc les humains [SRBB06]. L'utilisation d'infrarouges proches, qui ne nécessitent pas de dispositif particulier car la plupart des caméras y sont sensibles, est également envisagée dans [BFT⁺06]. Elle ne présente cependant d'intérêt que la nuit, en éclairant par ailleurs la scène avec une source infrarouge proche pour augmenter le contraste.

Des algorithmes de reconnaissance de forme sont ensuite appliqués dans les régions où ont été détectées ces primitives, pour reconnaître si elles appartiennent effectivement à un type d'obstacle connu.

L'arrière d'un véhicule étant très structuré, des modèles symboliques ont été proposés pour le représenter. L'obstacle éventuel détecté lors de la recherche de primitives est alors testé pour vérifier qu'il correspond à ce modèle [GE96]. Construire un tel modèle pour des piétons est nettement plus ardu. Plutôt qu'un modèle de piéton générique, Gavrilin *et al.* [GP99] proposent donc de construire une hiérarchie de modèles pour faciliter la recherche. Plus simplement, Bertozzi *et al.* obtiennent de bons résultats de reconnaissance de piétons dans des images infrarouge thermique, en se basant sur le rapport hauteur/largeur des obstacles, et sur la détection et la localisation de la tête de l'éventuel piéton [BBLR05].

Le nombre d'apparences possibles pour les obstacles est très grand, en particulier dans le cas des piétons. Cette caractéristique a poussé de nombreuses équipes à délaisser la représentation symbolique au profit de l'utilisation de méthodes par apprentissage sur un large échantillon d'images de piétons ou de véhicules, qui permettent la création automatique de modèles statistiques. Pour cela, il est nécessaire dans un premier temps de sélectionner les primitives discriminantes à exploiter. Pour les véhicules, les différents contours sont très marqués et souvent jugés suffisamment discriminants. Concernant les piétons, histogrammes de gradients orientés (HOG¹⁶) [SRBB06] et ondelettes de Haar sont le plus souvent utilisés [GSLP07]. L'apprentissage, puis la reconnais-

¹⁶Histogram of Oriented Gradients

sance sont ensuite réalisés par des méthodes de type SVM¹⁷ ou réseaux de neurones. L'utilisation préalable de techniques de "*Boosting*", comme AdaBoost ou ses variantes, permet d'améliorer l'apprentissage en tirant au mieux parti des attributs les plus discriminants, pour classer les cas difficiles.

Ces méthodes qui visent à valider l'appartenance d'un motif détecté à une classe d'obstacles recherchée nécessitent souvent une normalisation préalable de la taille du motif afin qu'il soit comparable au modèle. Cette étape suppose éventuellement de disposer d'une mesure de la distance qui sépare l'éventuel objet de la caméra, pour calculer un facteur d'homothétie. Pour cela, la détection de la surface de la route permet de situer la base de l'objet longitudinalement sur celle-ci et donc d'estimer sa distance [GE96] [GSLP07].

L'exploitation d'images successives fournies par la caméra embarquée est également une possibilité, car elle permet de calculer un flot optique. Tous les éléments immobiles de la scène sont alors, du point de vue du système de détection, animés d'un même mouvement. L'estimation de ce mouvement dominant permet donc de trouver les objets immobiles qui, une fois supprimés, laissent apparaître des pixels appartenant théoriquement aux objets mobiles. Ceux-ci peuvent alors être suivis pour assurer leur détection [YKN06]. L'estimation du mouvement dominant nécessite l'emploi d'une méthode robuste, car un grand nombre de pixels n'appartiennent pas à ce mouvement, ou appartiennent au fond de la scène mais sont trop lointains pour permettre une estimation précise. De nombreux auteurs font donc l'hypothèse d'une route plane, pour diminuer le nombre de paramètres à estimer et faciliter la mise en place d'une méthode robuste [SMS00]. Ces méthodes exploitant un calcul de flot optique sont limitées dans leur portée, en particulier dans la direction d'évolution du véhicule.

Plutôt que d'exploiter la composante temporelle pour réaliser la détection des obstacles, celle-ci peut être dédiée au suivi des obstacles précédemment détectés. Ainsi, Betke propose de créer un modèle d'un obstacle lors de sa première détection et de le rechercher ensuite dans les images successives. Cette recherche est réalisée par le biais d'une corrélation [BN98].

La combinaison de plusieurs approches permet enfin d'assurer une plus grande fiabilité à l'algorithme de détection en permettant de bénéficier d'informations complémentaires ou redondantes. [SLSW02] présente par exemple le système *FADÉ*, qui exploite la fusion des résultats d'algorithmes fondés sur la détection d'ombres, la détection des feux arrière, la détection de contours et la détection de symétrie.

1.1.3 La stéréovision

La stéréovision est une technique de vision fondée sur l'utilisation de plusieurs caméras (généralement deux, trois pour le système de Carnegie Mellon [Wil98]). En combinant leurs différents points de vue, cela permet d'accéder à des informations de profondeur difficilement accessibles lorsqu'une unique vue est utilisée. L'exploitation de cette information permet de bâtir des techniques de détection d'obstacles génériques et robustes. Selon de nombreux auteurs, ces techniques peuvent être rangées en deux catégories : les méthodes par appariements et les méthodes par rectification homographique.

¹⁷Support Vector Machine

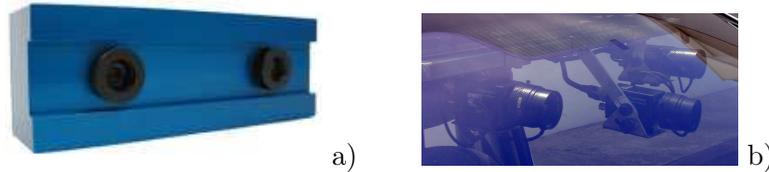


FIG. 1.5 – Deux types de systèmes stéréoscopiques..

a) Un capteur intégré. b) Des caméras embarquées dans un véhicule d'essais. Les deux caméras positionnées sur les bords forment un capteur stéréoscopique

Les approches par appariements consistent à mettre en correspondance des informations issues des différentes images du système stéréoscopique. En procédant ainsi pour chacun des pixels, cela permet de construire une image de disparité qui peut être interprétée comme une vue tri-dimensionnelle partielle de la scène observée. Connaissant la configuration du capteur, par calibrage par exemple, les informations de disparité que contient cette image peuvent être traduites en mesures métriques. C'est donc la segmentation de cette image de disparité qui permet de réaliser la détection des objets présents. Nous reviendrons sur le calcul de l'image de disparité au chapitre 2. Intéressons-nous ici aux méthodes proposées pour détecter les obstacles à partir de celle-ci.

La plupart des méthodes supposent de détecter dans un premier temps la surface de la route, ce qui permet par la suite de distinguer les objets situés au-dessus de celle-ci. Pour détecter cette surface, la majorité des approches la modélisent comme un plan (cette hypothèse est généralement connue sous le nom de "*monde plan*"). Si certains considèrent que celui-ci est fixe et connu par calibrage préalable [ZWH94], de nombreux auteurs choisissent de l'estimer continuellement, pour prendre en compte la dynamique du véhicule, qui influe nécessairement sur la position relative des caméras par rapport au sol. Cette estimation peut être réalisée à l'aide d'une régression planaire parmi les points de plus faible disparité de chaque ligne, l'optimisation de ce problème étant réalisée par une technique des moindres carrés [FJ00]. Cependant, la robustesse d'une telle approche n'est pas assurée dans le cas d'une scène complexe, qui contient beaucoup de pixels n'appartenant pas à la surface de la route. Pour résoudre cette problématique, Geronimo *et al.* [GSLP07] proposent de supprimer l'influence de ces "outliers" par l'utilisation de l'algorithme RANSAC¹⁸. Le plan finalement détecté est celui pour lequel le nombre d'"inliers" est maximum. Il est courant que les informations de disparité soient très bruitées. Plutôt que ces techniques itératives, des approches par histogramme, ou plus généralement par accumulation, permettent alors d'obtenir des estimations robustes au bruit. Dans cet esprit, la projection sur un histogramme polaire des points 3D situés à courte distance [NDF⁺04] fournit le tangage du système stéréoscopique relativement à la surface de la route. Un histogramme de courbure est ensuite construit avec les seuls points lointains pour trouver l'évolution de la courbure de la route, et reconstruire celle-ci comme une clothoïde. L'approche v-disparité proposée par Labayrade *et al.* [Lab03] permet quant à elle de réduire le problème de détection de plan à une détection de droite, en projetant avec accumulation les points de l'image de disparité le long des lignes. Le plan de la route est alors comme "vu de côté", et apparaît comme une droite, facilement détectée par une variante de la transformée de Hough (plus de précisions sur celle-ci sont données en annexe C). Outre sa simplicité de mise en œuvre, cette méthode a l'avantage de permettre de facilement

¹⁸RANdOm SAmpLe Consensus

s'abstraire de l'hypothèse "*monde plan*" en considérant la route comme une succession de plans, et en recherchant donc un ensemble de droites.

Si ces méthodes exploitant des cumuls sont très robustes à un bruit de type impulsionnel ou gaussien, elles sont directement mises à mal par des bruits fortement corrélés, qui forment des structures cohérentes dans l'espace de disparité.

Une fois détectée la surface de la route, les pixels de l'image de disparité sont classés comme appartenant ou non à celle-ci, par simple seuillage sur leur disparité. Les pixels de la route sont supprimés et les autres appartiennent donc soit à des objets soit à du bruit, souvent important.

L'étape suivante consiste à regrouper ces points de mesure pour former des objets, ou du moins trouver un rectangle pouvant englober ceux-ci. Généralement, les coordonnées images de ces pixels sont en premier lieu transformées en valeurs métriques, ce qui permet ensuite de les agréger en fonction de leur proximité [FJ00]. Avant cette agrégation, Nedeveschi *et al.* [NDF⁺04] proposent de ré-échantillonner ces points par une fonction qui assure que leur densité reste indépendante de leur distance au capteur. Dans sa thèse, Lemonde [Lem05b] propose plutôt de parcourir longitudinalement la représentation métrique de la scène pour trouver les intervalles où les points sont alignés verticalement, correspondant à des obstacles.

D'autres auteurs proposent de traiter ces points directement avec leurs coordonnées images, ce qui permet de juger de leur connexité [BCPZ06a], d'optimiser les algorithmes de parcours des points obstacles [TMRM02] ou encore d'initier une méthode de détection par monovision [TBM⁺06]. La représentation v -disparité est également utile à ce stade, car elle fait apparaître les obstacles verticaux comme des droites verticales, facilement détectables [Lab03].

Mis à part ces méthodes de détection génériques, il existe des méthodes qui visent à détecter un type obstacle donné, ce qui fournit un *a priori* fort sur ses dimensions. Pour détecter spécifiquement des piétons, [GSLP07] proposent de trouver les positions pour lesquelles l'un des cinq types de volumes prédéfinis, dont la taille varie entre $(0,75m * 0,75m * 1,5m)$ et $(0,95m * 0,95m * 1,9m)$, contient un maximum de points. Une fois les positions optimales trouvées, un classificateur est appliqué dans chacun de ces volumes pour vérifier qu'il s'agit bien de piétons.

Les approches par rectification homographique sont fondées sur une hypothèse de "*monde plan*". Dans ce cas, il suffit d'appliquer une transformation homographique à l'une des images, ou une transformation perspective inverse [BB98] aux deux, pour que l'apparence de la route soit semblable dans les deux images. La disparité sur la surface de celle-ci est alors nulle, et une différence entre les deux images, doublée d'un seuillage, fait donc directement apparaître les pixels n'appartenant pas à ce plan. Comme pour les techniques par appariements, il est ici nécessaire de connaître le positionnement du plan de la route relativement aux caméras. Cette connaissance ne peut pas être issue d'un calibrage préalable, car l'homographie à appliquer évolue avec l'assiette du véhicule. Elle peut en revanche être calculée à partir de la détection des marquages sur le sol, ce qui suppose leur existence et leur détection avec une grande précision.

Une fois réalisée la différence entre les images, les pixels restant doivent être agrégés en objets. Comme vu précédemment, ce regroupement est réalisé sur un critère de proximité ou de connexité. Dans [BB98], les auteurs remarquent que les obstacles apparaissent comme des triangles dans l'image de différence, détectés par l'utilisation d'histogrammes polaires.

Enfin, l'approche proposée par Williamson [Wil98] combine les caractéristiques des méthodes

par appariements et par rectification homographique. Les appariements sont réalisés pour la paire d'images brutes, et pour la paire dont une image est rectifiée. Les pixels sont jugés comme appartenant à la route si leur appariement est meilleur après rectification, et comme appartenant à des obstacles potentiels dans le cas contraire. L'utilisation d'un troisième point de vue confirme ensuite cette classification.

1.1.4 Les autres capteurs de vision tridimensionnelle

La stéréovision et le flot optique ne sont pas les seules techniques de vision permettant d'obtenir des informations 3D. Des capteurs optiques actifs, constitués d'un émetteur illuminant la scène par une lumière modulée et d'une matrice CCD recevant les réflexions de ces rayons sur la surface des objets, offrent également cette possibilité. La différence de phase entre l'onde émise et l'onde reçue fournit une estimation de la distance, tout comme pour les télémètres. Si ce type de capteur va très certainement prendre une place importante dans de futurs systèmes de détections d'obstacles, la faible portée et la faible résolution des modèles actuels les rendent encore peu adaptés à ces applications. Leur efficacité a néanmoins été prouvée pour de la reconnaissance de piétons [FDW⁺06].



FIG. 1.6 – *Caméra 3D exploitant une mesure de temps de vol.*

1.1.5 Les systèmes coopératifs

Une voie de recherche actuellement jugée prometteuse pour la détection d'obstacles et les aides à la conduite en général concerne les systèmes coopératifs. Cette appellation générique désigne des systèmes permettant aux différents acteurs d'une scène de "s'entraider" afin de fournir un gain en performance global. Ceux-ci peuvent donc être des dispositifs destinés à ce que l'obstacle à détecter aide à sa propre détection et éventuellement l'enrichisse d'informations supplémentaires. Des stratégies très variées existent en ce sens.

Des systèmes facilitant simplement la détection ont été proposés pour différents capteurs. Les dispositifs de vision peuvent, par exemple, être aidés en rendant les obstacles coopératifs plus visibles. Fabrizio [Fab04] propose d'équiper les obstacles de codes-barres visibles dans le proche infrarouge, et uniquement constitués de bandes adhésives comprenant des grains réflecteurs. Ils sont donc visibles à longue portée, et peuvent contenir une information codée. Aufrere *et al.* [AMC⁺00] équipent un véhicule à suivre d'amers, ici deux feux arrières et un phare sur le toit, pour en faciliter la détection et obtenir immédiatement une mesure de distance en vision monoculaire, la position relative des amers étant connues.

Le développement récent de LEDs¹⁹ très performantes favorise la conception de systèmes coopératifs. Un brevet déposé par Renault²⁰, exploite des phares de véhicules à LEDs, clignotant très rapidement. La soustraction d'images successives fournies par une caméra ayant une fréquence double de celle des lampes ne conserve alors que ces obstacles coopératifs. Cela est bien entendu vrai sous réserve que la fréquence de fonctionnement soit suffisamment élevée pour considérer que le mouvement global de la scène est nul entre deux instants successifs. La même stratégie de soustraction d'images peut être approfondie pour transmettre des informations en plus d'améliorer la détection. Les auteurs de [LP03] proposent un protocole dit "*A-on-A'-off*" qui permet de lire avec une caméra les motifs géométriques formés par les LEDs allumées parmi huit, en assurant une certaine robustesse de transmission.

Pour ne pas être trop limité en portée, Griffiths *et al.* [GLM⁺01] préfèrent un codage temporel de l'information, par des LEDs incluses dans les feux de freinage d'un véhicule. La réception est réalisée par un dispositif similaire au récepteur des signaux d'une télécommande. Ce type de récepteur est cependant limité, car il ne permet pas de localiser l'émetteur.

Parallèlement à ces efforts sur la vision coopérative, des chercheurs se sont penchés sur la possibilité d'améliorer la détection par radar. Cela est réalisé en embarquant des réflecteurs radars à base de trièdres (ou coins de cubes) ou en recouvrant les plaques d'immatriculation d'un revêtement réflecteur radar [GLM⁺01]. D'autres équipes de recherche visent plutôt au développement de radars coopératifs [FMDH00], qui échangent des informations plutôt que seulement attendre un écho.

L'utilisation des télécommunications inter-véhiculaires (V2V²¹) ou de l'infrastructure aux véhicules (I2V²²) constitue l'un des domaines les plus explorés actuellement pour le développement d'aides à la conduite. Un état de l'art des technologies exploitables peut être trouvé dans [Mou06].

Les télécommunications permettent de transmettre des avertissements de façon quasi-immédiate, ou du moins nettement plus rapidement qu'un système de détection ne pourrait les fournir, ce qui apporte un gain réel en terme de sécurité [MGL06]. Elles peuvent également fournir des informations en provenance de l'infrastructure, permettant par exemple d'avertir d'un danger situé au delà de la portée des autres systèmes perceptifs du véhicule.

Outre la transmission de ces données symboliques, l'émission de données numériques (vitesse, poids, dynamique, ...) par les différents acteurs d'une portion de réseau routier permet la construction de cartes dynamiques enrichies de l'environnement proche, par chacun des véhicules récepteurs. Celles-ci peuvent alors être utilisées pour des fonctions avancées d'aide à la conduite, comme la prévention des dangers en intersection [Amm07].

L'identification des différents objets détectés dans la scène est nécessaire pour exploiter les télécommunications V2V. En effet, pour qu'un message envoyé par un véhicule à un autre soit réellement exploitable, celui qui reçoit les informations a besoin d'identifier le véhicule émetteur. Pour bien comprendre cette problématique, la figure 1.7 donne l'exemple d'une portion de route sur laquelle roulent 3 véhicules. Le véhicule bleu est équipé en moyens de perception et détecte les deux autres protagonistes. Il reçoit également un message radio lui indiquant qu'un véhicule

¹⁹Light-Emitting Diode

²⁰C. Lavergne : "Système d'assistance à la conduite d'un véhicule automobile par détection de l'environnement", Brevet français No FR2893171

²¹Vehicle To Vehicle

²²Infrastructure To Vehicle

a détecté un obstacle et procède à un freinage d'urgence. Deux cas de figure sont alors possibles. Si le message provient du véhicule rouge, il doit être pris en compte pour aider le conducteur en l'avertissant ou en accompagnant sa manœuvre. En revanche s'il provient du véhicule jaune, il doit être pris en compte différemment, car l'obstacle n'est pas situé dans la voie de circulation.

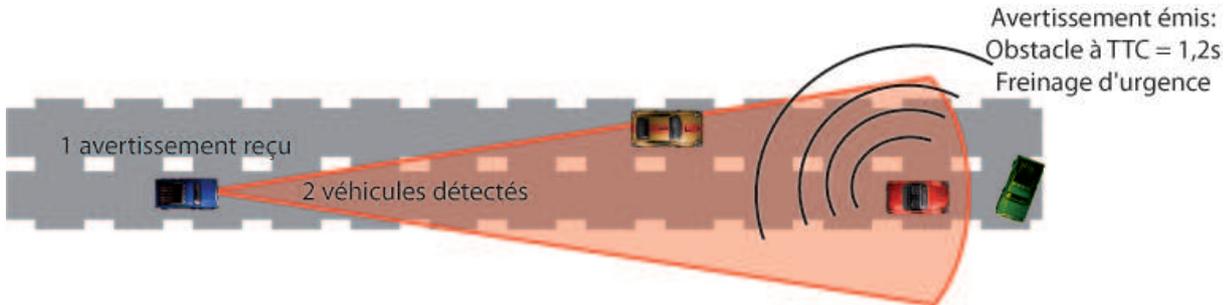


FIG. 1.7 – *Problème d'indétermination lors de l'utilisation de télécommunications V2V. le véhicule bleu détecte deux véhicules et reçoit un message d'arrêt d'urgence de l'un d'entre eux. Il n'est pas capable de déterminer quel véhicule a envoyé ce message et donc si la situation est dangereuse ou non.*

Cet exemple démontre que pour être pleinement efficaces, les communications V2V, qui ne sont pas directives, nécessitent l'aide d'un système supplémentaire de mesure du positionnement relatif. Cela peut être une localisation très précise des différents protagonistes de la scène dans un référentiel absolu. La localisation de l'émetteur est alors envoyée avec le message. Le récepteur étant lui aussi localisé dans le même référentiel, il peut vérifier à l'aide d'un système de cartographie si l'émetteur est sur sa voie de circulation et donc si le message le concerne. Cette solution ne tient cependant pas compte de la précision des systèmes de localisation actuellement disponibles pour des applications automobiles (type GPS²³). De plus, même dans le cas de systèmes très précis de type GPS différentiels, le fonctionnement n'est pas assuré dans de nombreuses situations comme les tunnels ou les canyons urbains. Un système d'identification de véhicules, fondé sur la vision, permet de palier cette faiblesse des télécommunications V2V. La figure 1.8 montre la même situation que la figure 1.7, mais dans le cas où le véhicule rouge est équipé d'un système qui l'identifie de façon unique comme étant N°3. Le véhicule bleu est capable de lire cet identifiant et de le localiser dans son repère propre. Il le compare alors à l'identifiant transmis dans le message et détermine ainsi que c'est bien le véhicule qui le précède qui a envoyé l'avertissement de freinage d'urgence.

Peu de méthodes ont été proposées pour réaliser une telle identification optique. La lecture de plaques d'immatriculation [AMAN04] ou de codes-barres coopératifs [Fab04] sont des solutions envisagées. Leur portée reste néanmoins fortement limitée par la résolution du capteur, puisqu'elles supposent la reconnaissance d'un motif spatial. Pour tirer parti de la grande portée des télécommunications, il est donc préférable de coder l'identifiant de façon temporelle. Takada [TFH97] propose pour cela un système qui combine deux émetteurs infrarouge et une trame de télécommunications, qui ont un protocole commun permettant de les associer. Nous présentons au chapitre 3 une technologie, proposée par von Arnim *et al.* dans [vAPBE07], fondée sur l'utilisation d'un unique émetteur et permettant une identification à grande portée.

²³Global Positioning System

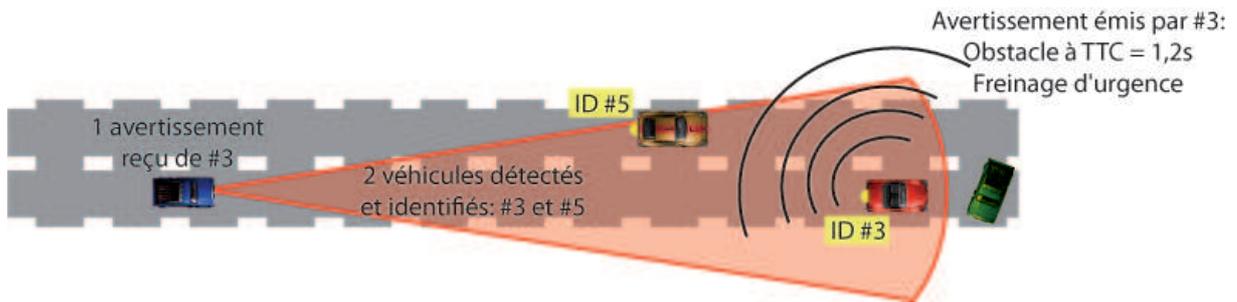


FIG. 1.8 – Utilisation d'un système d'identification optique pour supprimer l'indétermination lors de l'utilisation de télécommunications V2V. Le véhicule bleu détecte deux véhicules et reçoit un message d'arrêt d'urgence du véhicule No 3. Il identifie par ailleurs le véhicule présent devant lui comme étant le véhicule 3 et sait ainsi que le message le concerne.

1.1.6 Conclusion

Nous venons de présenter les capteurs généralement utilisés pour la détection d'obstacles, ainsi que les méthodes de traitement qui y sont associées. La figure 1.9 récapitule les domaines de fonctionnement idéaux de ces différents capteurs. En confrontant celle-ci à la figure 3 de l'introduction, il apparaît que le choix des capteurs à embarquer est commandé par les applications d'aide à la conduite qui sont conçues en aval. Une large gamme d'applications suppose donc l'emploi de plusieurs capteurs, mais les impératifs économiques prèchent plutôt pour un nombre minimal de capteurs embarqués. Il semble donc nécessaire de bâtir une architecture de détection qui couvre une plage de fonctionnement maximale, avec un nombre minimum de capteurs, tout en assurant une très grande robustesse de fonctionnement. Nous allons maintenant nous efforcer, en définissant une méthodologie de détection, de conserver ces impératifs à l'esprit et de les satisfaire au mieux.

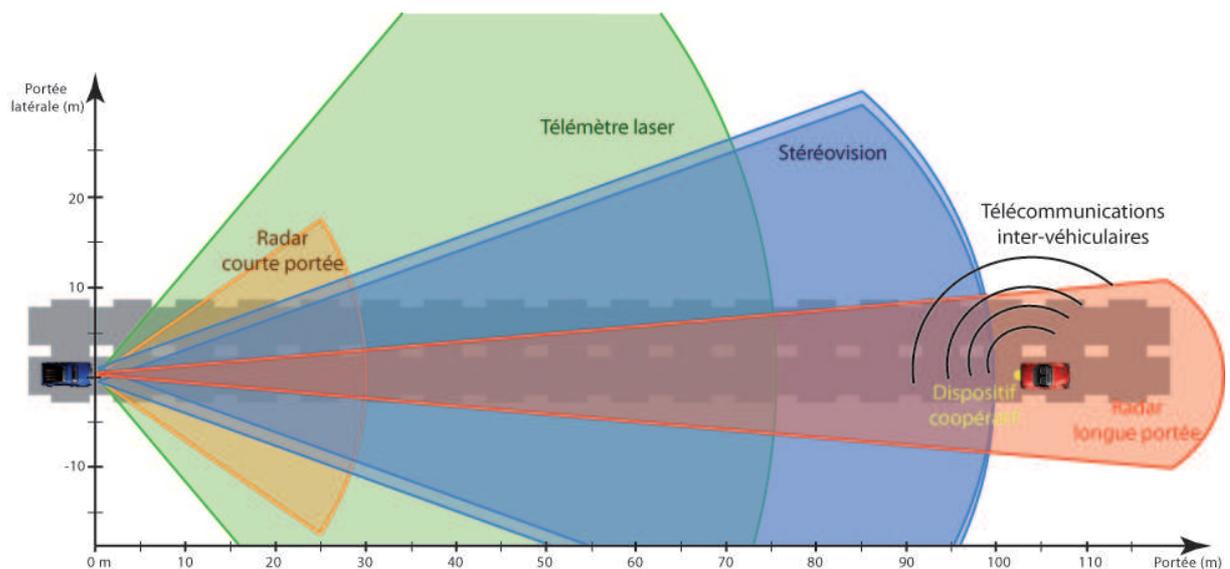


FIG. 1.9 – Domaines de fonctionnement des capteurs les plus fréquemment utilisés pour la détection d'obstacles.

1.2 Définition d'une méthodologie de détection

1.2.1 Les contraintes liées au contexte routier

Pour qu'un quelconque système actif d'assistance à la conduite présente un réel intérêt, il est nécessaire qu'il puisse percevoir et analyser un danger plus rapidement qu'un conducteur attentif ne le ferait lui-même. C'est donc une limite qui permet de définir la notion de temps réel pour notre application. Une borne maximale pour celle-ci peut être 300 ms, pour les applications de type pré-crash. Il est cependant préférable de s'imposer une valeur plus faible, qui permet d'observer plusieurs fois un obstacle pour valider sa détection. Cela est d'autant plus utile que, lorsque l'estimation de sa vitesse relative découle d'observations successives de sa position, la précision est directement liée à la fréquence d'acquisition et de traitement. Pour ces raisons, un temps de traitement compris entre 40 ms (25 Hz étant la fréquence de la plupart des caméras) et 100 ms pourra correspondre à une notion de temps réel.

La plupart des méthodes de détection réalisent un compromis entre les taux de *non détections* (obstacle présent mais non détecté) et de *fausses détections* (obstacle inexistant mais détecté). Nous avons vu en introduction que ce second type d'erreurs est à proscrire. Une contrainte essentielle consiste donc à garder quasiment nul le taux de *faux positifs* tout en visant à maximiser par ailleurs le taux de détections (ou *vrais positifs*).

Enfin, comme nous l'avons rappelé en conclusion de l'état de l'art sur les capteurs, les impératifs économiques imposent de maximiser le nombre d'applications possibles avec un nombre de capteurs embarqués minimal. Cette contrainte doit être prise également en compte pour la conception d'une méthodologie de détection viable.

1.2.2 La combinaison multi-capteurs

Compte tenu des exigences très fortes des applications routières en terme de robustesse de détection, en particulier lorsque le mode d'action est fortement intrusif, l'usage d'un capteur unique est souvent hasardeux. Pour cela, la combinaison des informations issues de plusieurs capteurs est une pratique souvent envisagée. Elle permet d'accéder à des mesures complémentaires ou à une redondance d'informations qui assure une plus grande fiabilité de fonctionnement.

De nombreuses équipes de recherche et développement ont travaillé sur des systèmes spécifiques exploitant un couple de capteurs. La plupart du temps, c'est un capteur de vision qui est associé à un capteur actif, ce qui permet de bénéficier d'informations très complémentaires. Sans être exhaustifs, nous pouvons citer des exemples de coopérations entre radar et vision monoculaire [SSO04][SLSW02], lidar et vision monoculaire [KBD05], ou encore lidar et stéréovision [LRA05]. La combinaison de plusieurs algorithmes de vision [BN98] ou de la vision monoculaire avec la stéréovision [TBM⁺06] sont des pistes qui ont également été explorées avec succès.

De façon plus générale, la fusion des données issues des différents capteurs peut être réalisée à des niveaux très différents. Ainsi, le système *FADE* [SLSW02] permet de fusionner directement les informations de bas niveau issues d'un radar et de quatre algorithmes de traitement d'image pour créer des hypothèses de cibles, qui sont ensuite validées par un algorithme d'inférence. D'autres auteurs proposent au contraire de réaliser la fusion sur des données symboliques de haut niveau, par exemple la présence d'un obstacle, fournie à l'issue des algorithmes de traitement des différents capteurs.

L'importance relative à accorder à des capteurs aux caractéristiques très différentes est également un problème qui se pose pour la fusion. Concernant cet aspect, Gruyer démontre dans [LRGA05] que pour la seule combinaison d'un lidar avec la stéréovision, la stratégie de combinaison permet de plus ou moins bien tirer parti des spécificités des deux capteurs. Il fait notamment la différence entre une stratégie de fusion redondante, qui place les deux capteurs à un niveau équivalent, ou une stratégie interdépendante, qui suppose qu'un capteur supervise la détection fournie par l'autre, en confirmant ses hypothèses de détection.

Etendre cette problématique au cas de plusieurs capteurs peut donc mener à un grand nombre de cas possibles. La fusion redondante peut alors consister à confirmer une détection si M capteurs parmi N l'ont détectée. Cela met les différents capteurs à un niveau équivalent. Il est également possible de prendre en compte les avantages et inconvénients des différents capteurs en construisant un indice de confiance sur l'existence d'un obstacle, à partir d'un indice de confiance fourni par chacun des capteurs et intégrant sa spécificité.

La fusion avec interdépendance est plus facilement généralisable au cas de N capteurs si l'on considère que la tâche de supervision est limitée à un unique capteur.

1.2.3 Méthodologie de combinaison proposée

Méthode de fusion

Dans cette thèse nous proposons une stratégie de fusion interdépendante, résumée par la figure 1.10. Elle repose donc sur deux étapes : la création d'hypothèses de détections par les différents capteurs embarqués, puis l'analyse et la validation de ces hypothèses par un capteur superviseur. Cette séparation en deux temps a notamment pour but de maximiser le rapport entre les taux de bonnes et de mauvaises détections. La première étape a en effet pour objectif de réaliser une détection surabondante pour ne pas manquer d'hypothèse importante, quand la seconde vise au contraire à assurer la fiabilité du système en minimisant le taux de fausses alarmes par une confirmation stricte. Le capteur impliqué dans cette dernière étape doit donc être à même de valider les détections de façon très robuste.

La stéréovision pour superviser la détection

Après la vue d'ensemble des capteurs impliqués dans la détection d'obstacles que nous avons proposée en début de chapitre, il semble assez naturel que la tâche de supervision des détections soit confiée à un capteur de vision. Ce type de capteur peut en effet assurer un fonctionnement robuste car il permet de disposer d'une grande quantité d'informations. D'autre part, il n'est pas sensible aux variations d'assiette du véhicule comme le serait un télémètre.

Nous avons vu que les méthodes associées à la vision artificielle sont nombreuses. La volonté d'une détection générique, qui ne soit pas tributaire d'un algorithme de reconnaissance de formes mais plutôt d'un nombre minimum d'hypothèses sur les obstacles à détecter, pousse à s'orienter vers la stéréovision. C'est donc le capteur qui sera utilisé dans cette thèse pour valider les hypothèses de détection fournies par les autres capteurs embarqués.

Confier ce rôle de validation à la stéréovision permet de limiter l'une des principales faiblesses de ce capteur : le temps de calcul. En effet, il n'y a nul besoin de traiter la totalité des images stéréoscopiques quand seule une petite partie de la scène contient des éléments d'intérêt. L'approche que nous proposons consiste donc à convertir les hypothèses de détection issues des

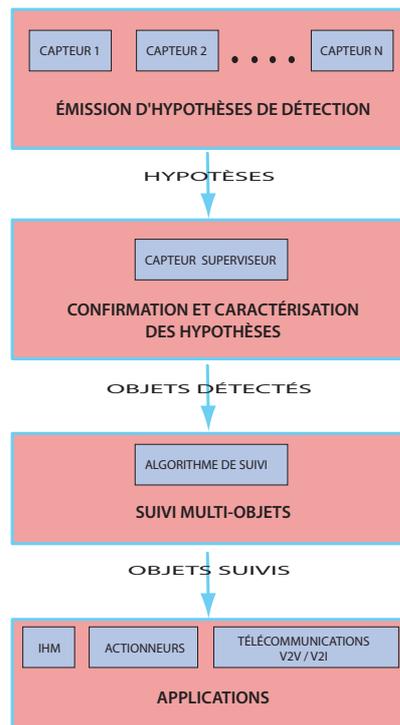


FIG. 1.10 – Vue d'ensemble de la méthodologie de détection par fusion interdépendante : les différents capteurs émettent des hypothèses qui sont ensuite confirmées et caractérisées par le capteur-superviseur.

différents capteurs en régions d'intérêt dans les images stéréoscopiques, afin de valider (ou rejeter) ces hypothèses par un traitement stéréoscopique limité à ces régions. L'intérêt d'une telle stratégie est double : d'une part elle permet d'utiliser des images d'assez haute résolution avec un temps de calcul acceptable, d'autre part l'aspect local de cette méthode permet de considérer que chaque zone d'intérêt ne contient qu'un seul objet potentiel, qui peut être caractérisé par un modèle simple.

Choix de capteurs pour la création d'hypothèses de détection

L'étape de création d'hypothèses revient donc à créer des zones d'intérêt dans les images stéréoscopiques. Tout type de capteur peut être utilisé pour cela. Dans cette thèse, nous nous sommes intéressés à la mise en œuvre de cette stratégie à travers trois capteurs qui diffèrent dans leur traitement, le type de données qu'ils fournissent, et leur portée. Cela vise à valider notre approche sur des cas très différents, et à permettre un fonctionnement sur une grande plage de portée.

La stéréovision faisant nécessairement parti du dispositif, elle peut être exploitée pour la création d'hypothèses de détections. Cependant, le temps de calcul nécessaire au traitement d'images complètes étant rédhibitoire, il est nécessaire de trouver une alternative. Nous proposons donc de dégrader la résolution des images stéréoscopiques pour cette étape de création de candidats. Cette stratégie n'est nullement une limitation, puisque l'étape de validation et de caractérisation des détections, à savoir la fin de la chaîne de traitement, peut bénéficier de la pleine résolution

grâce au traitement par zones d'intérêt.

Pour des applications de type mitigation de collision ou "stop & go", nous avons vu qu'il est nécessaire de disposer d'un capteur précis et réactif. Le lidar se prête très bien à ce type d'applications. Nous allons donc l'inclure dans l'implantation de notre système.

Enfin, pour des applications nécessitant une forte anticipation, il est nécessaire de disposer d'une grande portée de détection. Pour cela, nous proposons d'utiliser un capteur de vision coopératif, qui permet également d'identifier les véhicules équipés, offrant ainsi la perspective d'une détection enrichie par des informations issues des télécommunications inter-véhiculaires. Cette démarche est cohérente avec le besoin croissant d'anticipation des situations dangereuses pour les systèmes d'aide à la conduite de demain.

1.2.4 Choix d'un référentiel commun aux différents capteurs

Afin de permettre à ces différents capteurs de fournir des régions d'intérêt au système stéréoscopique, il est indispensable qu'ils puissent tous se référer à un même système de coordonnées. Cela est assez immédiat au sens où ils sont tout trois fixés de façon rigide au châssis du véhicule instrumenté. Nous pouvons donc utiliser un repère absolu R_a lié au véhicule et commun aux différents capteurs. La figure 1.11 illustre la géométrie d'une implantation du système complet de détection d'obstacles multi-capteurs. Le repère absolu est défini par le plan P_0 , d'équation $Z_a = 0$, qui correspond à la surface de la route supposée plane sous le véhicule, et par la projection verticale du centre de la base stéréoscopique sur ce plan, de coordonnées $(0, 0, 0)$.

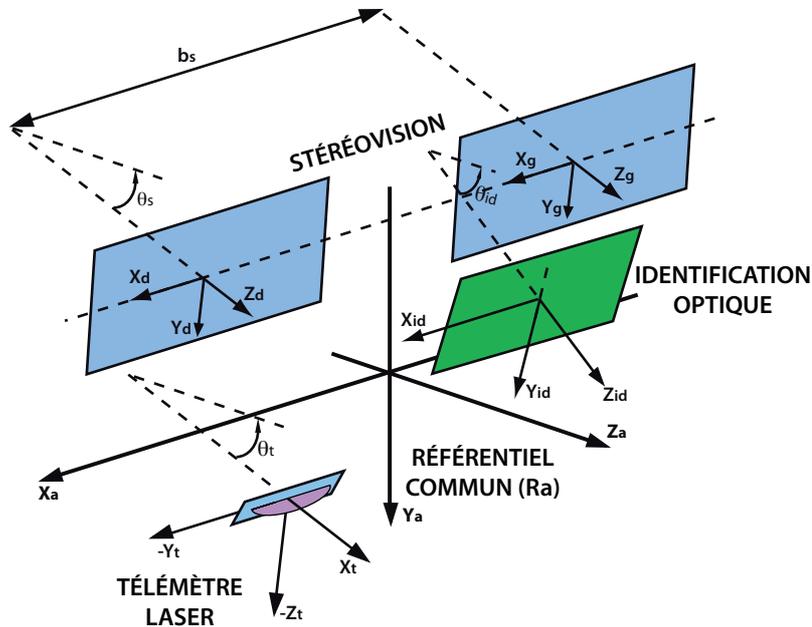


FIG. 1.11 – Les différents capteurs du système se réfèrent à un même système de coordonnées. Cela permet de les faire coopérer.

Afin que le plan P_0 corresponde effectivement toujours à la surface de la route à courte distance, celle-ci sera estimée à chaque instant par stéréovision, par application de l'approche

v-disparité. A partir de cette estimation, les valeurs de hauteur et de tangage du véhicule seront corrigées pour faire coïncider la surface de la route avec P_0 , selon la méthode expliquée plus en détails dans [LA03b].

1.3 Conclusion

Après un aperçu des principaux capteurs et méthodes de traitement utilisés pour la détection d'obstacles, nous avons proposé dans ce chapitre une méthodologie originale de détection multi-capteurs construite autour de deux étapes essentielles : la création d'hypothèses de détection par les différents capteurs impliqués et leur validation et caractérisation par un capteur dit "superviseur". Nous proposons également une implantation de cette méthodologie en confiant à un capteur stéréoscopique ce rôle de supervision. La création d'hypothèses est réalisée par l'utilisation de la stéréovision, de la télémétrie laser et de l'identification optique. La figure 1.12 récapitule cette implantation de notre stratégie de détection.

Compte tenu du rôle central de la stéréovision dans cette architecture, le chapitre suivant va être dédié aux traitements liés à ce capteur et à la définition d'un procédé de construction des points de mesure qui seront utilisés pour la détection.

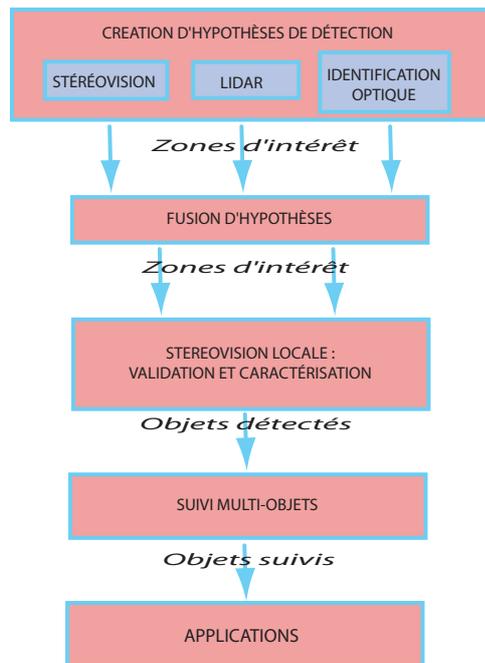


FIG. 1.12 – Mise en œuvre proposée de la stratégie de détection.

CONSTRUCTION DE POINTS DE MESURE PAR STÉRÉOVISION

Sommaire

2.1	Présentation du capteur stéréoscopique	30
2.1.1	Modèle d'un capteur de vision	30
2.1.2	Modèle d'un capteur stéréoscopique	34
2.2	Construction d'une carte de disparité	37
2.2.1	Principe	37
2.2.2	Méthodes de calcul	37
2.2.3	Principales améliorations des méthodes surfaciques	40
2.3	Détection et exploitation des plans de la scène	42
2.3.1	Modélisation de la scène	42
2.3.2	L'approche u-v-disparité	43
2.3.3	La surface de la route	44
2.3.4	Les plans verticaux	46
2.4	Vue d'ensemble algorithmique	48
2.4.1	Récapitulatif des méthodes retenues	48
2.4.2	Récapitulatif des filtrages utilisés	50
2.4.3	Algorithme final de construction des points de mesure	51
2.4.4	Remarques qualitatives sur les résultats	52
2.4.5	Exploitation du résultat : les points de mesure "obstacles"	53
2.5	Conclusion	54

AU cours du chapitre précédent, nous avons présenté les différents capteurs utilisés pour la détection d'obstacles dans le domaine automobile. Parmi ceux-ci, la stéréovision tient une place particulière, puisqu'il s'agit du capteur qui est à même de fournir l'analyse la plus approfondie de la scène observée. Cela provient d'une part du fait qu'il englobe un capteur de vision, et d'autre part de sa capacité, après traitement, à fournir des informations en trois dimensions sur une grande plage de portées.

Les données issues d'un capteur stéréoscopique se résument néanmoins à des paires successives d'images, nécessitant des traitements avant leur exploitation pour la détection. Ces traitements supposent souvent le calcul d'une image de disparité, assimilable à une représentation tri-dimensionnelle de la scène. Cependant, le calcul d'une telle image est long et peut s'avérer incompatible avec l'exigence d'un fonctionnement en temps réel, que nous avons mise en évidence au chapitre 1. L'objectif que nous nous sommes fixé est donc l'obtention d'un grand nombre de points de mesure exploitables pour la détection, plutôt qu'une représentation tridimensionnelle complète de la scène. Ces points de mesure sont positionnés dans un repère à trois dimensions et sont issus d'indices visuels présents sur les différents objets détectables de la scène. Ils sont obtenus à partir d'une carte de disparité éparsée qui a été filtrée.

Nous allons, au cours de ce chapitre, revenir plus en détail sur les étapes préliminaires de l'utilisation de la stéréovision pour la détection d'obstacles. Dans cette optique, nous allons tout d'abord présenter le modèle de capteur que nous considérerons par la suite, en décrivant le capteur de vision monoculaire puis en étendant cette description au capteur stéréoscopique.

Après cela, nous présenterons le processus de construction d'une image de disparité. Celle-ci contient les points de mesure du capteur stéréoscopique. Sa qualité est donc primordiale, ce qui justifie d'y consacrer un certain nombre d'efforts. L'étape de détection sera d'ailleurs d'autant plus facile à réaliser que les mesures sont peu bruitées. Les différentes équipes de recherche travaillant sur la stéréovision ont donc proposé un certain nombre de méthodes et d'améliorations possibles que nous présenterons, en discutant de leur intérêt pour le contexte routier.

Nous reviendrons ensuite sur l'approche *v*-disparité couramment utilisée pour la détection de plans dans l'espace de disparité. Ceci nous permettra de proposer des améliorations à l'étape de construction de la carte de disparité, en gardant en tête le gain qu'elles peuvent apporter sur le résultat final de détection.

Enfin, ce chapitre se terminera par un récapitulatif des différentes étapes de la méthode que nous avons choisi d'utiliser pour la construction des points de mesure exploités pour la détection d'obstacles.

2.1 Présentation du capteur stéréoscopique

2.1.1 Modèle d'un capteur de vision

L'outil de base des applications de vision par ordinateur est la caméra. Plusieurs modélisations en ont été proposées dans la littérature. Nous choisissons ici de présenter et d'utiliser la plus simple d'entre elles, le modèle en trou d'épingle ou *sténopé*. Cette description est inspirée de [HZ03] et [HM95] dans lesquels le lecteur pourra trouver un abondant complément d'informations.

Le modèle *sténopé*

Le modèle *sténopé* propose de représenter une caméra par un plan image et un centre optique C_0 situé à la distance focale f de ce plan (figure 2.1). La droite orthogonale au plan image et passant par C_0 est appelé axe optique de la caméra. Pour un point P de l'espace, le seul rayon lumineux issu de P et interagissant avec le capteur est le rayon porté par la droite (PC_0) . L'image de P est alors l'intersection de ce rayon avec le plan image de la caméra.

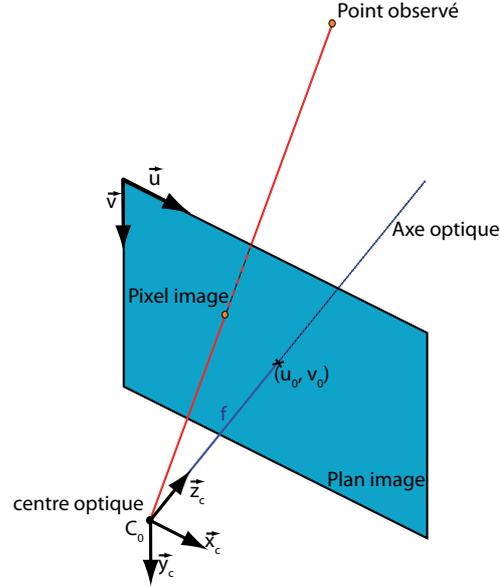


FIG. 2.1 – *Modèle de caméra en trou d'épingle ou sténopé.*

Soit un repère R_c lié à la caméra et centré sur C_0 . Le plan image est le plan d'équation $Z_c = f$. L'image du point $P(X_c^p, Y_c^p, Z_c^p)^T$ par la projection perspective sur ce plan est alors $p(fX_c^p/Z_c^p, fY_c^p/Z_c^p, f)^T$. Cette projection transforme un point de l'espace 3D en un point du plan image (2D). À l'inverse, un point du plan image $(X_c, Y_c)^T$ correspond à un rayon lumineux dans l'espace. Il peut alors être aisé de le représenter dans un système de coordonnées homogènes $(x_c, y_c, 1)^T = (X_c/Z_c, Y_c/Z_c, 1)^T$, qui tient intrinsèquement compte de cette propriété [HM95].

Par ailleurs, l'acquisition des images n'est pas réalisé par un plan continu et infini comme nous venons ici de le supposer implicitement. C'est une matrice de $N_{colonnes} \cdot N_{lignes}$ pixels rectangulaires qui est dévolue à cette tâche. Une autre transformation intervient alors, composée d'une homothétie, d'une translation et d'un échantillonnage. Afin d'exprimer cette transformation caméra/image, définissons donc un repère image $R_i(O_i, \vec{u}, \vec{v})$, lié à l'espace discret de la matrice d'acquisition. Appelons $p_0(u_0, v_0)$ l'intersection de l'axe optique avec le plan image, repérée dans R_i . Nommons également t_u et t_v la taille d'un pixel respectivement dans la direction de \vec{u} et de \vec{v} . Un point du plan image repéré par ses coordonnées homogènes $(x_c, y_c, 1)^T$ dans R_c s'exprime dans R_i sous la forme : $(u_0 + \frac{1}{t_u}x_c, v_0 + \frac{1}{t_v}x_c)^T$, soit la transformation :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \sim \begin{pmatrix} \frac{1}{t_u} & 0 & u_0 \\ 0 & \frac{1}{t_v} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix} \quad (2.1)$$

En combinant cette transformation avec la projection perspective de P sur le plan image (d'équation $Z_c = f$), il apparaît que le modèle de caméra *sténopé* transforme un point $(X_c^p, Y_c^p, Z_c^p)^T$ de l'espace en un couple : $(u, v) = (u_0 + \frac{f}{t_u} \frac{X_c^p}{Z_c^p}, v_0 + \frac{f}{t_v} \frac{Y_c^p}{Z_c^p})^T$.

Et en posant $\alpha_u = \frac{f}{t_u}$ et $\alpha_v = \frac{f}{t_v}$, cette transformation peut donc s'exprimer par :

$$Proj((X_c^p, Y_c^p, Z_c^p)^T) = (u_0 + \alpha_u \frac{X_c^p}{Z_c^p}, v_0 + \alpha_v \frac{Y_c^p}{Z_c^p})^T \quad (2.2)$$

Ce qui, en utilisant les coordonnées homogènes, peut finalement s'écrire sous la forme matricielle :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \sim \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix} \quad (2.3)$$

La matrice K_c :

$$K_c = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.4)$$

est dite matrice intrinsèque de la caméra. $u_0, v_0, \alpha_u, \alpha_v$ sont les paramètres intrinsèques. K_c suffit alors à décrire le modèle *sténopé* de la caméra utilisée et peut être mesurée par calibrage préalable [HM95].

Positionnement de la caméra dans l'espace

Nous venons de présenter le principe de la projection d'un point donné de l'espace observé sur la matrice d'une caméra. Celle-ci n'étant pas seule au monde mais incluse dans un environnement, il faut également prendre en compte son positionnement par rapport à cet environnement. Définissons donc un référentiel absolu $R_a(O_a, \vec{X}_a, \vec{Y}_a, \vec{Z}_a)$ dans lequel nous allons placer notre caméra (figure 2.2). Le positionnement de celle-ci est un changement de repère de R_a vers R_c , soit une composition de translations et de rotations suivant les 3 axes. Aucune dilatation n'intervient car la métrique des deux repères est identique.

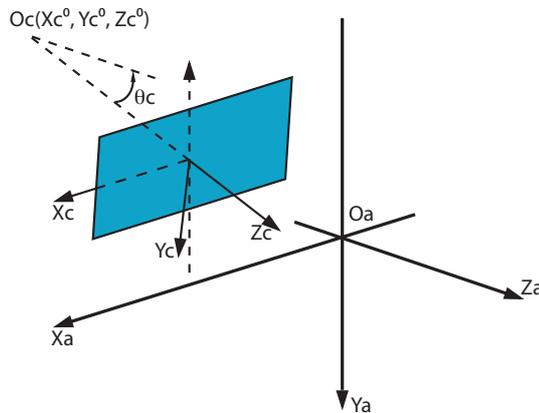


FIG. 2.2 – Positionnement d'une caméra dans le repère absolu.

Nous pouvons représenter ces transformations affines de façon matricielle. Les coordonnées d'un point dans R_a et R_c sont notées respectivement $(X_a, Y_a, Z_a, 1)^T$ et $(X_c, Y_c, Z_c, 1)^T$.

La translation de la caméra s'exprime alors sous la forme :

$$T_{ca} = \begin{pmatrix} X_c^0 \\ Y_c^0 \\ Z_c^0 \end{pmatrix} \quad (2.5)$$

La rotation autour de l'axe \vec{X}_a s'écrit quant à elle :

$$R_{\vec{X}_a} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_c & \sin \theta_c \\ 0 & -\sin \theta_c & \cos \theta_c \end{pmatrix} \quad (2.6)$$

Les rotations autour des axes \vec{Y}_a et \vec{Z}_a peuvent également être prises en compte. Par souci de simplicité, nous supposons que celles-ci sont nulles. La rotation se limite donc à :

$$R_{ca} = R_{\vec{X}_a} \quad (2.7)$$

Le changement de repère de R_c à R_a est alors représenté par la transformation :

$$P_{ca} = \begin{bmatrix} R_{ca} & T_{ca} \\ 0_3^T & 1 \end{bmatrix} \quad (2.8)$$

Si M est la représentation d'un point dans le repère caméra R_c et M' sa représentation dans le repère absolu R_a , la transformation s'exprime :

$$M' \sim P_{ca}M \quad (2.9)$$

qui s'inverse en :

$$M \sim P_{ac}M' \quad (2.10)$$

avec :

$$P_{ac} = \begin{bmatrix} R_{ca}^T & -R_{ca}^T T_{ca} \\ 0_3^T & 1 \end{bmatrix} \quad (2.11)$$

Les coefficients de la matrice P_{ac} sont dits paramètres extrinsèques de la caméra et décrivent son placement par rapport au repère R_a .

Equations du capteur de vision

En combinant les deux types de transformations qui viennent d'être présentées, nous pouvons maintenant expliciter les équations fondamentales de notre capteur :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \sim K_c P_{ac} \begin{pmatrix} X_a^p \\ Y_a^p \\ Z_a^p \\ 1 \end{pmatrix} \quad (2.12)$$

Il en résulte finalement le système d'équations :

$$\begin{cases} u = u_0 + \alpha_u \frac{X_a - X_c^0}{(Y_a - Y_c^0) \sin \theta_c + (Z_a - Z_c^0) \cos \theta_c} \\ v = v_0 + \alpha_v \frac{(Y_a - Y_c^0) \cos \theta_c - (Z_a - Z_c^0) \sin \theta_c}{(Y_a - Y_c^0) \sin \theta_c + (Z_a - Z_c^0) \cos \theta_c} \end{cases} \quad (2.13)$$

Ces équations ne sont pas inversibles parce que la notion de profondeur a été perdue lors de la projection perspective.

2.1.2 Modèle d'un capteur stéréoscopique

Si l'on utilise au moins deux capteurs de vision comme celui qui vient d'être présenté, cela permet d'obtenir un système stéréoscopique. Ce type de capteur possède alors l'indéniable avantage de fournir des informations à trois dimensions, en ajoutant au capteur de vision simple une notion de profondeur.

Principe de la stéréovision

Le principe de la stéréovision est illustré par la figure 2.3. Un pixel est théoriquement l'image d'un rayon lumineux sur le capteur de la caméra. Ce rayon peut être facilement déterminé parce qu'il est rectiligne et passe par deux points connus (le pixel et le centre optique de la caméra). En supposant que la seconde caméra du système stéréoscopique perçoit également la source S de ce rayon lumineux, il est possible de trouver un autre rayon issu de S . Nous disposons alors de deux droites se coupant en S et de la géométrie du capteur supposée connue (par calibrage par exemple), cela permet d'obtenir par triangulation la position de S dans l'espace [HZ03].

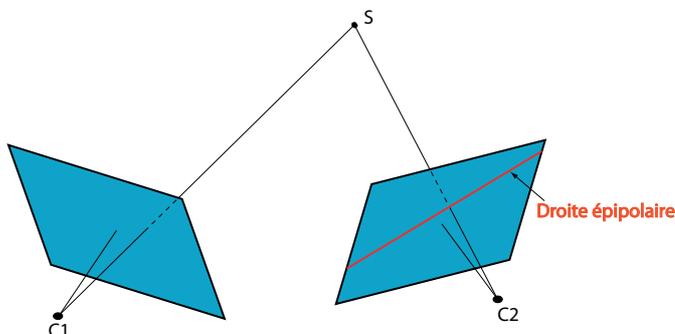


FIG. 2.3 – Principe de la stéréovision et de la géométrie épipolaire.

Mais si ce principe semble immédiat en théorie, sa mise en application pratique ne l'est pas. En particulier, trouver dans les deux images les pixels provenant de la même source (problème de l'appariement) constitue l'une des principales difficultés de la stéréovision, d'une part parce que ce problème est facilement source d'erreurs, d'autre part parce qu'il est très coûteux en temps de calcul. En particulier, il devient très complexe si l'on ne sait pas où rechercher dans la seconde image le correspondant d'un pixel de la première image. Heureusement, la connaissance de la configuration du capteur stéréoscopique simplifie grandement le problème. En effet la projection sur le plan image de la seconde caméra, du rayon lumineux engendrant un pixel sur la première caméra, est une droite. Sur celle-ci se trouve nécessairement l'image du point source. Pour chaque pixel, cette droite appelée ligne épipolaire nous fournit donc la direction unique de recherche de son correspondant. La figure 2.3 illustre cette propriété.

Le capteur stéréoscopique rectifié

Un cas particulier de géométrie épipolaire apparaît lorsque les plans image des deux caméras sont parfaitement alignés et parallèles. Dans ce cas les lignes épipolaires sont alignées et orientées suivant les lignes image horizontales. La recherche de correspondant se fait alors le long de ces mêmes lignes image, simplifiant largement le problème. La figure 2.4 montre une paire d'images stéréoscopiques avant et après l'étape appelée *rectification épipolaire* qui consiste à virtuellement réorienter les caméras pour obtenir un capteur stéréoscopique parfaitement rectifié. Il apparaît

alors que le processus d'appariement fournit, après rectification, la différence d'abscisse entre un point et son correspondant, appelée disparité. Celle-ci est inversement proportionnelle à la distance de la source au capteur.

Différents algorithmes ont été proposés pour calculer la transformation nécessaire à la rectification d'un capteur stéréoscopique, en estimant la matrice fondamentale du système. Cela est généralement réalisé par un calibrage préalable en observant un motif connu [Zha98], mais de nombreuses recherches portent sur l'automatisation de ce processus, sans *a priori* sur la scène observée. La matrice fondamentale est ensuite utilisée pour calculer des homographies à appliquer aux images issues du capteur mal configuré, afin de se rapporter au cas du capteur correctement rectifié [FTV00][LZ99]. En conséquence, nous pouvons donc nous placer dans le cas de figure d'une géométrie épipolaire parfaitement rectifiée sans que cela soit une hypothèse par trop restrictive.

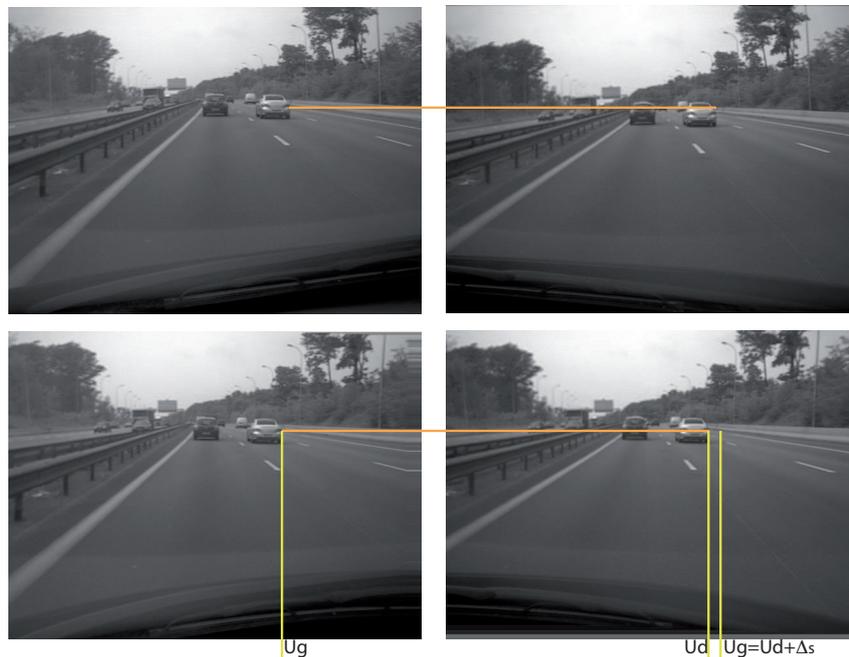


FIG. 2.4 – En haut : paire d'image non rectifiée. Les lignes images ne sont pas correctement alignées. En bas : la même paire après rectification de la géométrie épipolaire.

Un point (d'abscisse u_d) et son correspondant (d'abscisse u_g) se trouvent sur la même ligne. La différence d'abscisses, appelée disparité (Δ_s), est inversement proportionnelle à la distance de l'objet.

Positionnement du capteur dans l'espace

Tout comme pour le capteur de vision monoculaire présenté précédemment, le capteur stéréoscopique doit être situé par rapport à son environnement. La figure 2.5 illustre la configuration que nous avons retenue pour notre capteur stéréoscopique. L'espacement entre les axes optiques des deux caméras est appelé base stéréoscopique et sera noté b_s par la suite.

Nous considérerons que le capteur est placé de façon horizontale, avec le centre de sa base centré par rapport à R_a (i.e. $X_d^0 = -X_g^0 = b_s/2$, b_s étant la longueur de la base). Seule la rotation

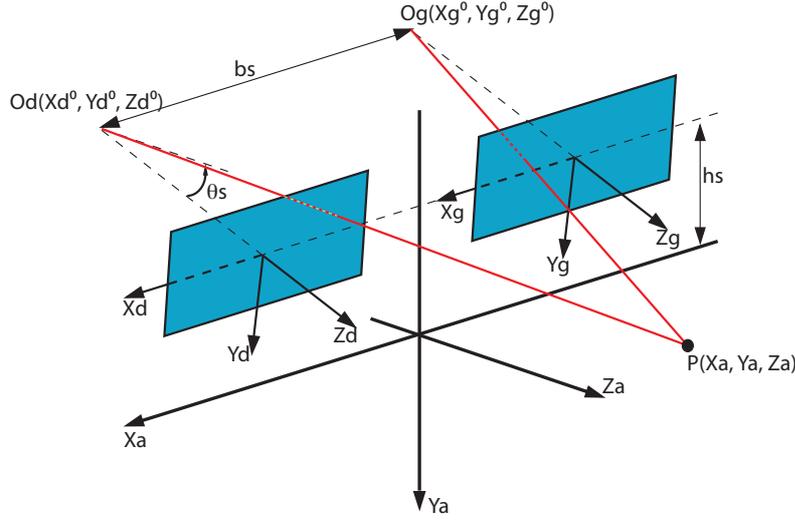


FIG. 2.5 – Positionnement du capteur stéréoscopique dans un repère absolu.

autour de l'axe X_a (caractérisée par un angle *tilt* θ_s) et les translations suivant Y_a et X_a sont donc prises en compte.

Equations du capteur de stéréovision

Les paramètres intrinsèques et extrinsèques du capteur étant fixés, les équations fondamentales du capteur en découlent immédiatement. En appliquant le système d'équations (2.13) aux caméras gauche et droite, les grandeurs images peuvent donc être calculées par :

$$\begin{cases} u_i &= u_0 + \alpha_u \frac{X_a - X_i^0}{(Y_a - Y_i^0) \sin \theta_s + (Z_a - Z_i^0) \cos \theta_s} \\ v_i &= v_0 + \alpha_v \frac{(Y_a - Y_i^0) \cos \theta_s - (Z_a - Z_i^0) \sin \theta_s}{(Y_a - Y_i^0) \sin \theta_s + (Z_a - Z_i^0) \cos \theta_s} \end{cases} \quad (2.14)$$

avec $i = d$ ou $i = g$. En posant :

$$\begin{cases} Y_d^0 = Y_g^0 = Y_s^0 \\ X_d^0 = -X_g^0 = \frac{b_s}{2} \\ Z_d^0 = Z_g^0 = Z_s^0 \end{cases} \quad (2.15)$$

on obtient finalement le système d'équation fondamental de notre capteur :

$$\begin{cases} u_d &= u_0 + \alpha_u \frac{X_a - b_s/2}{(Y_a - Y_s^0) \sin \theta_s + (Z_a - Z_s^0) \cos \theta_s} \\ u_g &= u_0 + \alpha_u \frac{X_a + b_s/2}{(Y_a - Y_s^0) \sin \theta_s + (Z_a - Z_s^0) \cos \theta_s} \\ v &= v_0 + \alpha_v \frac{(Y_a - Y_s^0) \cos \theta_s - (Z_a - Z_s^0) \sin \theta_s}{(Y_a - Y_s^0) \sin \theta_s + (Z_a - Z_s^0) \cos \theta_s} \\ \Delta_s &= \alpha_u \frac{b_s}{(Y_a - Y_s^0) \sin \theta_s + (Z_a - Z_s^0) \cos \theta_s} \end{cases} \quad (2.16)$$

$\Delta_s = u_g - u_d$ est la disparité du pixel considéré, $v = v_g = v_d$ son ordonnée.

Il apparaît alors qu'aucune dimension n'est perdue dans cette transformation et que celle-ci est bijective. Le système est donc inversible, ce qui permet d'obtenir l'équation de rétro-projection :

$$\begin{cases} X_a &= \frac{b_s}{2} + \frac{b_s(u_d - u_0)}{\Delta_s} \\ Y_a &= Y_s^0 + \frac{\alpha_u b_s((v - v_0) \cos \theta_s + \alpha_v \sin \theta_s)}{\Delta_s} \\ Z_a &= Z_s^0 + \frac{\alpha_u b_s(\alpha_v \cos \theta_s - (v - v_0) \sin \theta_s)}{\Delta_s} \end{cases} \quad (2.17)$$

L'espace de coordonnées E_Δ lié aux images stéréoscopiques et repéré par $R_\Delta = (\Omega, \vec{u}_d, \vec{v}, \vec{\Delta}_s)$ est donc un espace à trois dimensions isomorphe à l'espace 3D E_a repéré par R_a . Il est donc équivalent de travailler dans l'un ou l'autre des espaces. Ce sont les besoins d'une application donnée qui justifient alors de choisir l'un des deux.

Notons enfin que, sauf mention du contraire, nous considérerons désormais que les pixels des caméras sont carrés, ce qui entraîne : $\alpha_u = \alpha_v = \alpha$.

2.2 Construction d'une carte de disparité

2.2.1 Principe

Étant donnée une paire d'images issues d'un capteur stéréoscopique rectifié, l'appariement de tout ou partie des pixels permet d'obtenir pour chacun d'entre eux une mesure de disparité. Le report de ces mesures sur une image, appelée carte de disparité, fournit une vue tridimensionnelle partielle de la scène observée. La construction de cette image de disparité est donc généralement une étape essentielle du traitement d'images stéréoscopiques et de nombreuses approches ont été proposées pour cela dans la littérature. Beaucoup sont recensées et comparées dans [Cha05] et [SS02a]. Nous allons en présenter quelques-unes. Auparavant, afin de choisir la technique la mieux adaptée à notre problème, il convient de mettre en évidence les différents aspects qui caractérisent une *bonne* image de disparité pour la détection d'obstacles routiers. Il s'agit de :

- la densité de l'image de disparité, caractérisée par le pourcentage de pixels appariés,
- le pourcentage de mauvais appariements réalisés,
- l'indépendance des erreurs, les méthodes de recherche de plans étant sensibles aux erreurs d'appariement fortement corrélées, puisque celles-ci peuvent apparaître comme un objet fantôme,
- le comportement dans les zones de semi-occultations, quand un objet n'est visible que sur l'une des caméras,
- le temps de calcul, qui impose un compromis sur les autres caractéristiques.

2.2.2 Méthodes de calcul

Le calcul de la carte de disparité se décompose en plusieurs étapes. Tout d'abord, pour chaque pixel à appairer, un certain nombre de candidats potentiels est retenu. Pour chacune de ces **hypothèses d'appariement**, un **coût de mise en correspondance** est ensuite calculé. Enfin, la dernière étape est l'**optimisation** d'un critère, afin d'attribuer l'hypothèse d'appariement la plus plausible à chacun des pixels.

Cette étape d'**optimisation** peut être réalisée de façon locale, en considérant chaque pixel indépendamment, ou par une approche globale, qui assure une cohérence entre les appariements

obtenus pour chaque pixel. Dans ce cas, l'appariement global est ramené à un problème de minimisation d'énergie sur l'image, la fonction d'énergie globale comprenant un terme mesurant la qualité de l'appariement global et un terme de lissage. De nombreuses approches permettent de réaliser ce type d'optimisation, comme les recuits simulés, les algorithmes génétiques, les graph-cuts, etc [MS07]. Les cartes de disparité obtenues sont denses et respectent des contraintes de type ordre ou proximité, mais au prix d'un temps de calcul considérable vis à vis des puissances de calcul aujourd'hui disponibles. Elles sont donc peu adaptées à un fonctionnement en *temps réel*, indispensable pour la détection d'obstacles routiers.

La programmation dynamique, initialement introduite pour décoder des messages lors de transmissions entachées d'erreurs, est très adaptée au traitement de séquences. Elle a donc permis de mettre au point des techniques d'appariements semi-globales, visant à optimiser le coût de mise en correspondance par lignes. Ce type d'algorithme suppose qu'une ligne de l'image de gauche est identique à la ligne de droite, moyennant des actions de suppression, ajout et interversion d'éléments. En attribuant alors un coût à ces différentes actions, la programmation dynamique permet de trouver la séquence de pixels minimisant un coût total de mise en correspondance de la ligne. Cela fournit alors immédiatement une ligne de la carte de disparité. Malheureusement, lorsqu'un tel algorithme produit des erreurs, celles-ci se retrouvent souvent dans une séquence d'erreurs et sont par conséquent fortement corrélées.

Pour les méthodes dites par corrélation, l'optimisation est locale, car indépendante pour chacun des pixels : la disparité attribuée à un pixel est celle qui permet de minimiser (ou maximiser) le coût pour cet unique pixel. C'est la stratégie du gagnant qui raffle tout (ou WTA pour Winner Take All) [SS02a]. Cette approche présente les avantages d'un temps de calcul raisonnable et d'erreurs d'appariements peu corrélées en raison de son aspect local. En revanche, la carte de disparité résultante n'est généralement pas dense. En effet, les pixels étant traités indépendamment, chacun d'eux doit disposer dans son voisinage proche de suffisamment d'informations discriminantes pour que la mise en correspondance s'effectue correctement. Cette hypothèse n'est généralement pas valide dans les zones de l'image faiblement texturées ou sur les contours horizontaux.

Lhuiller *et al* proposent dans [LQ02] une façon d'obtenir une carte de disparité semi-dense en propageant la disparité de proche en proche à partir d'un petit nombre d'appariements considérés comme corrects (germes initiaux), suivant un algorithme inspiré d'une croissance de région. Seule une petite variation de disparité est tolérée à chaque pas. Le nombre de cas testés est donc considérablement réduit par rapport à une méthode par corrélation classique. Ainsi, le temps de calcul ne dépend que du nombre d'appariements finaux n et la complexité est en $O(n \log(n))$ alors qu'elle est en $O(n \cdot \Delta_{Max})$ pour obtenir une carte dense avec une méthode WTA, Δ_{Max} étant la valeur maximale testée pour la disparité. De plus, cela permet de propager une valeur de disparité le long des contours horizontaux. Le compromis entre le temps de calcul et la densité de l'image de disparité est donc très intéressant, mais les erreurs que celle-ci contient peuvent être fortement corrélées, car propagées.

Différents coûts de mise en correspondance d'un couple de pixels ont été proposées, les résultats pouvant différer sensiblement suivant le critère retenu. Si ce coût peut correspondre à une simple différence d'intensité entre les deux pixels, la plupart des approches supposent l'agrégation du coût sur un proche voisinage. Ces méthodes sont alors dites surfaciques et sont fondées sur l'hypothèse que les deux images sont localement semblables autour des points à ap-

parier. Comme le rappelle Devernais dans sa thèse [Dev97], cette hypothèse est vraie pour les projections des points provenant de surfaces parallèles aux plans images (dites fronto-parallèles), localement continues et respectant localement une contrainte Lambertienne (*i.e.* l'intensité de projection dans une image est indépendante du point de vue).

Parmi les principaux scores utilisés, se trouve SAD²⁴ qui mesure la distance (en valeur absolue) entre l'intensité des pixels du voisinage sur l'image de droite et de leurs homologues dans l'image de gauche. Ces distances sont ensuite sommées sur la totalité de la fenêtre. Très similaire, le score SSD²⁵ repose quant à lui sur la distance euclidienne entre les vecteurs d'intensité des deux fenêtres à comparer. Ces deux scores sont des mesures de dissimilarité et valent théoriquement 0 à la disparité optimale. Ils offrent d'ailleurs des performances assez comparables, et le choix de l'un ou de l'autre repose essentiellement sur le coût calculatoire que représente une valeur absolue comparativement à une élévation au carré pour l'architecture matérielle sur laquelle l'algorithme est amené à tourner.

Les formulations théoriques des scores SSD et SAD sont sensibles aux différences d'illumination entre les deux caméras. Cette faiblesse peut être corrigée en recentrant les intensités des points d'une fenêtre sur la moyenne de luminance de celle-ci, donnant les critères ZSSD²⁶ et ZSAD²⁷. Le critère ZNSSD²⁸ est une variante du critère ZSSD, mais normalisé par l'écart-type des intensités des deux fenêtres, de façon à prendre également en compte une éventuelle différence de gain entre les deux caméras.

Le score ZNCC²⁹ est le produit scalaire des vecteurs d'intensité recentrés sur leurs moyennes, normalisé par l'écart type de l'intensité des deux fenêtres. Ce score mesurant une similarité, la décision finale sur la disparité d'un pixel n'est pas prise pour sa valeur minimale comme pour les critères précédents, mais pour le score de corrélation maximal (le plus proche de 1).

Afin de limiter les problèmes liés aux occultations et aux variations de gain entre les deux images Zabih et Woodfill ont proposé dans [ZW94] une mesure de similarité fondée sur une transformation locale non paramétrique, la transformée de Census, qui caractérise la répartition locale d'intensité des pixels. Le principe consiste à construire pour chaque pixel P un vecteur de bits. Chacun de ces bits représente un pixel du voisinage de P , qui vaut 1 si son intensité est inférieure à l'intensité de P , et 0 sinon. Une fois la transformée appliquée aux deux images de la paire stéréoscopique, la mise en correspondance est effectuée en réalisant une corrélation au sens de la distance de Hamming, c'est-à-dire en comptant le nombre de bits différents entre les transformées des points à associer.

La transformée de Census compare donc des motifs et est en conséquence bien moins sensible aux occultations que les méthodes classiques.

Des critères beaucoup plus complexes, attribuant à chaque pixel un ensemble de caractéristiques locales, comme le permet par exemple le descripteur SIFT³⁰ [Low04] ou son homologue

²⁴Sum of Absolute Differences

²⁵Sum of Squared Differences

²⁶Zero-mean Sum of Squared Differences

²⁷Zero-mean Sum of Absolute Differences

²⁸Zero-mean Normalized Sum of Squared Differences

²⁹Zero-mean Normalized Cross Correlation

³⁰Scale Invariant Feature Tracker

rapide SURF³¹ [BTG06] sont utilisés pour des problèmes de mise en correspondance. Leur utilisation en temps réel se limite cependant à un petit nombre de points caractéristiques et les exclut donc des méthodes utilisables pour le calcul d'une image de disparité.

L'ensemble des hypothèses d'appariements pour un pixel donné correspond normalement à l'ensemble des pixels de même ordonnée et d'abscisse supérieure ou égale à l'abscisse dudit pixel, un système stéréoscopique rectifié ne permettant pas l'existence de disparité négative. Diminuer la taille de cet ensemble d'hypothèses est un bon moyen d'économiser du temps de calcul tout en réduisant le risque d'erreurs d'appariements. La méthode de propagation de la disparité est un exemple d'algorithme tirant parti de cette stratégie.

Vérifier que le signe du gradient horizontal d'intensité est bien le même pour les deux pixels à appairer permet de diviser par 2 le nombre d'hypothèses mais ne peut être réalisé que si la valeur absolue de ce gradient est suffisamment importante pour éviter toute ambiguïté avec le bruit présent dans l'image.

Une autre méthode efficace pour réduire le nombre d'hypothèses d'appariements consiste à décomposer ce processus en deux passes. Dans un premier temps, une carte de disparité très frustrée et éparse est calculée le plus rapidement possible, puis utilisée pour détecter la route et les plans fronto-parallèles. Dans un second temps, la carte de disparité est recalculée en ne testant que les hypothèses d'appariements correspondant à ces plans [LA03a]. Dans [HLPA06] Hautière *et al.* proposent d'utiliser une approche similaire, mais en exploitant l'algorithme de propagation de la disparité lors de la seconde passe, les directions de propagation étant contraintes par le résultat de la détection de plans. La carte de disparité résultante est finalement beaucoup plus dense et contient des contours horizontaux, difficiles à obtenir avec une stratégie WTA normale.

Enfin, une méthode classique de réduction du nombre d'hypothèses est l'approche multi-résolutions. [Lem05a] et [KKO⁺06] proposent pour cela une approche pyramidale. La carte de disparité est tout d'abord calculée pour une faible résolution, un petit nombre d'hypothèses étant alors évaluées. Celle-ci est ensuite propagée vers la résolution supérieure, en affinant les résultats de mise en correspondance. Ainsi, seules 3 disparités sont évaluées pour chaque pixel lors de cette descente dans la pyramide. Le processus peut être ensuite réitéré jusqu'à la résolution maximale.

2.2.3 Principales améliorations des méthodes surfaciques

Parce qu'elle est la plus efficace en temps de calcul et qu'elle permet d'assurer une certaine indépendance des erreurs réalisées, c'est la stratégie par corrélation surfacique, la décision étant prise selon le principe WTA, qui est le plus souvent utilisée pour la détection d'obstacles routiers. De nombreux auteurs s'accordent sur cette conclusion. Mais comme cela est illustré dans [SS02a], les cartes de disparité construites ainsi sont entachées d'un grand nombre d'erreurs. Diverses propositions ont donc été faites pour améliorer cette approche.

Tout d'abord, plusieurs techniques permettent de supprimer des faux appariements. Le plus simple consiste à simplement seuiller le score de mise en correspondance de façon à ne garder que les appariements que l'on pourrait considérer comme "forts". Ce type de seuil est cependant très

³¹Speeded Up Robust Feature

sensible à régler et peut amener à supprimer un grand nombre de bons appariements plutôt que des erreurs, en particulier dans les zones peu texturées. Plus efficace est le rejet des appariements ambigus, *i.e.* pour lesquels le second meilleur score est proche du premier, en valeur relative [HIG02].

La vérification croisée des appariements permet également de supprimer des erreurs, en particulier dans les zones de semi-occultations [EW02]. Cette technique consiste à vérifier que les résultats sont bien les mêmes lorsque l'on prend pour référence l'image de droite de la paire pour chercher les correspondants dans l'image de gauche et lorsque l'on procède dans le sens contraire. En outre, si l'on dispose dans le matériel de calcul d'un espace mémoire suffisant, cette méthode présente un surcoût calculatoire très faible. En effet, le stockage en mémoire de tous les scores de mise en correspondance d'une ligne donnée évite d'avoir à les recalculer lors de l'étape de vérification croisée : le coût d'appariement d'un pixel (u_d, v) de l'image de droite avec le pixel (u_g, v) de l'image de gauche est symétrique : $Cout(u_d, \Delta, v) = Cout(u_g, -\Delta, v)$ quand $u_g = u_d + \Delta$.

Le compromis entre la densité de la carte de disparité et le taux de bons appariements passe également par le choix des pixels à appairer. En effet, un pixel situé dans une région peu texturée de l'image contient peu d'information discriminante et, par conséquent, peut plus facilement être source d'erreurs lors du processus d'appariements qu'un pixel situé sur un contour image marqué. Pour cela, de nombreux auteurs conseillent de détecter les points à appairer dans un premier temps. Cela est réalisé à l'aide d'une mesure de valeur de texture [LQ02] ou encore du calcul du gradient horizontal d'intensité des pixels par le passage d'un masque $[-1, 1]$ ou $[-1, 0, 1]$ sur les images [Lab03]. Le temps de calcul de l'image de disparité est alors fortement réduit. Cependant, lors de l'implantation de l'algorithme de corrélation sur un matériel dédié [Kon97] ou en utilisant des instructions SIMD³² [vdMG06], il est plus efficace de calculer la totalité de l'image de disparité pour assurer un traitement systématique. C'est en post-traitement que sont détectés les pixels significatifs en utilisant un opérateur d'intérêt, comme l'opérateur de Moravec [Mor77].

Le gain le plus important en matière de qualité de la carte de disparité peut être obtenu en agissant sur le voisinage de corrélation. Généralement rectangulaire, sa taille a une influence importante, comme cela est décrit dans l'annexe B. Une fenêtre trop petite rend le procédé d'appariement très sensible au bruit et aux ambiguïtés. En revanche, une fenêtre trop grande peut couvrir la surface de plusieurs objets situés à des profondeurs différentes, allant ainsi à l'encontre de l'hypothèse de présence d'un seul objet continu. Cet effet est très présent aux frontières des objets, faussant les dimensions de ceux-ci. Plusieurs pistes ont été explorées pour résoudre ce dilemme. Par exemple, [KO94] et [Vek02] présentent des algorithmes visant à adapter automatiquement la fenêtre de corrélation aux surfaces à appairer. Plutôt que ces techniques au coût calculatoire élevé, d'autres auteurs [FRT97][HIG02] proposent l'utilisation de plusieurs fenêtres rectangulaires légèrement décalées et la sélection de celles qui donnent le meilleur score, reconstituant ainsi virtuellement une fenêtre non rectangulaire adaptée. Dans [YK05], un résultat similaire à l'utilisation d'une fenêtre adaptée est obtenu en utilisant une fenêtre rectangulaire dont l'influence de chaque pixel est pondérée en fonction de sa proximité au pixel central, en termes de distance et de couleur.

Par ailleurs, lorsque l'hypothèse des surfaces fronto-parallelles n'est pas respectée (typiquement sur la surface de la route), l'emploi d'une grande fenêtre complique le processus d'apparie-

³²Single Instruction Multiple Data

ment. En effet, si l'on utilise une fenêtre de corrélation ayant une certaine hauteur, la notion de disparité optimale pour ce voisinage n'a plus vraiment de sens, car elle est différente pour chaque ligne. Lefebvre *et al.* considèrent alors que l'usage de fenêtres unidimensionnelles (*i.e.* de hauteur 1 pixel) permet de s'abstraire de ce problème de distorsion perspective [LAC06]. Une approche totalement différente est proposée par Williamson [Wil98], qui applique une homographie à l'une des images en amont de la corrélation pour rectifier la perspective, comme cela est illustré sur la figure 2.6. Cette technique suppose cependant l'hypothèse d'un *monde plan* et la connaissance précise du positionnement des caméras par rapport à ce plan.

Plutôt que d'appliquer une homographie aux images, il est équivalent (au mode d'interpolation près) d'appliquer une déformation aux fenêtres. Dans sa thèse, Devernais démontre la possibilité de déformer les voisinages de corrélation en fonction des dérivés directionnelles de la disparité [Dev97].

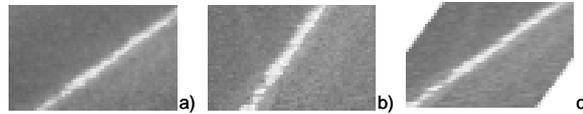


FIG. 2.6 – Observation d'un détail de la route.

a) Image de droite b) Image de gauche c) Image de gauche après rectification homographique.

Filtrer la carte de disparité est ensuite un bon moyen de supprimer une partie du bruit. Dans [BCPZ06a] par exemple, il est proposé d'appliquer un filtrage aux données pour supprimer tous les pixels de la carte de disparité qui n'appartiennent pas à un alignement vertical. Un filtrage médian peut également supprimer efficacement du bruit impulsif, mais son usage est limité par l'aspect éparsé de la carte de disparité.

Enfin, il peut être intéressant de filtrer les images stéréoscopiques en amont du processus d'appariement afin de faciliter celui-ci. Par exemple, dans [Wil98] l'auteur propose d'appliquer un filtrage LOG³³, qui lisse l'image puis réhausse les contours. En outre, comme dans tout problème de traitement d'image, il est classique d'appliquer un filtrage aux images à traiter en fonction du type de bruit que celles-ci contiennent, filtrage gaussien pour un bruit blanc, médian pour un bruit impulsif, etc. Un simple filtrage gaussien peut être particulièrement bénéfique au processus d'appariement si les images sont bruitées.

Nous reviendrons au paragraphe 2.4 sur les choix que nous avons arrêté pour notre système. Avant cela et pour justifier ces choix, nous allons présenter l'intérêt que présentent certains types de plans pour le traitement des scènes routières.

2.3 Détection et exploitation des plans de la scène

2.3.1 Modélisation de la scène

Nous allons au cours de cette thèse considérer que tous les objets à détecter peuvent être considérés comme des surfaces localement planes, ce qui est plutôt réaliste vis à vis des scènes routières à traiter et du type d'objets qu'elles contiennent (véhicules, piétons, signalisation, ...). Nous

³³Laplacian of Gaussian

pouvons d'ailleurs noter deux types d'orientations principales pour ces plans : quasi-verticaux quand il s'agit d'obstacles et quasi-horizontaux pour la route.

Nous considérerons que la route est une succession de plans de dévers nul. Les obstacles apparaîtront comme des portions de plans verticaux fronto-parallèles (voitures, panneaux, piétons, ...) ou non (barrières de sécurité, murs d'immeubles, ...).

Cette modélisation a cependant ses limites. Ainsi, un piéton peut vraisemblablement être considéré comme plan lorsqu'il est loin du dispositif d'acquisition, car l'erreur sur sa planarité est faible par rapport à la distance qui le sépare du capteur. En revanche, lorsqu'il est proche, cette hypothèse n'est plus vérifiée. L'incertitude doit alors être prise en compte par une notion d'"épaisseur" du plan.

Mis à part cette limitation, nous nous retrouvons donc à devoir détecter des plans dans l'espace observé. E_Δ étant isomorphe à E_a , cette détection peut être opérée indifféremment dans l'un ou l'autre des espaces. En outre, la géométrie projective assure qu'un plan dans un repère monde se transforme en plan dans l'espace de disparité [HZ03]. Ce sont donc bien des plans que l'on recherche, dans E_a ou dans E_Δ .

Ainsi, si l'on définit un plan dans R_a par :

$$A_0(X_a - X_s^0) + A_1(Y_a - Y_s^0) + A_2(Z_a - Z_s^0) + A_3 = 0 \quad (2.18)$$

il est représenté dans R_Δ par une équation de la forme :

$$a_0(u_d - u_0) + a_1(v - v_0) + a_2\Delta_s + a_3 = 0 \quad (2.19)$$

avec :

$$\begin{cases} a_0 &= A_0 \\ a_1 &= A_1 \cos \theta_s - A_2 \sin \theta_s \\ a_2 &= \frac{A_3}{b_s} \\ a_3 &= \alpha(A_1 \sin \theta_s + A_2 \cos \theta_s) \end{cases} \quad (2.20)$$

2.3.2 L'approche u-v-disparité

Il existe différentes façons de détecter un plan dans E_Δ . Pour réaliser cette opération à moindre coût, Labayrade *et al.* ont proposé de projeter les points de l'espace de disparité suivant les "bonnes" directions de recherche, avec accumulation [Lab03]. Les plans orientés suivant la direction de projection apparaissent alors comme des droites et sont ainsi détectables à l'aide d'un simple extracteur de droites. Les pixels qui n'appartiennent pas aux plans ainsi orientés ne s'accumulent pas et ne perturbent donc pas la détection. En accord avec la modélisation que nous venons de proposer, plusieurs directions de projection présentent un intérêt particulier, sous l'hypothèse que θ_s est faible :

- La projection le long de l'axe \vec{u} , fournit une vue de côté de la scène qui fait apparaître les plans fronto-parallèles et la surface de la route. C'est l'approche *v-disparité*, du nom du plan de projection.
- De la même façon, la projection le long de l'axe \vec{v} crée une vue de dessus de la scène dans le plan u-disparité, mettant en évidence les plans verticaux.
- L'utilisation de projections suivant des directions proches de \vec{v} permet de gérer le cas des routes ayant du dévers ou le cas du roulis du véhicule détecteur.

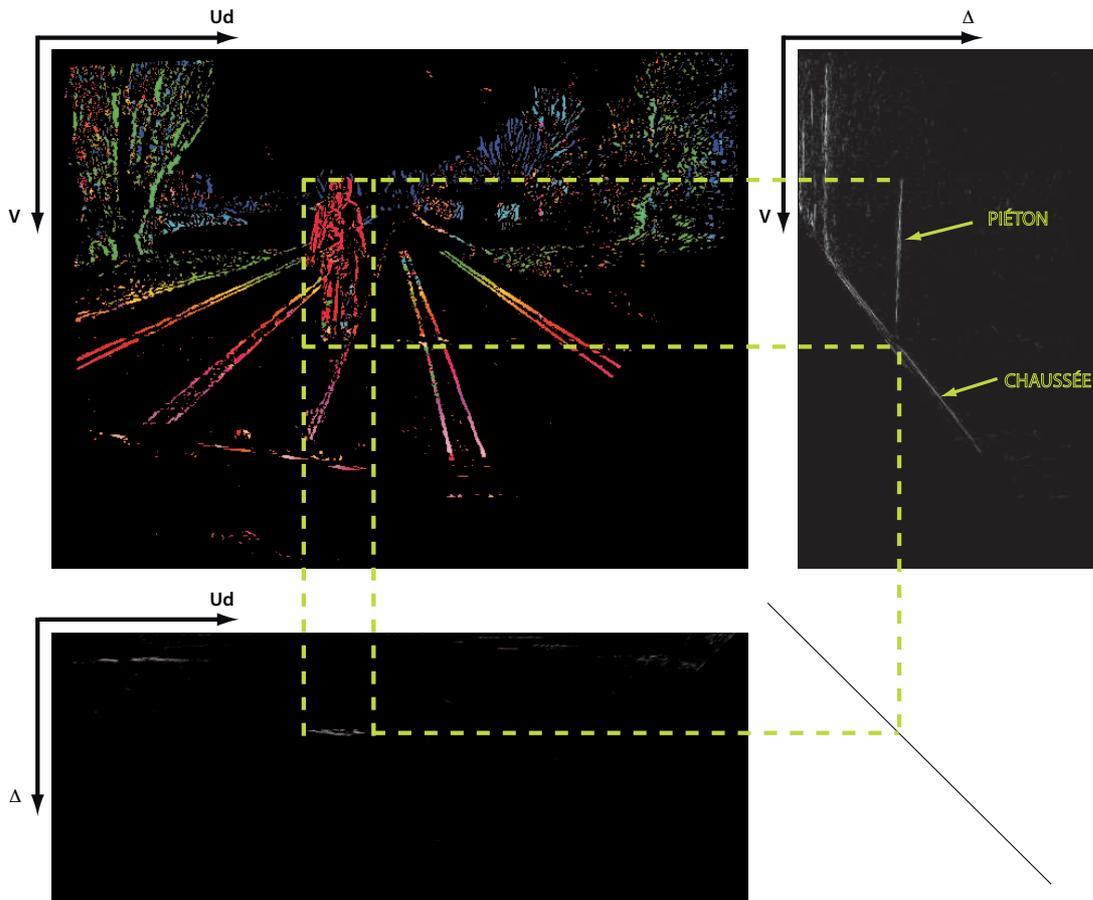


FIG. 2.7 – Paire stéréoscopique et images *v*-disparité et *u*-disparité correspondantes.

La figure 2.7 illustre la projection cumulative d'une image de disparité sur les plans *v*-disparité et *u*-disparité. La route et les obstacles fronto-parallèles sont respectivement visibles dans l'image *v*-disparité comme des segments de droites obliques et verticales. Dans la représentation *u*-disparité, seuls les obstacles apparaissent, car ils s'accumulent verticalement. Ces deux représentations suffisent donc à détecter le profil longitudinal de la route et les plans verticaux. Nous allons maintenant étudier plus en détails les deux types de plans correspondants à ces surfaces.

2.3.3 La surface de la route

Détection *v*-disparité

La détection des surfaces planes de la route est effectuée en détectant les droites correspondantes dans l'image *v*-disparité, à l'aide d'une variante en niveaux de gris de la transformée de Hough. Ce procédé est décrit en annexe C. La détection du plan de la route dans les premiers mètres devant le véhicule permet de ré-estimer le positionnement du capteur stéréo par rapport au sol et donc d'obtenir une mesure de la variation de tangage et de hauteur qu'a subie le véhicule [LA03b]. Les valeurs de θ_s et Y_s^0 sont donc mises à jour à chaque image pour compenser cette

variation. Par la suite, le plan de la route dans les premiers mètres pourra donc toujours être considéré comme étant le plan d'équation $Y_a = 0$.

Exploitation de la distorsion perspective

Comme nous l'avons déjà remarqué au paragraphe 2.2.3, l'observation des images stéréoscopiques laisse entrevoir que la distorsion perspective sur la surface de la route est très importante (figure 2.6). Cela est préjudiciable à la corrélation et ceci, d'autant plus que la fenêtre de corrélation est grande dans la direction de \vec{v} .

La méthode que propose Williamson, consistant à appliquer une homographie à l'une des images est bien adaptée à ce problème, d'autant qu'elle permet en plus de séparer route et obstacles [Wil98]. Cependant, en considérant un monde plan, elle nécessite de connaître l'homographie qui s'applique à la surface de la route lors de la projection perspective, celle-ci dépendant du tangage du véhicule. Or cette homographie correspond à la transformation de plan à plan qui a lieu lors de la projection du plan de la route dans E_Δ . Trouver ce plan dans l'espace de disparité permet donc de retrouver les paramètres de la transformation.

D'après l'équation (2.17), la relation entre l'ordonnée image et la disparité sur le plan d'équation $Y_a = 0$ est donnée par :

$$\Delta_s = -\frac{b_s((v - v_0) \cos \theta_s + \alpha_v \sin \theta_s)}{Y_s^0} \quad (2.21)$$

Un déplacement selon \vec{v} le long de ce plan se caractérise donc par une variation de la disparité :

$$\frac{\partial \Delta_s}{\partial v} = -\frac{b_s \cos \theta_s}{Y_s^0} = p_r \quad (2.22)$$

Une fenêtre de corrélation standard, présentant une certaine hauteur, ne permet donc pas de trouver la disparité optimale lors de l'appariement de pixels sur la surface de la route, puisqu'elle contient des lignes de disparité différentes. En revanche, incliner la fenêtre de corrélation permet de compenser cette variation de disparité.

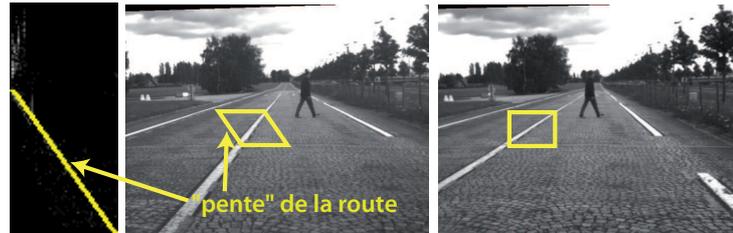


FIG. 2.8 – Mise en correspondance d'une fenêtre rectangulaire à droite avec une fenêtre inclinée à gauche pour s'adapter à la mise en correspondance des pixels de la route.

Nous proposons donc de mettre en correspondance une fenêtre rectangulaire dans une image avec une fenêtre inclinée d'une pente p_r dans l'autre image (figure 2.8) pour corriger la distorsion perspective lors de la phase d'appariement. La pente de la fenêtre inclinée est directement liée à l'assiette du véhicule, et varie donc au cours du temps. Néanmoins, elle reste toujours proche de sa valeur initiale p_0 (lorsque le véhicule est au repos). Celle-ci peut donc être utilisée pour calculer une première carte de disparité, permettant d'estimer plus précisément le profil de la route, et donc l'inclinaison réelle que doit avoir la fenêtre de corrélation pour traiter au mieux la surface de la route. Cette méthode est alors facilement extensible à un profil de route non plan : la représentation de celui-ci dans le plan v-disparité fournit localement une pente en fonction de

l'ordonnée.

D'un point de vue algorithmique, la corrélation est réalisée avec les deux types de fenêtres. Seule celle qui fournit le meilleur résultat est retenue. C'est théoriquement la fenêtre classique qui donne le meilleur score lorsque le pixel à apparier est situé sur un objet vertical, et la fenêtre inclinée lorsque ce pixel est sur la surface de la route.

Cette technique permet également de séparer les points obstacles des points appartenant à la surface de la route dès le calcul de l'image de disparité, sans avoir à seuiller celle-ci *a posteriori* sur un critère de disparité. Elle présente en outre un réel intérêt pour l'estimation du profil de la route à partir de l'image v-disparité. En effet, dans [SPLA07], Soquet *et al.* ont mis en évidence une limitation de l'utilisation de la transformée de Hough en niveaux de gris pour extraire le profil globale de la surface de la route dans cette image : la présence d'obstacles comprenant un grand nombre de points peut fausser l'estimation en créant un point attracteur (figure 2.9-a), en particulier lorsque peu d'indices visuels sont présents sur la surface de la route. Nous proposons donc de supprimer ce type de points en ne projetant sur l'image v-disparité que les pixels répondant mieux à la fenêtre de corrélation adaptée à la surface de la route. En pratique, cela permet de fiabiliser l'estimation de la surface de la route en supprimant l'influence des obstacles, comme le montre la figure figure 2.9-b.

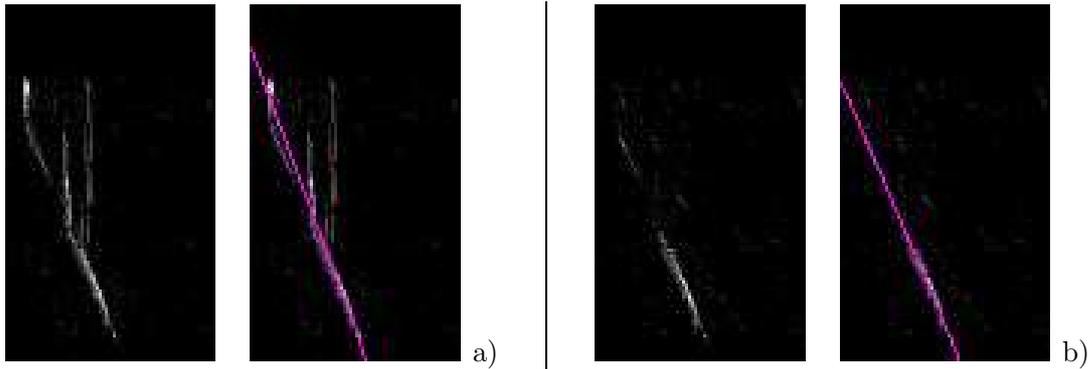


FIG. 2.9 – *Fiabilisation de l'extraction du plan de la route dans l'image v-disparité.*

a) *Un grand nombre de pixels accumulés sur un obstacle "attirent" la droite et faussent la détection.* b) *Seuls les appariements correspondant à une hypothèse "route" sont projetés dans le plan v-disparité. La détection est alors réalisée correctement.*

2.3.4 Les plans verticaux

Nous avons vu que les plans verticaux font partie des éléments de la scène à détecter. D'une manière générale, ceux-ci s'accroissent sur l'image u-disparité. La figure 2.10 illustre ce phénomène. Ainsi, piétons, barrière, voiture et murs des immeubles apparaissent clairement sur cette image.

Les portions de plans fronto-parallèles s'accroissent identiquement sur les images v-disparité et u-disparité. Il est alors théoriquement aisé de déterminer les dimensions de l'objet correspondant puisque ses limites latérales sont trouvées dans l'image u-disparité et ses limites

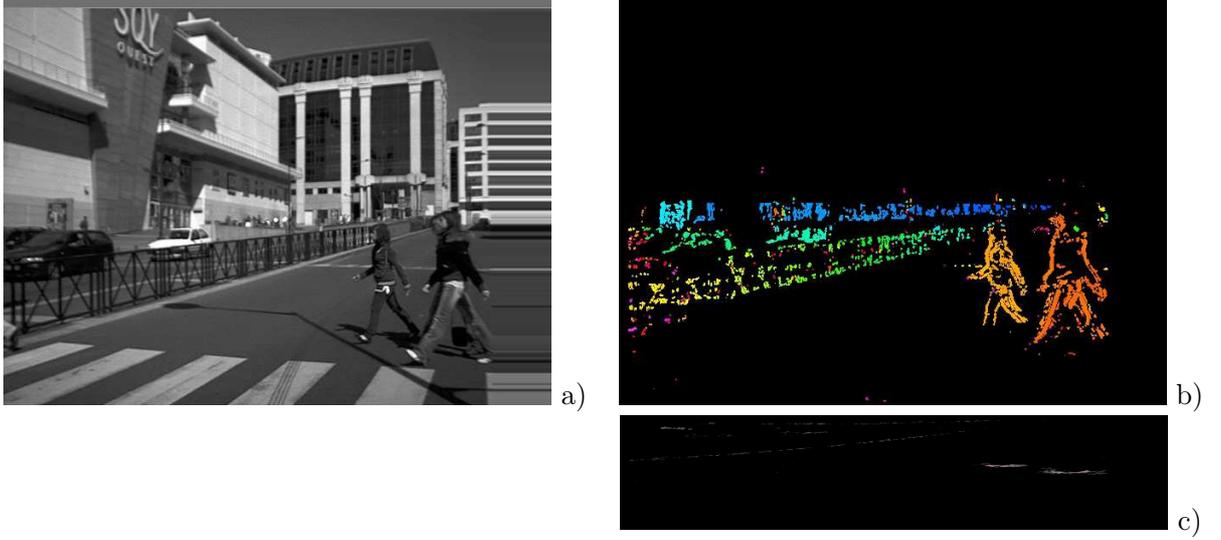


FIG. 2.10 – *Plans verticaux vus en u-disparité. Les piétons apparaissent comme fronto-parallèles, pas la barrière.*

verticales dans l'image v-disparité.

Les portions de plans non fronto-parallèles posent plus de problèmes. En effet, elles constituent un continuum de disparités et d'abscisses qui apparaissent comme un bruit de fond lorsque la scène est observée suivant la direction de \vec{u}_d ou $\vec{\Delta}$. C'est le cas de la barrière présente sur l'image 2.10. Un tel plan vertical est représenté dans R_a par une équation :

$$A_0(X_a - X_s^0) + A_2(Z_a - Z_s^0) + A_3 = 0 \quad (2.23)$$

Son aspect frontal est caractérisé par la dérivée partielle $\frac{\partial Z_a}{\partial X_a}$, qui fournit l'angle que forme le plan avec le vecteur directeur \vec{X}_a , par :

$$\varphi = \arctan\left(\frac{\partial Z_a}{\partial X_a}\right) \quad (2.24)$$

Par ailleurs, nous avons vu que ce plan peut être représenté dans R_Δ par un plan dont les coefficients sont donnés par l'équation 2.20. L'orientation du plan réel est donc liée à son équation dans l'espace de disparité par :

$$\varphi = \arctan\left(\frac{\alpha a_0}{a_3 \cos \theta_s - a_1 \alpha \sin \theta_s}\right) \quad (2.25)$$

Il suffit donc de caractériser le plan dans l'espace de disparité pour obtenir son orientation dans R_a . Celui-ci se projetant sur une droite dans le plan u-disparité, il est facilement détecté par l'utilisation de la transformée de Hough en niveau de gris dans le plan $U_d - \Delta_s$.

Cela conduit finalement à l'algorithme suivant :

1. Projection de l'image de disparité sur le plan u-disparité,
2. Recherche, dans l'espace de Hough associé, des plans verticaux P_i présents dans l'image et calcul de leur orientation réelle φ_i . Cela permet de construire l'ensemble N des plans verticaux non fronto-parallèles : $N = \{P_i/\varphi_i > \varphi_0\}$, φ_0 étant un seuil arbitraire,
3. Calcul d'un histogramme de disparité pour construire l'ensemble F des plans fronto-parallèles,
4. Pour chaque pixel p_j de l'image de disparité, vérification de son appartenance à un éventuel plan à supprimer ou à un plan frontal. p_j est alors supprimé si $p_j \in (N \cap \overline{F})$.

Le résultat de cet algorithme appliqué à l'image de disparité de la figure 2.10 est représenté sur la figure 2.11. La barrière a globalement bien été supprimée, sans que les pixels appartenant à des obstacles n'aient été touchés.



FIG. 2.11 – Filtrage des plans non fronto-parallèles. La barrière a bien été supprimée de l'image de disparité.

2.4 Vue d'ensemble algorithmique

Nous venons de voir un grand nombre de méthodes de calcul de carte de disparité et d'améliorations de celles-ci. Nous avons également proposé deux améliorations adaptées au contexte routier : la corrélation à l'aide d'une double fenêtre, prenant en compte différemment les objets verticaux et horizontaux, et une méthode de suppression des plans verticaux non fronto-parallèles. Récapitulons maintenant les méthodes que nous souhaitons utiliser pour notre système, afin de définir précisément un algorithme de construction des données immédiatement exploitables pour la détection.

2.4.1 Récapitulatif des méthodes retenues

Le choix des pixels à appairer est réalisé en seuillant le résultat de la convolution des images d'entrée avec un masque de gradient horizontal centré $[-1, 0, 1]$. Le seuil est calculé en

fonction de la variance des images d'entrée afin de s'adapter automatiquement aux variations de contraste dans celles-ci au cours du temps. Sa valeur est fixée à 7,5 % de variance moyenne des deux images de la paire. Baisser ce seuil augmente fortement le temps de calcul.

Le signe de ce gradient est par la suite pris en compte afin de réduire le nombre d'hypothèses d'appariement.

La décision sur les appariements est faite par une approche WTA afin que les appariements soient indépendants et que le bruit sur l'image de disparité puisse être considéré comme majoritairement décorrélé. Ce choix se justifie également par la contrainte *temps réel* que nous impose le contexte applicatif.

Le coût de mise en correspondance est calculé en utilisant le critère ZSSD. Après comparaison sur des images issues du simulateur SiVIC, il apparaît comme un compromis très intéressant entre temps de calcul et qualité des appariements. L'expression analytique de ce critère s'écrit :

$$ZSSD(u, \Delta, v) = \sum_{i,j} ((I_d(u+i, v+j) - \overline{I}_d(u, v)) - (I_g(u+\Delta+i, v+j) - \overline{I}_g(u+\Delta, v)))^2 \quad (2.26)$$

$I_d(u, v)$ et $I_g(u, v)$ représentent respectivement l'intensité d'un pixel de coordonnées (u, v) dans les images droite et gauche. $\overline{I}_d(u, v)$ (resp $\overline{I}_g(u, v)$) est la valeur moyenne d'intensité de la fenêtre centrée en (u, v) dans l'image de droite (resp gauche).

La fenêtre de corrélation utilisée est rectangulaire ou inclinée, pour tester les hypothèses d'appariements sur la route ou sur des objets verticaux. Elle mesure 7 pixels de largeur sur 7 pixels de hauteur, de façon à assurer un bon compromis entre les performances d'appariement et le temps de calcul. L'annexe B donne plus de détails sur l'influence de la fenêtre de corrélation.

Les hypothèses d'appariements sont sélectionnées par une approche en deux passes, inspirée de la méthode proposée dans [LA03a]. La première passe consiste à calculer une carte de disparité le plus rapidement possible, en considérant l'assiette du véhicule au repos, puis en extraire le véritable profil longitudinal de la route. Cette démarche permet d'une part de trouver la pente réelle de la fenêtre de corrélation inclinée qui est utilisée au cours de la seconde passe, et d'autre part de diminuer la taille de l'espace de recherche de disparité en définissant trois plages de disparités pour chaque ligne, comme cela est illustré sur l'image 2.12.

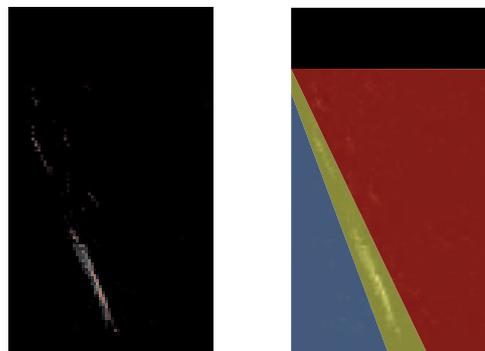


FIG. 2.12 – *Trois plages de disparité correspondant à trois types d'hypothèses d'appariements : sous la route (bleu), route ou obstacle (jaune), obstacle uniquement.*

Ces trois plages de disparité sont définies en autorisant une tolérance τ_r (en centimètres) autour du profil global de route détecté. Les droites frontières sur l'image v-disparité sont donc définies comme les projections suivant \vec{u}_d des plans $y = \tau_r$ et $y = -\tau_r$.

Les plages de disparité donnent également le type d'hypothèse d'appariements correspondant :

1. En dessous du profil de la route (bleu), se trouvent les disparités impossibles à atteindre, ou situées dans des trous. Elles sont rejetées de la liste des hypothèses d'appariements.
2. Autour du profil de la route (jaune), les disparités peuvent correspondre à un obstacle ou à la surface de la route.
3. Au dessus du profil de la route (rouge), seuls des obstacles peuvent être positionnés à de telles disparités.

Dans une seconde passe, la carte de disparité peut donc être calculée avec beaucoup plus de justesse en testant les hypothèses d'appariements autour du profil de la route avec deux types de fenêtre (inclinée ou non) et les hypothèses d'appariements correspondant à des obstacles avec une fenêtre droite. Au final, associée à une approche WTA, cela permet à la fois de calculer plus facilement la disparité des pixels sur la route et de prendre une décision sur la nature (route ou obstacle) de chaque pixel en fonction de la fenêtre de corrélation qui a permis d'obtenir le meilleur score. Cette stratégie permet d'améliorer grandement la qualité de la carte de disparité ainsi calculée en utilisant des fenêtres de corrélation plus grandes. Par ailleurs, le fait qu'en sortie de cet algorithme chaque pixel soit classé comme route ou obstacle permet de séparer ces deux classes dans les images à traiter. D'une part, une image de disparité "obstacle" est créée, qui ne contient que des points obstacles. Cette image est utilisée pour l'étape de détection proprement dite et permet de ne pas perturber cette étape avec des pixels appartenant à la surface de la route. D'autre part, la projection des seuls points "route" sur une image v-disparité permet de fiabiliser la détection du profil précis de la route.

2.4.2 Récapitulatif des filtrages utilisés

Un grand nombre d'erreurs d'appariements est rejeté par l'utilisation de la vérification croisée. Les appariements ambigus sont quant à eux rejetés systématiquement en n'acceptant que les pixels dont le rapport entre les deux meilleurs coûts de mise en correspondance est grand. Si C_1 et C_2 représentent les deux meilleurs scores, un appariement est considéré comme n'étant pas ambiguë si $\frac{C_1 - C_2}{C_1} > 5\%$

Les pixels dont la hauteur réelle est supérieure à la hauteur maximale de détection sont supprimés de la carte de disparité obstacles, afin qu'ils ne perturbent pas inutilement cette détection. La hauteur de filtrage est directement liée à l'application envisagée.

La carte de disparité obstacle est filtrée sur la base de l'hypothèse que les surfaces observées sont localement continues. Cela signifie qu'un pixel d'une disparité donnée doit nécessairement avoir dans son proche voisinage au moins un autre pixel dont la disparité est similaire (différence inférieure à 1). Ce type de filtrage morphologique permet de supprimer en grande partie le bruit impulsionnel créé par les erreurs d'appariements présentes dans l'image.

Les plans non-fronto-parallèles de l'image de disparité obstacles sont supprimés en utilisant l'algorithme présenté plus haut.

2.4.3 Algorithme final de construction des points de mesure

Finalement, le processus de construction des images pour la détection peut être résumé par la figure 2.13.

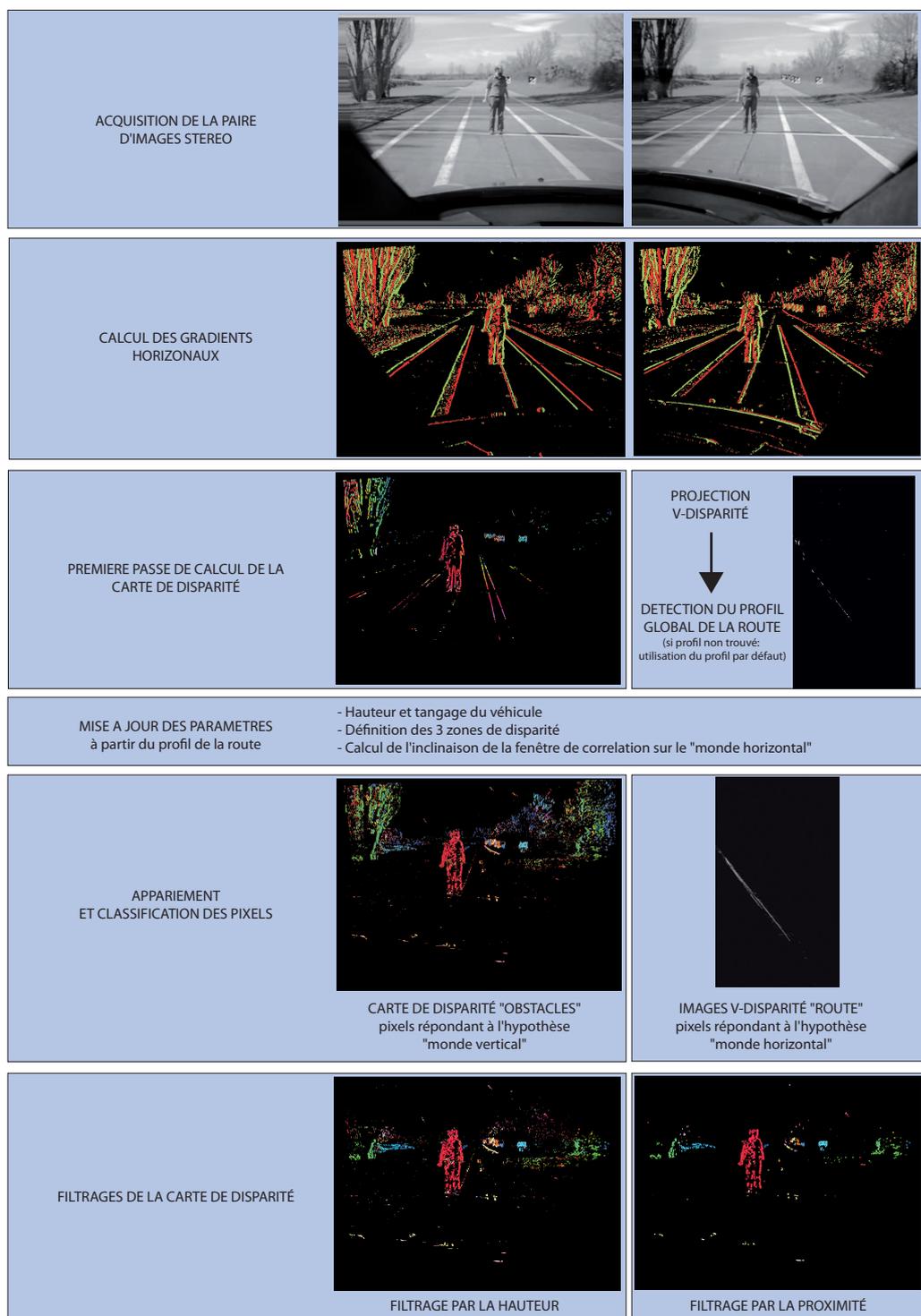


FIG. 2.13 – Vue d'ensemble de l'algorithme de calcul des images de disparité obstacles et v-disparité route.

2.4.4 Remarques qualitatives sur les résultats

Maintenant que nous avons présenté un algorithme complet de construction d'une image de disparité obstacle, observons rapidement la qualité des résultats. Pour cela, nous proposons d'examiner les cartes de disparité construites pour des paires d'images issues de contextes routiers assez variés.

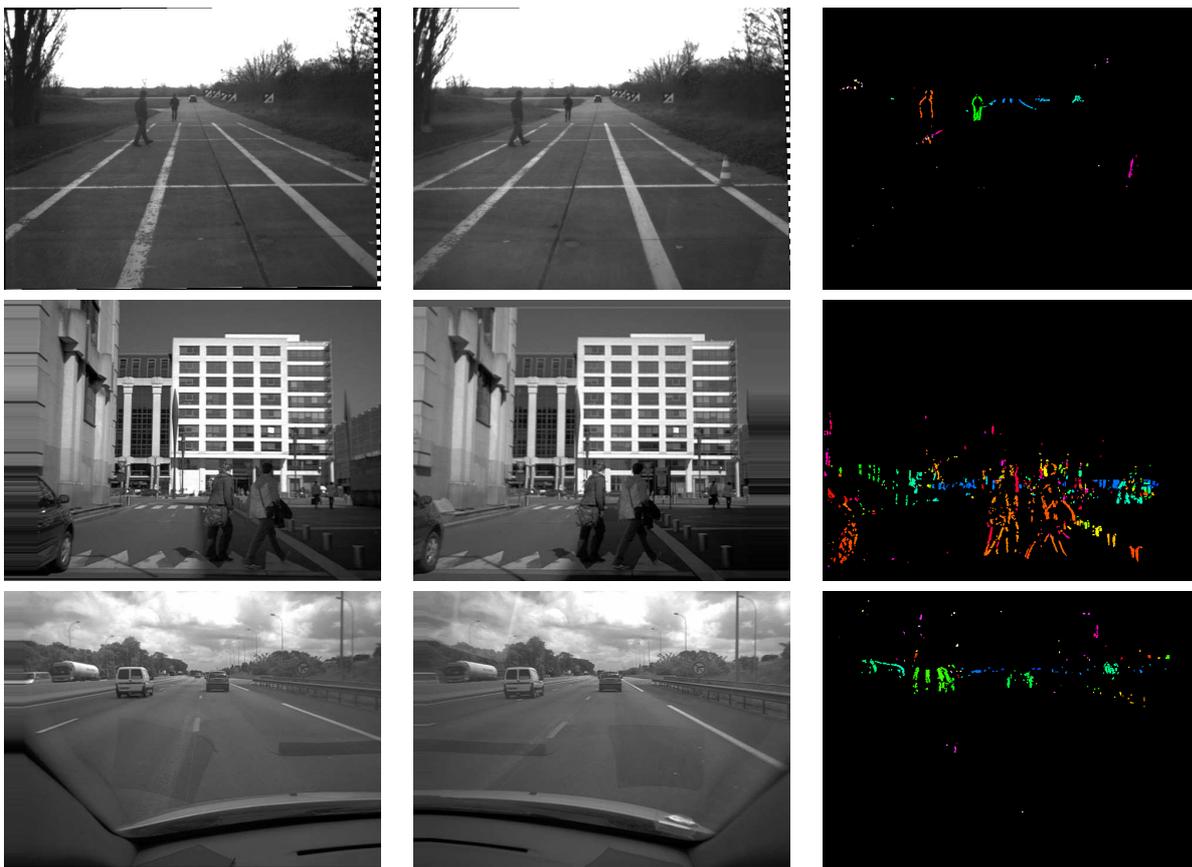


FIG. 2.14 – Paires d'images stéréoscopiques et cartes de disparité obstacles associées.

La figure 2.14 représente une scène simple acquise sur le site de calibrage du LIVIC. Les objets présents (deux piétons, un plot, une voiture et des cibles de calibrage) se retrouvent pour la plupart sur l'image de disparité. On pourra néanmoins remarquer que les jambes du piéton de gauche sont devant un marquage et présentent en conséquence un aspect trop différent selon le point de vue pour être correctement appariées. Le plot n'est qu'à moitié présent car sa partie droite n'est pas dans le champ de vue commun des deux caméras. La première mire ainsi que la voiture sont bien appariées. En revanche, la route n'est pas plane, ce qui fait apparaître ses bords lointains dans la carte de disparité obstacle. En dehors des obstacles, nous pouvons noter la présence de quelques erreurs. En particulier derrière le piéton de droite, un marquage semi-occulté est faussement apparié, créant des erreurs corrélées. Les branchages, texture très aléatoire, sont également source d'erreurs.

La seconde paire d'images présente une scène urbaine plus complexe. On pourra remarquer

en particulier la présence d'un grand nombre de motifs répétitifs, sources d'un bruit caractérisé par des pixels de disparité élevée. Le caractère local de la méthode d'appariement utilisée ne permet pas de prendre en compte ce type d'erreurs, à moins d'utiliser des fenêtres de corrélation démesurément grandes. En outre, compte tenu de la présence d'un bâtiment très texturé en arrière plan, les contours des piétons ont un aspect sensiblement différent entre les deux images, illustrant à nouveau la difficulté de mise en correspondance en présence de semi-occultations.

La troisième paire illustre, quant à elle, le comportement de l'algorithme dans un contexte autoroutier. Les différents objets présents dans la scène (véhicules et panneaux) se retrouvent correctement sur la carte de disparité. En revanche, les pixels appariés se trouvent uniquement sur les contours verticaux de ces objets. Ainsi, les voitures apparaissent comme séparées en plusieurs objets fins et verticaux plutôt que comme des objets compacts. Cela pourra induire une sur-segmentation dans la suite du processus. En dehors de ce point noir, il est intéressant de noter que la différence d'illumination entre les deux images ne perturbe globalement pas l'appariement. C'est le centrage sur la moyenne du critère ZSSD qui assure cette robustesse.

2.4.5 Exploitation du résultat : les points de mesure "obstacles"

La carte de disparité obstacle ainsi créée ne doit pas être vue uniquement comme une image. Elle représente en effet une liste de points que nous appellerons par la suite "points obstacles". Ceux-ci sont véritablement les points de mesure issus du capteur stéréoscopique et sont directement utilisables pour la détection. Ils peuvent indifféremment être représentés dans l'espace de disparité E_{Δ} ou dans l'espace réel E_a . La figure 2.15 représente l'ensemble des points de l'image de disparité de la figure 2.13, représentés dans ces deux espaces.

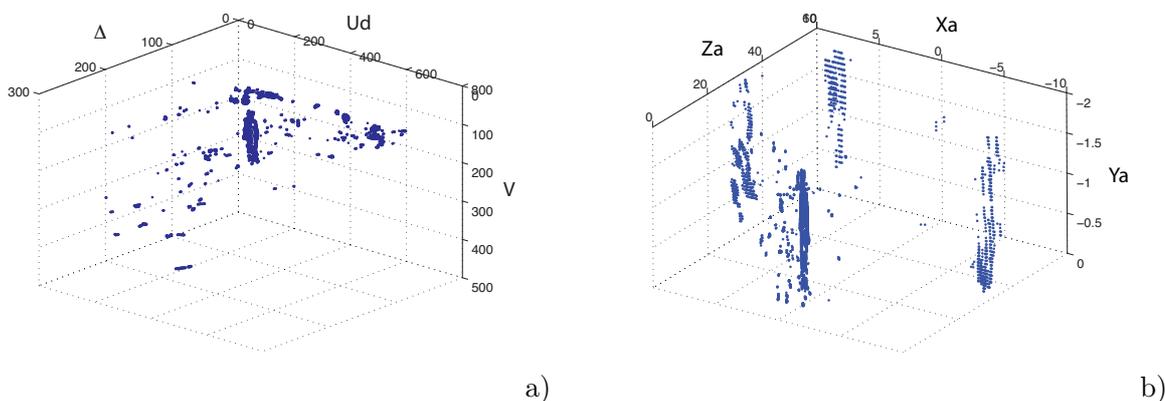


FIG. 2.15 – Représentation d'une carte de disparité comme un ensemble de points représentés dans R_{Δ} et dans R_a .

Les deux représentations sont équivalentes mais leur exploitation diffère.

Le choix de l'un ou l'autre de ces espaces dépend de l'utilisation qui doit être faite des points obstacles. Cette double représentation permet donc d'utiliser des méthodes de traitement très variées. La représentation image présente l'intérêt de permettre de travailler directement avec

les données issues du capteur. Elle conserve la connexité des points obstacles ainsi que leur précision. En revanche, les proportions des objets ne sont pas indépendantes de leur distance au capteur. Toutes les grandeurs métriques utilisées (taille des obstacles par exemple) doivent donc être converties en fonction de la distance.

A l'inverse, la représentation cartésienne permet d'exploiter directement ce type de grandeurs. Elle est cependant construite en transformant les informations images issues de la carte de disparité, avec leur précision intrinsèque. Au loin, les points obstacles sont donc entachés d'une erreur très importante. Cette représentation est de plus très dépendante de la discrétisation réalisée lors de la transformation et deux pixels connexes dans l'image originale peuvent apparaître séparés de plusieurs mètres, une erreur d'un demi-pixel en disparité étant couramment admise.

2.5 Conclusion

Nous avons présenté au cours de ce chapitre le principe de la stéréovision telle qu'elle sera utilisée dans la suite. En particulier, nous avons retenu parmi les nombreuses méthodes existantes celles qui s'adaptent le mieux au contexte routier auquel le système que nous proposons est destiné. De plus, nous avons proposé quelques améliorations des méthodes existantes ainsi qu'un algorithme complet de création des données utiles à la détection d'obstacles. La quantité de bruit présent dans ces images a ainsi été fortement diminuée, sans pour autant être complètement supprimée. En revanche, il reste encore des faiblesses à l'approche proposée. En particulier, la faible densité de la carte de disparité et le fait que les pixels appariés se limitent aux contours verticaux est un point faible qui aura de fortes implications sur l'étape de détection. En effet, un objet présent dans la scène apparaissant vide en son centre a un aspect similaire à deux objets côtes à côtes.

Nous allons maintenant voir comment ce type de points de mesure peut être utilisé au sein du système complet de détection d'obstacles que nous proposons, en les traitant dans leur ensemble (chapitre 3) ou localement (chapitre 4).

CRÉATION D'HYPOTHÈSES DE DÉTECTION

Sommaire

3.1	Définition des zones d'intérêt	56
3.2	Création d'hypothèses de détection par stéréovision	58
3.2.1	Création des points de mesure obstacles	58
3.2.2	Segmentation de l'espace de disparité	58
3.2.3	Intérêt de la suppression des plans non fronto-parallèles	59
3.2.4	Influence de la discrétisation des histogrammes	60
3.2.5	Construction des zones d'intérêt	60
3.2.6	Respect de la contrainte temps-réel	61
3.2.7	Résultats	62
3.3	Création d'hypothèses de détection par télémétrie laser	64
3.3.1	Modèle d'un télémètre laser à balayage 2D ou lidar	64
3.3.2	Données fournies par le lidar	65
3.3.3	Filtrage des données brutes	66
3.3.4	Construction des cibles lidar	67
3.3.5	Suivi des détections lidar	68
3.3.6	Construction des zones d'intérêt	69
3.3.7	Résultats	69
3.4	Création d'hypothèses de détection par identification optique	70
3.4.1	Présentation du capteur	70
3.4.2	Algorithme de traitement du récepteur	72
3.4.3	Construction des zones d'intérêt	74
3.4.4	Résultats	77
3.5	Perspectives : Association des différents capteurs	78
3.5.1	Combinaison de deux capteurs	78
3.5.2	Combinaison de plus de deux capteurs	80
3.5.3	Zones d'intérêt résultantes	80
3.6	Conclusion	80

LORS de sa tâche de conduite, le conducteur observe l'environnement dans lequel son véhicule évolue. Son regard est susceptible d'être furtivement attiré par les différents objets de la scène présentant un intérêt ou un risque éventuel. A cette phase de détection, succède une étape d'analyse plus approfondie au cours de laquelle le conducteur va vérifier qu'il "a bien vu" et évaluer le risque potentiel que représente chacun de ces objets. Ce mode de perception, dit bottom-up (de bas en haut), est l'une des stratégies de focalisation de l'attention utilisée par les primates [IKN98].

D'une façon similaire, un véhicule "intelligent" devrait être à même de percevoir rapidement les hypothétiques objets sur lesquels il va devoir mobiliser son attention, en l'occurrence sa puissance de calcul. Partant de cette idée, nous avons proposé au chapitre 1 une méthodologie de détection des objets qui suppose de localiser dans un premiers temps des régions d'intérêt, pour ensuite procéder à une analyse plus approfondie de ces zones par stéréovision. Lors de cette première phase de recherche de candidats, il est nécessaire de n'exclure aucune hypothèse, quitte à ce que certaines d'entre elles soient ensuite infirmées. En revanche, la seconde phase se doit d'être particulièrement discriminante, pour ne valider que l'existence d'objets réels.

Nous allons dans ce chapitre présenter l'étape de création d'hypothèses de détection. Dans un premier temps, nous allons définir la sortie attendue de celle-ci : une liste de volumes d'intérêt dans l'espace de disparité E_{Δ} . La création d'hypothèses sera ensuite abordée à travers les trois capteurs considérés dans cette thèse : un capteur stéréoscopique, qui fournit une information tridimensionnelle sur une grande plage de portée, un télémètre laser, précis et rapide mais n'observant qu'un plan et un capteur coopératif d'identification optique, permettant une détection robuste à grande distance et éventuellement enrichie d'informations provenant de télécommunications. Pour autant, ces choix ne restreignent pas la généralité de l'approche vis à vis de l'utilisation d'autres capteurs.

3.1 Définition des zones d'intérêt

La forme de régions d'intérêt la plus couramment utilisée dans les algorithmes de traitement d'image est un rectangle dont les côtés sont parallèles aux bords de l'image. Cette géométrie, très intuitive, est très largement exploitée par la communauté du traitement d'image, notamment dans des bibliothèques célèbres comme OpenCV³⁴. Cette popularité provient de la grande simplicité d'utilisation informatique de telles régions d'intérêt puisque le parcours des pixels s'effectue de la même façon que dans une image entière. En l'occurrence, la méthode de construction d'une carte de disparité présentée au chapitre précédent est compatible avec cette forme. Par ailleurs, de plus en plus de modèles de caméras de vision industrielle permettent désormais d'adresser dynamiquement de façon matérielle ce type de zones d'intérêt. Par exemple, dans le cadre du projet de recherche européen TRACKSS³⁵ sur les capteurs coopératifs, l'équipementier automobile BOSCH développe ce type de caméras.

L'espace de travail que nous considérons dans notre approche (E_{Δ}) est tri-dimensionnel. Il est donc nécessaire d'étendre la géométrie rectangulaire afin de l'adapter à notre problématique, en y incluant une notion de profondeur. Nous proposons pour cela un parallépipède rectangle

³⁴INTEL Open Computer Vision Library

³⁵Technologies for Road Advanced Cooperation Knowledge Sharing Sensors

fronto-parallèle aux caméras, défini par :

- un intervalle horizontal $[u_{min}, u_{max}]$
- un intervalle vertical $[v_{min}, v_{max}]$
- un intervalle de disparité $[\Delta_{min}, \Delta_{max}]$

Cette définition des zones d'intérêt permet, d'une part de limiter la charge calculatoire en réduisant la plage de disparités à explorer, d'autre part de séparer des objets connexes sur des images en 2 dimensions alors qu'ils sont situés à des profondeurs différentes.

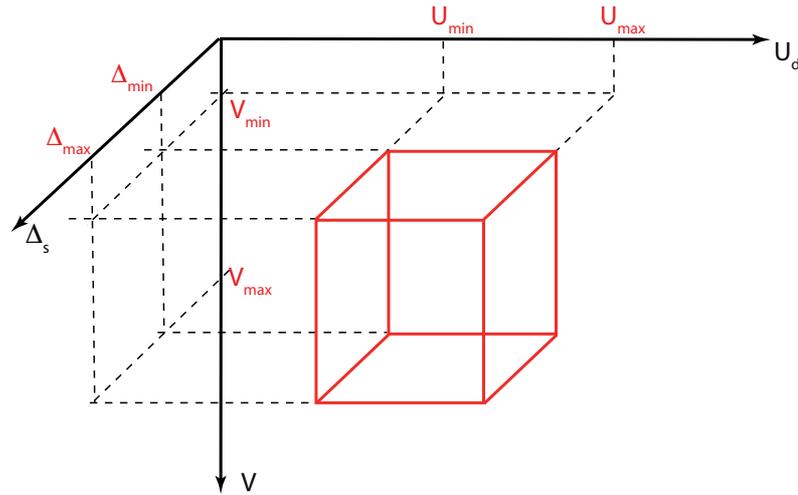


FIG. 3.1 – Définition des zones d'intérêt utilisées pour la suite.

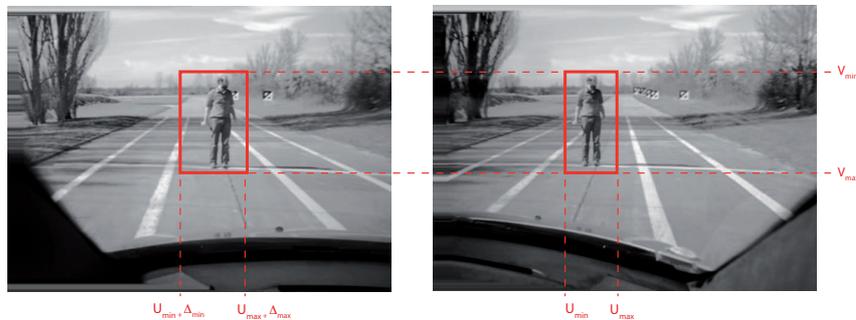


FIG. 3.2 – Représentation d'une zone d'intérêt dans une paire d'images stéréoscopiques.

Nous allons maintenant décliner notre méthodologie de création d'hypothèses pour les trois capteurs considérés dans cette thèse, en détaillant le processus de construction de zones d'intérêt pour chacun d'entre eux.

3.2 Création d'hypothèses de détection par stéréovision

La première approche que nous allons aborder est la création d'hypothèses de détection par stéréovision. Pour cette étape, la problématique consiste à déterminer les régions d'images comportant un grand nombre de points de mesure correspondant à des obstacles, par segmentation de l'espace de disparité.

3.2.1 Création des points de mesure obstacles

L'algorithme de détection débute par la construction des points de mesure correspondant à des obstacles, que nous appellerons "points obstacles". Pour des explications détaillées sur ce processus, le lecteur est invité à se reporter au chapitre 2.

3.2.2 Segmentation de l'espace de disparité

Pour détecter les obstacles potentiels, les points obstacles doivent être regroupés en classes. Afin de respecter un temps de calcul compatible avec les contraintes du contexte automobile, nous avons exclu les techniques d'optimisations globale sur toute l'image ou les algorithmes de classification récursifs, type K-Mean. De plus, la présence parmi les points obstacles d'un nombre important d'erreurs incite à procéder par des méthodes d'accumulation, pour éviter de traiter chaque pixel individuellement et favoriser les structures cohérentes.

Pour répondre à ces contraintes, l'idée que nous proposons consiste à construire de façon récursive des histogrammes à segmenter, afin de toujours se rapporter à un problème à une dimension. La possibilité de travailler directement sur des données "image" prend ici son intérêt. En effet, les accumulations sont toujours réalisées selon les axes principaux de R_Δ . Il suffit donc de parcourir des lignes et colonnes, plutôt qu'une liste complète de points à chaque opération. En outre, cette technique prend implicitement en compte la proximité des points de mesures. Cette propriété se révèle importante à grande distance. Deux pixels voisins dans E_Δ peuvent en effet être séparés de plusieurs mètres après le changement de système de coordonnées pour R_a , en raison de la discrétisation de l'espace de disparité.

Le processus de segmentation est récapitulé sur la figure 3.3. La construction des volumes d'intérêt est réalisée en séparant les pixels selon leur profondeur, puis leur positionnement latéral, puis leur hauteur. Pour cela, une image u-disparité est construite à partir de la carte de disparité. Une accumulation le long de ses lignes fournit ensuite l'histogramme de disparité de la scène. Celui-ci est seuillé pour séparer des classes. Pour chaque tranche de disparité \mathcal{C}_i ainsi créée, la portion de l'image u-disparité correspondante est ensuite projetée le long des colonnes. Du seuillage de ce nouvel histogramme naît un découpage horizontal de l'espace.

A ce stade du traitement, le plan $(u - \Delta)$ est partitionné en rectangles $\mathcal{C}_{i,j}$, définis par les couples $[u_{min}^j, u_{max}^j]$ et $[\Delta_{min}^i, \Delta_{max}^i]$. Il reste donc à les positionner verticalement pour achever de segmenter E_Δ . Pour cela, un histogramme de la variable v est construit pour chaque classe $\mathcal{C}_{i,j}$, à partir des points obstacles qu'elle contient. A nouveau, un seuillage permet d'effectuer le découpage, donnant des intervalles de type $[v_{min}^k, v_{max}^k]$. En accord avec le contexte routier, v_{max}^k est recalculé de façon à ce que la base de la face avant du volume repose sur le sol :

$$v_{max}^k = v_0 - \frac{\alpha_v}{\alpha_u} \frac{Y_s^0 \Delta_{max}^i}{b_s \cos \theta_s} - \alpha_v \tan \theta_s \quad (3.1)$$

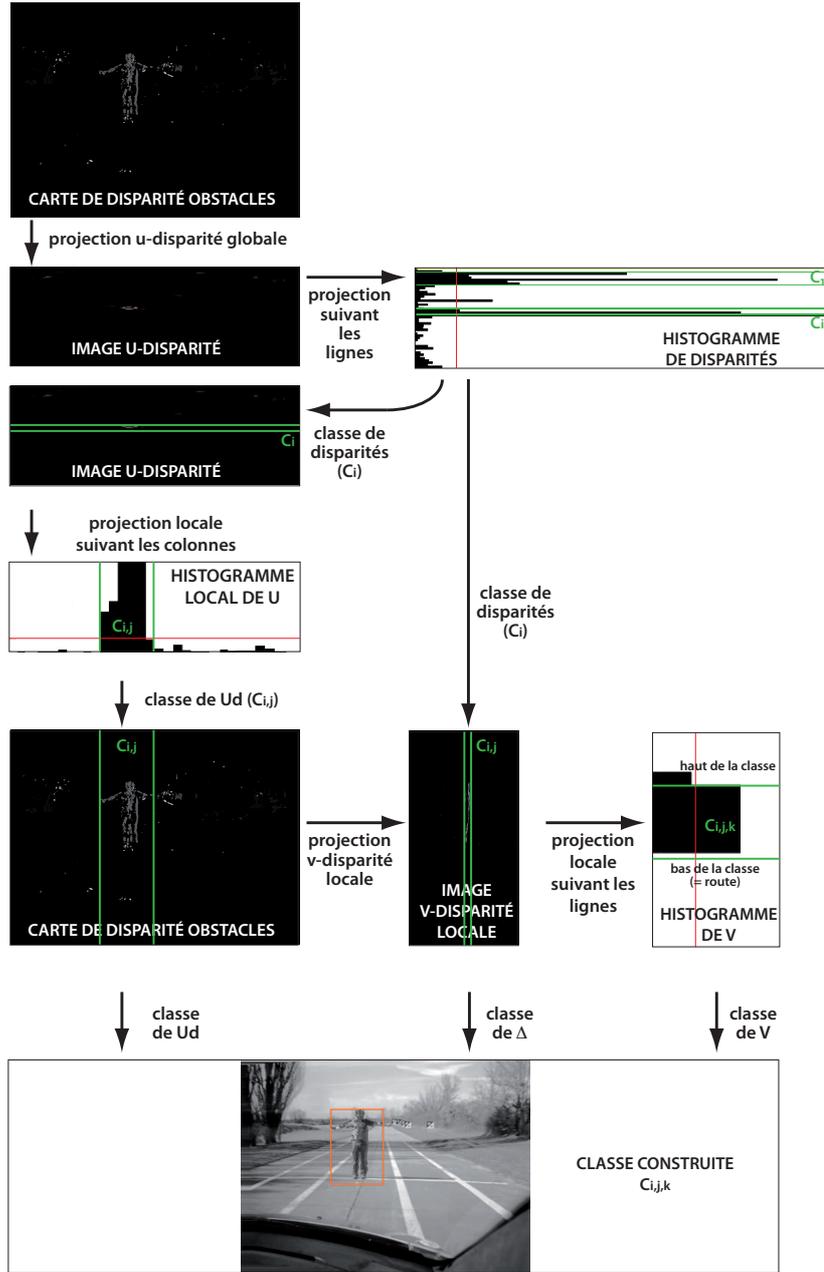


FIG. 3.3 – *Algorithme de segmentation de l'espace de disparité : de l'image de disparité à la construction d'une classe obstacle.*

La séparation est alors terminée, car chaque classe $\mathcal{C}_{i,j,k}$, dispose de son triplé d'intervalles :

$$\mathcal{C}_{i,j,k} = \{[u_{min}^j, u_{max}^j], [\Delta_{min}^i, \Delta_{max}^i], [v_{min}^k, v_{max}^k]\}. \quad (3.2)$$

3.2.3 Intérêt de la suppression des plans non fronto-parallèles

L'algorithme de détection par stéréovision que nous proposons suppose que les obstacles soient fronto-parallèles, puisque la recherche se fait en premier lieu le long de l'axe $\vec{\Delta}$. Si tel n'est

pas le cas, et qu'un objet présentant sur sa longueur un continuum de disparités se trouve dans la scène, celui-ci provoquera une large classe de disparités lors de la première segmentation, ou une large classe d'abscisses lors de la seconde. Il se peut alors qu'à la fin du processus, la séparation soit incomplète. C'est pour éviter ce type d'erreurs que nous avons proposé au paragraphe 2.3.4 une méthode de suppression des pixels appartenant à des plans non-fronto-parallèles, permettant de filtrer les données afin qu'elles soient conformes aux hypothèses de l'algorithme de détection. L'effet sur le résultat de la détection est visible sur la figure 3.4.

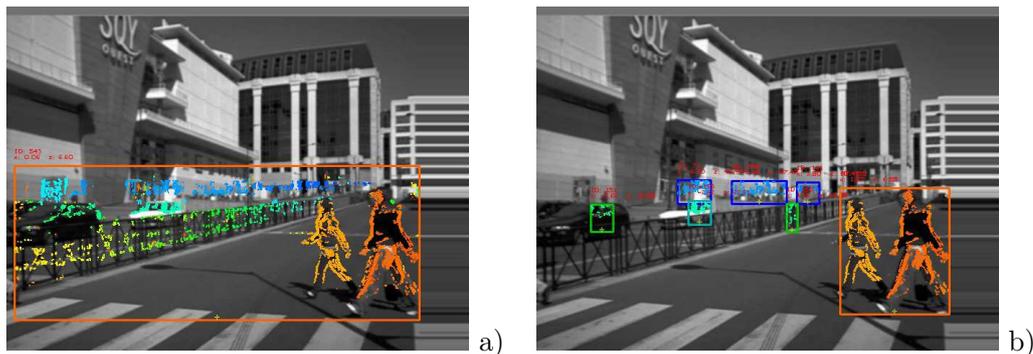


FIG. 3.4 – a) La présence d'une barrière oblique forme un continuum de disparité très préjudiciable à la détection. b) Sa suppression facilite la segmentation.

3.2.4 Influence de la discrétisation des histogrammes

La taille des classes d'accumulation des histogrammes (que nous nommerons désormais *bins* de leur appellation anglaise, afin d'éviter toute confusion avec les classes de détections) joue un rôle important sur le résultat de la détection. Ces *bins* constituent en effet le moyen de regrouper dans une même classe des pixels plus ou moins voisins. Ainsi, le choix de *bins* larges peut conduire à sous-segmenter l'espace, mais peut également permettre de rassembler tous les pixels d'un même obstacle (par exemple une voiture) dans un même objet, malgré l'absence de contours horizontaux. A l'inverse, des *bins* trop fins conduisent à une sur-segmentation qui rend l'interprétation et l'exploitation des résultats difficile. Comme souvent en traitement d'images, un équilibre doit être trouvé entre la précision, la stabilité et le type d'objets recherchés.

La figure 3.5 illustre l'influence de la taille des *bins* sur le résultat de la segmentation. Il apparaît que pour détecter des véhicules sur autoroute, il est préférable d'utiliser des *bins* larges, qui compensent l'absence de points obstacles sur les contours horizontaux.

3.2.5 Construction des zones d'intérêt

A partir des différentes classes issues de l'algorithme de segmentation, la construction des régions d'intérêt est immédiate, la géométrie des classes étant identique à la définition que nous avons proposée pour les zones d'intérêt. Cela dit, si les *bins* mesurent plus d'un pixel de largeur, il n'est pas impossible que le volume résultant d'une détection soit légèrement surdimensionné par rapport aux points obstacles qu'il contient. Il est donc indispensable de terminer la segmentation par la mise à jour des bords des volumes pour qu'ils englobent strictement leurs points de mesure. Cet ajustement est réalisé en considérant les valeurs extrêmes des coordonnées dans R_{Δ} des points obstacles contenus dans la zone d'intérêt.

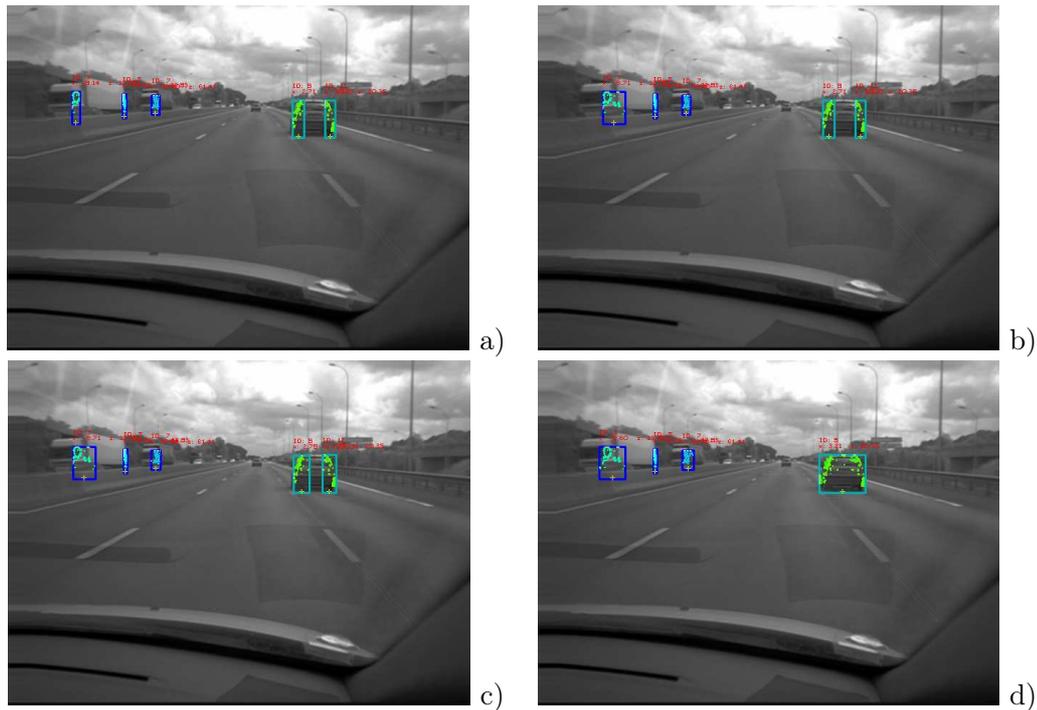


FIG. 3.5 – Influence de la taille des bins sur le résultat de la segmentation. Bins de 5 (a), 10 (b), 15 (c) et 20 (d) pixels pour l'histogramme en u . La portée de détection est ici limitée à 30 mètres dans le cadre du projet DO30.

3.2.6 Respect de la contrainte temps-réel

Dans le cas où une image de disparité pleine résolution peut être calculée en temps réel, l'approche proposée s'applique directement. Cependant, compte tenu des puissances de calcul usuelles, cela est rarement possible. La détection doit alors être effectuée sur des images de plus faible résolution, afin de conserver un temps de calcul raisonnable. La précision et la robustesse de la détection ne sont en aucune manière des éléments essentiels à ce stade de l'algorithme, car elles sont assurées par l'étape de caractérisation et de confirmation des hypothèses. La vitesse d'exécution et un taux de détection élevé sont en revanche primordiaux.

Si l'on retient cette stratégie, la taille des images d'entrée doit être réduite par sous-échantillonnage. Pour une réduction d'un facteur N (entier) d'une image de dimension $n_{lignes} \times n_{colonnes}$, l'image obtenue comporte $\mathbb{E}(n_{lignes}/N) \times \mathbb{E}(n_{colonnes}/N)$ pixels, \mathbb{E} représentant la partie entière. La réduction pourrait être effectuée par simple ré-échantillonnage en gardant un pixel sur N dans l'image d'origine. Cependant cette méthode provoque un repliement de spectre (ou *aliasing*) puisqu'elle ne satisfait pas la condition d'échantillonnage posée par Shanon [Sha49] lorsque l'image contient des fréquences élevées. Il est alors nécessaire de filtrer l'image d'origine par un filtre passe bas avant échantillonnage pour corriger ce problème. Un filtre gaussien utilisé pour réduire le bruit des images (Annexe A) peut remplir ce rôle. Dans le cas contraire, la réduction des images par moyennage supprime cette limitation : un pixel de l'image réduite est la moyenne des



FIG. 3.6 – Réduction d'un facteur 2 d'une paire d'images stéréoscopique par moyennage, et image de disparité obstacle résultante.

intensités de $N * N$ pixels correspondants dans l'image originale :

$$Im_{reduite}(u, v) = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (Im_{original}((u + i) * N, (v + j) * N))}{N^2} \quad (3.3)$$

Cette formulation prend intrinsèquement en compte la contrainte de Shanon puisque le calcul de moyenne tient lieu de filtrage passe bas.

La figure 3.6 illustre l'effet d'une telle réduction d'un facteur 2 sur une paire d'images stéréoscopiques, et montre la carte de disparité obstacles construite à partir des images réduites. C'est sur cette image que s'applique alors l'algorithme de détection.

Réduire la résolution s'avère particulièrement efficace pour économiser du temps de calcul. Cela ne diminue pas seulement le nombre de pixels à apparier mais également la plage de disparité explorée. Ainsi le traitement complet, qui comporte le calcul des points de mesure et la segmentation de E_{Δ} , nécessite en moyenne 343 ms par image³⁶ lorsqu'il est appliqué à une séquence d'images VGA (séquence "mixte" présentée au chapitre 5). Traiter ces images au quart de leur résolution native ne nécessite en moyenne que 71 ms.

3.2.7 Résultats

La figure 3.7 montre les résultats obtenus en appliquant la méthode de segmentation que nous proposons sur les images de disparité présentées au chapitre précédent. La méthode est globalement peu sensible aux erreurs présentes parmi les points obstacles si celles-ci sont indépendantes. Il apparaît en revanche que dans le cas de la scène urbaine, les motifs répétitifs des

³⁶temps mesuré sur un ordinateur portable équipé d'un processeur core2duo cadencé à 2,0 GHz

immeubles induisent des erreurs corrélées auxquelles l'algorithme est sensible. Des hypothèses fausses de détections sont alors générées et devront être supprimées lors de la phase de validation présentée au chapitre 4. Enfin, il est notable qu'aucune détection évidente, telle qu'un gros objet proche, n'est manquée lors de cette phase de création d'hypothèses.

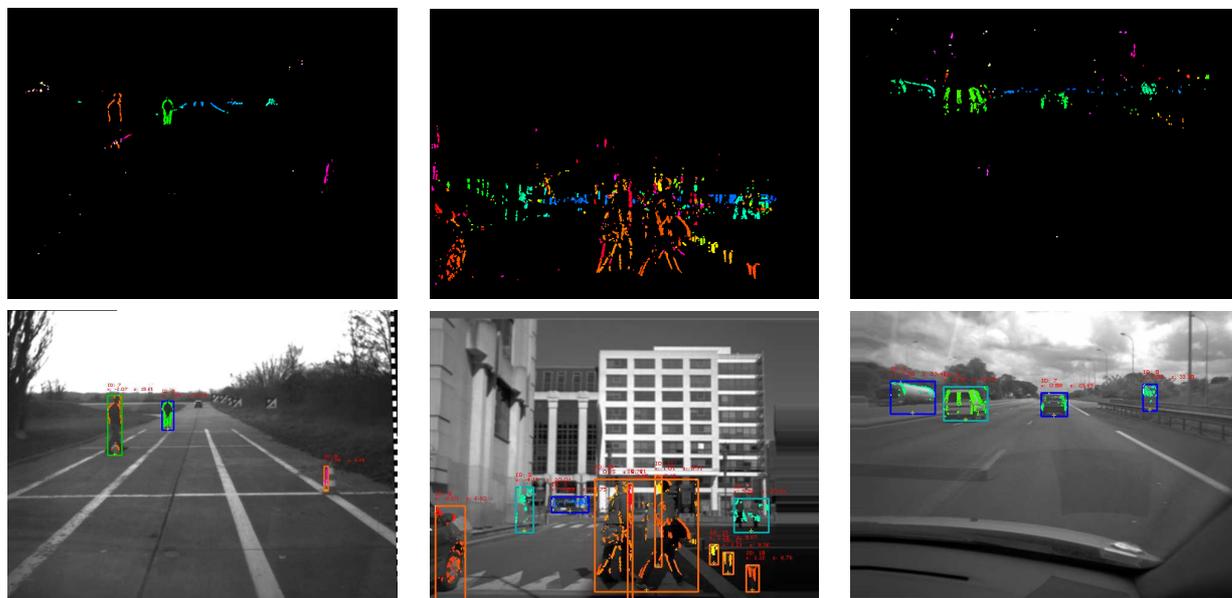


FIG. 3.7 – *Hypothèses de détection fournies par la stéréovision.*

3.3 Création d'hypothèses de détection par télémétrie laser

Nous venons de décrire une méthode de création de zones d'intérêt à partir d'images stéréoscopiques. Dans le cas où notre véhicule est également équipé d'un télémètre laser à balayage (ou lidar), celui-ci peut être utilisé pour créer des hypothèses de détection, et donc pour construire de façon rapide et précise des zones d'intérêt pour la stéréovision. Dans cette section, nous allons présenter le fonctionnement d'un tel télémètre, puis rappeler la méthode de détection d'obstacles proposée par Gruyer *et al.* [LRGA05] et que nous avons utilisée. Nous verrons enfin comment ces détections sont converties en zones d'intérêt.

3.3.1 Modèle d'un télémètre laser à balayage 2D ou lidar

Présentation du capteur

Pour réaliser une mesure de distance, un télémètre laser à temps de vol émet une impulsion lumineuse et estime le temps mis par la lumière pour revenir vers le capteur après avoir heurté un obstacle. Nous appellerons "impact" chaque point de mesure ainsi obtenu.

A titre d'exemple pour notre application, c'est un laser à balayage 2D que nous utilisons. Celui-ci balaye une portion de plan de l'espace de façon discrète et mesure une distance à l'objet le plus proche pour chaque angle d'incidence. Il fournit donc pour chaque balayage une carte de distance polaire sur son plan de mesure. D'un point de vue technologique, le balayage est réalisé en dirigeant l'émetteur laser vers un miroir rotatif actionné par un moteur pas à pas. Ce type de système induit une imprécision de mesure longitudinale et angulaire car le faisceau laser est divergeant. Un impact n'est donc pas ponctuel, mais caractérisé dans le plan de balayage par un secteur de couronne. Cependant, comme cela est rappelé dans [Gru99], il est courant de modéliser cette incertitude par une ellipse englobante, définie par une matrice de variance-covariance, plus facilement exploitable.

Pour pouvoir exploiter le résultat du traitement de ces données laser dans notre application

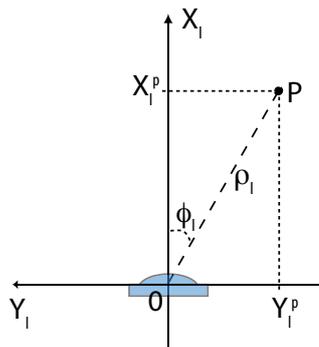


FIG. 3.8 – Repères polaire et cartésien liés au télémètre laser.

de détection d'obstacles, il doit être représenté dans un repère cartésien. Sur la figure 3.8 sont représentés un repère polaire $R_l^p(\rho_l, \phi_l)$ et un repère cartésien $R_l^c(X_l, Y_l)$ tout deux liés au lidar. La transformation permettant de passer de R_l^p à R_l^c est donnée par le système :

$$\begin{cases} X_l &= \rho_l \cos \phi_l \\ Y_l &= \rho_l \sin \phi_l \end{cases} \quad (3.4)$$

Positionnement du capteur dans l'espace

Tout comme les capteurs de vision, les données issues du télémètre doivent pouvoir être exprimées dans un repère absolu, lié au véhicule instrumenté. Le capteur étant positionné en rotation et en translation simples par rapport à ce repère (figure 3.9), le changement de repère de R_l à R_a peut être représenté sous la forme :

$$P_{la} = R_a^l T_l R_l^l \quad (3.5)$$

R_l^l représente une composition de rotations de $\pi/2$ autour des axes Y_l puis Z_l . T_l est une translation de X_l^0, Y_l^0, Z_l^0 . R_a^l est une rotation autour de l'axe X_a correspondant au tangage θ_l du télémètre relativement à R_a .

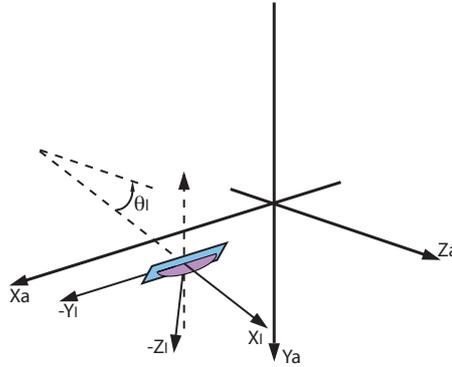


FIG. 3.9 – Positionnement du télémètre dans un repère absolu.

Le repère choisi correspond à un repère couramment utilisé pour les télémètres.

Expression des données lidar dans R_a

Le système d'équations lié à l'équation matricielle ci-dessus permet finalement d'exprimer les données du télémètre laser dans R_a :

$$\begin{cases} X_a &= X_l^0 - Y_l \\ Y_a &= Y_l^0 + X_l \sin \theta_l \\ Z_a &= Z_l^0 + X_l \cos \theta_l \end{cases} \quad (3.6)$$

3.3.2 Données fournies par le lidar

Le lidar fournit, pour chaque mesure, une liste d'impacts représentant la distance à l'objet le plus proche rencontré par le rayon du laser pour toutes les positions angulaires de celui-ci. Ces impacts peuvent être représentés dans le repère absolu R_a lié au châssis du véhicule. Pour exploiter les données issues du lidar en utilisant la stéréovision, il est aussi essentiel de pouvoir les représenter dans le repère R_Δ lié aux images stéréoscopiques. En utilisant les équations (2.16)

et (3.6), la projection des données lidar dans l'espace de disparité se traduit par le système :

$$\begin{cases} u_d &= u_0 + \alpha_u \frac{X_l^0 - Y_l - b_s/2}{(Y_l^0 - Y_s^0 + X_l \sin \theta_l) \sin \theta_s + (Z_l^0 - Z_s^0 + X_l \cos \theta_l) \cos \theta_s} \\ v &= v_0 + \alpha_v \frac{(Y_l^0 - Y_s^0 + X_l \sin \theta_l) \cos \theta_s - (Z_l^0 - Z_s^0 + X_l \cos \theta_l) \sin \theta_s}{(Y_l^0 - Y_s^0 + X_l \sin \theta_l) \sin \theta_s + (Z_l^0 - Z_s^0 + X_l \cos \theta_l) \cos \theta_s} \\ \Delta_s &= \alpha_u \frac{b_s}{(Y_l^0 - Y_s^0 + X_l \sin \theta_l) \sin \theta_s + (Z_l^0 - Z_s^0 + X_l \cos \theta_l) \cos \theta_s} \end{cases} \quad (3.7)$$

Les données lidar peuvent donc être représentées indifféremment dans R_a par une vue de dessus du plan $Y_a = 0$ (figure 3.10-a) ou projetées dans l'image de droite de la paire stéréoscopique (figure 3.10-b).

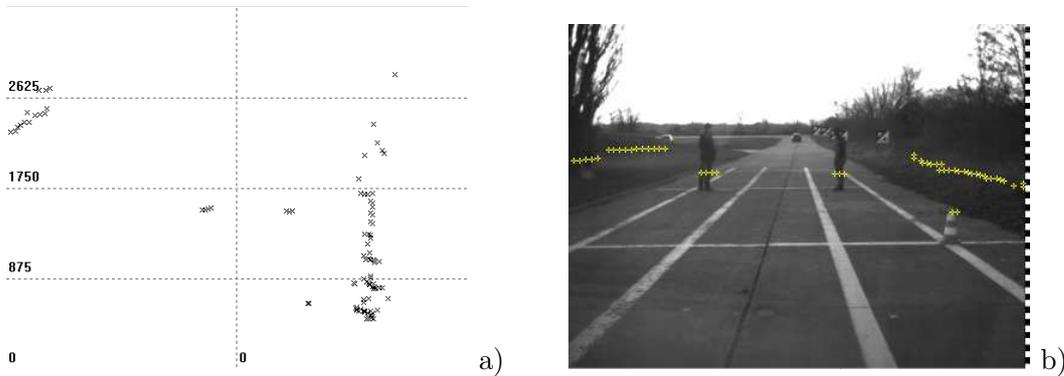


FIG. 3.10 – Impacts d'une trame lidar représentés en vue de dessus dans le repère cartésien R_t (a) et projetés dans l'image de droite de la paire stéréoscopique (b).

3.3.3 Filtrage des données brutes

Avant de les utiliser pour la détection d'obstacles, les impacts laser peuvent être filtrés de façon à ne conserver que ceux qui présentent un réel intérêt. L'objectif d'un tel filtrage est la limitation du temps de calcul nécessaire au traitement ultérieur des cibles laser et la suppression d'un grand nombre de fausses détections.

Filtrage latéral

Tout d'abord, l'utilisation d'une zone de filtrage latérale délimitant la zone de risque potentielle permet de supprimer des impacts laser inutiles. La définition de cette zone dépend fortement de l'application envisagée pour la détection d'obstacles. Elle peut alors être estimée dynamiquement à l'aide d'un algorithme de détection de marquages [DLLC05] ou estimée à l'aide des capteurs proprioceptifs du véhicule [LRA05]. En l'absence de l'un ou l'autre de ces systèmes, une zone de filtrage statique, correspondant au champ de vue du système stéréoscopique peut être utilisée avantageusement sans pour autant limiter le champ des applications du système.

Filtrage par la hauteur

D'autre part, un inconvénient majeur de l'utilisation d'un lidar réside dans sa grande sensibilité aux variations d'assiette du véhicule ou de la pente de la route. En cas de fort tangage ou

d'une géométrie de route non plane, le plan laser peut en effet couper la surface de la chaussée, voyant celle-ci comme un obstacle. A l'inverse, ce tangage peut également empêcher des détections si le plan laser passe au dessus de l'obstacle. Cela est problématique car des hypothèses de détections sont manquées.

L'altitude des impacts est donc un attribut important pour repérer des fausses détections. Connaissant la surface de la route, donnée par stéréovision, et l'altitude des impacts laser, il est aisé de supprimer ceux qui sont situés sur la route. En considérant que la surface de la chaussée est représentée dans E_Δ par une fonction f_{route} telle que, pour tout point (u_{dr}, v_r, Δ_r) de la route :

$$v_{route} = f_{route}(\Delta) \quad (3.8)$$

et sachant que la position (u_{dl}, v_l, Δ_l) d'un impact laser dans l'espace de disparité est donnée par l'équation (3.7), le filtrage par la hauteur revient à supprimer les impacts tels que :

$$v_l \leq f_{route}(\Delta_l) \quad (3.9)$$

La figure 3.11 illustre le résultat d'un tel filtrage appliqué aux données laser.

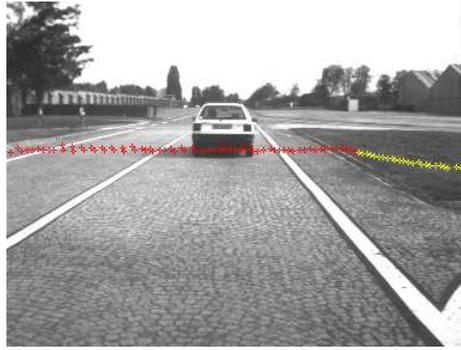


FIG. 3.11 – Filtrage des impacts laser par l'utilisation d'un profil de route estimé par stéréovision. Les impacts rouges sont trouvés comme appartenant à la surface de la route et sont supprimés pour la détection. Les impacts jaunes, situés sur une butte, sont considérés comme n'appartenant pas à cette surface et sont conservés.

3.3.4 Construction des cibles lidar

Après l'étape de filtrage, les impacts laser sont groupés de façon à construire des cibles lidar qui représentent les objets détectés. Nous avons utilisé la méthode présentée par Gruyer et Royère [LRGA05], qui proposent d'agglomérer les impacts par un algorithme de classification dynamique non supervisée. Pour cela, une distance permettant de mesurer l'appartenance d'un point X à une classe existante de centre μ est définie par :

$$D = \frac{\sqrt{(X - \mu)(X - \mu)^T}}{\sqrt{(X_\mu - \mu)(X_\mu - \mu)^T + (X_X - X)(X_X - X)^T}} \quad (3.10)$$

où X_μ et X_X représentent respectivement les points frontières de la classe centrée sur μ et de la modélisation de la mesure par une ellipse, calculés à l'aide des matrices de covariance de la classe et de la mesure. Une valeur de D supérieure à 1 signifie que la mesure X n'appartient pas

à la classe testée. Une valeur égale à 0 signifie au contraire qu'il y a appartenance sans ambiguïté.

Considérant cette mesure de distance, le principe de l'algorithme consiste à considérer le premier impact de la liste comme première classe détectée, puis à vérifier pour chaque impact s'il appartient ou non à une classe déjà existante, en suivant l'ordre du balayage du télémètre. Lorsque l'impact appartient effectivement à une classe donnée, celle-ci est remise à jour. Sinon, une nouvelle classe est créée, centrée sur ce nouveau point.

Lorsque tous les impacts ont été classés, une phase de regroupement de classes permet de fusionner les classes proches, suivant le même critère de distance. Les détails de la méthode sont récapitulés dans [LRGA05].

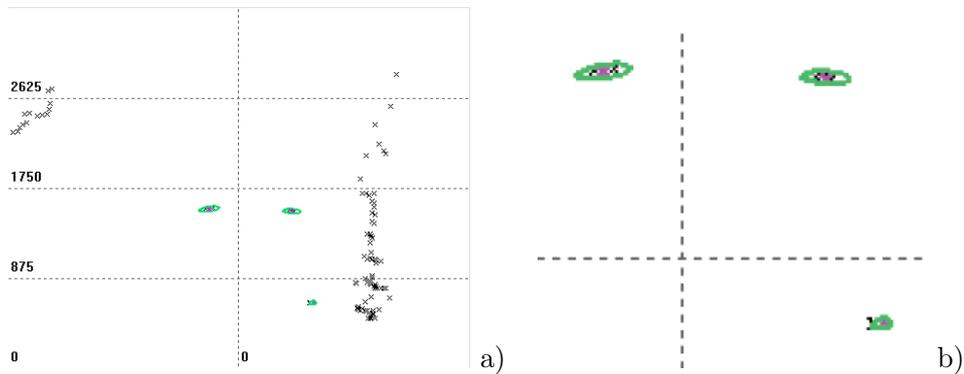


FIG. 3.12 – Résultats de la détection par le lidar.

a) Les résultats sont modélisés par des ellipses dans le plan de balayage. Les impacts situés sur les bords de l'image ne sont pas pris en compte dans la détection car ils sont situés hors de la zone de filtrage latérale. b) Zoom sur les obstacles détectés.

3.3.5 Suivi des détections lidar

Nous avons vu que la détection par le lidar est très sensible au tangage du véhicule instrumenté. En situation de conduite, les pertes de détection sont nombreuses lorsque le plan laser passe au-dessus ou au-dessous d'un obstacle. Suivre les détections au cours du temps est une stratégie permettant de palier à cet inconvénient. En effet, la prédiction de la position des objets par l'utilisation d'un filtre estimateur permet de combler les pertes de détections momentanées, assurant un taux de détection élevé lors de la création d'hypothèses. En outre, ce processus fournit une estimation de leur vitesse, indispensable pour le calcul ultérieur d'un TTC.

Appliquer un algorithme de suivi directement sur les données lidar offre l'opportunité de bénéficier d'une précision élevée et indépendante de la distance, contrairement aux données stéréoscopiques, justifiant alors pleinement cette stratégie. A cette fin, les auteurs de [LRGA05] proposent un algorithme de suivi multi-objets, fondé sur la théorie de l'évidence de Dempster et Shaffer, et bien adapté au suivi des détections lidar. Nous avons utilisé celui-ci comme une brique de la détection par télémétrie laser, pour assurer une continuité des détections. Néanmoins, sa description serait trop longue et hors propos dans cette thèse. Le lecteur intéressé par un complément d'informations pourra donc se référer à [Gru99], puis [Roy02].

3.3.6 Construction des zones d'intérêt

En sortie de l'algorithme de détection, le système dispose d'un ensemble de pistes laser 2D, caractérisées par un centre et une matrice de covariance. Cela permet de les représenter, dans le plan $Y_A = 0$, par une ellipse correspondant à la troncation d'une gaussienne 2D. Cette ellipse est ensuite transformée en un parallélogramme fronto-parallèle. Comme illustré sur la figure 3.13, Z_{pres} , X_{gauche} et X_{droite} sont directement fournis par les dimensions de l'ellipse. Z_{loin} et Y_{haut} sont ensuite construits à partir d'une connaissance *a priori* sur les caractéristiques des obstacles ou les spécifications du système. La base du parallélépipède est positionnée sur le sol. Une fois le volume calculé, il est projeté dans l'espace de disparité. La zone d'intérêt correspondante est alors le plus petit parallélépipède rectangle conforme à la définition du paragraphe 3.1 qui englobe cette projection, comme l'illustre la figure 3.13-b.

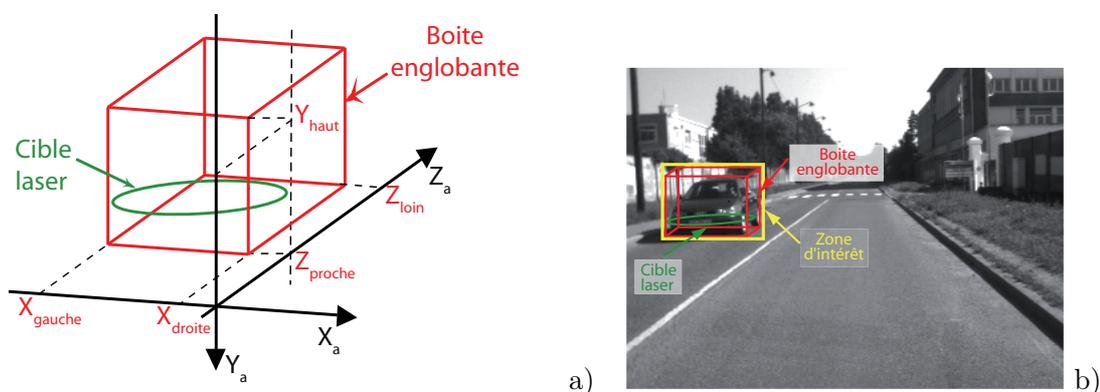


FIG. 3.13 – Construction d'une zone d'intérêt dans les images stéréoscopiques à partir d'une détection laser.

3.3.7 Résultats

La figure 3.14 présente des zones d'intérêt construites à partir de données lidar. Le positionnement est précis, mais les dimensions *a priori* les rendent peu adaptées aux petits obstacles (cela est visible sur le plot). Une caractérisation 3D des détections est alors nécessaire. D'autre part, si la portée théorique est assez élevée (75 m), elle est plus faible en pratique car les variations d'assiette du véhicule empêchent régulièrement l'intersection du plan de balayage avec les objets présents à grande distance. Le suivi peut compenser ces pertes en partie.

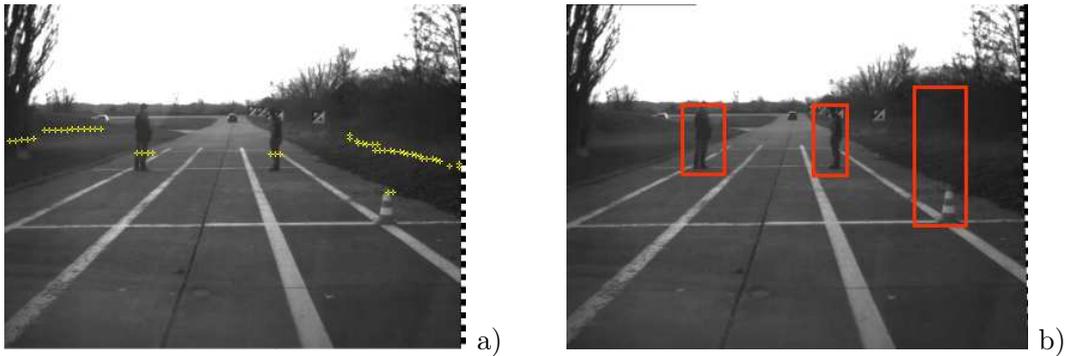


FIG. 3.14 – Zones d'intérêt générées par le lidar.

3.4 Création d'hypothèses de détection par identification optique

Dans cette section, nous présentons une troisième approche de la création d'hypothèses de détection : l'identification optique des objets. Le capteur présenté ici [vAPBE07] s'inscrit dans le domaine des télécommunications inter-véhiculaires (V2V pour "vehicle to vehicle" telecommunications), comme cela a été abordé au chapitre 1. Néanmoins, il peut être considéré comme un dispositif coopératif pour la détection d'obstacles, puisqu'il suppose que l'objet à détecter est lui-même équipé d'une partie du système qui va lui permettre d'être détecté. Nous allons maintenant présenter plus en détails ce capteur et proposer une méthodologie pour assurer son utilisation pour la création d'hypothèses de détection.

3.4.1 Présentation du capteur

Le capteur d'identification optique que nous présentons ici est destiné à identifier des véhicules, ou plus généralement des objets, dans une scène routière et à les localiser dans le référentiel du véhicule suiveur. Afin d'assurer une portée maximale au capteur et de ne pas être sensibles à des occultations partielles, le choix a été fait de ne pas chercher à détecter un motif spatial. La solution proposée repose donc sur un signal d'identification codé temporellement par une lampe émettant dans le proche infrarouge et sur son décodage par un récepteur optique. Le capteur est donc constitué de deux parties distinctes : un émetteur placé sur les objets à identifier et un récepteur embarqué dans le véhicule suiveur.



FIG. 3.15 – Prototype de lampe émettrice.

L'émetteur

L'émetteur (figure 3.15) est composée d'un ensemble de LEDs rayonnant dans le proche infrarouge (880 nm), d'un circuit de commande générant le signal identifiant et d'un circuit d'alimentation fournissant la puissance électrique nécessaire au bon fonctionnement de la lampe.

Le signal identifiant est une trame de communication émise de façon cyclique et continue. Cette trame est codée de façon binaire, le 1 logique correspondant à l'état allumé de la lampe et le 0 logique à l'état éteint. Ainsi, la lampe clignote pour émettre un message, suivant un protocole de communication pré-défini, présenté sur la figure 3.16. Pour assurer un décodage correct de l'identifiant, un code de début de trame est placé en tête de celle-ci (1101, sur notre prototype). Ensuite se trouvent les données d'identification puis, pour limiter le risque d'erreur de décodage du message et de fausse détection, un 0 suivi d'un bit de parité terminant la trame.



FIG. 3.16 – *Protocole d'émission de la trame d'identification.*

Le récepteur

La partie réceptrice du dispositif est une caméra sensible au rayonnement infrarouge proche. La plupart des caméras CCD ou CMOS traditionnelles sont sensibles à cette plage de longueur d'onde. Une telle camera acquière donc des images de la scène routière en avant du véhicule équipé et un algorithme de traitement d'images est utilisé pour localiser et décoder les signaux d'identification émis par les objets coopératifs présents. Afin que le processus ne soit pas perturbé par l'image d'objets ne rayonnant pas dans le proche infrarouge, l'optique de la caméra est équipée d'un filtre passe bande centré sur la longueur d'onde 880 nm. La figure 3.17 présente une image de scène routière acquise sans et avec ce filtre. Le filtre n'est visiblement pas parfait puisqu'il laisse passer une quantité non négligeable de lumière visible. D'autre part, il apparaît que de nombreux éléments de la scène rayonnent dans le proche infrarouge, comme le ciel, la végétation ou les phares des véhicules. Enfin, différents objets reflètent la lumière infrarouge du soleil, apparaissant alors dans l'image du récepteur comme des sources de forte intensité.

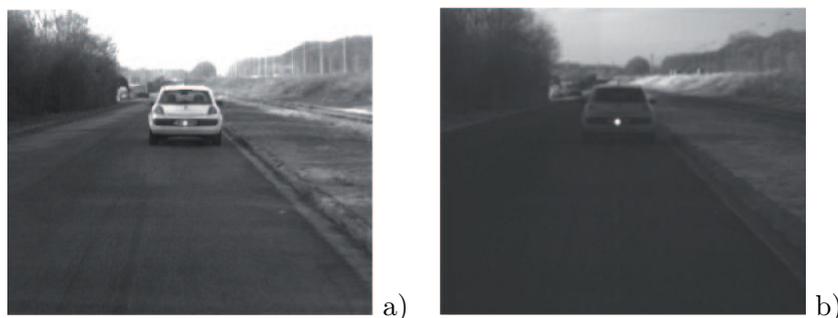


FIG. 3.17 – *Image d'une scène routière acquise avec la caméra utilisée comme récepteur sans (a) et avec (b) le filtre passe bande 880 nm.*

Les émetteurs ne sont donc pas les seuls objets de la scène à être perçus par la caméra réceptrice. Pour concevoir un algorithme de traitement qui permette de décoder les messages

émis, il est nécessaire de mettre en évidence les caractéristiques discriminantes des émetteurs. Ceux-ci seront définis comme des régions de l'image caractérisés par :

- une forme circulaire
- une forte luminosité
- le respect du protocole de communication
- un mouvement faible entre deux images successives, soit un mouvement relatif des véhicules faible devant la fréquence de traitement.

Ces hypothèses sont suffisantes au bon fonctionnement du système et ont permis de mettre au point l'algorithme de décodage présenté plus loin.

Enfin, notons que la fréquence d'acquisition de la caméra est nécessairement supérieure au double de la fréquence d'émission de la lampe, afin de remplir la condition d'échantillonnage formulée par Shanon [Sha49]. Cette contrainte implique de privilégier la vitesse d'acquisition sur la résolution de la caméra utilisée, comme expliqué en annexe E.

3.4.2 Algorithme de traitement du récepteur

Le traitement des images issues du récepteur s'effectue en deux étapes. Tout d'abord, un traitement de bas niveau permet d'extraire les primitives correspondant aux objets pouvant potentiellement être des lampes émettrices. Une algorithmes de plus haut niveau intervient ensuite pour suivre ces primitives au cours du temps, décoder les trames émises et rejeter les fausses détections.

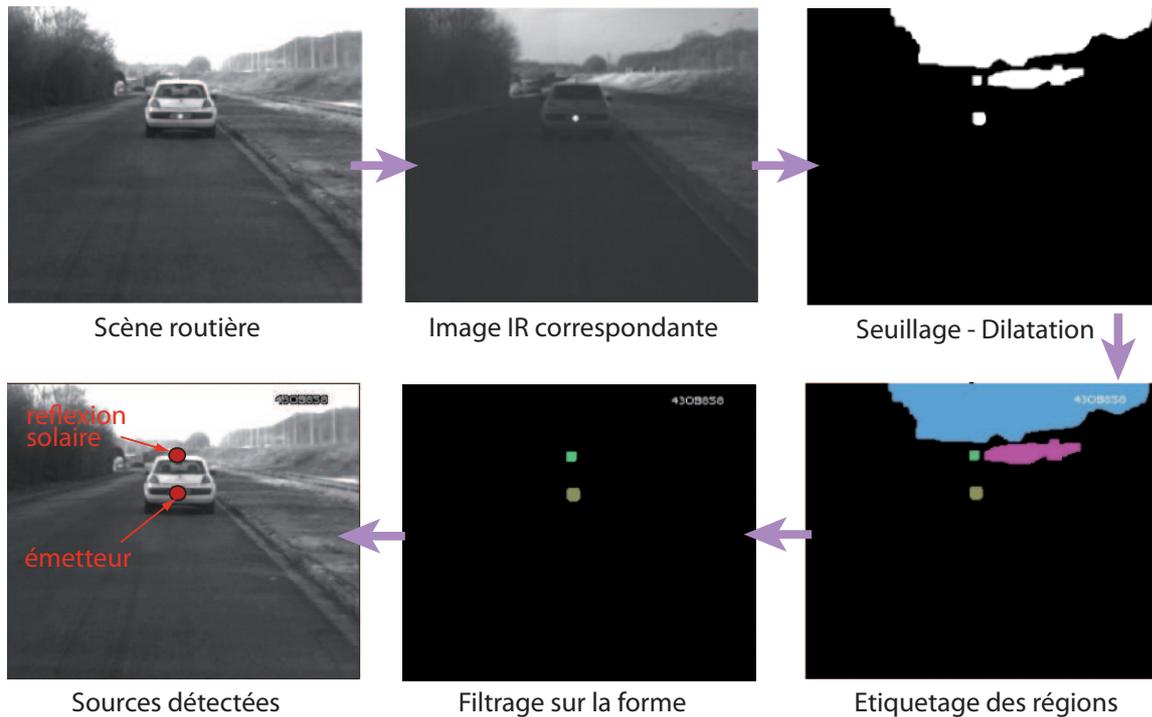


FIG. 3.18 – Algorithme de bas niveau de traitement des images infrarouge.

Traitement bas niveau des images

La figure 3.18 présente le processus d'extraction des primitives des images infrarouge.

1. L'image infrarouge est rendue binaire par un seuillage simple. Seules les sources IR de forte puissance sont conservées.
2. Une dilatation morphologique permet de regrouper les régions fragmentées.
3. Un algorithme de recherche de composantes connexes permet de séparer et d'étiqueter les différentes régions 8-connexes de l'image.
4. Chacune des régions ainsi formées représente un émetteur potentiel. Un critère de dimension permet alors de rejeter celles qui sont trop grandes pour être créées par une lampe, le ciel par exemple.
5. Un test de rotondité permet également de rejeter de mauvais candidats. Soient N_i^{reg} le nombre de pixels contenu dans la région i et N_i^{circ} la surface en pixels du cercle circonscrit à la région i . Une région i est conservée si et seulement si le rapport N_i^{reg}/N_i^{circ} est proche de 1.

Nous appellerons cibles IR les régions qui ont été conservées en sortie de cet algorithme. Certains objets comme des phares de véhicules ou des réflexions solaires sont validées lors de cette étape mais seront rejetés par la suite parce qu'ils ne respectent pas le protocole de communication décrit plus haut.

Suivi des primitives extraites

Une fois les primitives extraites, elles sont suivies au cours du temps de façon à décoder le message émis et à rejeter les détections qui ne correspondent pas à un message cohérent. Ce suivi est fondé sur une prédiction de la position de la cible dans l'image suivante, en supposant que son mouvement entre deux images successives est faible, hypothèse d'autant plus facilement vérifiée que la fréquence de la caméra est élevée. Ce suivi ne fait pas intervenir de modèle d'évolution, car aucune notion de positionnement dans l'espace n'est disponible à ce niveau de l'algorithme. La cible à l'instant $t+1$ est recherchée dans un voisinage de sa position à un instant t . Si aucune cible ne s'y trouve, l'état logique mesuré de la piste en cours est 0 et le voisinage est dilaté pour l'image $t+2$. Si au contraire une cible s'y trouve, l'état logique est 1 et le voisinage est centré sur la position de la cible puis réduit à sa taille originale. Au fur et à mesure que le processus de suivi est itéré, les états logiques des instants successifs sont stockés dans une file de type FIFO³⁷.

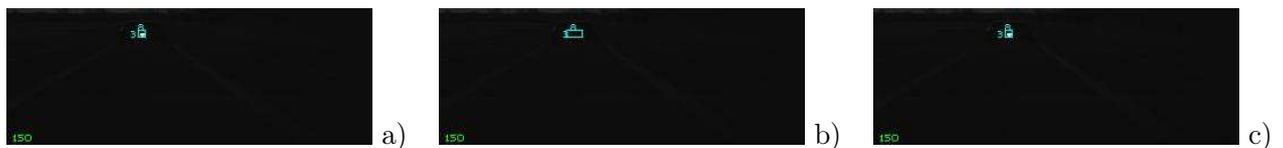


FIG. 3.19 – Suivi d'une lampe infrarouge.

Lorsque celle-ci est éteinte, le voisinage de recherche se dilate ((b) 8 dilatations successives). Il est recentré sur la lampe lorsque celle-ci est de nouveau allumée (c).

Décodage et filtrage des trames reçues

La file de bits d'une piste est analysée en permanence comme un message. Parallèlement, une mesure de confiance est associée à cette piste et évolue en fonction du résultat du décodage.

³⁷First In First Out

Cette valeur diminue lorsque le protocole de communication n'est pas respecté et augmente dans le cas contraire. Une piste est alors considérée comme correcte si et seulement si sa confiance est maximale. Dans ce cas, son identifiant et sa position dans l'image y sont attachés pour être utilisés plus tard.

Cette étape permet de supprimer les fausses pistes créées par des réflexions solaires ou par des feux de véhicules ou de signalisation, parce que ces lumières ne clignotent pas ou de façon aléatoire ou encore de façon incohérente vis à vis du protocole (clignotants de voiture par exemple). Des exemples de pistes correctes et incorrectes sont donnés sur la figure 3.20.

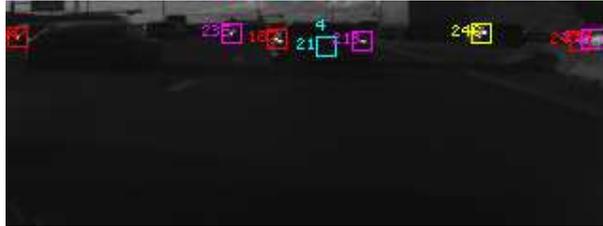


FIG. 3.20 – Une piste correcte étiquetée 4 (cyan), et des pistes incorrectes (autres couleurs) justement invalidées par l'algorithme de décodage/filtrage. L'aspect de l'image est du à la réduction de sa résolution verticale pour accroître la fréquence d'acquisition (détails en annexe E).

3.4.3 Construction des zones d'intérêt

Une fois qu'une cible infrarouge est considérée comme correcte, elle doit être positionnée dans le repère absolu R_a de façon à être exploitable par les autres capteurs équipant le véhicule, et en particulier le capteur stéréoscopique. Après cette étape de localisation, la zone d'intérêt est construite de façon similaire à celle utilisée pour le lidar, en utilisant des connaissances *a priori* sur le véhicule équipé. Celles-ci sont la hauteur de la lampe, la largeur du véhicule et sa hauteur, comme indiqué sur la figure 3.21. Ces deux derniers paramètres sont néanmoins surévalués de façon à compenser d'éventuelles incertitudes de localisation.

La principale difficulté réside dans la rétroprojection de l'identifiant dans R_a . Nous proposons

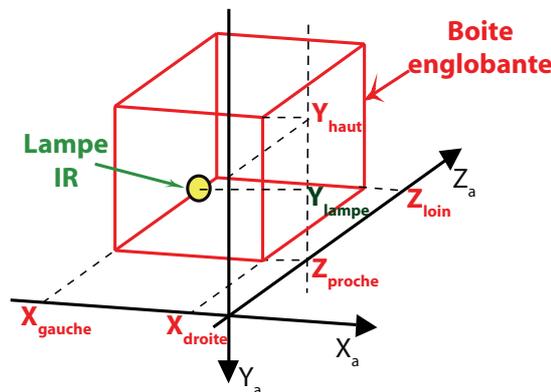


FIG. 3.21 – Construction d'une zone d'intérêt dans les images stéréoscopiques à partir de l'identification et la localisation d'un émetteur infrarouge.

pour cela deux méthodes. La première considère l'utilisation du système d'identification tel qu'il

vient d'être décrit. La seconde suppose que deux caméras sont utilisées pour le décodage, ce qui permet d'exploiter une information de disparité.

Utilisation de la hauteur de la lampe

Le problème de localisation peut être vu comme une projection inverse du pixel image correspondant au centre de la cible vers le repère R_a . Cependant, comme cela a été vu au paragraphe 2.1.1, la projection perspective de l'objet observé sur la surface de la caméra a fait perdre une dimension. Les deux seules coordonnées image (u_{id}, v_{id}) sont équivalentes à un rayon lumineux mais ne sont pas suffisantes pour résoudre ce problème à trois inconnues. Il faut donc ajouter une nouvelle contrainte au système. Nous avons décidé de supprimer un degré de liberté sur la position de la lampe émettrice. La hauteur de celle-ci sera donc considérée comme connue. Cette contrainte n'est pas trop restrictive car la lampe est fixée une fois pour toutes sur l'objet à identifier. Il n'est de plus pas irréaliste de considérer qu'elle peut transmettre sa hauteur dans la trame de communication infrarouge ou dans une communication radio. En revanche, cela suppose une hypothèse de "monde plan" qui n'est vérifiée qu'à courte portée et peut donc se révéler problématique.

Pour calculer les équations de la localisation de l'identifiant, nous allons exploiter les équations du capteur de vision tel qu'il est défini au chapitre 2.1.1. Appelons $M_L^a = (X_L^a, Y_L^a, Z_L^a)^T$ et $M_L^c = (X_L^c, Y_L^c, Z_L^c)^T$ les représentations du point central de la lampe respectivement dans R_a et dans le repère R_c lié à la caméra réceptrice. L'équation (2.10), permet d'exprimer les coordonnées dans M_L^a en fonction de M_L^c , ce qui permet en particulier d'explicitier :

$$\begin{cases} X_L^a &= X_L^c + X_c^0 \\ Y_L^a &= Y_L^c \cos \theta_c + Z_L^c \sin \theta_c + Y_c^0 \\ Z_L^a &= -Y_L^c \sin \theta_c + Z_L^c \cos \theta_c + Z_c^0 \end{cases} \quad (3.11)$$

la hauteur Y_L^a étant supposée fixe et connue. La deuxième équation du système (3.11) peut aussi s'écrire :

$$\frac{Y_L^a - Y_c^0}{Z_L^c} = \frac{Y_L^c}{Z_L^c} \cos \theta_c + \sin \theta_c \quad (3.12)$$

Par ailleurs, d'après l'équation (2.3), la projection sur le plan image se caractérise par :

$$\begin{cases} \frac{X_L^c}{Z_L^c} &= \frac{u_{id} - u_0}{\alpha_u} \\ \frac{Y_L^c}{Z_L^c} &= \frac{v_{id} - v_0}{\alpha_v} \end{cases} \quad (3.13)$$

ce qui permet d'obtenir :

$$Z_L^c = \frac{\alpha_v (Y_L^a - Y_c^0)}{(v_{id} - v_0) \cos \theta_c + \alpha_v \sin \theta_c} \quad (3.14)$$

En combinant cette équation avec l'équation (3.11), il est finalement possible d'explicitier :

$$\begin{cases} X_L^a &= X_c^0 + \frac{\alpha_v (Y_L^a - Y_c^0)(u_{id} - u_0)}{\alpha_u (v_{id} - v_0) \cos \theta_c + \alpha_v \sin \theta_c} \\ Z_L^a &= Z_c^0 + \frac{(Y_L^a - Y_c^0)(\alpha_v \cos \theta_c - (v_{id} - v_0) \sin \theta_c)}{(v_{id} - v_0) \cos \theta_c + \alpha_v \sin \theta_c} \end{cases} \quad (3.15)$$

X_c^0 et Z_c^0 sont fixés et donnés par le positionnement de la caméra réceptrice dans R_a . En revanche, la dynamique liée à l'écrasement des suspensions alors que le véhicule roule entraîne

une variation non négligeable de Y_c^0 et surtout de θ_c . Ces deux paramètres doivent donc être estimés en temps réel pour assurer une localisation correcte de l'identifiant.

Pour cela, n'importe quel capteur de tangage peut être utilisé. Dans notre cas, nous disposons d'une estimation du profil proche de la route dans l'image v-disparité, qui permet d'accéder à des estimations de Y_c^0 et θ_c .

Cette approche est limitée au cas d'une route plane, ce qui peut entraîner une erreur de positionnement importante à grande distance. Pour régler ce problème, l'exploitation de l'image v-disparité permet d'obtenir un profil de route constitué d'une série de plans, comme proposé dans [Lab03]. Pour chacun de ces plans, le calcul de la localisation est réalisé comme cela vient d'être présenté. La localisation correcte est alors celle dont le positionnement longitudinal correspond effectivement à la portion de plan qui a permis de calculer cette localisation.

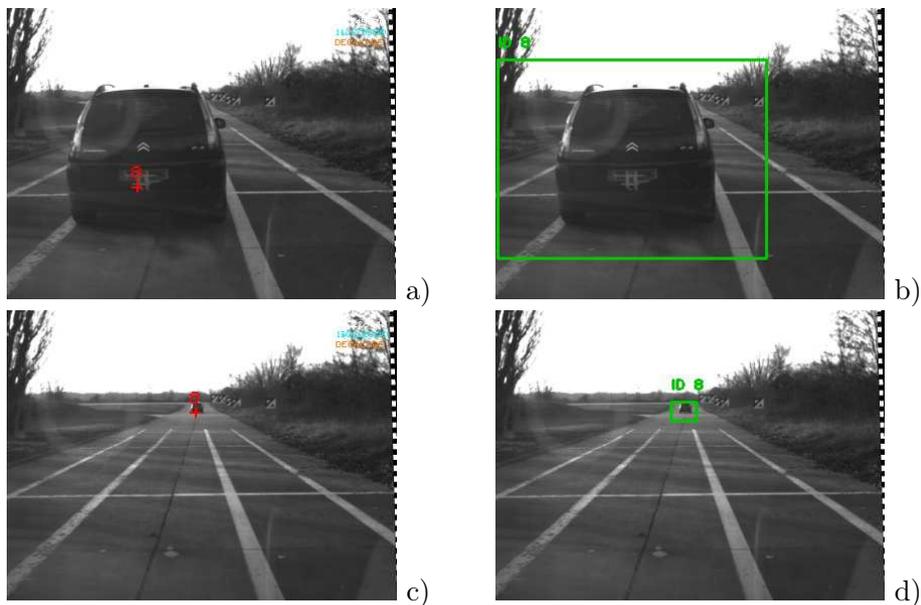


FIG. 3.22 – Identifiants localisés dans l'image de droite de la paire stéréoscopique en utilisant la hauteur connue de la lampe (a,c) et zones d'intérêt correspondantes (b,d).

L'erreur de localisation à grande distance (c,d) est due au non respect de l'hypothèse de monde plan.

Utilisation de l'identification en stéréovision

Une autre façon de résoudre ce problème de la route non plane consiste à ne pas utiliser une seule caméra pour la détection et le décodage des identifiants, mais deux caméras formant un système stéréoscopique. L'information de localisation est alors directement estimée par triangulation, sans utiliser de connaissances sur la hauteur de la lampe ou le profil de la route.

La principale problème de la stéréovision, à savoir l'appariement, est trivial lorsque l'on utilise deux caméras pour l'identification. En effet, si ces deux capteurs procèdent indépendamment aux étapes de détection et de décodage, ils fournissent chacun une liste d'identifiants positionnés dans leurs images respectives. De fait, les primitives à appairer sont directement les identifiants

localisés, ce qui laisse peu de possibilité d'erreurs, compte tenu de la robustesse du décodage.

Une fois qu'un identifiant donné est localisé dans les images gauche et droite (respectivement en (u_{id}^g, v_{id}^g) et (u_{id}^d, v_{id}^d)), le calcul de sa position dans R_a est immédiatement donné par les équations de la stéréovision explicités au chapitre 2. En considérant que la base stéréoscopique du système d'identification est parallèle à l'axe X_a et que son centre est positionné en $(X_{id}^0, Y_{id}^0, Z_{id}^0)$, la position de l'identifiant dans R_a est :

$$\begin{cases} X_a = X_{id}^0 + \frac{b_{id}}{2} + \frac{b_{id}(u_{id}^d - u_{id}^g)}{u_{id}^g - u_{id}^d} \\ Y_a = Y_{id}^0 + \frac{\alpha_u}{\alpha_v} \frac{b_{id}((\frac{v_{id}^d + v_{id}^g}{2} - v_{id}^0) \cos \theta_{id} + \alpha_v \sin \theta_{id})}{\Delta_{id}} \\ Z_a = Z_{id}^0 + \frac{\alpha_u}{\alpha_v} \frac{b_{id}(\alpha_v \cos \theta_{id} - (\frac{v_{id}^d + v_{id}^g}{2} - v_{id}^0) \sin \theta_{id})}{\Delta_{id}} \end{cases} \quad (3.16)$$

Le terme $\frac{v_{id}^d + v_{id}^g}{2}$ permet de prendre en compte une ordonnée moyenne afin de résoudre le cas où, en raison de la taille de la lampe, $v_{id}^d \neq v_{id}^g$.



FIG. 3.23 – Identifiants localisés dans l'image de droite de la paire stéréoscopique en utilisant l'identification en stéréovision (a,c) et zones d'intérêt correspondantes (b,d).

Avec cette méthode, la précision de localisation n'est plus directement liée à la résolution verticale comme dans le cas précédent, mais elle dépend alors plutôt de la résolution horizontale du capteur puisque l'essentiel de l'information provient alors de la notion de disparité.

3.4.4 Résultats

Les figures 3.22 et 3.23 illustrent les résultats de création de zones d'intérêt par identification optique avec les deux méthodes de localisation proposées. Les régions obtenues sont volontaire-

ment surdimensionnées, pour s'assurer qu'elle contiennent bien les objets identifiés. Ce surdimensionnement sera compensé par l'usage de la stéréovision dans les régions d'intérêt, qui permet de caractériser plus finement les obstacles.

Nous pouvons noter que le capteur utilisé permet la détection des obstacles à grande portée (>100 m), mais que la précision de la localisation à cette distance semble faible lorsqu'une unique caméra est utilisée. L'hypothèse du monde plan et la faible résolution verticale de la caméra utilisée en sont les principales explications. Nous aborderons cet aspect plus en détail au chapitre 5.

3.5 Perspectives : Association des différents capteurs

Nous venons de présenter des méthodes de détection d'obstacles utilisant trois capteurs différents. Les détections qui en résultent peuvent ensuite être directement utilisées, sous forme de zones d'intérêt, par un système stéréoscopique. Cette exploitation peut être réalisée indépendamment pour les différents capteurs, mais il est préférable de fusionner leurs détections avant le calcul dans les zones d'intérêt. En effet, comme cela est visible sur la figure 3.24, traiter toutes les hypothèses brutes implique de calculer plusieurs fois certaines zones de la scène qui se superposent. Pour assurer un fonctionnement optimal à notre système, l'étape de fusion est donc nécessaire, mais elle soulève divers problèmes que nous allons évoquer rapidement. Pour cette raison, elle sort du cadre générale de cette thèse et n'a donc été traitée que de façon partielle. Voyons rapidement les points importants de l'étape d'association et la solution que nous proposons.

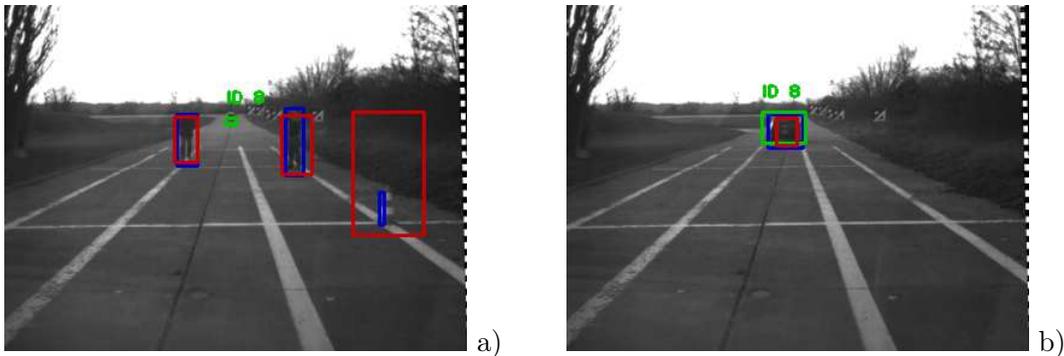


FIG. 3.24 – Résultats des zones détectées par les trois capteurs embarqués.

En rouge, les candidats créés par le lidar, en bleu les candidats issus de la stéréovision et en vert les candidats proposés par l'identification optique.

3.5.1 Combinaison de deux capteurs

Dans un premier temps, nous allons nous intéresser au cas le plus simple de deux capteurs. L'étape de combinaison consiste alors à identifier les candidats issus des deux capteurs qui correspondent à un même objet. Cela sous entend de mesurer une notion de proximité entre des candidats qui sont chacun caractérisés par un volume repéré dans E_{Δ} . Nous proposons d'utiliser deux critères pour juger de cette proximité. Le premier est fondé sur l'utilisation de données métriques, quand le second se focalise sur la surface au sol commune aux détections. Appelons

H_i^1 et H_j^2 , respectivement la i -ème hypothèse fournie par le capteur 1 et la j -ème hypothèse fournie par le capteur 2.

Distance métrique :

Le premier critère que nous proposons d'exploiter consiste à simplement mesurer la distance euclidienne qui sépare la projection sur le plan $Y_a = 0$ des centres de gravité. Si ceux-ci sont respectivement notés (X_i^1, Z_i^1) et (X_j^2, Z_j^2) , le critère à évaluer est simplement :

$$Dist_{i,j}^{1,2} = \sqrt{(X_i^1 - X_j^2)^2 + (Z_i^1 - Z_j^2)^2} \quad (3.17)$$

Surface commune :

Le second critère que nous avons retenu est la surface commune des projections sur le plan $(u_d - \Delta_s)$ des zones d'intérêt correspondant aux deux hypothèses. Soit $S_{u\Delta}(H)$ la mesure d'une telle surface pour l'hypothèses H . Nous proposons de mesurer la proximité entre H_i^1 et H_j^2 par :

$$Prox_{i,j}^{1,2} = \frac{S_{u\Delta}(H_i^1 \cap H_j^2)}{S_{u\Delta}(\min(H_i^1, H_j^2))} \quad (3.18)$$

Cet indice est nul en cas de non-recouvrement et vaut 1 si l'un des volumes est complètement inclus dans l'autre. La combinaison est donc réalisé pour les couples de candidats qui présentent un indice de proximité élevé (supérieur à un seuil s).

Le premier critère permet de fusionner deux hypothèses proches sans pour autant qu'il n'y ait recouvrement, ce que le second ne permet pas d'obtenir. En revanche, lorsqu'une petite cible est incluse dans une grande, elle peut néanmoins se trouver éloignée de son centre de gravité, et ne peut alors pas être fusionnée sur la base du premier critère. Ces deux approches sont donc complémentaires, d'autant que l'une utilise des données cartésiennes quand l'autre exploite des données image.

Stratégie de combinaison :

La fusion des hypothèses issues de deux capteurs est simplement réalisée si au moins l'un des critères ci-dessus est rempli (distance faible ou surface commune importante). Cependant, il peut y avoir un doute si un même candidat du capteur 1 est considéré comme proche de deux candidats du capteur 2. Dans ce cas, les trois sont rassemblés.

Il existe des stratégies beaucoup plus évoluées de combinaison multi-capteurs qui permettent une gestion plus fine du doute sur les associations. Ainsi, nous avons proposé de prendre une décision immédiatement sur le correspondant éventuel d'un candidat issu du premier capteur, avant de vérifier la consistance de cette association lorsque la combinaison est réalisée dans l'autre sens. Pour obtenir une combinaison plus fine, il serait préférable de calculer des indices numériques de correspondance en procédant dans les deux sens, avant de fusionner ces informations puis de prendre une décision en dernier lieu.

3.5.2 Combinaison de plus de deux capteurs

Le problème de la combinaison est grandement complexifié lors de l'utilisation de plus de deux capteurs. Par souci de simplicité, nous avons choisi une stratégie "gloutonne" qui fusionne des hypothèses tant que cela est possible.

3.5.3 Zones d'intérêt résultantes

Une zone d'intérêt construite par la combinaison d'hypothèses issues de plusieurs capteurs doit conserver la géométrie que nous avons définie au paragraphe 3.1. De plus, afin d'exploiter autant que possible la spécificité des différents capteurs, les dimensions d'une telle zone d'intérêt sont héritées des mesures les plus précises possibles. Lorsque cela est possible, la largeur proviendra donc préférentiellement d'une hypothèse fournie par le lidar, et la hauteur sera celle donnée par la stéréovision. La figure 3.25 illustre le résultat d'une telle association réalisée sur des cibles précédemment créées par les trois capteurs que nous avons envisagés (figure 3.24).

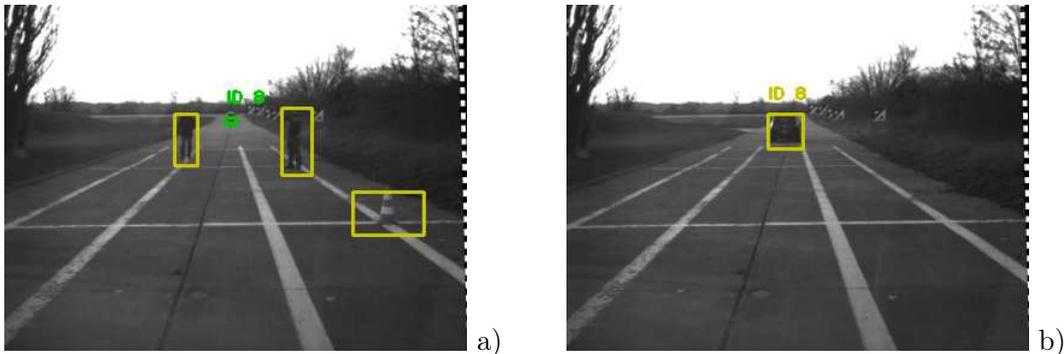


FIG. 3.25 – Les zones d'intérêt de la figure 3.24 ont été combinées (en jaune).

Toutes les informations associées à un candidat avant la combinaison (identifiant détecté, vitesse estimée par le lidar, ...) sont transmises au candidat nouvellement créé par la combinaison.

3.6 Conclusion

Ce chapitre a été l'occasion de proposer une méthodologie de création d'hypothèses de détection à travers trois capteurs très différents. Le résultat obtenu à chaque fois est un ensemble de zones d'intérêt positionnées dans l'espace E_{Δ} . La combinaison de ces différentes zones d'intérêt a également été brièvement présentée. Celle-ci met en évidence l'intérêt d'utiliser plusieurs capteurs complémentaires en parallèle. A titre d'illustration, les images de la figure 3.26 montrent une séquence au cours de laquelle la continuité de la détection d'un véhicule est assurée par ce biais.

L'approche que nous venons de présenter n'est en aucun cas limitée aux trois seuls capteurs que nous avons évoqués. Il serait envisageable en effet d'exploiter des résultats de détection d'un radar ou d'un algorithme de vision monoculaire.

Nous allons, au cours du chapitre suivant, présenter comment sont exploitées les zones d'intérêt désormais créées pour valider et caractériser les détections.



FIG. 3.26 – Détection multi-capteurs lors du suivi d'un véhicule.

a) Le véhicule est détecté par le lidar et par la stéréovision. b) En raison du tangage du véhicule équipé et de la distance importante à l'obstacle (50 m), le lidar perd la détection. La stéréovision assure la continuité. c) Le lidar détecte à nouveau le véhicule.

VALIDATION ET CARACTÉRISATION DES HYPOTHÈSES DE DÉTECTION

Sommaire

4.1	Obtention des données nécessaires	84
4.1.1	Construction des points de mesure par stéréovision locale	84
4.1.2	Intérêt de l'approche locale	85
4.1.3	Portée, précision et temps de calcul : des contraintes antinomiques	85
4.1.4	L'approche subpixelique	86
4.1.5	Approche proposée : l'utilisation d'un zoom numérique local	87
4.2	Validation des hypothèses de détections par stéréovision	90
4.2.1	Principales sources d'erreurs	90
4.2.2	Critère <i>points obstacles</i>	91
4.2.3	Critère <i>alignement principal</i>	92
4.2.4	Critère <i>hauteur de la base de l'obstacle</i>	93
4.3	Mesure des caractéristiques statiques des détections	94
4.3.1	Mesure de la géométrie	94
4.3.2	Mesure de la position	94
4.3.3	Mesure de l'orientation	95
4.4	Estimation de caractéristiques dynamiques	96
4.4.1	Nécessité d'un suivi temporel	96
4.4.2	Principe de l'algorithme de suivi	96
4.4.3	Modèle d'évolution des objets détectés	97
4.4.4	Observation et association	97
4.4.5	Résultats	98
4.5	Résumé de l'Approche et Conclusion	99

APRÈS l'étape de création d'hypothèses de détection à l'aide de divers capteurs qui a été présentée au chapitre 3, nous proposons d'assigner à un capteur stéréoscopique les tâches de validation et de caractérisation de ces détections, conformément à la stratégie de combinaison proposée au chapitre 1. La première tâche vise à assurer une robustesse accrue à la détection. La seconde est destinée à associer des informations de position, géométrie et vitesse aux objets détectés.

Dans un premier temps, nous présenterons la méthode d'obtention des données 3D nécessaires au traitement des hypothèses de détection. Celle-ci repose sur la méthode de calcul de "points obstacles" présentée au chapitre 2, appliquée dans chacune des zones d'intérêt. Nous verrons que le caractère local de cette approche présente un intérêt fort. Le calcul des points obstacles étant très exigeant en temps de calcul, nous proposerons par ailleurs une stratégie permettant d'optimiser celui-ci sans pour autant nuire à la précision des données ainsi construites. Celle-ci consiste à mettre en adéquation la résolution de chaque zone d'intérêt avec sa distance au capteur, par l'application d'un zoom numérique variable.

Dans un second temps, nous proposons trois critères de validation des hypothèses de détection, fondés sur trois caractéristiques essentielles des obstacles que nous recherchons, en accord avec la définition que nous avons fournie en introduction. Nous supposons en effet que les objets à détecter ont une surface visible de taille minimum, qu'ils peuvent être considérés comme quasi-verticaux et que leur base touche le sol.

A l'issue de cette étape de confirmation, les hypothèses de détection considérées comme valides sont mesurées pour permettre leur utilisation ultérieure. Des mesures de position et géométrie peuvent ainsi être obtenues à chaque image. En revanche, la stéréovision ne permettant pas de mesurer directement leur vitesse, une étape de suivi est indispensable pour estimer cette caractéristique.

Enfin, nous concluons ce chapitre par le récapitulatif des différents éléments de méthodes qui ont été proposés.

4.1 Obtention des données nécessaires

4.1.1 Construction des points de mesure par stéréovision locale

L'évaluation des critères de validation des hypothèses, puis leur caractérisation nécessite de disposer de données 3D sur les objets à traiter. Dans cette optique, nous proposons d'exploiter les images stéréoscopiques dans chacune des zones d'intérêt fournies par les différents capteurs pour construire des "points obstacles". Néanmoins, si le processus de construction d'hypothèses par stéréovision a permis de calculer de tels points de mesure pour la totalité de la scène, aucun calcul supplémentaire n'est nécessaire, puisqu'il suffit de considérer les points présents dans chacune des zones d'intérêt. A l'inverse, si de telles données ne sont pas disponibles, ou à une résolution trop faible, un calcul des points de mesure doit être réalisé pour chacune des hypothèses. Le traitement exécuté repose alors sur la méthode de calcul en deux passes présentée au chapitre 2. Il suppose donc en premier lieu l'estimation du profil longitudinal de la route, réalisée sur la totalité de l'image. La seconde passe est ensuite réalisée localement dans chacune des régions d'intérêt. A l'issue de cette étape, chacune d'entre elles dispose donc d'une liste de points de mesure représentés dans l'espace de disparité. A titre d'exemple, la figure 4.1 présente une scène comportant une hypothèse de détection, et les points obstacles correspondant à celle-ci,

représentés sous la forme d'une image de disparité et de sa projection sur les plans u-disparité et v-disparité. C'est le traitement de ces trois images qui permet la validation puis la caractérisation de l'hypothèse.

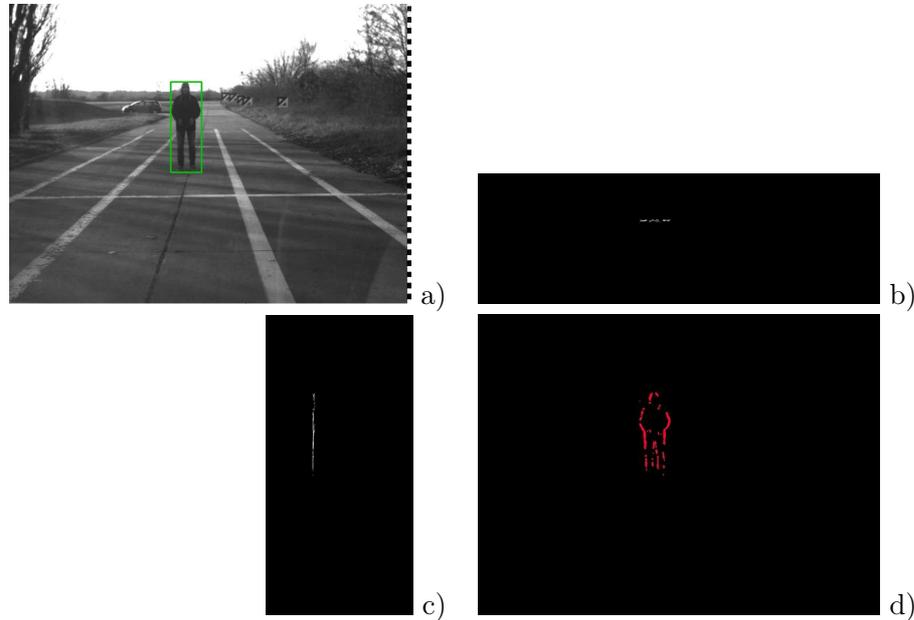


FIG. 4.1 – Scène comportant un candidat pour la détection (a), et données 3D associées représentées sous la forme d'images de disparité (d), u-disparité (b), et v-disparité (c).

4.1.2 Intérêt de l'approche locale

L'aspect local de cette approche permet de considérer que les points obstacles utilisés lors du traitement d'une hypothèse appartiennent tous à un même objet. L'évaluation de critères de validation n'est en conséquence pas perturbée par les objets avoisinants. De plus, pour l'estimation d'une orientation, il n'est alors pas nécessaire de faire appel à une méthode de détection et caractérisation comme la transformée de Hough. Une régression linéaire sur les points obstacles est suffisante, voire plus performante lorsqu'un objet n'est pas parfaitement plan.

4.1.3 Portée, précision et temps de calcul : des contraintes antinomiques

Nous avons mis en évidence en introduction le besoin d'une portée ou d'une précision élevées pour certaines applications. La configuration du système stéréoscopique étant fixée par les spécifications, comme cela est présenté en annexe A, c'est en agissant sur la résolution des images que ces performances peuvent être atteintes. Cependant, la contrainte forte de fonctionnement temps réel ne permet pas d'utiliser des résolutions particulièrement élevées. Calculer une carte de disparité avec une résolution VGA ou plus demande effectivement beaucoup de ressources. Certaines architectures matérielles dédiées le permettent, mais pour un coût élevé. L'approche par zones d'intérêt que nous présentons dans cette thèse permet de répondre en partie à la problématique du temps de calcul. Il reste cependant des cas où la résolution des

images utilisées peut ne pas être suffisante. En particulier, pour les obstacles lointains plusieurs problèmes interviennent :

- La précision de la reconstruction est très faible. Ainsi, l’erreur moyenne d’un demi-pixel induite par l’échantillonnage des images et le fait d’utiliser des disparités entières provoque une erreur métrique importante.
- Les objets sont petits par rapport à la fenêtre de corrélation : le risque que des objets de différentes disparités soient présents dans la même fenêtre est alors augmenté, induisant d’éventuelles erreurs dans la carte de disparité ou simplement l’impossibilité de trouver l’appariement correct.
- Le nombre de pixels dans les régions d’intérêt devient faible. Les erreurs sur la carte de disparité ont alors une influence accrue.
- La plage de valeurs de disparité d’une cible devient faible. La distinction entre obstacles et surface de la route est d’autant plus ardue.

Outre ces problématiques liées à la longue portée, il existe également certains problèmes à courte portée. En particulier :

- La discrétisation très précise de l’espace à courte portée lors de la transformation projective qui permet la formation des images stéréoscopiques peut amener à perdre la notion de connexité sur l’image d’un objet proche. Un objet qui aurait pu être assimilé à un plan vertical de loin ne l’est donc plus forcément de près. Sa détection par la recherche d’un alignement vertical dans le plan v -disparité peut alors devenir problématique.
- Une grande partie du temps de calcul est alloué au traitement des objets proches qui apparaissent particulièrement gros dans les images. Pour ceux-ci la précision de mesure peut être excessivement élevée par rapport aux besoins des applications.

Ces aspects justifient l’usage d’un traitement local et indépendant pour les différents objets, mais rappellent que certains compromis doivent donc être envisagés entre le temps de calcul, la portée et la précision.

4.1.4 L’approche subpixélique

Plutôt que de faire des compromis trop importants sur la précision, une autre possibilité consiste à tirer plus d’informations du système utilisé. Pour cela, une approche dite subpixélique, qui vise à calculer des valeurs de disparité non entières pour gagner en précision, est envisageable. Diverses méthodes allant dans ce sens ont été proposées dans la communauté.

L’approche la plus intuitive pour résoudre ce problème consiste à sur-échantillonner avec interpolation les images stéréoscopiques puis à réaliser les appariements de manière habituelle. Ainsi, un ré-échantillonnage d’un facteur F_s permet théoriquement de calculer des disparités avec une précision de $1/F_s$ pixel. Cela peut être réalisé en gardant l’image de référence telle qu’elle est et en n’interpolant que la seconde image de la paire (dans [BT98], Birchfield et Tomasi exploitent cette technique pour rendre l’appariement indépendant de l’échantillonnage) ou en appliquant l’interpolation aux deux images [SS02b]. Cependant, cette technique peut être très coûteuse en temps de calcul si elle est appliquée à la totalité de l’image, car elle sous-entend un nombre de calculs d’appariements accru. D’autre part, dans [SBS00] les auteurs mettent en évidence une erreur systématique induite par une telle interpolation et incitent à utiliser une interpolation d’ordre élevée (au moins cubique) pour limiter cet effet.

Ces techniques permettent d’obtenir une précision subpixélique directement à l’issue du processus d’appariement. Il existe également des stratégies d’affinement des résultats a posteriori. Par exemple, Chambon envisage la possibilité d’interpoler les valeurs de disparité déjà calculées plutôt que les images sources [Cha05]. Cela permet d’obtenir une carte de disparité lissée, com-

prenant des valeurs non entières, mais nécessite une carte de disparité initialement dense.

Il est également souvent proposé d'interpoler les scores de corrélation autour de l'optimum à l'aide d'une fonction parabole ou fonction toit [Dev97]. La position du minimum de la courbe (pour le critère ZSSD) fournit alors une valeur de disparité subpixelique. Il n'est cependant pas prouvé que la courbe des scores de corrélation suit véritablement une parabole et n'est pas purement aléatoire. En outre, les auteurs de [SO01] prouvent l'existence d'une erreur systématique lorsqu'une telle interpolation est utilisée.

Plutôt que de rechercher à optimiser l'appariement de façon subpixelique, une autre solution consiste à calculer une disparité entière puis à trouver la position subpixelique des points appariés dans leurs images respectives. C'est ce qui est proposé dans [NOD⁺06] : les points considérés appartenant à des contours verticaux, la recherche du véritable passage par zéro du Laplacien d'intensité, par interpolation de celui-ci, fournit la position subpixelique du contour. Il suffit d'appliquer ceci pour trouver l'abscisse réelle des deux points appariés puis de les soustraire l'une à l'autre pour obtenir la disparité subpixelique.

4.1.5 Approche proposée : l'utilisation d'un zoom numérique local

Principe

L'approche subpixelique permet donc d'augmenter la précision que l'on peut obtenir d'une paire d'images données. Nous avons pour notre part choisi d'utiliser une approche par ré-échantillonnage des images stéréo. Si cette méthode peut paraître coûteuse, elle présente néanmoins quelques aspects bien adaptés à notre problématique. En effet, au delà du gain en précision, elle répond précisément aux problématiques concernant les obstacles lointains, évoquées au paragraphe 4.1.3. D'une part, elle permet d'augmenter le nombre de points appariés, d'autre part, l'augmentation significative de la taille des objets observés rend ceux-ci suffisamment grands devant la taille de la fenêtre de corrélation pour faciliter le processus d'appariement.

C'est l'approche par régions d'intérêt qui permet d'exploiter une telle méthode de zoom numérique, en réduisant la surcharge calculatoire qu'elle induit. De surcroît, cela permet d'ajuster le taux de sur-échantillonnage pour chaque région, offrant la possibilité d'économiser du temps de calcul en dégradant la résolution dans les régions proches.

La figure 4.2 illustre le mode de fonctionnement que nous avons donc décidé d'exploiter. Il s'agit d'utiliser un zoom numérique variable dans chacune des zones d'intérêt de façon à rapporter tous les objets à une taille équivalente, indépendamment de leur distance au capteur. Contrairement aux autres approches subpixeliques, cela permet d'adapter précisément la précision de traitement à chacun des objets.

Calcul du facteur de grossissement

Afin de calculer le facteur de grossissement à appliquer à une région d'intérêt, considérons un objet de largeur L , parallèle aux plans image. Ses extrémités $P1(X_{a,1}, Y_a, Z_a)$ et $P2(X_{a,2}, Y_a, Z_a)$ sont projetées respectivement en $p1(u_{d,1}, \Delta_s, v)$ et $p2(u_{d,2}, \Delta_s, v)$. La dérivation de l'équation de retro-projection de la stéréovision (2.17), à Δ_s et v constants, donne :

$$dX_a = \frac{b_s}{\Delta_s} du_d \quad (4.1)$$

ce qui assure que la largeur de l'objet dans les images, selon la direction \vec{u}_d est donnée par :

$$l = \int_{X_{a,1}}^{X_{a,2}} \frac{\Delta_s}{b_s} dX_a = \frac{\Delta_s(X_{a,2} - X_{a,1})}{b_s} = \frac{L\Delta_s}{b_s} \quad (4.2)$$

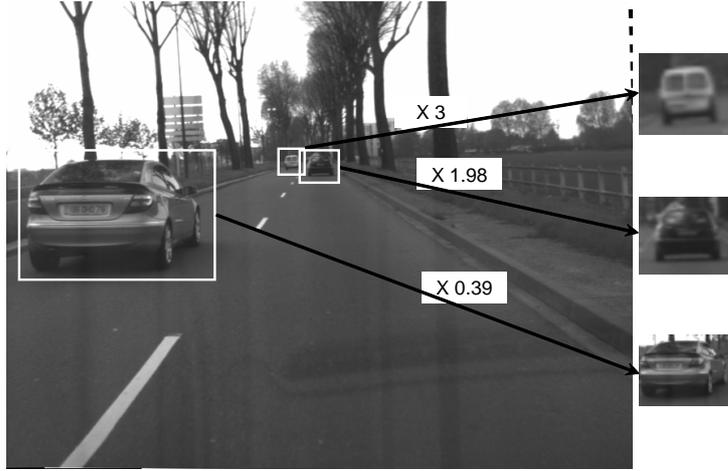


FIG. 4.2 – Principe du zoom numérique variable.

Différents facteurs de zoom sont appliqués pour différents obstacles pour obtenir des tailles d'objets indépendantes de leurs distances.

Avec l'utilisation d'un zoom numérique de facteur F_z , elle devient :

$$l = F_z \frac{L \Delta_s}{b_s} \quad (4.3)$$

Si l'on définit par ailleurs un facteur d'échelle K_s par :

$$K_s = \frac{\text{Taille dans l'image (pixels)}}{\text{Taille réelle (m)}} \quad (4.4)$$

soit pour notre objet :

$$K_s = \frac{l}{L} \quad (4.5)$$

Le facteur de grossissement vaut finalement :

$$F_z = K_s \frac{b_s}{\Delta_s} \quad (4.6)$$

Pour appliquer celui-ci à une zone d'intérêt candidate à la détection, il convient de prendre sa disparité la plus élevée pour calculer le facteur de grossissement correspondant. Après application du zoom, la nouvelle valeur de disparité est alors :

$$\Delta_{new} = F_z * \Delta_s = K_s * b \quad (4.7)$$

Dans l'espace de disparité, l'utilisation du zoom numérique revient donc toujours à rapporter le plan frontal d'une cible à la disparité idéale $\Delta_i = bK_s$.

En pratique, des limitations doivent être appliquées au facteur de zoom. Celui-ci ne peut en effet pas atteindre de valeurs trop importantes pour des raisons évidentes de temps de calcul, d'autant que pour un grossissement élevé l'image devient trop *floue* pour être correctement exploitable. D'autre part, la taille des objets proches ne doit pas être réduite excessivement, sous peine de rendre difficile leur détection, normalement triviale. La limitation du facteur de zoom est réalisée par un seuillage haut et bas.

Problématiques relatives à la mise en œuvre

L'implantation d'un tel zoom numérique local est réalisée par l'application d'une homographie à l'ensemble des pixels d'une zone d'intérêt, de façon à construire une imagerie de la région rééchantillonnée. Considérant les coordonnées du point en haut à gauche de la cible considérée (définies sur la figure 3.1), l'homographie est donnée par :

$$H_z = \begin{pmatrix} 1 & 0 & -u_{min}^{cible} \\ 0 & 1 & -v_{min}^{cible} \\ 0 & 0 & 1/F_z \end{pmatrix}$$

Celle-ci est combinée à une interpolation bi-cubique de façon à limiter l'erreur systématique évoquée dans [SBS00]. Dans le paragraphe 2.1.2 nous avons vu par ailleurs qu'une autre homographie est déjà appliquée aux images originales pour réaliser de façon logicielle la rectification épipolaire du système stéréoscopique. Avec l'utilisation du zoom numérique, chaque zone d'intérêt subit donc deux interpolations successives, apportant alors une dégradation supplémentaire aux images. Il est donc préférable de combiner les deux homographies en une seule à appliquer à l'image originale pour limiter cet effet, comme le montre la figure 4.3.

Si H_r représente l'homographie de rectification appliquée à l'une des images stéréo, la combinaison des deux homographies s'exprime par un simple produit matriciel :

$$H_t = H_z * H_r \quad (4.8)$$

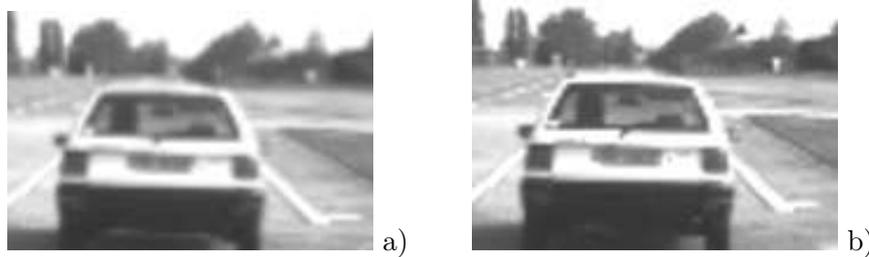


FIG. 4.3 – a) Dégradation d'un détail d'image stéréoscopique par l'application de deux homographies successives (une rectification et un zoom de facteur 2), et (b) comparaison au cas où ces homographies ont été combinées en une seule.

Si des attribut sont calculés à partir d'une cible zoomée (dimension, position, ...), les attributs réels peuvent être retrouvés en y appliquant la transformation H_z^{-1} aux mesures.

Limitations

Si l'utilisation du zoom numérique semble bien se prêter à l'augmentation de la portée de la stéréovision, cela introduit quelques problèmes supplémentaires.

D'une part, le fait d'appliquer un zoom numérique aux images stéréoscopiques rend le système plus sensible aux erreurs de rectification entre les deux caméras. D'autre part, une autre limite du zoom numérique survient dans les zones peu texturées des images originales, qui deviennent d'autant plus homogènes que le facteur de zoom est important.

4.2 Validation des hypothèses de détections par stéréovision

4.2.1 Principales sources d'erreurs

Les méthodes de création d'hypothèses de détection proposées au chapitre précédent ont pour objectif d'assurer un taux de détection le plus élevé possible, au risque de créer des fausses détections. Conformément à la stratégie de fusion proposée au chapitre 1, cet aspect est contrebalancé par l'utilisation d'une stratégie de validation *a posteriori* des hypothèses en utilisant localement le capteur stéréoscopique. Afin de définir cette stratégie, il est nécessaire de comprendre les différentes sources d'erreurs couramment rencontrées. Les images de la figure 4.4 présentent des cas caractéristiques. Celles-ci sont principalement dues à :

- a) l'intersection du plan laser avec la surface de la route,
- b) une erreur de suivi lors de la détection par le lidar, lorsque le comportement d'un objet observé diffère fortement du modèle d'évolution utilisé,
- c) la mauvaise estimation de la surface de la route, conduisant à la présence de points obstacles sur celle-ci,
- d) des erreurs d'appariement causées par la mauvaise qualité des images stéréo,
- e) des erreurs d'appariement causées par la présence de motifs répétitifs de grande taille comparativement à la taille de la fenêtre de corrélation,
- f) la présence d'une source infrarouge dont le scintillement est cohérent avec le protocole de communication établi au chapitre 3.

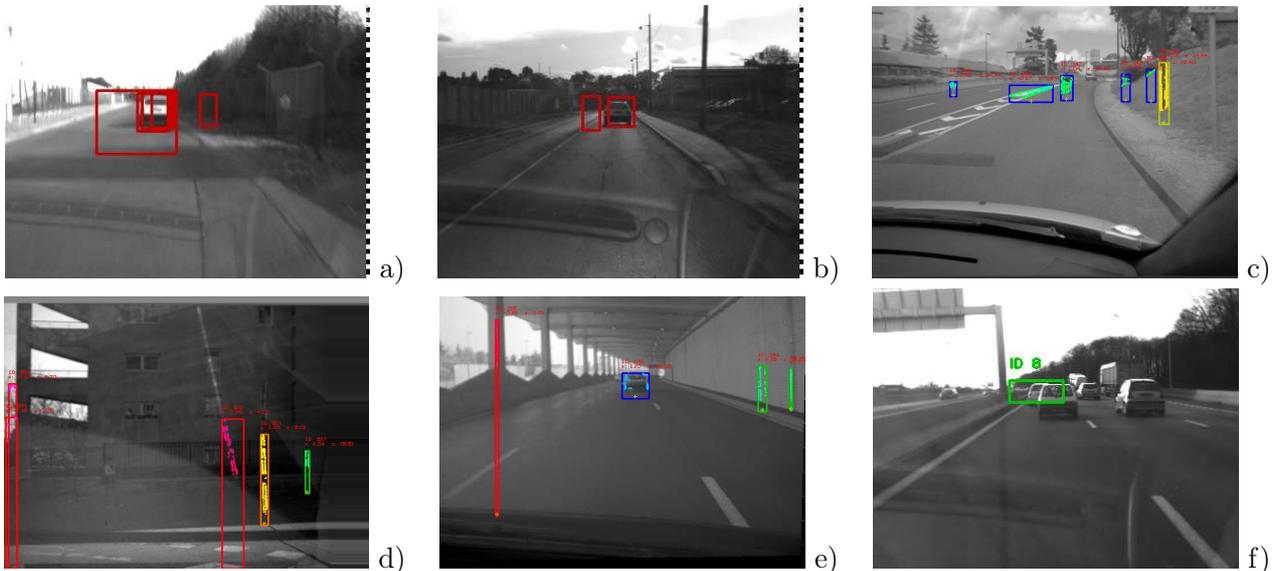


FIG. 4.4 – Exemples de fausses détections.

(a) intersection du plan laser avec la route, (b) erreur de prédiction d'une piste lidar, (c) mauvaise estimation de la surface de la route, (d) faux appariements dus à la mauvaise qualité des images, (e) mauvais appariements dus à la présence de motif répétitifs grands devant la fenêtre de corrélation, (f) mauvaise identification d'une source infrarouge comme étant un émetteur d'identification optique, due au reflet d'un émetteur sur la carrosserie d'un autre véhicule.

Nous proposons trois critères pour rejeter ces hypothèses invalides.

4.2.2 Critère *points obstacles*

L'approche la plus naturelle consiste à vérifier que des points de mesure "obstacles" se trouvent effectivement sur l'hypothèse de détection à valider. Le critère de validation consiste donc à dénombrer les points obstacles présents dans la zone d'intérêt, puis comparer le résultat à un seuil. Deux points de vue distincts peuvent être envisagés pour choisir ce seuil.

D'un point de vue image, si l'on considère que tous les pixels contiennent une information équivalente, un nombre fixe \mathcal{S}_i permet de s'assurer qu'il ne s'agit pas de bruit.

A l'inverse, en considérant une reconstruction métrique, tous les pixels ne sont pas équivalents. En effet, un pixel représente la projection d'une surface donnée sur la matrice d'une caméra, les dimensions de cette surface étant directement liées à la distance séparant l'objet observé du capteur, et donc à la disparité dudit pixel. Les objets apparaissent donc d'autant plus petits qu'ils sont lointains.

Afin d'établir un seuil sur la base de ce constat, nous proposons de mesurer la surface observée à partir des points obstacles qui la composent. Calculons la surface réelle élémentaire représentée par un pixel d'une image. Pour cela, considérons la fronto-parallèle (*i.e.* Δ est constant). Les variations de X_a , Y_a et Z_a lors d'un faible déplacement le long de l'image de cette surface peuvent s'exprimer sous la forme :

$$\begin{cases} dX_a &= \frac{\partial X_a}{\partial u_d} du_d + \frac{\partial X_a}{\partial v} dv + \frac{\partial X_a}{\partial \Delta} d\Delta \\ dY_a &= \frac{\partial Y_a}{\partial u_d} du_d + \frac{\partial Y_a}{\partial v} dv + \frac{\partial Y_a}{\partial \Delta} d\Delta \\ dZ_a &= \frac{\partial Z_a}{\partial u_d} du_d + \frac{\partial Z_a}{\partial v} dv + \frac{\partial Z_a}{\partial \Delta} d\Delta \end{cases} \quad (4.9)$$

ce qui, combiné à l'équation de passage de R_Δ à R_a (2.17), donne :

$$\begin{cases} dX_a &= \frac{b_s}{\Delta_s} du_d \\ dY_a &= \frac{b_s \cos \theta_s}{\Delta_s} dv \\ dZ_a &= \frac{-b_s \sin \theta_s}{\Delta_s} dv \end{cases} \quad (4.10)$$

En utilisant l'approximation selon laquelle les obstacles fronto-parallèles sont aussi verticaux (valide pour des valeurs faibles de θ_s), dZ_a est négligeable. La surface équivalente à la projection d'un pixel p_i de disparité Δ_s^i est alors :

$$Surf(p_i) = \left(\frac{b_s}{\Delta_s^i}\right)^2 \cos \theta_s \quad (4.11)$$

Dès lors, le critère de confirmation d'un candidat \mathcal{C} par l'hypothèse que la surface observée est importante revient à vérifier que :

$$\sum_{p_i \in \mathcal{C}} \left(\frac{b_s}{\Delta_s^i}\right)^2 \cos \theta_s > \mathcal{S}_m = \text{surface minimale} \quad (4.12)$$

le seuil \mathcal{S}_m correspondant physiquement à la surface observée minimale permettant de valider une hypothèse.

Cela permet donc de fixer un seuil constant par sa valeur métrique, donnant lieu à un seuil image exprimé en nombre de pixels et dépendant de la disparité maximale de la cible :

$$\mathcal{S}_m^i = \frac{\mathcal{S}_m}{\cos \theta_s} \left(\frac{\Delta_s^{max}}{b_s}\right)^2 \quad (4.13)$$

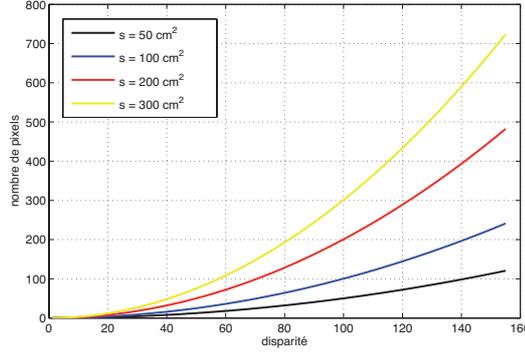


FIG. 4.5 – Valeur du seuil de confirmation en pixels en fonction de la surface exposée minimum et de la disparité des pixels (pour $b_s = 1$ m) dans le cas d'un obstacle fronto-parallèle.

Son évolution en fonction de la disparité de la cible est illustrée sur la figure 4.5. A grande portée, ce seuil permet de confirmer des cibles paraissant petites dans l'image mais correspondant à de grands objets. En revanche, il trouve ses limites à courte portée, car les cartes de disparité locales exploitées pour la confirmation sont très éparées et la surface visible peut être inférieure à la surface réelle de l'objet observé. De fait, il est préférable de coupler cette approche avec l'approche image évoquée plus haut pour définir le seuil final comme :

$$\mathcal{S}_p = \min(\mathcal{S}_m^i, \mathcal{S}_i) \quad (4.14)$$

Le critère *points obstacles* permet de rejeter de nombreuses fausses détections dues à des erreurs d'appariements ou à la détection de la route par le lidar. Il supprime également les objets présentant trop peu d'indices visuels pour être considérés comme visibles sans ambiguïtés par la stéréovision. Il est cependant un critère très local au sens où il considère chaque pixel indépendamment et est en conséquence peu robuste à un grand nombre d'erreurs d'appariements.

4.2.3 Critère *alignement principal*

Une autre façon de valider les hypothèses consiste à exploiter la quasi-verticalité des obstacles, alors que la route est quasi-horizontale. Nous proposons donc de rechercher suivant quelle direction sont alignés les points obstacles de la zone d'intérêt à confirmer. Pour cela, l'image de disparité locale de la cible est projetée dans le plan v-disparité. Une régression linéaire est ensuite réalisée selon la méthode des moindres carrés, afin de trouver l'orientation globale de la répartition des points obstacles. Les caractéristiques de la droite trouvée permettent de confirmer la détection ou non.

Considérons que l'objet observé est assimilable à un plan parallèle à la base stéréoscopique, d'équation :

$$A_2(Z_a - Z_s^0) + A_1(Y_a - Y_s^0) + A_3 = 0 \quad (4.15)$$

D'après l'équation (2.20), il est représenté dans l'espace de disparité par un plan d'équation :

$$\Delta_s = \frac{b_s A_2}{A_3} ((v - v_0)(a \cos \theta_s + \sin \theta_s) + \alpha(a \sin \theta_s - \cos \theta_s)) \quad (4.16)$$

avec $a = \frac{dZ_a}{dY_a}$, la pente du plan objet dans R_a , valeur que nous cherchons à exprimer. Le plan représenté dans R_Δ est projeté sous la forme d'une droite dans le plan v-disparité. Celle-ci a

pour pente :

$$p_0 = \frac{d\Delta_s}{dv} = \Delta_0 \frac{a \cos \theta_s + \sin \theta_s}{\alpha(a \sin \theta_s - \cos \theta_s)} \quad (4.17)$$

en posant $\Delta_0 = \Delta_s(v = v_o)$. Inverser cette équation permet donc d'exprimer a en fonction des caractéristiques (p_0, Δ_0) de la droite trouvée dans le plan v-disparité :

$$\frac{dZ_a}{dY_a} = \frac{\alpha p_0 \cos \theta_s + \Delta_0 \sin \theta_s}{\alpha p_0 \sin \theta_s - \Delta_0 \cos \theta_s} \quad (4.18)$$

Un obstacle sera alors confirmé si :

$$\frac{dZ_a}{dY_a} < \mathcal{S}_a = \text{pente maximale} \quad (4.19)$$

Contrairement au précédent, le critère *alignement principal* repose sur une approche globale au sein de la région d'intérêt (recherche d'alignement). Il est donc sensible à la cohérence entre les pixels et de ce fait plus robuste face aux erreurs d'appariements. Il est particulièrement dédié à la suppression des hypothèses de détections situées sur la surface de la route. Ces hypothèses résultent généralement de l'intersection du plan lidar avec la chaussée, ou de la présence de points obstacles sur celle-ci, causée par une mauvaise estimation lors de la première passe de l'algorithme de construction des données 3D. Des erreurs dues au suivi ou à de faux appariements sont également bien rejetées par ce critère.

4.2.4 Critère hauteur de la base de l'obstacle

Un type bien particulier de fausses détections par stéréovision apparaît dans les scènes présentant de nombreuses structures répétitives, notamment les scènes urbaines. Des faux appariements fortement corrélés font alors apparaître un objet fantôme proche du véhicule. Cela est d'autant plus gênant que les critères de validation présentés ci-dessus, exploitant l'approche v-disparité, supposent que les erreurs d'appariements ne sont pas corrélées et sont donc facilement mis en défaut par ce type de fausses détections.

Parmi les erreurs de ce type, celles qui s'avèrent problématiques ont lieu lorsque la valeur de disparité des pixels considérés est surévaluée. En effet, dans le cas d'une sous-évaluation, l'hypothèse de détection se situe au delà de l'objet réel, et correspond en conséquence plutôt à un cas de non-détection. Étudions l'influence d'une telle surestimation de la disparité sur le positionnement des points correspondants le long de l'axe Y_a . Pour cela, appelons (u_d, v, Δ_1) et (X_1, Y_1, Z_1) les coordonnées d'un pixel représenté respectivement dans les repères R_a et R_Δ , et (u_d, v, Δ_2) et (X_2, Y_2, Z_2) ses coordonnées lorsqu'une erreur a été produite sur l'estimation de la disparité. L'équation de reconstruction (2.17) permet de calculer l'erreur produite le long de l'axe Y_a :

$$Y_2 - Y_1 = \frac{\Delta_1 - \Delta_2}{\Delta_1 \Delta_2} b_s ((v - v_o) \cos \theta_s + \alpha \sin \theta_s) \quad (4.20)$$

qui peut être simplifiée :

$$Y_2 - Y_1 = \frac{\Delta_1 - \Delta_2}{\Delta_2} (Y_1 - Y_s^0) \quad (4.21)$$

En posant $Y_1 = 0$, pour considérer que le pixel en question est situé à la base d'un obstacle, on obtient :

$$Y_2 = Y_s^0 \left(1 - \frac{\Delta_1}{\Delta_2}\right) \quad (4.22)$$

Cette valeur correspond donc à la hauteur minimale de la base d'une hypothèse de détection résultant d'une erreur d'estimation de la disparité. Cette hauteur peut être importante lorsque la disparité est fortement surévaluée et peut donner l'impression que l'objet flotte sans support au sol. Le critère de validation consiste donc à mesurer l'altitude du point obstacle le plus bas de l'hypothèse de détection, et vérifier que cette altitude est suffisamment faible, soit :

$$\max_{\text{pointsobstacles}}(Y_a) < S_h = \text{hauteur minimale} \quad (4.23)$$

Ce critère est très utile lors de l'utilisation de méthodes d'appariements locales, exploitant une fenêtre de corrélation de petite taille, qui n'est pas capable d'estimer la disparité correctement sur des objets d'apparence répétitive de taille importante. Il peut en revanche causer le rejet des détections pertinentes, comme un chargement dépassant de l'arrière d'un camion s'il n'apparaît pas comme appartenant au camion.

4.3 Mesure des caractéristiques statiques des détections

Après la validation des détections, il convient de les caractériser par des mesures de position et de géométrie. Celles-ci sont les dimensions et positions des objets détectés et doivent être représentées en grandeurs métriques. Voyons comment ces grandeurs sont directement déduites des points obstacles.

4.3.1 Mesure de la géométrie

Un objet détecté est représenté par sa zone d'intérêt, dont la définition a été donnée au paragraphe 3.1. Une telle représentation n'est pas exactement équivalente à une géométrie simple dans l'espace cartésien E_a , puisqu'elle est équivalente à un cylindre à base trapezoidal non régulière. Elle est de fait difficilement exploitable par un système d'aide à la conduite. Nous préférons une représentation simple sous la forme d'un parallélépipède rectangle fronto-parallèle dans E_a . Celle-ci n'est donc pas déduite de la géométrie de la zone d'intérêt mais des points obstacles contenus dans cette dernière. Ces points de mesure permettent en effet d'obtenir les valeurs extrêmes suivant les trois axes de R_a et fournissent donc des intervalles $[X_a^{min}, X_a^{max}]$, $[Y_a^{min}, Y_a^{max}]$ et $[Z_a^{min}, Z_a^{max}]$, directement équivalents à des mesures de largeur, hauteur et profondeur de l'objet détecté. Notons cependant que Y_a^{max} est toujours pris égal à zéro de façon à placer la base de l'objet sur le sol.

Le résultat obtenu est illustré par la figure 4.6. A gauche, le rectangle englobant représente le parallélépipède correspondant au résultat de la détection dans R_Δ . Sur l'image de droite, est projeté le parallélépipède correspondant à la caractérisation de la détection dans R_a , construit à partir des points de mesure.

4.3.2 Mesure de la position

Il est nécessaire d'accompagner chacune des détections d'une mesure de position dans le repère cartésien R_a . Plusieurs stratégies peuvent être envisageables pour cela : utiliser le point de mesure le plus proche, utiliser le centre de gravité de l'objet ou de ses points de mesure, ou encore utiliser un point fixe lié à la géométrie. C'est la solution que nous proposons de retenir en exploitant la projection au sol du centre de la face avant du volume obstacle, comme l'illustre la figure 4.7.

Le choix de la face avant permet de disposer de la distance minimale à l'obstacle. La projection au sol résulte de l'hypothèse selon laquelle les obstacles sont censés toucher la surface de la route,

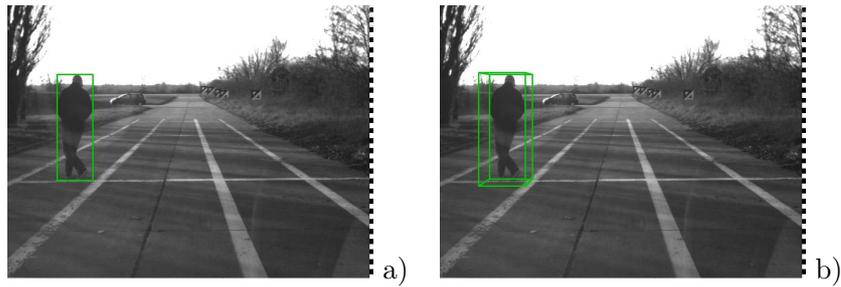


FIG. 4.6 – Détection représentée par sa zone d'intérêt (a) et son volume réel fronto-parallèle (b).

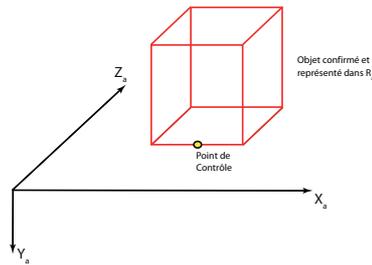


FIG. 4.7 – Définition du point de contrôle utilisé pour la mesure de position.

en conséquence de quoi la composante Y_a de la mesure de position ne présente pas d'intérêt fort. Ce point, que nous appellerons point de contrôle, est finalement défini par :

$$\begin{cases} X_a^p &= (X_a^{min} + X_a^{max})/2 \\ Y_a^p &= 0 \\ Z_a^p &= Z_a^{min} \end{cases} \quad (4.24)$$

Cette représentation, est stable pour les objets rigides, mais atteint ses limites dès lors que l'objet détecté est déformable, comme nous le verrons au cours de la prochaine section.

4.3.3 Mesure de l'orientation

Il est également possible d'accéder à une notion d'orientation. Pour cela, il faut revenir aux points obstacles et considérer leur positionnement dans le plan u-disparité. Une régression linéaire, par exemple, peut fournir l'orientation du nuage de points dans ce plan. Connaissant celle-ci, l'orientation réelle est directement fournie par l'équation (2.25).

La figure 4.8 donne l'exemple du calcul d'orientation d'un carton positionné de façon oblique par rapport au système stéréoscopique.

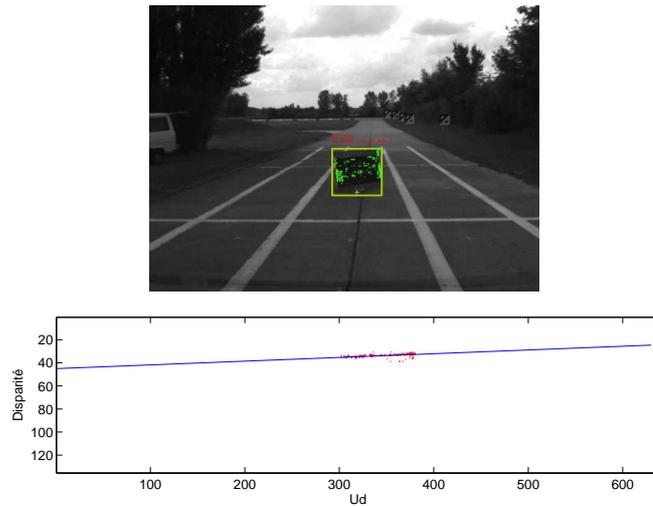


FIG. 4.8 – Mesure de l'orientation d'une cible d'après ses points obstacles.

Le carton est positionné de façon à former un angle de 45 degrés par rapport à l'axe X_a . La droite passant par les points obstacles au sens des moindres carrés a pour équation $\Delta_s = 34,02 - 0,032u_d$, ce qui correspond à un angle de 39 degrés par rapport à l'axe X_a .

4.4 Estimation de caractéristiques dynamiques

4.4.1 Nécessité d'un suivi temporel

Afin de calculer un temps à collision, que nous avons présenté en introduction comme étant un indicateur indispensable pour le développement de nombreuses aides à la conduite, il est nécessaire de procéder à un suivi des détections au cours du temps, les capteurs utilisés dans cette thèse n'étant pas à même de fournir directement une mesure de vitesse relative. Nous avons donc développé un algorithme de suivi multi-objets, qui présente l'intérêt de prendre en compte la nature des données stéréoscopiques traitées.

4.4.2 Principe de l'algorithme de suivi

L'algorithme de suivi que nous proposons est fondé sur une approche classique. Celle-ci repose sur 3 étapes distinctes [BP99] :

- la prédiction de la position future de l'objet à partir des observations et estimations précédentes,
- l'association de cette prédiction avec l'objet nouvellement observé qui semble être le même, cette étape assurant également la gestion de l'apparition et de la disparition d'objets,
- la correction de l'estimation sur la base de cette nouvelle observation.

Les phases d'estimation et prédiction sont réalisées par un filtre estimateur, le plus souvent un filtre de Kalman ou l'une de ses variantes non linéaires, type EKF³⁸. Divers auteurs y préfèrent l'utilisation d'un filtre particulière.

L'étape d'association a aussi donné lieu à différents travaux. Si dans la majorité des cas elle

³⁸Extended Kalman Filter

peut être réalisée facilement, en prenant l'objet le plus proche, certaines situations donnent lieu à des ambiguïtés. Gruyer propose dans sa thèse [Gru99] une méthode permettant de gérer ces ambiguïtés en utilisant la théorie de l'évidence.

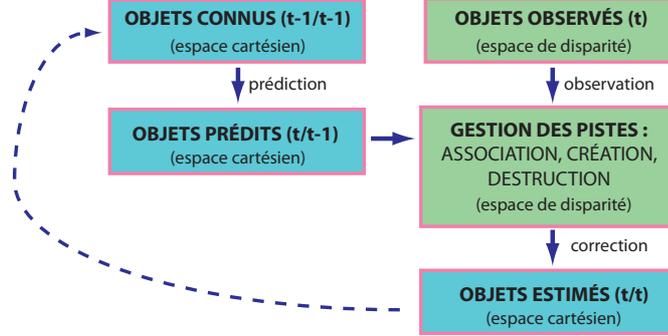


FIG. 4.9 – Principe de l'algorithme de suivi appliqué aux données stéréoscopiques.

4.4.3 Modèle d'évolution des objets détectés

Le modèle d'évolution des objets est un modèle linéaire simple, à vitesse constante. Il considère que l'objet suivi est toujours situé sur la surface de la route ($Y_a = 0$) et s'applique donc au point de contrôle que nous avons présenté précédemment. Le vecteur d'état du système étant $\mathbf{X} = (X_a, Z_a, V_{X_a}, V_{Z_a})^T$, l'évolution est régie par le système 4.25.

$$\begin{cases} X_a^{k+1} &= X_a^k + V_x^k \cdot dt \\ Z_a^{k+1} &= Z_a^k + V_z^k \cdot dt \\ V_{X_a}^{k+1} &= V_{X_a}^k \\ V_{Z_a}^{k+1} &= V_{Z_a}^k \end{cases} \quad (4.25)$$

4.4.4 Observation et association

Observation dans l'espace cartésien

L'approche la plus directe qui peut être envisagée pour l'observation consiste à exploiter une mesure représentée dans le plan ($Y_a = 0$) de R_a . Appliquée à des données lidar, cette méthode est bien adaptée, pour peu que l'on considère que le bruit de mesure est blanc, car la précision des données fournies par le lidar est indépendante de la distance mesurée. En revanche, la discrétisation de l'espace de disparité, non constante par rapport à E_a , pose problème si l'on désire appliquer cette stratégie aux données stéréoscopiques. En effet, l'imprécision sur les données stéréoscopiques est directement liée à la mesure elle-même : un demi pixel d'erreur correspond à quelques centimètres lorsque l'objet est près, et à plusieurs mètres lorsqu'il est éloigné du capteur. En raison de ce phénomène, un déplacement constant dans R_a , perçu comme tel par le lidar, donne lieu à un déplacement par palier à partir des données stéréoscopiques, ces paliers étant liés à la profondeur. Le fonctionnement correct d'un filtre de Kalman linéaire n'est par conséquent pas assuré pour des données stéréoscopiques converties en mesures métriques.

Observation dans l'espace de disparité

En revanche, le bruit de mesure est indépendant de la distance lorsque l'on travaille dans l'espace de disparité. Dès lors, il est préférable de procéder à l'observation directement dans cet

espace. L'état du système étant représenté dans E_a , l'observation est effectuée à travers l'équation de changement de repère (2.17). Celle-ci n'étant pas linéaire, c'est un filtre de Kalman étendu (EKF) qui est considéré. L'observation est donc linéarisée en entrée autour de sa valeur, par l'utilisation de la Jacobienne du système d'observation.

Association

L'association est réalisée entre la prédiction de l'état du système et l'observation. Elle est effectuée dans R_Δ . Pour cela, la position prédite $\hat{\mathbf{X}}_{k/k-1}$ de la piste en cours est projeté dans l'espace de disparité en utilisant le système d'équation (2.16), où θ_s et Y_s^0 sont fournis par l'estimation du profil de route à l'instant k , afin de tenir compte des variations d'assiette du véhicule. La dimension de la piste prédite à l'instant k est de plus dilatée par rapport à son estimation $\hat{\mathbf{X}}_{k-1/k-1}$ à l'instant $k-1$, de façon à tenir compte des incertitudes sur la prédiction. Si σ_x^2 et σ_z^2 sont les valeurs propres de la matrice de covariance de prédiction, la dilatation consiste à ajouter $3\sigma_x$ à la largeur de la cible et $3\sigma_z$ à sa profondeur.

L'association multi-objets entre l'ensemble des pistes prédites et des objets détectés est ensuite réalisée suivant les mêmes critères que l'association multi-capteurs présentée en perspective du chapitre 3.

4.4.5 Résultats

La méthode de suivi proposée permet, en l'état, de suivre correctement des objets ayant des trajectoires rectilignes pendant quelques secondes et d'estimer une vitesse relative. Les trajectoires plus complexes sont difficilement prises en compte par le modèle d'évolution linéaire utilisé. Cela est en deçà des performances qui peuvent être attendues pour le contexte routier. Une raison de cette faiblesse est la faible stabilité du point de contrôle utilisé. Si l'on prend l'exemple d'un piéton traversant la route, la position du point de contrôle relativement au corps du piéton varie lorsque celui-ci marche à cause de la position des bras et des jambes (figure 4.10). L'évolution de la position de ce point peut alors apparaître comme saccadée alors que le piéton marche à vitesse constante, comme illustré par la figure 4.11. L'amélioration future de la méthode passe donc par le remplacement du point de contrôle actuellement utilisé par une mesure qui soit insensible aux déformations des objets détectés.

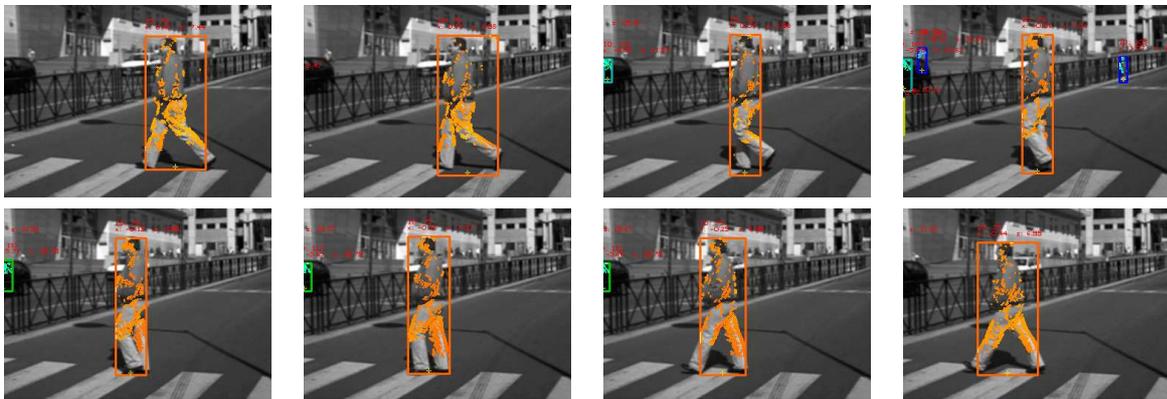


FIG. 4.10 – Limites du point de contrôle utilisé pour modéliser la position d'un objet déformable.

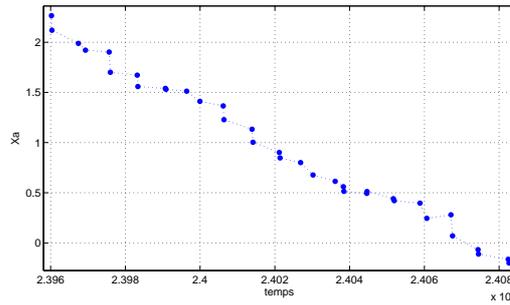


FIG. 4.11 – *Abscisse mesurée pour la position du piéton de la figure 4.10 qui traverse à vitesse constante. La déformation due à l'écartement et au resserrement des jambes laisse entrevoir des alternances de position fixe et de sauts, pouvant mesurer jusqu'à 20 cm.*

Une fois amélioré, le suivi des objets offre des perspectives d'amélioration de la précision et de la robustesse des détections. Mis à part l'estimation de la vitesse, il peut en effet être exploité pour :

- filtrer les mesures acquises aux instants successifs, pour construire une estimation débruitée de la position ou de la géométrie,
- assurer une continuité dans la détection en prédisant la position d'un objet sur la base des observations précédentes, lorsque celui-ci n'est momentanément pas détecté,
- permettre un fonctionnement robuste de l'application en construisant un critère de confiance sur la présence d'un obstacle en fonction du nombre et de la qualité des observations successives de celui-ci.

4.5 Résumé de l'Approche et Conclusion

Nous avons présenté dans ce chapitre une stratégie de validation des détections issues de différents capteurs. Cette approche permet, par l'utilisation de la stéréovision de façon locale, d'analyser finement les détections préalablement réalisées, et d'évaluer trois critères de validité qui assurent que les hypothèses qui ont été posées sur la nature des obstacles sont respectées. Le caractère local de la méthode proposée permet, entre autres, d'assurer une utilisation efficace du temps de calcul disponible en limitant les traitements à des zones d'intérêt. Pour le cas où cela ne serait pas suffisamment efficace, en particulier pour l'utilisation d'images de très grande résolution, nous proposons également l'exploitation d'un zoom numérique local qui optimise, pour chaque région, sa résolution et en conséquence les ressources nécessaires à son traitement. Par ailleurs, les points de mesure fournis par la stéréovision locale permettent de caractériser les détections en leur associant des mesures de position et de géométrie. En outre, l'ajout d'une étape de suivi permet de plus d'obtenir une information de vitesse, indispensable au calcul d'un TTC.

La figure 4.12 offre une vue synoptique de l'ensemble de ces étapes de validation et de caractérisation, et récapitule ainsi les différentes tâches intervenant, ainsi que leur ordonnancement. Nous allons dans le chapitre suivant évaluer la validité de l'approche sur des cas réels de situations routières.

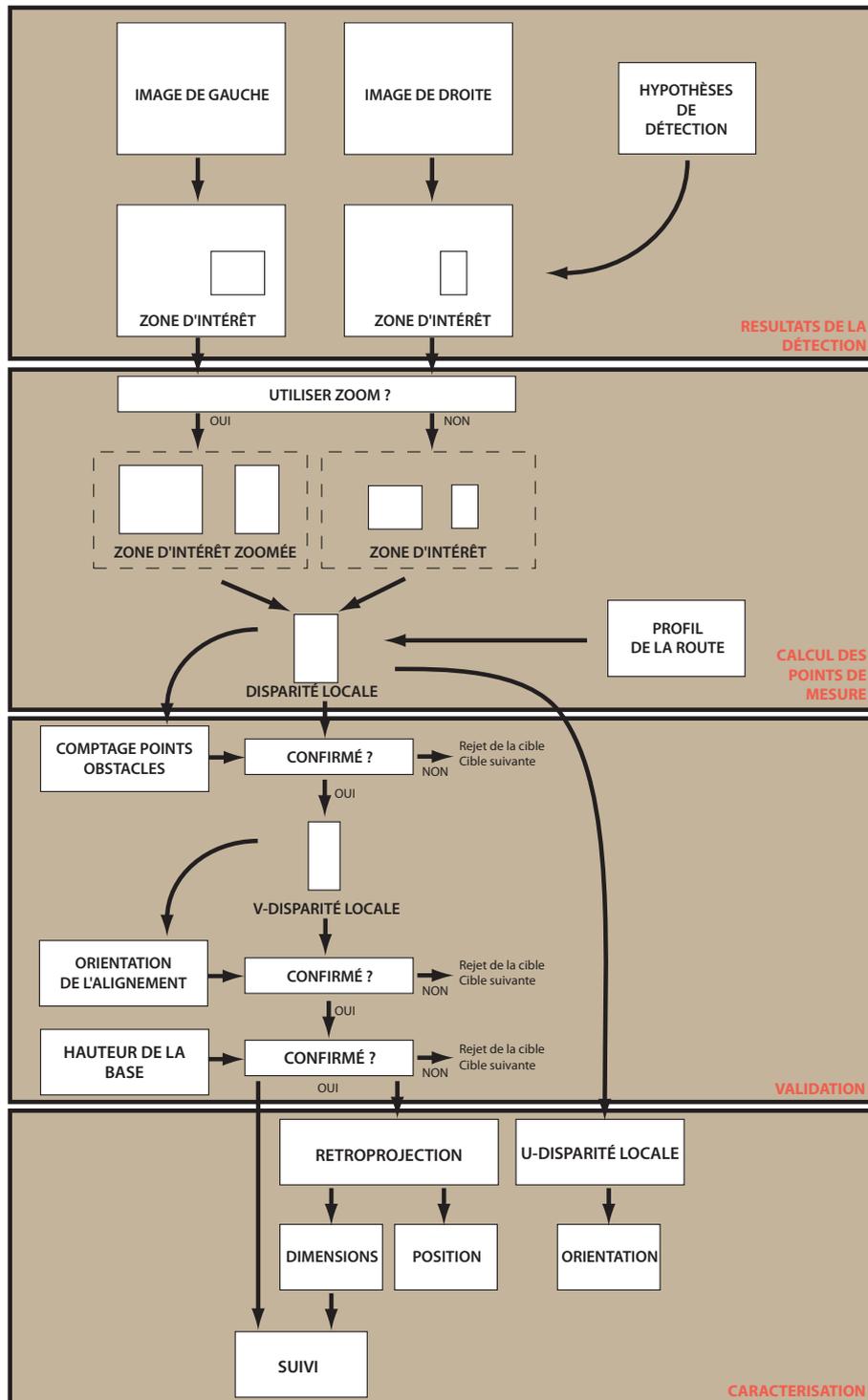


FIG. 4.12 – Vue d'ensemble de l'algorithme de confirmation des détections.

EVALUATION EXPÉRIMENTALE ET APPLICATIONS

Sommaire

5.1	Capteurs utilisés	102
5.1.1	Moyens expérimentaux du LIVIC	102
5.1.2	Capteurs stéréoscopiques	103
5.1.3	Télémètre laser	104
5.1.4	Capteur d'identification optique	104
5.1.5	Calibrage et configuration des différents capteurs	104
5.2	Evaluation des performances de détection en situation routière . .	104
5.2.1	Performances de détection par stéréovision	105
5.2.2	Performances de détection par télémétrie laser	112
5.2.3	Performances de détection par identification optique	114
5.3	Evaluation statique des performances	115
5.3.1	Evaluation comparée de la précision de création d'hypothèses	115
5.3.2	Evaluation de la caractérisation d'un piéton	118
5.3.3	Evaluation de l'intérêt du zoom numérique	120
5.4	Applications aux aides à la conduite	121
5.4.1	Détection d'obstacles par stéréovision multi-résolution	121
5.4.2	Détection par stéréovision et télémétrie laser : de la mitigation de collisions au "stop & go" et suivi automatique	123
5.4.3	Enrichissement de la détection par perception coopérative	126
5.5	Conclusion	127

LE chapitre 1 a permis, après un rapide état de l'art des capteurs et des méthodes utilisés pour la détection d'obstacles, de positionner notre travail en proposant une méthodologie de détection qui suppose une coopération entre différents capteurs. Celle-ci repose sur deux étapes essentielles : la création d'hypothèses de détection, puis la validation et la caractérisation *a posteriori* de celles-ci. La création d'hypothèses a été abordée au cours du chapitre 3, à travers l'utilisation d'un capteur stéréoscopique, d'un télémètre laser à balayage 2D, et d'un capteur coopératif d'identification optique. Le chapitre 4, quant à lui, était dévolu à l'étape de validation et de caractérisation des obstacles par la stéréovision, utilisée localement dans des zones d'intérêt.

Ce chapitre est consacré à l'évaluation des différentes briques qui viennent d'être proposées et d'une façon générale de la méthodologie de détection proposée. Cette évaluation ne se veut pas exhaustive, mais vise à valider la démarche sur des cas concrets. Elle repose à la fois sur la mesure d'indicateurs objectifs de performances, comme un taux de fausses alarmes ou une portée, et sur l'appréciation subjective de la qualité globale de la détection. Pour ce dernier point, l'attention sera portée sur l'utilisation de données réelles correspondant à des cas classiques de conduite.

Avant de rentrer dans les détails de l'évaluation proprement dite, nous allons détailler les caractéristiques des capteurs utilisés.

Nous proposerons ensuite une étude des performances de détection pour chacune des trois techniques de création d'hypothèses que nous avons abordées au chapitre 3. Cette étude est réalisée sur des séquences réelles d'autoroute, ville et route nationale. Celle-ci permettra également de juger de l'efficacité de la validation des hypothèses par stéréovision.

Après cela, nous comparerons de façon quantitative les portées et précisions obtenues lors de la détection avec les différents capteurs. Ceci est réalisé par des mesures statiques sur le banc de calibrage du site de Satory.

Enfin, nous proposerons trois mises en application de la méthode proposée. La première vise à réaliser un système de détections d'obstacles par stéréovision seule, en visant à optimiser le temps de calcul nécessaire. La seconde consiste à utiliser le couple lidar/stéréovision dans des applications d'aide à la conduite. Le système de mitigation de collision existant au LIVIC est ainsi étendu pour permettre également des modes de fonctionnement de type "stop & go" et suivi automatisé. Enfin, nous proposons de mettre en œuvre l'identification optique pour faire coopérer la détection d'obstacles avec les télécommunications inter-véhiculaires, et ainsi illustrer pleinement le concept de coopération entre véhicules "intelligents".

5.1 Capteurs utilisés

La plupart des données utilisées ont été obtenues à partir des véhicules instrumentés du LIVIC. D'autres ont été fournies par des partenaires dans le cadre de projets de recherche.

5.1.1 Moyens expérimentaux du LIVIC

Parmi ses moyens expérimentaux, le LIVIC dispose de plusieurs véhicules instrumentés, équipés de moyens de perception, de calcul et d'action. La figure 5.1 présente trois d'entre eux : VIPER (Véhicule Instrumenté pour la Perception de l'Environnement Routier), CARLLA (Contrôleur d'Assistance Routière Longitudinale et LATérale) et LOLA (LONgitudinal-LATéral), tous trois

caméra	résolution	technologie	focale	α_u	α_v	fréquence
SONY	768*576	CCD	8,5 mm	1100	1100	12,5 Hz
SONY	384*288	CCD	8,5 mm	550	550	25 Hz
Cypress SMaL	640*480	CMOS	6 mm	850	850	30 Hz
Jai Pulnix TM-6740	640*480	CCD	12 mm	1620	1620	200 Hz
Jai Pulnix TM-6740	320*120	CCD	12 mm	810	405	514 Hz

TAB. 5.1 – Caractéristiques des différentes caméras utilisées dans cette thèse.

dédiés à la perception et au contrôle automatique. Nous avons utilisé pour cette thèse VIPER et CARLLA, qui ont des configurations de capteurs relativement proches. Ces deux véhicules sont équipés d'une part d'un PC et d'un ensemble d'actionneurs dédiés au contrôle du véhicule, et d'autre part de nombreux capteurs associés à un second PC pour traiter les tâches de perception.



FIG. 5.1 – Trois des véhicules expérimentaux du LIVIC. De gauche à droite : VIPER, CARLLA, LOLA.

Le LIVIC dispose également d'un accès à des pistes d'essais pour ses tests. Certaines des séquences de données exploitées ici ont été produites sur ces pistes.

Enfin, le simulateur SIVIC peut être utilisé pour créer des bases de test associées à une vérité terrain.

5.1.2 Capteurs stéréoscopiques

Si certains systèmes stéréoscopiques sont fournis complets et assemblés, nous nous sommes plutôt intéressé, dans cette thèse, à l'utilisation d'un système "fait maison" à partir de deux caméras et d'un PC pour le traitement. Cela permet d'envisager différents types de caméras et de positionnement. Dès lors, de nombreuses possibilités de configuration sont envisageables, selon que l'on modifie le type de capteurs, leur tailles, les optiques ou la base stéréoscopique. L'annexe A présente l'influence de ces différents paramètres.

La possibilité d'exploiter plusieurs véhicules et l'accès à d'autres bases de données dans le cadre de projets nous a permis d'évaluer les performances de la méthode proposée pour des configurations assez différentes, qui constituent des hypothèses plus ou moins réalistes d'intégration future par des constructeurs. Celles-ci sont présentées dans le tableau 5.2. Les paramètres des caméras sont pour leur part reportés dans le tableau 5.1. Ces différentes configurations ont été définies pour disposer d'un angle de vue supérieur à 40 degrés, d'une portée minimale de 3 m et d'une erreur sur la distance relative à 30 m qui soit inférieure à 2 m.

TAB. 5.2 – Configuration du capteur stéréoscopique pour les différents véhicules ayant servi à l'évaluation.

Véhicule utilisé	modèle de caméras	base stéréoscopique	tangage au repos	hauteur au repos
VIPER	SONY (PAL)	1,03 m	5 deg	1,35 m
VIPER	SONY (1/4PAL)	1,03 m	5 deg	1,35 m
CARLLA	SMaL	1,04 m	7 deg	1,4 m
Renault	SMaL	0,65 m	-5 deg	1,3 m
PSA	SMaL	0,50 m	8 deg	1,4 m

5.1.3 Télémètre laser

Le lidar utilisé, équipe VIPER et CARLLA, est un SICK LMS 210. Il balaye un plan sur un angle de vue de 100 degrés en fournissant 201 mesures de distance toutes les 16 ms. Sa précision, de l'ordre de quelques centimètres, est indépendante de la distance. Enfin, sa portée est de 75 m pour les objets les plus réfléchissants.

5.1.4 Capteur d'identification optique

Le récepteur utilisé pour l'identification optique est une caméra PULNIX TM-6740, équipée d'une optique de 12 mm et d'un filtre passe bande centré sur 880 nm et fonctionnant à une fréquence de 514 Hz. Les choix qui ont conduit à l'utilisation d'une telle caméra sont précisés dans l'annexe E.

La lampe émettrice est configurée de façon à émettre 210 bits par seconde et transmettre des messages de 16 bits de longueur.

5.1.5 Calibrage et configuration des différents capteurs

Afin de pouvoir comparer les résultats obtenus par les trois capteurs, il est indispensable que leur position respective par rapport au référentiel R_a soit connue. Pour cela, une étape préalable de calibrage et de configuration est nécessaire. Celle-ci est décrite dans l'annexe D.

5.2 Evaluation des performances de détection en situation routière

Dans un premier temps, nous allons juger du fonctionnement du système proposé en conditions réelles. Ceci est réalisé à travers l'étude de scènes routières, acquises pour la plupart sur route ouverte. Nous proposons pour cette étude de mesurer deux critères objectifs, à savoir les taux de détection et de fausses alarmes lors de la détection des différents objets présents. Définissons donc ces deux critères :

- **Le taux de détection** correspond au nombre d'objets correctement détectés rapporté au nombre d'objets détectables présents dans la scène. Précisons qu'un véhicule présent dans N images successives est considéré pour ce critère comme N objets différents.

- **Le taux de fausses alarmes** est le rapport entre le nombre de fausses détections réalisées et le nombre d’images traitées dans la séquence.

Ces critères seront évalués sans, puis avec l’étape de validation des hypothèses de détection, afin de juger de la pertinence de l’approche proposée. D’autre part, cette évaluation est réalisée indépendamment pour chacun des capteurs, la plupart des séquences d’évaluation n’étant pas acquises en utilisant conjointement les trois capteurs proposés.

En outre, et en l’absence de vérité terrain associée aux séquences de test, nous proposons de compléter cette évaluation par des impressions subjectives sur les performances obtenues avec la méthode proposée.

5.2.1 Performances de détection par stéréovision

Présentation des séquences

Afin d’évaluer le fonctionnement de la stéréovision dans un grand nombre de situations de conduite, nous avons retenu trois séquences de test correspondant à des contextes variés et réalistes. En outre, chacune d’entre elles a été acquise avec une configuration donnée du capteur stéréoscopique, ce qui permet de vérifier que la fiabilité de fonctionnement du système reste relativement indépendante de celle-ci.

Les trois séquences sont illustrées par les figures 5.2, 5.3 et 5.4, et la configuration du capteur pour chacune d’entre elles est reportée dans le tableau 5.2.

La première séquence est issue du projet DO30 et a été fournie par le constructeur PSA. Elle illustre 6 minutes de roulage sur autoroute et couvre la plupart des situations de conduite dans ce contexte. Elle permet notamment d’observer : des véhicules de gabarit différents (voitures, camions, motos), des objets non frontaux (barrières de sécurité, parois latérales des camions), des manœuvres de suivi et de dépassement, et des masquages d’objets. Malheureusement, en raison d’un problème électrique lors de l’acquisition, certaines images sont fortement déformées. De plus, la géométrie épipolaire n’est pas parfaitement rectifiée au niveau des objets les plus proches. De fait, les conditions de fonctionnement de l’algorithme ne sont pas optimales. Les objets jugés détectables dans cette séquence sont les véhicules présents sur la voie du véhicule instrumenté et sur les autres voies, pour une distance variant de 0 à 50 m.

La seconde séquence a été fournie par Renault pour les besoins du projet LOVE. Elle correspond à un contexte urbain très complexe pour les algorithmes de stéréovision. En effet, une grande variété d’objets est présente : piétons, voitures, motos, barrières, plots, dos d’ânes... En outre, les principales difficultés sont la présence d’un grand nombre d’objets très répétitifs, en particulier les façades d’immeubles, et de nombreux masquages partiels. Les performances sont évaluées pour la détection de piétons présents dans notre voie de circulation et sur la totalité de l’image, pour une portée inférieure à 50 m.

La troisième séquence, quant à elle, a été acquise à l’aide de CARLLA. Son contexte est majoritairement de la route nationale, ce qui implique de nombreux objets autour de la route (panneaux, végétation, véhicules...), ainsi que des situations de croisements, ronds points, ou insertion. La séquence comprend également une partie d’autoroute et de ville. Elle se distingue par ailleurs des deux premières séquences par l’utilisation d’une large base stéréoscopique (1,04 m). Compte tenu du nombre très important d’objets présents dans cette séquence, nous n’avons considéré pour l’évaluation du taux de détection que les objets présents directement en avant de

notre véhicule, sur une portée de 50 m.

Paramétrage de l'algorithme

Pour que l'évaluation à travers ces différentes séquences soit rigoureuse, il est nécessaire que l'algorithme soit paramétré à l'identique pour chacune d'elles. Seuls les paramètres de configuration du capteur sont donc modifiés. Les paramètres relatifs à la construction des points de mesure ont été fixés à la fin du chapitre 2. Les paramètres à fixer sont donc la taille des *bins* des histogrammes et les seuils associés pour la détections. Ils ont été choisis de façon à assurer une détection d'obstacles générique. Les *bins* utilisés mesurent donc 3 pixels suivant l'axe $\vec{\Delta}$, 50 suivant l'axe \vec{u}_d et 10 suivant \vec{v} .

Carte de disparité

Sur les figures 5.2, 5.3 et 5.4 sont présentées les cartes de disparités obstacles obtenues pour les trois séquences. Celles-ci sont très éparées. En particulier, le filtrage des points obstacles a permis de supprimer tous les pixels isolés. Les pixels restants forment pour la plupart des regroupement autour des indices visuels verticaux présents sur les objets. Dans le cas des piétons, les vêtements, ombres et traits du visage permettent alors d'obtenir une certaine densité de points sur la silhouette. Sur les voitures, en revanche, même si un grand nombre d'éléments permettent de réaliser des appariements (roues, phares, plaque d'immatriculation, ...), de grandes zones uniformes sont également présentes. De près, les voitures apparaissent donc souvent comme plusieurs objets verticaux.

Les réels contours verticaux des objets sont présents pour leur majorité. Cependant, lorsqu'au niveau de ces contours, se trouve en arrière plan une texture plus marquée que sur l'objet, et différente selon le point de vue, les points obstacles sont absents. Ce phénomène peut être observé pour le piéton en blanc passant devant une façade d'immeuble sur la seconde paire d'images de la figure 5.3, alors que les appariements sont réussis sur le piéton portant un pull rayé. Le bas des objets est quant à lui amputé des 20 cm de tolérance donnée autour du profil de la route.

Certains faux appariements persistent malgré les filtrages. C'est en particulier le cas pour des motifs répétitifs de grande taille, comparée à la fenêtre de corrélation, comme les immeubles de la séquence urbaine ou les poteaux bordant l'autoroute sur la figure 5.2. Ils peuvent alors former de grandes zones pouvant être interprétées comme des objets.

Enfin, les plans non fronto-parallèles sont rejetés avec succès lorsqu'ils sont longs ou très marqués. La barrière de la figure 5.3 illustre l'exemple d'un plan correctement supprimé, profitant pleinement à la qualité de détection. En revanche, quelques marques sur une glissière d'autoroute en béton ne suffisent pas à détecter celle-ci comme un plan.

Estimation de la surface de la route

Pour sa détection, la route est considérée comme un plan sur les 25 m en avant du véhicule. Elle est bien détectée lorsque des marquages, ombres ou autres indices visuels sont présents sur sa surface. Dans le cas contraire, un profil par défaut est utilisé.

La détection est rendue stable par la stratégie qui consiste à n'utiliser que des pixels répondant mieux à la fenêtre de corrélation inclinée qu'à la fenêtre droite. En effet, sans cette méthode, l'estimation est parfois faussée par la présence d'un grand nombre de points obstacles.

Il demeure cependant encore de rares situation dans lesquelles le profil estimé fluctue anormalement, par exemple en ville lorsqu'il y a ambiguïté avec un trottoir ou une barrière basse et horizontale. Dans la séquence d'autoroute, la mauvaise qualité des images conduit également à

une telle fluctuation.

C'est pour des distances plus importantes que cette méthode présente ses limites, puisque l'hypothèse du monde plan est alors souvent fautive. La reconstruction plus précise de la route est donc l'une des perspectives d'améliorations les plus immédiates à considérer pour profiter d'une portée élevée, comme cela a été présenté dans [Lab03] ou [Lem05b].

Détection

Les résultats de détection sont très cohérents avec les cartes de disparité obtenues. Les remarques qui ont pu être faites lors de l'étude de ces images se retrouvent donc dans la détection. Comme le montrent les tableaux 5.3, 5.4 et 5.5, de bonnes performances de détection sont atteintes. Tous les objets présents dans la carte de disparité sont effectivement détectés lors de la création des candidats. Cela assure un bon taux de détection mais fait apparaître un certain nombre d'erreurs dues à de faux appariements ou à une mauvaise estimation du profil de la route. A l'inverse, les objets sur lesquels les appariements ont globalement échoué ne sont pas détectés, et le taux de détection n'atteint donc pas les 100%.

Si l'on essaie de modifier les seuils de détection pour supprimer les erreurs, le taux de détection baisse rapidement. Espérer obtenir un taux d'erreur nul après l'étape de segmentation entraîne en conséquence une perte de détection très importante. C'est là tout l'intérêt de l'étape de confirmation qui permet de conserver un taux de détection élevé en rejetant efficacement la plupart des erreurs. Quelques fausses détections ne sont cependant pas rejetées, et correspondent à des cas plus difficiles, lorsque 2 types d'erreurs sont présentes dans la même zone d'intérêt (par exemple de faux appariements corrélés et des points obstacles sur la route). Notons tout de même que les erreurs sont pour la plupart ponctuelles et ne se réalisent donc pas sur des images successives.

Les résultats obtenus semblent donc confirmer que la stratégie de détection proposée est bien adaptée à la problématique de l'environnement routier. Elle permet de plus de compenser partiellement la faiblesse du modèle de route utilisé. En effet, la séquence mixte permet de constater que lorsqu'une surface de route longitudinalement concave provoque de fausses détections, celles-ci sont bien rejetées par le critère de l'alignement. Cela est visible pour la seconde paire d'images de la figure 5.4

TAB. 5.3 – Performances de la détection par stéréovision pour la séquence "autoroute". Le taux de détection est meilleur sur les voies latérales car il correspond à des dépassements de véhicules, et donc à une portée de détection plus faible.

nombre d'images	Fausses détections		Voie de circulation		Autres voies	
	rejetées	confirmées	détections	détections confirmées	détection	détections confirmées
10662	5,4%	0,07%	85,5%	85,1%	94,2%	90,0%

Concernant **la segmentation de l'espace**, l'algorithme fonctionne correctement pour les piétons puisque, mis à part les cas où les appariements ne sont pas possibles, les boîtes englobantes sont relativement bien positionnées et dimensionnées sur ceux-ci. Il en est de même pour les deux roues, sauf lorsqu'ils ont une trajectoire de biais par rapport à celle du véhicule intrumenté, puisqu'ils sont alors éventuellement supprimés comme plans non fronto parallèles.

La plus grosse difficulté concerne les voitures, souvent scindées en deux objets, en raison du faible nombre de points obstacles présents au centre de celles-ci (fig 5.3). Parfois même, ce peut également être deux demies voitures côte à côte qui sont rassemblées. Cette problématique est particulièrement visible de près, les véhicules étant généralement complètement détectés à plus grande distance. Une plus grande densité de la carte de disparité, ou du moins des appariements sur les contours horizontaux, est nécessaire pour résoudre ce problème tout en restant générique vis-à-vis des obstacles à détecter.

TAB. 5.4 – Performances de la détection par stéréovision pour la séquence "ville".

nombre d'images	Fausses détections		Piétons dans toute l'image		Piétons sur notre voie	
	rejetées	confirmées	détections	détections confirmées	détection	détections confirmées
4080	5,4%	0,1%	70,5%	68,2%	92,0%	91,9%

La continuité des détections a également son importance. Il n'y a pas en effet d'intérêt à détecter un objet sur une seule image, puis le perdre pour la suite de la séquence. Dans les séquences de test, un objet détecté dans notre voie est en général bien détecté dans les images suivantes. En pratique, le principal problème de continuité concerne moins la détection que la géométrie des objets. Nous avons vu que, de près, une voiture pouvait être perçue comme deux ou trois objets (les deux bords et la plaque d'immatriculation), et de loin comme un objet unique. A une distance intermédiaire, il y a donc une plage de distance sur laquelle la géométrie de la voiture peut fluctuer entre ces deux représentations.

TAB. 5.5 – Performances de la détection par stéréovision pour la séquence "mixte".

nombre d'images	Fausses détections		Voie de circulation	
	rejetées	confirmées	détections	détections confirmées
5524	3,6%	0,3%	96,5%	95,1%

Enfin, **le suivi des détections** permet d'estimer des vitesses pour les objets, mais l'étape d'association n'est pas assez performante pour maintenir un suivi pendant une longue durée. C'est en particulier la géométrie des cibles qui en est responsable. D'une part, lorsqu'un véhicule est séparé en deux objets, puis de nouveau aggloméré, puis séparé à nouveau, son suivi n'est pas possible correctement avec la technique proposée. D'autre part, nous avons vu au chapitre 4 que la déformation d'un piéton dans son déplacement peut nuire au suivi. Ces aspects mettent en évidence le travail restant à réaliser en perspective de cette thèse pour améliorer le suivi.

5.2. Evaluation des performances de détection en situation routière



FIG. 5.2 – Résultats obtenus par stéréovision pour la séquence "autoroute".
De haut en bas : images de droite, images de gauche, images de disparité, zones d'intérêt détectées, détections confirmées.

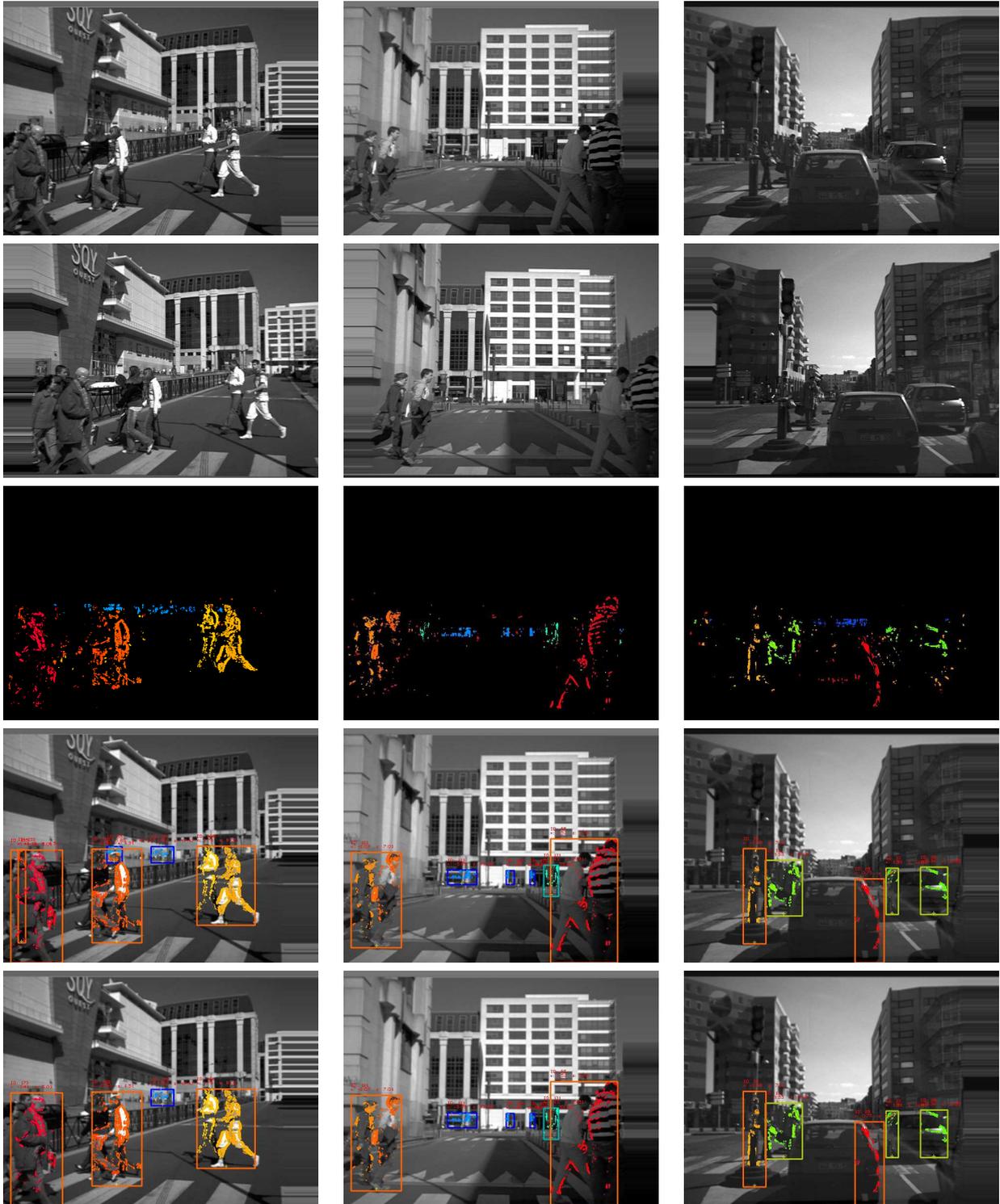


FIG. 5.3 – Résultats obtenus par stéréovision pour la séquence "ville".
De haut en bas : images de droite, images de gauche, images de disparité, zones d'intérêt détectées, détections confirmées.

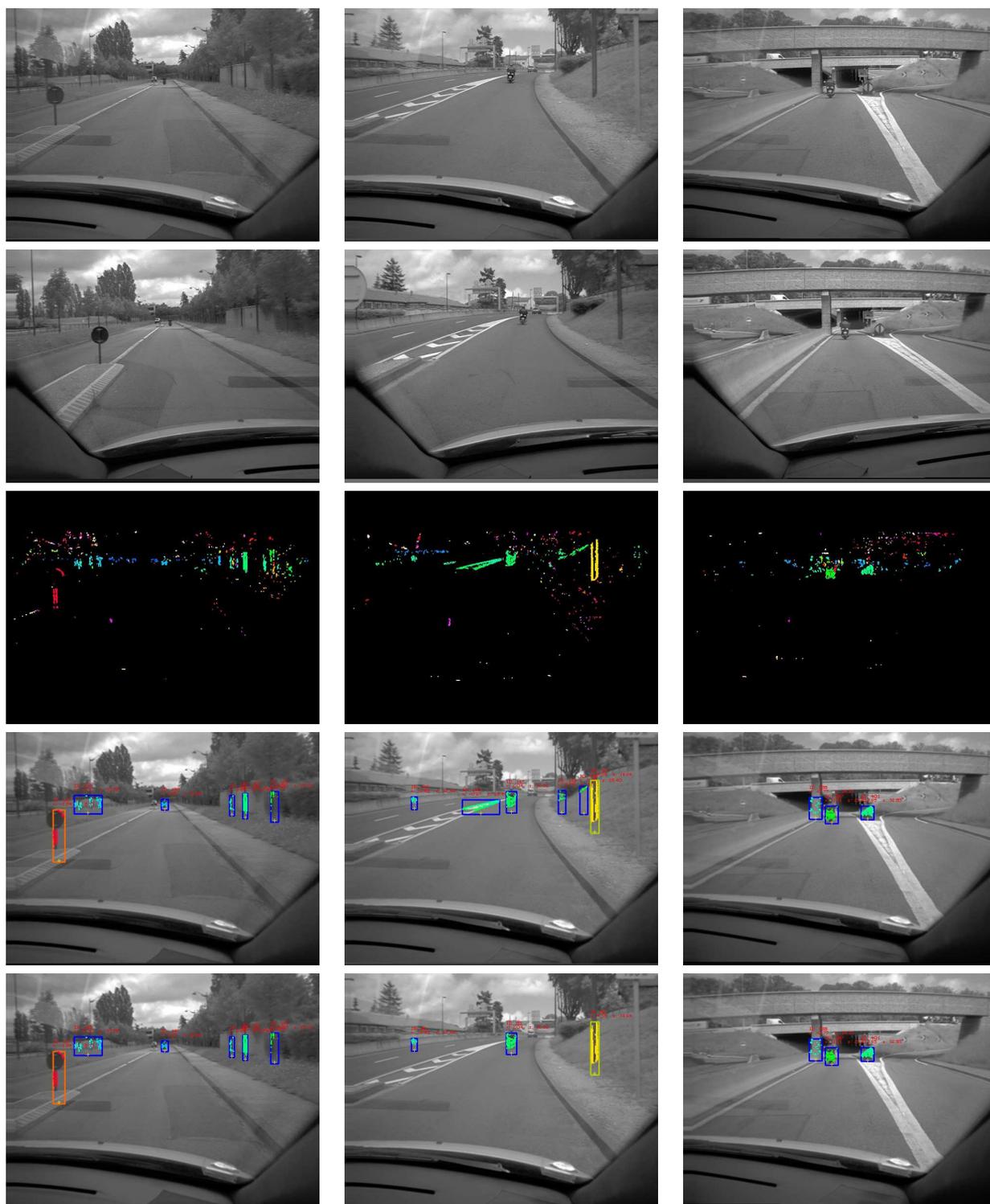


FIG. 5.4 – Résultats obtenus par stéréovision pour la séquence "mixte".
De haut en bas : images de droite, images de gauche, images de disparité, zones d'intérêt détectées, détections confirmées.

5.2.2 Performances de détection par télémétrie laser

Présentation des séquences

Nous allons maintenant nous intéresser brièvement aux résultats de détection à l'aide du lidar. Pour cela, nous avons choisi deux séquences, acquises avec VIPER, et illustrées par la figure 5.5.

La première a été enregistrée lors du suivi automatisé d'un véhicule sur la piste routière du site de Satory.

La seconde séquence a été réalisée en roulant sur une aire de parking riche en creux et en bosses, de façon à ce que la variation de tangage du véhicule soit importante. Le lidar crée ainsi un très grand nombre de fausses détections, nous permettant d'évaluer assez facilement l'étape de validation des détections.

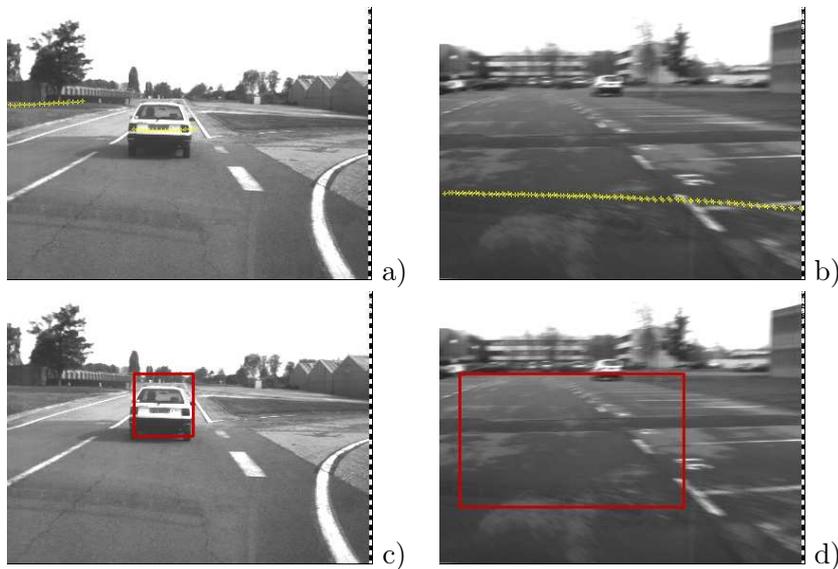


FIG. 5.5 – Illustration des deux séquences de test utilisées pour la détection lidar.

a) impacts projetés dans l'image de droite de la paire stéréoscopique. b) zones d'intérêt résultantes.

Résultats

Continuité de la détection : L'étude de la première séquence met clairement en évidence la faible constance de la détection par le lidar en roulage, si l'on n'utilise pas de suivi sur les détections. Le plan de détection du laser a en effet de fortes chances de passer au dessus ou au dessous de l'obstacle, ne permettant donc pas de le détecter. La figure 5.6 illustre cela en représentant l'évolution de la distance mesurée à l'obstacle en fonction du temps. Il est notable que la courbe initialement très discontinue, est rendue plus continue par l'étape de prédiction intégrée dans le suivi des objets. En revanche, des erreurs de prédictions peuvent apparaître, qui doivent être rejetées lors de l'étape de validation.

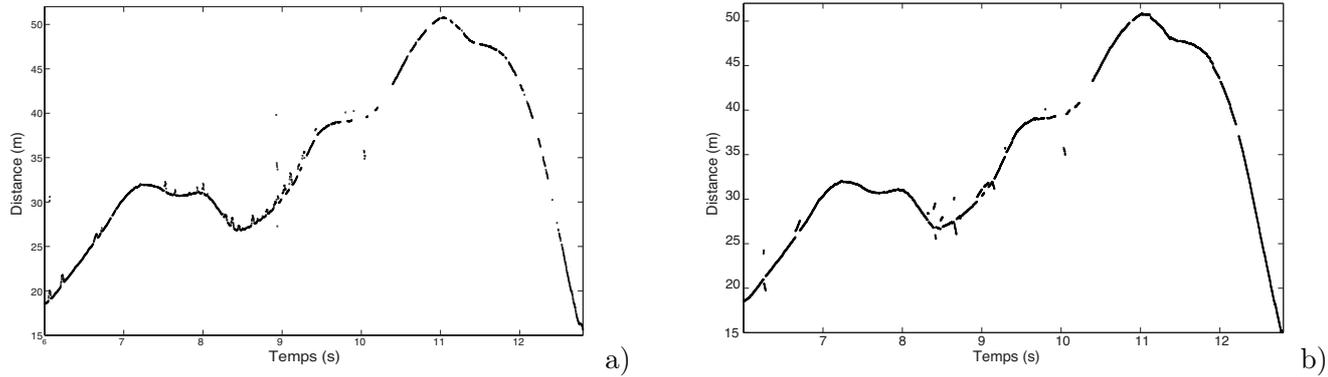


FIG. 5.6 – Intérêt du suivi lors de la création d'hypothèses par le lidar.

a) Distance mesurée par le télémètre lors du suivi d'un véhicule. La discontinuité de la courbe, en particulier à grande distance, illustre le besoin d'un suivi des objets au cours du temps. b) Distance estimée lors du couplage du télémètre avec un algorithme de suivi. La plupart des discontinuités sont alors comblées par la prédiction de la position de l'obstacle.

Rejet des fausses détections : Afin de vérifier l'efficacité de l'étape de validation des hypothèses fournies par le lidar, la seconde séquence contient un très grand nombre de fausses détections dues au fort tangage du véhicule. Le tableau 5.6 récapitule les résultats du traitement de cette séquence de 7032 images. Seuls les critères *points obstacles* et *alignement principal* sont pris en compte pour cette évaluation, car le critère de la *hauteur de la base* n'est pas pertinent pour la validation d'hypothèses de détection issues du lidar.

TAB. 5.6 – Nombre de fausses détections pour la séquence "parking", avec le seul lidar, et après validation par stéréovision.

	hypothèses non confirmées	<i>Points Obstacles</i>	<i>Alignement Principal</i>
Fausse alarmes (Taux)	781 (11,1%)	3 (0,04%)	10 (0,14%)

Tout comme pour le traitement de zones d'intérêt fournies par la stéréovision, l'étape de confirmation prouve son efficacité sur les données lidar, les fausses détections étant presque toutes rejetées. La plupart des erreurs restantes correspondent pour la plupart à des cas ambigus, comme la détection de la route juste devant un obstacle.

5.2.3 Performances de détection par identification optique

Présentation des séquences

Afin d'évaluer le fonctionnement de l'identification optique en situation réelle, nous avons traité une séquence de 16 minutes de roulage sur autoroute, en suivant un véhicule équipé d'un émetteur infrarouge. Le temps est beau et le trafic assez chargé. Les images de la figure 5.7 illustrent trois instants de cette séquence.

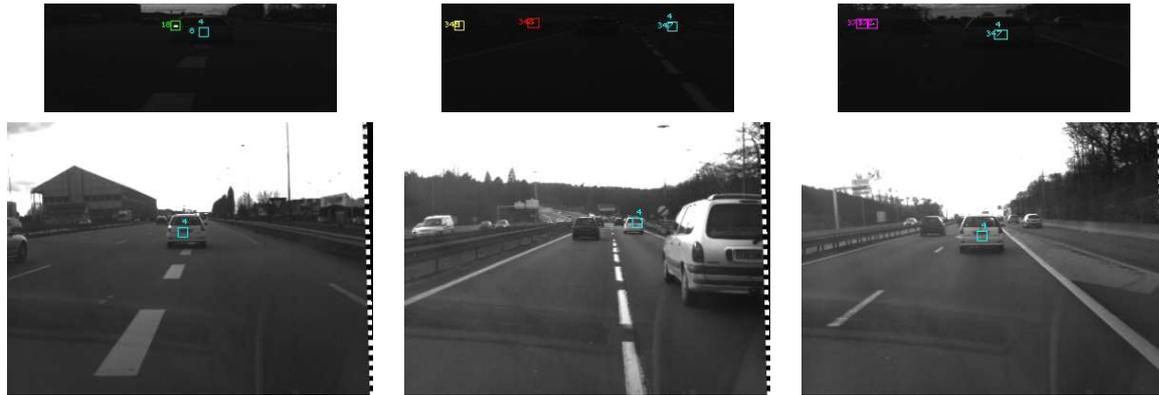


FIG. 5.7 – Illustration de la séquence de test utilisée pour la détection par identification optique. Dans les images infrarouge, seuls les rectangles cyan correspondent à des codes cohérents.

Résultats

Le tableau 5.7 regroupe les résultats de détection obtenus lors du traitement de cette séquence. Parmi le très grand nombre de sources infrarouge détectées, autres que la lampe, cinq ont été perçues comme des émetteurs respectant le protocole de communication et ont donc donné lieu à une fausse détection. La très courte durée de celles-ci, comparée à la fréquence de traitement de la stéréovision, fait que seules deux d'entre elles ont été considérées comme des hypothèses de détection.

La détection de la lampe est quant à elle continue et assure un taux de détection maximal, sauf

TAB. 5.7 – Fausses détections lors d'un scénario de roulage sur autoroute.

Distance parcourue	21 km
Fausses détections	5
Durée moyenne d'une fausses détection	28 ms
Taux de fausses détections (nb/minute)	0.31
Taux de fausses détections (nb/km)	0.24

en cas de masquage ou d'éblouissement. Lorsque l'émetteur apparaît à nouveau, l'identification est obtenue avec une latence moyenne de 100 ms.

Les résultats sont donc encourageants, d'autant que l'allongement du code identifiant peut diminuer le nombre d'erreurs. Néanmoins, un type d'erreurs pose particulièrement problème. Il

est illustré par les images de la figure 5.8. Dans ce cas, l'émetteur se reflète sur la carrosserie d'un autre véhicule, faisant apparaître le reflet comme une source cohérente. Deux identifiants identiques sont donc détectés dans l'image, ce qui est théoriquement impossible.

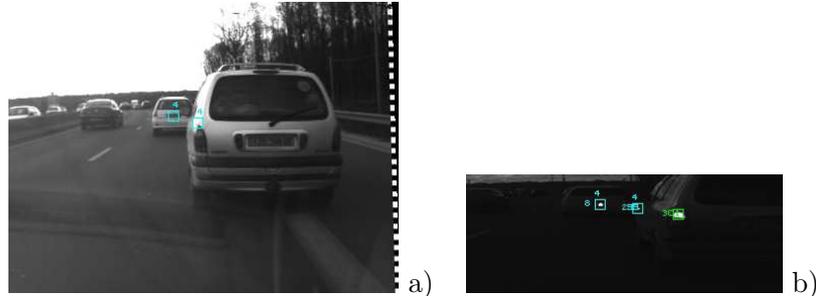


FIG. 5.8 – Lorsque la lampe émettrice se reflète sur une surface réfléchissante, ici la carrosserie d'un autre véhicule, le reflet peut être vu comme une autre source transmettant le même code.

L'étude qualitative de la localisation de l'identifiant consiste à observer la projection de celui-ci dans l'image de droite de la paire stéréoscopique. Il doit alors être localisé au centre du véhicule émetteur. L'observation démontre qu'il est presque toujours sur la face arrière du véhicule. Mais à grande portée, lorsque la route n'est pas plane, il est parfois mal positionné, comme le montre l'image centrale de la figure 5.7.

5.3 Evaluation statique des performances

Outre l'aspect qualitatif des performances qui vient d'être évoqué à travers des situations réelles de conduite, nous avons tenu à proposer également une approche plus quantitative de leur évaluation. Pour cela, plusieurs séquences ont été réalisées en plaçant le véhicule VIPER dans une position fixe sur le banc de calibrage du site de Satory, et en procédant à la détection d'objets connus (voiture et piétons) à des positions connues. Ainsi, il est possible d'évaluer et de comparer les portées et précisions obtenues avec les différents capteurs mis en œuvre. Nous avons également jugé de la qualité de la caractérisation des détections par stéréovision en s'intéressant au cas particulier d'un piéton.

5.3.1 Evaluation comparée de la précision de création d'hypothèses

Portée et positionnement longitudinal

Dans un premier temps, nous avons tenu à comparer la portée et la précision longitudinale de la création d'hypothèses de détection par les trois capteurs proposés. Pour cela, un véhicule équipé d'un topomètre s'écarte le long d'une ligne droite, tout en transmettant la distance qu'il parcourt à VIPER par le biais d'une connexion Wifi. De cette manière est construite une référence de la distance séparant la cible du véhicule instrumenté.

Sur les graphiques de la figure 5.9 sont respectivement présentées les distances mesurées après détection par le lidar, la stéréovision, et l'identification optique, en fonction de cette distance de référence. Il apparaît que le lidar permet d'obtenir une mesure très proche de la référence, et cela sur toute sa gamme de portée.

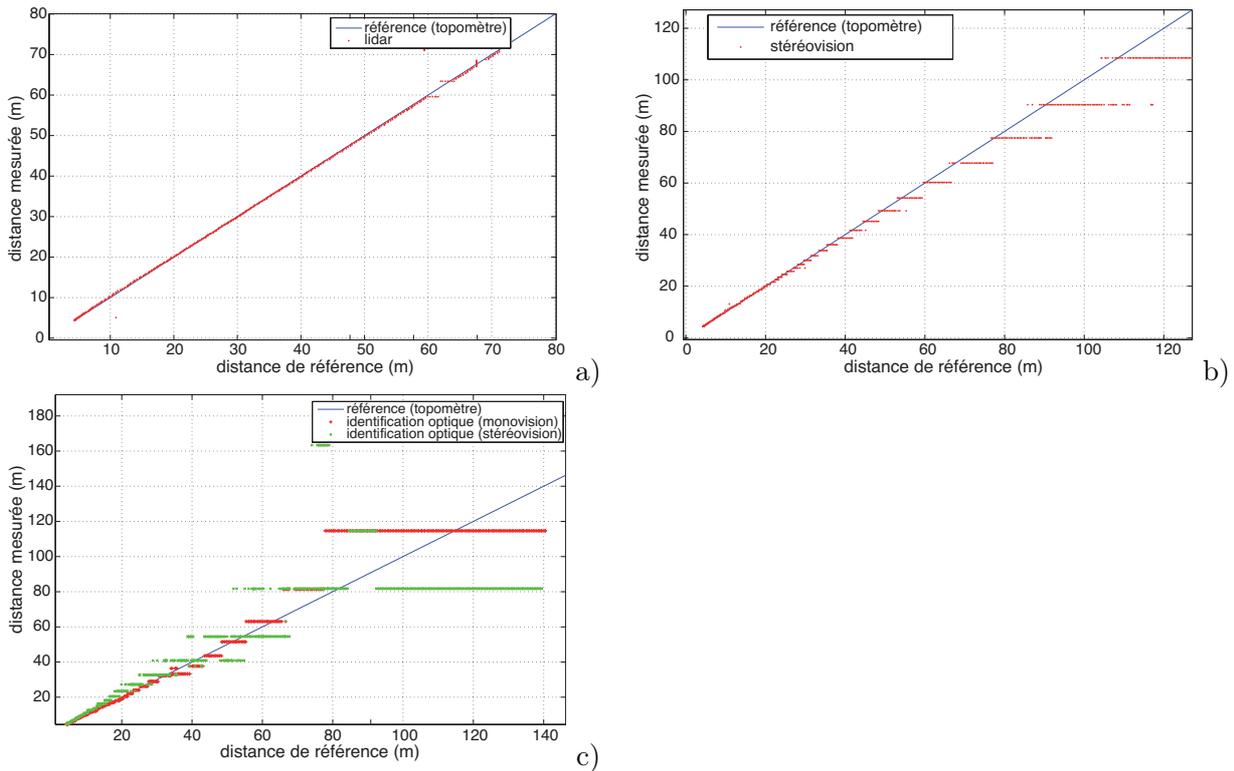


FIG. 5.9 – Comparaison des mesures de distance obtenues par chacun des trois capteurs avec une référence donnée par un topomètre.

La stéréovision perd en précision à mesure que la cible s'écarte. La discrétisation des mesures sur des valeurs entières de pixels est alors bien visible sous la forme de paliers. La mesure demeure néanmoins très stable. Au delà de 100 m, la mesure est bloquée, car la détection est limitée par une montée de la route.

Enfin, la mesure obtenue par localisation de l'identifiant optique est assez précise à courte distance, mais se révèle d'une faible précision au delà de 60 m. La faible résolution du capteur utilisé est en partie responsable de ce phénomène, tout comme l'utilisation de valeurs entières pour mesurer la position des identifiants dans les images infrarouge. Il est, d'autre part, à noter que la piste d'essais n'est plus plane au delà de 80 m, ce qui fausse le résultat de la localisation par la hauteur de la lampe. Contrairement à nos attentes, l'identification en stéréovision ne permet pas de résoudre cette limitation avec la résolution utilisée. Une plus grande résolution horizontale serait nécessaire pour tirer un bénéfice de cette méthode. En dehors de ces problèmes de précision, il est notable que l'identification optique permet d'atteindre une portée bien supérieure aux autres capteurs. Comme le prouve le tableau 5.8, celle-ci peut atteindre 400 m par temps couvert (le contraste entre la lampe et les autres sources infrarouge est alors maximum).

TAB. 5.8 – Portée d'identification.

Portée d'identification - temps couvert (<i>m</i>)	400
Portée d'identification - temps ensoleillé (<i>m</i>)	110

Positionnement latéral

Les graphiques de la figure 5.10 présentent la mesure de positionnement latéral de la cible détectée pour les différents capteurs.

Le premier (figure 5.10-a) présente le positionnement mesuré par le lidar. Pour ce capteur, la mesure latérale devient imprécise à grande distance, contrairement à la mesure longitudinale. En effet, compte tenu de la résolution angulaire, à 50 m l'écart latéral entre deux impacts est de 44 cm. Selon les impacts renvoyés par l'arrière du véhicule, le positionnement est donc variable. Ceci explique la fluctuation entre deux valeurs qui s'observe sur la courbe et dont l'amplitude croît avec la distance. Néanmoins, à défaut d'une véritable mesure de référence, nous allons utiliser cette mesure par le lidar comme comparaison avec les deux autres capteurs.

Dans le cas de la stéréovision (figure 5.10-b), une incertitude supérieure à 1 m est visible à courte distance, selon que l'un ou l'autre des côtés de la voiture soit pris en compte. Cela confirme les fluctuations de détections que nous avons soulignées au cours de l'étude de la séquence auto-routière et explique la dispersion des points à ce niveau de la courbe. A plus grande distance, ce phénomène devient anecdotique, mais la mesure s'écarte de l'encadrement de la mesure fournie par le lidar. La non-validité de l'hypothèse d'un "monde plan" explique cette erreur.

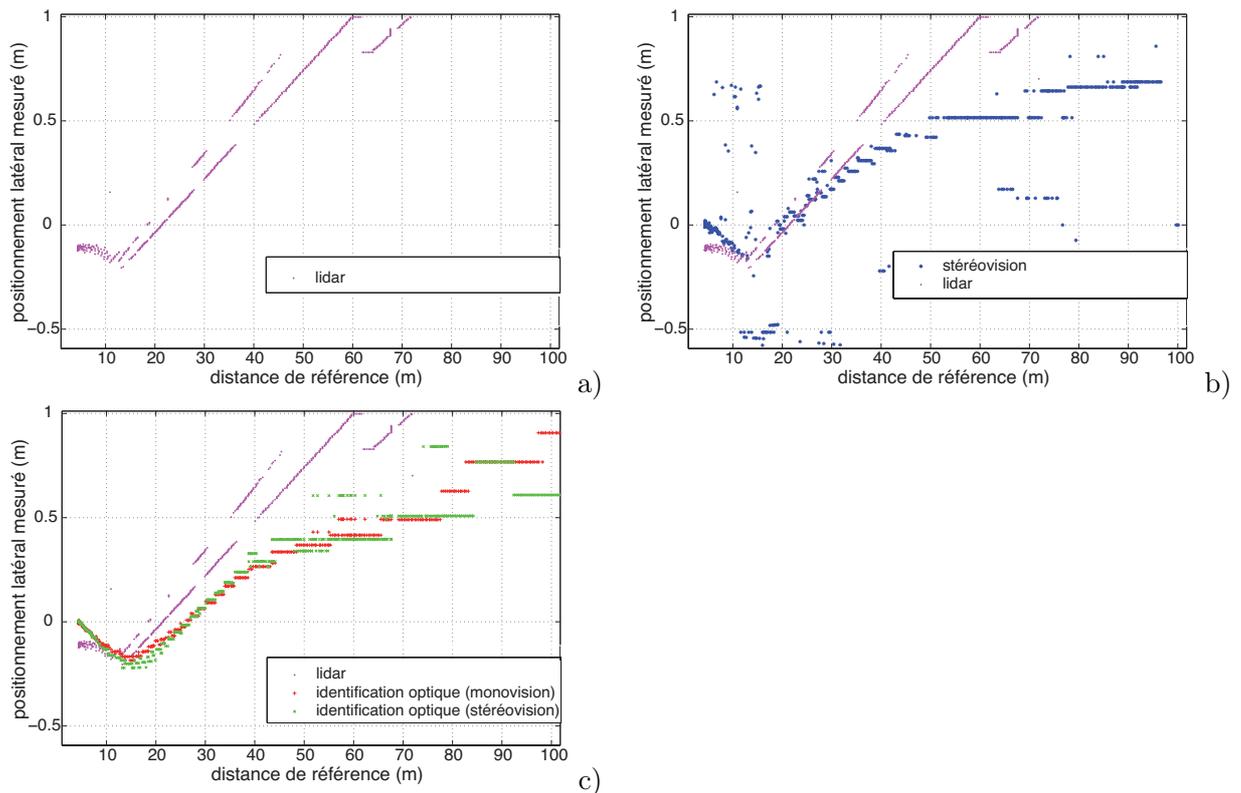


FIG. 5.10 – Positionnement latéral mesuré par les différents capteurs lors de l'éloignement d'un véhicule.

Comme le montre la figure 5.10-c, l'identification optique n'est pas entachée d'une telle va-

riation des mesures entre plusieurs valeurs, car une seule lampe est présente sur le véhicule. L'imprécision liée à la discrétisation des mesures apparaît néanmoins avec l'éloignement du véhicule, de même qu'une erreur due au "monde non plan". Il est cependant remarquable que l'erreur commise reste inférieure à 50 cm en deçà de 60 m, par rapport au lidar. Cela signifie qu'il est possible de reconnaître la voie de circulation sur laquelle se trouve un véhicule identifié à cette distance, ce qui présente un intérêt certain pour l'application qui sera proposée au paragraphe 5.4.3.

5.3.2 Evaluation de la caractérisation d'un piéton

Nous avons également voulu évaluer les performances de l'étape de caractérisation par stéréovision, à travers le cas particulier d'un piéton. Pour cela, en l'absence de capteur permettant de fournir une vérité terrain, le piéton s'est déplacé sur le banc de calibrage du site de Satory, en s'arrêtant en des points connus. Ces points correspondent, pour des distances de 3, 5, 10, 15, 20, 25 et 30 m, à des positionnements latéraux de -2,5, -1, 0, 1 et 2,5 m. La figure 5.11 illustre le trajet réalisé en fournissant, pour chaque position mesurée, le nombre d'instant de mesures correspondant. Les points d'arrêt apparaissent donc comme des pics. Le point (-2,5, 35) correspond à la fin du parcours du piéton.

Cela nous a permis d'évaluer les mesures de position et de géométrie fournies lors de la caractérisation par stéréovision.

Positionnement

Les positions mesurées à l'aide du lidar, à titre de comparaison, sont proches des résultats attendus. Les imprécisions latérales sont bien inférieures à la largeur du piéton lui-même et sont donc acceptables. Tout comme cela était visible pour la mesure de portée, l'imprécision longitudinale est inférieure à 10 cm, et est indépendante de la distance séparant le piéton du capteur.

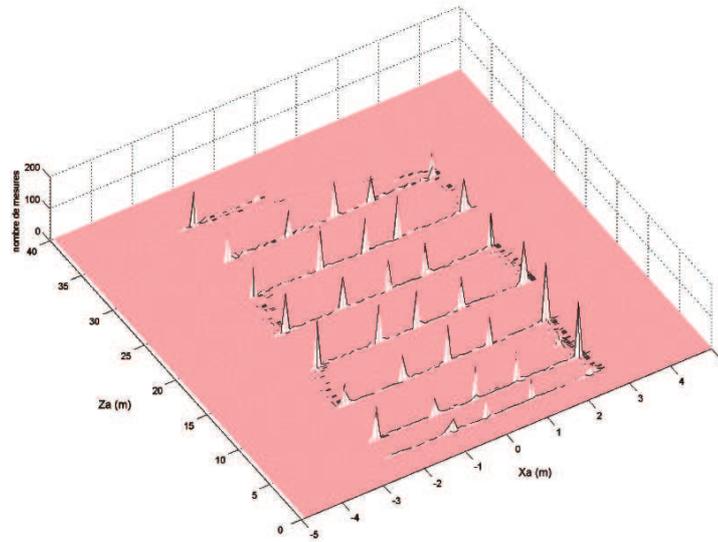
Les résultats obtenus par stéréovision sont différents sur plusieurs points. Tout d'abord, nous pouvons remarquer que l'angle de vue du capteur est plus faible que celui du lidar. Ainsi, les points (-2,5 ; 5) et (2,5 ; 5) sont absents, car ils n'appartiennent pas au champ commun des deux caméras. Ce dernier point donne d'ailleurs lieu à de faux appariements, qui placent un obstacle en (2,3 ; 7,3). Sur la ligne $Z_a = 3 \text{ m}$, les détections sont imprécises et peu nombreuses car seuls les bords du piétons sont détectés. Une erreur latérale d'environ 0,3 m est alors présente. Concernant la précision longitudinale, la discrétisation des pixels est de plus en plus visible à mesure que le piéton s'éloigne. Pour une distance donnée, la mesure oscille donc entre 2 ou 3 valeurs, comme cela est visible sur la première image de la figure 5.12.

Géométrie

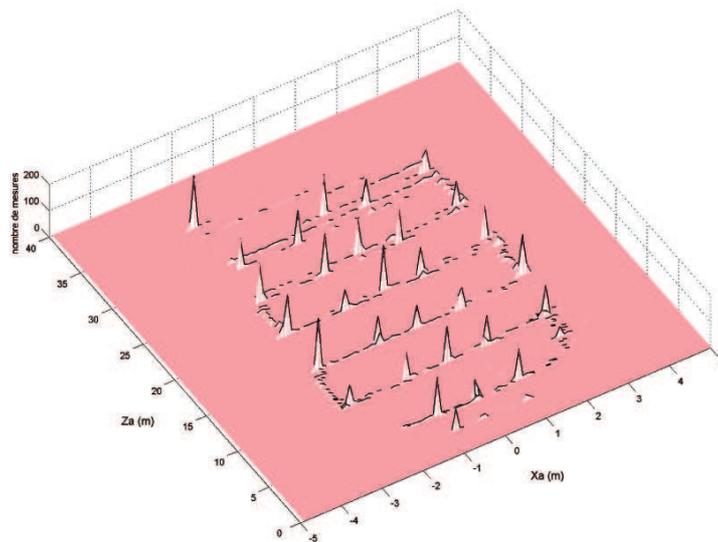
Mis à part la précision de positionnement des objets détectés, nous nous sommes également attachés à juger de la caractérisation de la géométrie du piéton par stéréovision. Pour cela, nous avons calculé pour chaque distance les valeurs moyennes de hauteur, largeur et épaisseur fournies par l'algorithme de stéréovision. Les résultats sont reportés sur la figure 5.12. Le piéton qui a participé à l'expérience mesure $186 \text{ cm} \times 50 \text{ cm} \times 30 \text{ cm}$.

A nouveau, la précision des résultats n'est pas concluante en deçà de 5 m, le piéton n'étant pas complètement détecté alors. En revanche, au delà la largeur et la hauteur sont assez proches des valeurs attendues et varient peu alors que le piéton s'éloigne du capteur.

La mesure d'épaisseur est quant à elle complètement erronée et augmente avec la distance. Cet



a)



b)

FIG. 5.11 – *Parcours d'un piéton qui s'arrête sur différents points du site de calibrage.*
 a) *histogramme du nombre d'instantés de mesure en fonction de la position du piéton mesurée par le lidar.* b) *le même histogramme obtenu à partir des détections par stéréovision.*

aspect était prévisible, car le piéton n'étant pas caractérisé avec une précision inférieure au pixel, son épaisseur ne peut être inférieure à 1 le long de l'axe Δ_s de R_{Δ} . L'épaisseur mesurée est donc toujours supérieure à distance métrique séparant deux valeurs entières successives de disparité.

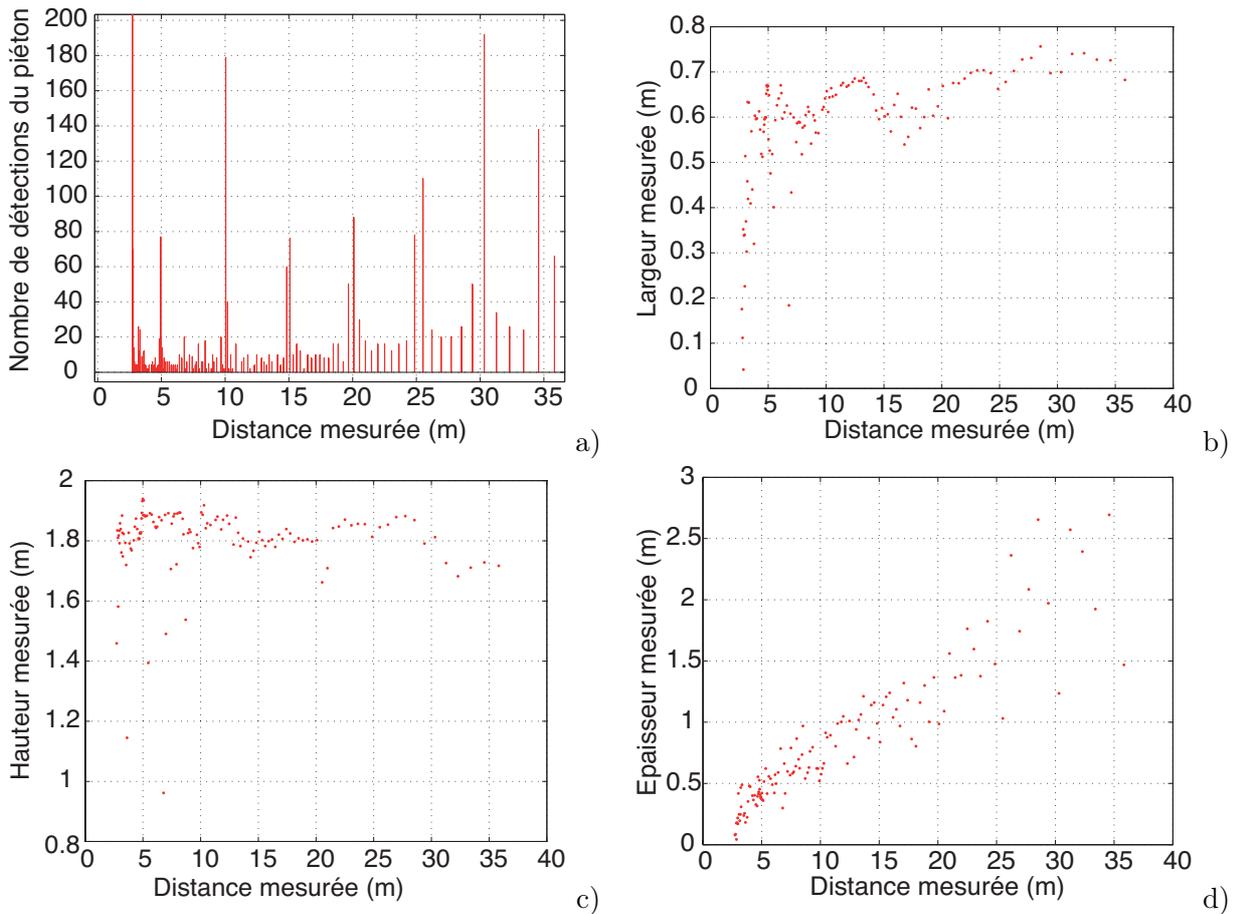


FIG. 5.12 – Caractérisation d'un piéton se déplaçant sur des points connus du banc de calibration.

5.3.3 Evaluation de l'intérêt du zoom numérique

Nous avons proposé au chapitre 4 l'éventuelle utilisation d'un zoom numérique variable dans les régions d'intérêt afin d'assurer une plus grande portée de détection et un gain en temps de calcul à courte portée. Afin d'évaluer l'intérêt d'une telle approche, nous avons placé VIPER dans une position fixe et fait s'éloigner un véhicule à détecter selon une ligne droite. Nous avons alors mesuré la portée de confirmation par stéréovision. Les images stéréoscopiques utilisées ont une résolution quart de PAL (384*288) et le facteur de zoom maximal autorisé est 2, pour assurer un temps de calcul réaliste, qui reste toujours inférieur à 40 ms. Les résultats obtenus sont reportés dans le tableau 5.9.

A l'observation de ces résultats, il est notable que la faible résolution du système stéréoscopique limite la portée d'ensemble du dispositif. En effet, si le lidar est capable d'observer le véhicule jusqu'à sa limite de portée, la confirmation ne dépasse pas 58 m dans le meilleurs des cas. L'utilisation du zoom numérique augmente clairement la portée de confirmation (+80,8% pour le premier critère), et le système n'est plus limité par la résolution de la stéréovision, mais par la portée du lidar. Ce gain est essentiellement dû à l'augmentation du nombre de points de mesure, la plus grande taille des objets par rapport à la fenêtre de corrélation et, dans le cas du critère de l'alignement, la plus facile séparation entre la route et les obstacles, par l'augmentation

TAB. 5.9 – Influence du zoom numérique sur la portée de confirmation du lidar.

	Lidar	Points Obstacles	Alignement Principal
Sans zoom	74 m	31 m	58 m
Avec zoom	74 m	56 m	74 m

de la dynamique en disparité dans la zone d'intérêt considérée.

5.4 Applications aux aides à la conduite

Après ces différents tests qui ont permis de valider le principe de la méthode proposée, nous avons voulu apporter également une validation par la mise en œuvre de celle-ci dans des applications fonctionnelles, embarquées dans les véhicules instrumentés du LIVIC. Les applications proposées sont de trois ordres. Tout d'abord, nous présentons un système complet de détection d'obstacles par stéréovision seule, qui a été développé dans le cadre des projets LOVE et DO30. Plutôt qu'une application directement exploitable par un conducteur, il s'agit donc d'une réalisation de l'architecture proposée pour répondre à la problématique de la détection d'obstacles, en donnant beaucoup d'importance au temps de calcul et à la fiabilité.

Nous nous intéressons ensuite à la conception d'un système de "stop & go" et de suivi automatisé, exploitant le couple lidar/stéréovision. Celui-ci est une évolution du système de mitigation de collision présenté dans [LRA05]. Nous proposerons donc également la stratégie retenue pour intégrer ces deux systèmes dans un unique dispositif d'aide à la conduite.

Enfin, nous mettrons en pratique l'identification optique dans le cadre d'une coopération entre véhicules, pour enrichir la détection d'obstacles par des informations issues de communications inter-véhiculaires.

Toutes ces applications ont été testées dans les véhicules expérimentaux du LIVIC et ont été implantées à l'aide de l'environnement de programmation par flux RTMaps [Ste01].

5.4.1 Détection d'obstacles par stéréovision multi-résolution

Nous avons développé un système de détection d'obstacles par stéréovision, fondé sur la méthodologie présentée dans cette thèse, en visant à assurer un temps de calcul raisonnable. Pour cela, nous proposons l'utilisation d'une approche exploitant trois résolutions. Les images originales étant au format VGA (640*480 pixels), les trois résolutions utilisées sont 1/16 VGA, 1/4 VGA et VGA.

L'idée de l'approche en trois résolutions que nous avons envisagée est de combiner celles-ci avec les trois étapes principales dont nous avons besoin pour la détection. Elle permet de plus d'affiner progressivement l'estimation de la surface de la route par des calculs successifs.

Basse résolution : le profil de la route

La basse résolution est exploitée pour l'estimation du profil de la route et ainsi la correction des valeurs de tangage et de hauteur du véhicule. Cette étape sert donc de première passe de l'algorithme de calcul des points de mesure pour la résolution supérieure.

Moyenne résolution : Les zones d'intérêt

La moyenne résolution a pour vocation d'assurer l'étape de création d'hypothèses de détection. Les points de mesure sont donc calculés sur la totalité de l'image, qui est ensuite segmentée en utilisant l'algorithme présenté au chapitre 3. Celui-ci fournit donc une liste d'hypothèses de détection, sous la forme de régions d'intérêt dans les images haute résolution. De plus, l'image v-disparité route créée lors du calcul des points obstacles permet d'affiner l'estimation fournie par la basse résolution.

Haute résolution : Confirmation, affinement, suivi

Les hypothèses de détection sont validées et caractérisées en pleine résolution. Celle-ci permet également de procéder au suivi des détections.

Filtrage des résultats

Les détections finales peuvent éventuellement être comparées à une zone de filtrage latérale, fournie par la prédiction de la trajectoire future du véhicule à partir des données proprioceptives du véhicule instrumenté, comme l'illustre la figure 5.13.



FIG. 5.13 – Zone de filtrage latérale utilisée dans le projet LOVE.

Les rectangles en traits gras figurent les détections.

Temps de calcul

La fiabilité de cette application de détection par stéréovision est directement liée aux résultats évoqués en début de ce chapitre. En revanche, l'un des principaux intérêts de notre nouvelle approche est un gain en temps de calcul par rapport à une approche classique. Celui-ci a été évalué en moyennant le temps de calcul obtenu sur une séquence de 700 images, acquises avec une base de 1 m, et comportant des obstacles de différentes tailles (piétons et véhicules). Les résultats sont mesurés avec un Pentium 4-C cadencé à 2,4 GHz et reportés dans le tableau 5.10.

A titre de comparaison, ceux-ci sont comparés au temps nécessaire au calcul des mêmes images de disparité en pleine résolution, sans considérer de détection. Notre approche reste donc très efficace car elle intègre la totalité de l'algorithme en un temps inférieur à celui nécessaire au calcul d'une image de disparité.

TAB. 5.10 – *Temps de calculs obtenus par l'application de détection d'obstacles utilisant 3 résolutions.*

Basse résolution	14,3 ms
Moyenne résolution	94,6 ms
Haute résolution	77,9 ms
Suivi des détections	6,7 ms
Temps de calcul de l'approche en 3 résolutions	193,5 ms
Calcul d'une carte de disparité VGA complète	492,4 ms

5.4.2 Détection par stéréovision et télémétrie laser : de la mitigation de collisions au "stop & go" et suivi automatique

Développé au cours du projet ARCOS³⁹ et fondé sur l'utilisation conjointe de la télémétrie laser et de la stéréovision, un système de mitigation de collision a été mis en place dans les véhicules expérimentaux du LIVIC. Celui-ci est décrit et évalué dans [LRA05]. Ses bonnes performances y sont démontrées sur des scénarii contrôlés. Par ailleurs, un test réalisé sur différents types de routes ouvertes n'a laissé apparaître que 7 fausses détections pour 1000 km de route.

Dans le cadre de cette thèse, la portée de ce système a été étendue par l'utilisation du traitement par zones d'intérêt et du zoom numérique présenté au chapitre 4. Cette amélioration a permis d'étendre également le champ applicatif du système en permettant son utilisation au sein de systèmes de "stop & go" et de suivi automatique à moyenne vitesse, grâce à l'adjonction d'un contrôleur longitudinal flou. Dans une telle application, le but est l'asservissement du véhicule à un TTC constant par rapport au véhicule précédent, sans que le conducteur n'agisse sur les commandes d'accélération et de freinage. Il est primordial que la commande soit confortable pour les passagers, sans accoups, et que la perception soit exempte de fausses détections.

Nous allons ici présenter les différentes composantes de ce système, puis la combinaison de celui-ci avec l'application de mitigation de collision déjà existante.

Architecture de perception pour le "stop & go"

L'architecture de détection proposée repose sur la coopération entre le lidar et la stéréovision, selon la méthodologie présentée dans cette thèse. La stéréovision est exploitée dans un premier temps pour estimer le profil de la route à courte portée. Celui-ci est utilisé pour filtrer les impacts laser résultants de l'intersection du plan de détection du lidar et de la surface de la route. Les impacts non éliminés sont agglomérés en objets que nous appellerons cibles. Le suivi de celles-ci au cours du temps permet d'obtenir des objets estimés appelés pistes. Les pistes situées hors d'une zone de filtrage latérale sont supprimées, pour ne conserver que celles situées dans la

³⁹<http://www.arcos2004.com/>

trajectoire prédite du véhicule, qui sont ensuite transformées en zones d'intérêt pour la validation par stéréovision. L'utilisation d'un zoom numérique variable à cette étape de l'algorithme assure que la confirmation par stéréovision ne limite pas la portée du système et que le fonctionnement temps réel est assuré. Enfin, un indice de confiance est associé à chacune des pistes, de façon à ce qu'une détection ne soit donc fournie à l'algorithme de contrôle que si celle-ci a été observée plusieurs fois.

Contrôle longitudinal du véhicule

A partir des détections caractérisées par une estimation de distance et de vitesse relative, un algorithme de commande floue permet de calculer les consignes d'accélération et de freinage à communiquer aux actionneurs. Cet algorithme prend deux variables en entrée : l'inter-distance (m) et la vitesse relative ($m.s^{-1}$) de la cible la plus proche. La sortie est une consigne d'accélération/décélération du véhicule ($m.s^{-2}$).

En entrée, les deux variables numériques sont immédiatement traduites en variables symboliques, chacune répartie sur 7 valeurs (*Très Près, Près, Moyen Près, Idéale, Moyen Loin, Loin, Très Loin* pour l'interdistance et *Rapproche Vite, Rapproche, Rapproche Lentement, Égale, Éloigne Lentement, Éloigne, Éloigne Vite* pour la vitesse relative). Afin d'éviter les transitions trop brusques entre ces différentes valeurs, la conversion est réalisée à l'aide de sous-ensembles flous. La combinaison de ces entrées est ensuite réalisée à travers une matrice $7*7$, contenant les règles correspondant aux différents cas possibles, et permet de déterminer la consigne adaptée, également sur une échelle floue (*Freiner Très Fort, Freiner Fort, Freiner, Rien, Accélérer, Accélérer Fort, Accélérer Très Fort*). Cette consigne est enfin convertie en valeur numérique pour être transmise aux actionneurs du véhicule.

L'avantage principal de ce type de commande est sa souplesse et son confort en utilisation courante. Elle est néanmoins très réactive dans les cas critiques et peut produire des freinages violents si nécessaire.

Gestion des pertes de détection

Lors de la mise en œuvre du système, son fonctionnement s'est vu dégradé par un nombre important de pertes de détections, dues à la grande sensibilité du lidar au tangage. Ainsi, lors d'un freinage prolongé, ou lorsque le véhicule suivi est à grande portée (pour un suivi automatique à 100 km/h par exemple), la perte de détection peut être trop longue pour être compensée par une prédiction. Dès lors, des heuristiques de haut niveau ont été conçues pour résoudre ce défaut. Entre autres, lorsqu'une piste vient d'être perdue, la consigne est maintenue ou non selon que le véhicule était en accélération ou en freinage, de façon à assurer une sécurité maximale :

- si le véhicule procédait à un freinage, celui-ci est maintenu,
- si le véhicule était en phase d'accélération, la consigne d'accélération est annulée, de façon à maintenir une vitesse constante.

L'utilisation de telles stratégies permet de compenser jusqu'à 50% en temps des pertes de pistes, en assurant le maintien de la consigne de vitesse jusqu'à la reprise de la détection. De plus, lorsque le véhicule suivi freine fortement, cela assure que le suiveur freine en conséquence. En effet, sans une telle stratégie, le freinage du véhicule instrumenté provoque un tangage qui induit une perte de détection, entraînant une commande d'accélération. Sur la totalité d'une manœuvre de freinage, cela provoque des saccades brutales.

été testé avec succès dans des contextes variés ("stop & go" à basse vitesse, suivi automatique jusqu'à 100 km/h, maintient d'un temps à collision à 2 s, 1,5 s, 1,2 s) sur les pistes de Satory.

5.4.3 Enrichissement de la détection par perception coopérative

Dans la perspective d'une infrastructure routière sur laquelle les différents acteurs coopèrent et se transmettent des informations, le système d'identification optique que nous avons présenté au chapitre 3 permet de faire le lien entre les systèmes de perception de l'environnement immédiat (vision, télémétrie, ...) et les télécommunications inter-véhiculaires (V2V) ou véhicule-infrastructure (V2I). Nous avons mis en œuvre ce concept au sein d'un système de perception coopérative, combinant stéréovision, télémétrie laser, identification optique, télécommunications V2V et infrastructure communicante.

Considérons une portion de route comprenant plusieurs véhicules équipés de moyens de communications V2V (dans notre cas, des modems Wifi), et le véhicule instrumenté. Celui-ci reçoit des messages radio, les trie pour déterminer ceux qui présentent un réel intérêt, puis les associe à d'éventuelles détections qu'il a pu réaliser par le biais de ses capteurs extéroceptifs. Nous allons présenter la façon selon laquelle ce concept peut s'appliquer à l'anticipation d'un freinage d'urgence grâce aux télécommunications.

Réception de messages radio

Parmi tous les messages V2V que le véhicule instrumenté est à même de recevoir en permanence, il lui faut déterminer ceux qui sont pertinents. Cette tâche est réalisée par l'exploitation conjointe de l'identification optique et des capteurs proprioceptifs. Ces derniers sont, en effet, en mesure de fournir une zone de filtrage latérale correspondant à la trajectoire future estimée du véhicule. Nous considérerons donc qu'un message est pertinent s'il a été émis depuis cette portion de l'espace, condition vérifiée si une lampe émettant le même code identifiant que celui compris dans le message reçu a été détectée et localisée dans cette zone. Les différentes actions sont donc réalisées selon la séquence suivante :

1. détection et localisation des émetteurs d'identification,
2. en parallèle, réception des messages V2V et I2V,
3. association des messages aux émetteurs correspondants s'ils ont été détectés,
4. comparaison de la position des messages localisés avec la zone de filtrage,
5. traitement des messages pertinents.

Traitement des informations reçues

Lorsqu'un message est jugé pertinent, il est exploité pour l'aide à la conduite. Cette tâche peut être réalisée de nombreuses façons. Comme nous l'avons vu en introduction, cela peut donner lieu à une remontée d'information au conducteur ou à des ordres émis vers les actionneurs du véhicule.

Dans le cadre de la collaboration liant le LCPC à l'équipementier Bosch au sein du projet TRACKSS, cette exploitation consiste à fournir une image du véhicule émetteur à une IHM⁴⁰, par le moyen d'une caméra capable d'adresser dynamiquement des zones d'intérêt.

Une démonstration réalisée au LIVIC a également permis de démontrer la possibilité de commander le freinage d'urgence du véhicule récepteur par des messages radio, ce qui permet d'être

⁴⁰Interface Homme Machine

plus réactif qu'avec un système de perception fondé sur la vision.

Une autre exploitation envisageable est le transfert de données numériques, comme la vitesse du véhicule émetteur, pour améliorer l'estimation de la vitesse relative.

Combinaison avec les résultats de détection

Il est cependant hasardeux de concevoir des systèmes très intrusifs sur le seul couple V2V / identification, car celui-ci ne permet pas de fournir des informations précises d'inter-distance. Pour cette raison, il est nettement préférable de le faire coopérer avec des systèmes perceptifs complémentaires, en associant les informations provenant de ces différents capteurs pour obtenir une information enrichie. L'association des messages avec les éventuelles détections est réalisée par la méthode proposée en perspective du chapitre 3. En revanche, si aucune détection n'est fournie par les capteurs extéroceptifs, une mesure de distance est réalisée en caractérisant par stéréovision locale la zone d'intérêt correspondant à l'émetteur. La figure 5.15 illustre l'avantage qui peut être tiré de l'utilisation conjointe d'un lidar avec le système de réception et traitement des messages : un message d'arrêt d'urgence est détecté et précisément localisé.

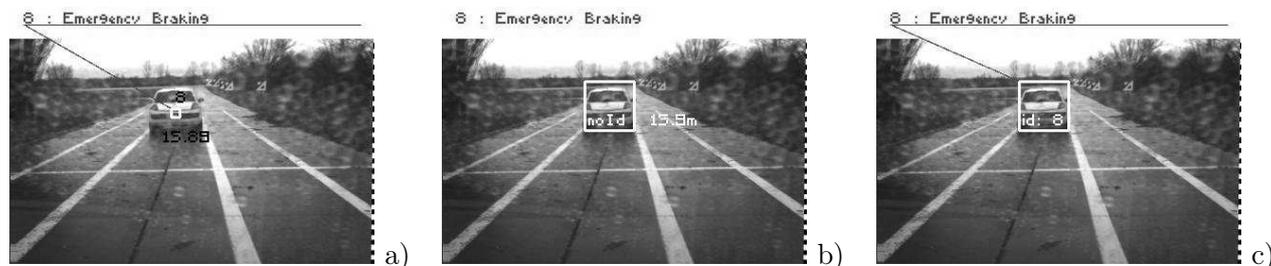


FIG. 5.15 – Détection enrichie par communications V2V

a) Un objet est détecté grâce au lidar et un message est reçu du véhicule 8. b) Un objet étiqueté 8 est perçu et localisé. c) Les deux détections sont associées : c'est le véhicule sur notre voie qui a émis le message.

Informations provenant de l'infrastructure

Une infrastructure routière coopérative peut également participer aux aides à la conduite embarquées. Partant de cette idée, nous nous sommes intéressés à la possibilité de transmettre les informations liées à la signalisation verticale. Pour cela, des panneaux ont été équipés d'émetteurs infrarouge transmettant un identifiant correspondant à leur nature (stop, limitation, feu rouge,...). Le véhicule est alors capable de lire ces informations et de les localiser bien plus précisément que ne le permettrait un système de localisation GPS doublé d'une cartographie. L'information est alors transmise au conducteur à travers une IHM qui affiche une imagerie du panneau identifié.

5.5 Conclusion

Ce chapitre a été consacré à l'évaluation des différentes briques élémentaires et de l'architecture complète de détection proposée dans cette thèse. L'étude des séquences réelles a permis de constater que, si les performances ne sont pas encore à la hauteur d'un système industrialisable,

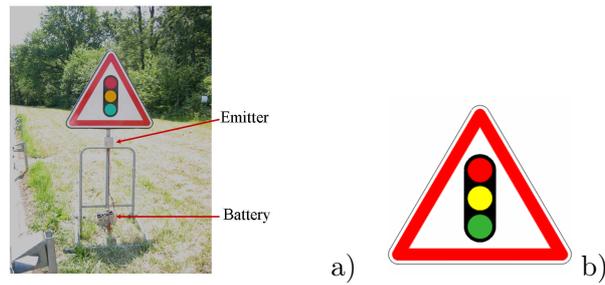


FIG. 5.16 – Panneau de signalisation équipé d’un émetteur infrarouge, et imagerie affichée sur l’interface conducteur lorsque ce panneau est identifié.

l’approche proposée, exploitant détection surabondante puis confirmation, semble adaptée à la problématique de l’environnement routier. Elle permet en effet de rejeter la grande majorité des erreurs, tout en conservant un taux de détection élevé.

Suite à cette évaluation, la stéréovision apparaît comme un capteur performant pour la détection des obstacles, fournissant des détections robustes et bien dimensionnées pour les piétons et les objets texturés. Certains aspects restent néanmoins à améliorer. La densité de la carte de disparité et l’utilisation d’une géométrie de route plus précise sont les améliorations futures qui devraient apporter les gains les plus importants.

Le télémètre laser est performant en termes de vitesse et de précision, mais demeure peu robuste. Nous avons montré dans ce chapitre que son utilisation conjointe avec la stéréovision, selon la stratégie proposée, le rend suffisamment performant pour être utilisé efficacement au sein de systèmes d’aide à la conduite.

Enfin, nous avons démontré l’utilité du système d’identification optique présenté dans cette thèse. La précision des mesures est satisfaisante à courte portée mais devient insuffisante à longue portée. A l’avenir, il pourrait être intéressant d’augmenter sa résolution. Néanmoins, nous avons vu qu’en deçà de 60 m sa précision latérale le rend tout à fait exploitable pour une application de détection coopérative utilisant des communications V2V.

Au delà de l’évaluation de la méthode de détection, nous avons également proposé trois applications qui démontrent l’intérêt pratique du travail réalisé.

Conclusion Générale

Conclusion

AU cours de cette thèse, nous avons proposé une méthodologie de détection d'obstacles multi-capteurs s'adressant au contexte routier. Nous avons décrit son principe, proposé une implantation par un choix spécifique de capteurs, évalué sa validité et enfin proposé des mises en application fonctionnelles dans les véhicules expérimentaux du LIVIC.

La méthodologie de détection proposée est fondée sur une coopération interdépendante entre plusieurs capteurs. Ainsi, les différents capteurs impliqués fournissent des hypothèses de détection, qui sont ensuite converties en régions d'intérêt dans les images stéréoscopiques. La stéréovision est alors utilisée pour valider et caractériser ces hypothèses. Cette approche permet de tirer parti des avantages de la stéréovision en termes de robustesse et de champ de vue, tout en conservant un temps de calcul raisonnable. L'évaluation expérimentale sur des cas de conduite réels a par ailleurs démontré l'efficacité de cette stratégie, puisqu'elle permet de conserver un taux de détection élevé, tout en faisant chuter le nombre de fausses détections.

La mise en œuvre de cette méthodologie a été proposée à travers trois capteurs de type différents, afin de souligner le caractère générique de l'approche proposée et de couvrir un grand champ de vue. Ainsi, nous nous sommes intéressés à l'intégration d'un télémètre laser, d'un capteur stéréoscopique et d'un capteur coopératif d'identification optique. Pour chacun d'entre eux, nous avons proposé une méthode permettant d'obtenir des volumes d'intérêt dans l'espace de disparité. Pour la stéréovision, cela est réalisé par segmentation d'une image de disparité. Pour le lidar et l'identification optique, c'est la localisation des détections dans le repère du véhicule qui permet une telle coopération.

En outre, nous nous sommes intéressés au traitement des images stéréoscopiques, pour obtenir des points de mesure fiables. La méthode que nous avons retenue pour cela s'est avérée suffisamment efficace pour traiter les étapes de caractérisation et de validation des détections. Cependant, la densité des images de disparité qu'elle permet de fournir s'est avérée insuffisante pour assurer une détection continue des véhicules dans leur intégralité.

Lors du traitement des images stéréoscopiques, nous avons également proposé d'exploiter les caractéristiques des plans horizontaux et verticaux de la scène. L'utilisation d'une fenêtre de corrélation inclinable a eu pour effet de rendre plus fiable l'estimation du profil de la route et de simplifier la séparation entre les pixels obstacles et ceux appartenant à la surface de la route. La détection des plans verticaux obliques a permis de simplifier le processus de segmentation de l'espace de disparité.

Enfin, la mise en application de la méthodologie proposée a permis d'intégrer celle-ci dans trois applications très différentes. Les deux premières entrent dans le cadre classique de la détection d'obstacles routiers et prouvent la possibilité d'exploiter la méthode proposée pour la conception d'aides à la conduite. En revanche, la troisième explore le concept de route coopérative. En permettant l'utilisation conjointe de télécommunications V2V et de la détection d'obstacles, elle permet d'envisager un grand nombre d'applications futures aux aides à la conduite, avec une réactivité et un niveau d'anticipation accrus.

Finalement, l'approche proposée répond aux objectifs évoqués en introduction. La généralité de la détection est assurée par le choix des méthodes de traitement des données télémétriques

et stéréoscopiques, qui ne nécessitent ni *a priori* fort, ni reconnaissance de forme. Le domaine de fonctionnement du dispositif complet est effectivement assez vaste, car il couvre une plage de portée allant de 2,5 m à plus de 100 m, pour un angle de vue de 40 degrés. Enfin l'exigence d'un temps de calcul réaliste a été une contrainte forte pour le choix des méthodes. L'exploitation de la stéréovision en temps réel (moins de 40 ms) pour la validation des détections est possible sur des images de petite résolution, grâce à l'approche par zones d'intérêt. Elle a également été rendue possible sur des images de plus grande résolution, par l'utilisation d'un zoom numérique variable. La création d'hypothèses de détection par le lidar ou l'identification optique fonctionnent également en temps réel. Seule la création d'hypothèses par stéréovision nécessite plus de 40 ms. Cependant, la réduction de la résolution des images pour cette étape permet d'obtenir des temps de calculs moyens de l'ordre de 70 ms, ce qui souligne le réalisme de l'approche.

Perspectives

Malgré les résultats encourageants obtenus, de nombreuses pistes restent à explorer avant d'envisager la mise en œuvre des méthodes proposées dans le contexte automobile réel. Nous allons présenter certaines des perspectives d'évolution qui s'inscrivent dans la suite du travail réalisé dans cette thèse.

Concernant la stéréovision, l'une des problématiques qui ouvre les plus forte perspectives d'amélioration est l'augmentation de la densité de la carte de disparité. La segmentation de l'espace de disparité comme l'évaluation des critères de validation tireraient pleinement profit d'une telle augmentation du nombre de points de mesure. Celle-ci est néanmoins un point complexe à résoudre, car elle suppose de réaliser des appariements sur des régions de l'image contenant peu d'indices visuels discriminants, tout en évitant de créer des erreurs qui soient fortement corrélées. Cependant, comme nous l'avons présenté au chapitre 2, un grand nombre de recherches sont réalisées en ce sens et des voies restent à explorer. En particulier, la capacité de propager des appariements sur les contours horizontaux [HLPA06] est prometteuse et pourrait être approfondie.

La reconstruction plus précise de la surface de la route, en s'affranchissant de l'hypothèse de "monde plan", est pour sa part indispensable pour augmenter la portée de détection des dispositifs existants. Celle-ci passe dans un premier temps par l'utilisation d'un modèle longitudinal plus précis (succession de plans, polynôme, chloïde, ...), puis par une modélisation tridimensionnelle qui prenne également en compte le profil transversal. Une telle reconstruction implique néanmoins la présence d'appariements sur la surface de la route, soulignant à nouveau la nécessité de l'augmentation de la densité des cartes de disparité utilisées pour obtenir des points de mesure. En revanche, la possibilité de supprimer les points obstacles pour estimer le profil de la route, que nous avons proposée au chapitre 2, devrait permettre de s'orienter plus efficacement vers ces perspectives.

Enfin, la mise en place de critères supplémentaires de validation des hypothèses de détection devrait permettre d'augmenter encore la fiabilité de la méthode que nous avons proposée.

Les perspectives d'évolution de l'identification optique concernent en premier lieu l'amélioration de la précision de localisation des lampes identifiées, notamment par l'augmentation de la résolution du capteur utilisé. Compte tenu de la fréquence de traitement nécessaire, cela suppose d'accroître encore l'efficacité calculatoire du dispositif existant.

Au delà de cette problématique de la précision, de nombreuses voies de recherche sont ouvertes

sur les applications possibles de la coopération entre détection d'obstacles et télécommunications. La transmission de données numériques, comme des mesures de vitesse, pourraient enrichir la détection. La transmission de données symboliques permettrait d'aider le conducteur ou les dispositifs d'assistance à anticiper des dangers. Une telle coopération inter-véhiculaire permet un gain de réactivité dans la perception de phénomènes comme le freinage d'un véhicule, mais également de percevoir des dangers qui ne seraient autrement pas visibles : obstacles masqués par l'infrastructure ou par des conditions météorologiques dégradées, présence de verglas sur la route,...

Le lidar est une technologie de détection d'obstacles assez mature et largement utilisée. Il s'est bien intégré à notre système de détection. Une autre piste à explorer serait l'utilisation d'un radar longue portée pour créer des hypothèses de détection, suivant un mode de création des régions d'intérêt similaire à celui qui a été proposé pour le lidar.

Le suivi tirerait également bénéfice d'une attention particulière. Il nécessiterait en effet l'utilisation de points à suivre qui soient plus stables au cours du temps que le point de contrôle que nous avons utilisé. Le suivi simultané de plusieurs points d'intérêt pourrait conduire à une telle stabilisation, s'il existe effectivement un nombre minimum de points fortement caractérisés qui se retrouvent d'une image à la suivante. Dès lors que le suivi fonctionne de façon suffisamment robuste, la qualité globale de détection peut être améliorée en conséquence. En effet, cela permet de ne valider que les détections observées à plusieurs instants successifs ou encore d'assurer la continuité de détection par l'exploitation d'une estimation lorsqu'aucune observation n'est présente.

Enfin, la garantie d'un fonctionnement optimal et continu d'un tel système de détection ne peut être obtenue que grâce à une méthode rigoureuse de calibrage et de configuration des capteurs, individuellement et collectivement. L'automatisation de celle-ci, à commencer par l'autodiagnostic de la qualité du calibrage, devra être résolue avant une intégration future dans des véhicules commercialisés.

DES CAMÉRAS AU SYSTÈME STÉRÉOSCOPIQUE

Il existe une grande variété de caméras utilisables pour concevoir un système stéréoscopique tel que celui utilisé dans cette thèse. La principale contrainte est la nécessité d'une synchronisation de l'acquisition entre les deux caméras du système. En dehors de cet aspect, un grand nombre de paramètres doivent être fixés, comme le type de capteur, ses paramètres intrinsèques ou encore la base stéréoscopique. Le but de cette annexe est la présentation rapide de ces variables et de leur influence sur le fonctionnement de l'algorithme de détection. Il est essentiel de bien appréhender cet aspect si l'on souhaite s'assurer du fonctionnement de l'algorithme de détection pour la majorité des configurations possibles, ce qui était l'un des objectifs pratiques de cette thèse.

A.1 Technologie du capteur vidéo

Le capteur d'une caméra standard pouvant être caractérisée par un modèle sténopé (présenté au chapitre 2) est constituée d'une matrice de pixels capables chacun de mesurer une intensité lumineuse. Il existe principalement deux types de technologies de capteurs vidéos : CCD⁴¹ et CMOS⁴². Nous n'allons pas aborder ici leur fonctionnement précis, mais plutôt la qualité de leurs images. La figure A.1 présente une scène routière acquise par l'un et l'autre de ces capteurs. Il est notable que l'image CMOS est nettement plus bruitée que l'image CCD. La technologie étant plus récente, elle n'a certainement pas obtenu le même niveau de maturité, et gageons que ce défaut sera corrigé dans un futur proche. En revanche, cela impose un filtrage des images comme préalable aux algorithmes de traitement. A titre d'exemple, la figure A.2 présente l'image de droite d'une paire stéréoscopique acquise par un capteur CMOS, et l'image de disparité calculée selon la méthode que nous avons proposée, puis le même couple d'images lorsqu'un filtrage gaussien (la taille du masque étant $3 * 3$ pixels) est appliqué préalablement. Celui-ci réduit substantiellement le niveau de bruit sur l'image ainsi que sur la carte de disparité. Il introduit cependant un flou qui peut perturber la phase d'appariement, en rendant les indices visuels moins discriminantes.

La technologie CMOS demeure néanmoins une voie d'avenir, car outre un prix de production relativement faible, elle présente l'avantage de fournir une dynamique d'intensité lumineuse nettement supérieure à la technologie CCD. Par le biais d'une fonction de réponse en intensité dite *lin-log*, car linéaire dans les faibles valeurs et logarithmique dans les hautes intensités, elle fournit jusqu'à 120 dB de dynamique contre environ 60 dB pour une caméra CCD linéaire classique.

⁴¹Coupled Charge Device

⁴²Complementary Metal Oxide Semiconductor

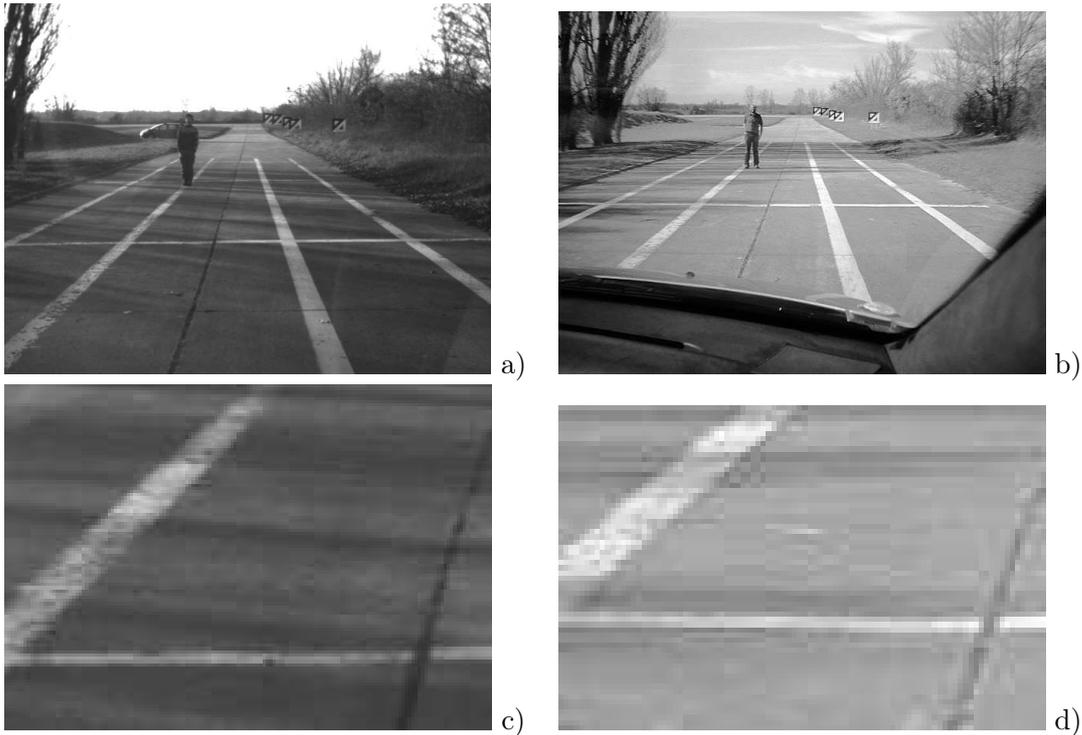


FIG. A.1 – Les deux principales technologies de caméras.

a) une image acquise avec un capteur CCD (caméra SONY), b) avec un capteur CMOS (caméra SMAL). c et d : zoom sur un détail des images a et b. L'image fournie par le capteur CMOS est nettement plus bruitée.

Dans un contexte routier, cet aspect est essentiel lors de l'entrée ou la sortie d'un tunnel, ou pour éviter un éventuel éblouissement par les phares des autres véhicules.

A.2 Les paramètres intrinsèques du capteur

L'angle de vue et la qualité de la reconstruction 3D sont directement liés aux paramètres intrinsèques des caméras, et donc à la résolution du capteur et au rapport entre la distance focale et la taille d'un pixel. Après la technologie de la matrice, c'est la taille de celle-ci qui doit être fixée, en même temps que sa résolution puisque ces deux paramètres sont directement figés par le modèle de caméra retenu. La nécessité d'un angle de vue minimum, imposé par l'application (voir l'introduction générale), fournit alors la longueur focale de l'objectif à utiliser. La relation entre cet angle Φ , la focale f et la largeur de la matrice L_m est donnée par :

$$\Phi = 2 * \operatorname{atan}\left(\frac{L_m}{2 * f}\right) \quad (\text{A.1})$$

Pour un angle donné, la diminution de la taille de la matrice impose donc une diminution de la focale de la lentille. Cet aspect n'est pas anodin, car si une petite matrice est synonyme de bas coût, une faible distance focale peut induire une forte distorsion optique de l'image. Cela est visible sur la figure A.1 : les deux caméras ont un angle de vue équivalent, mais la matrice CMOS étant plus petite, un objectif de focale 6 mm (contre 8,5 mm) est utilisé. Le pourtour

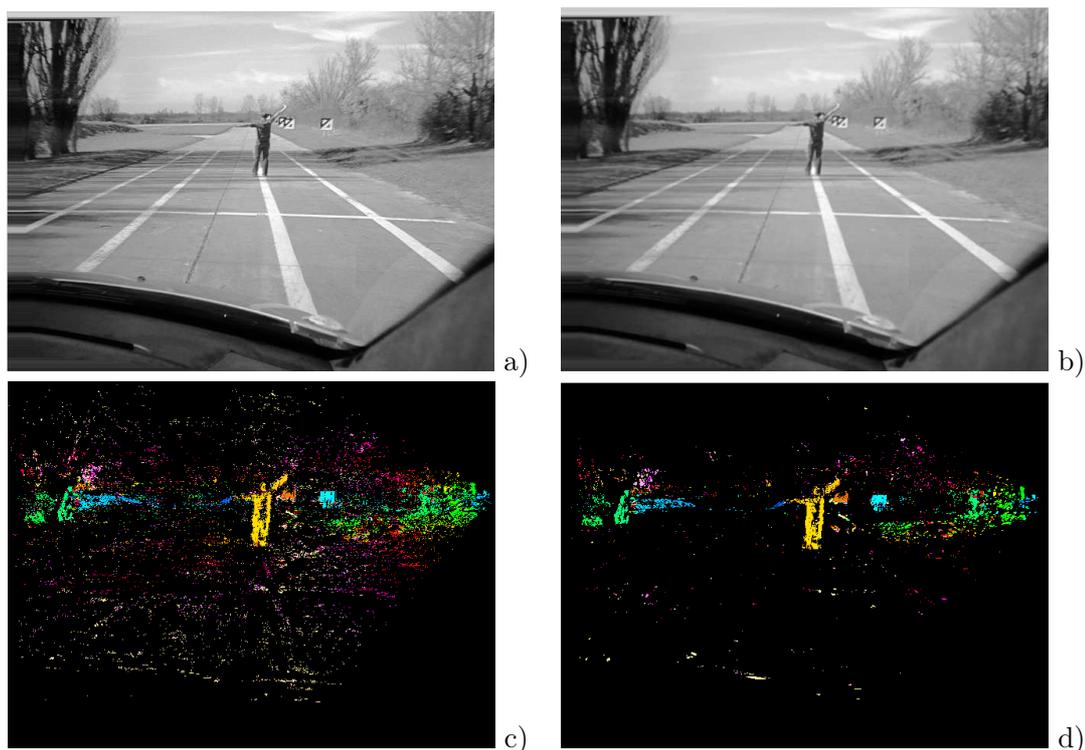


FIG. A.2 – a) Une image de droite d'une paire stéréoscopique acquise avec des caméras CMOS. c) La carte de disparité résultante. Le bruit caractéristique des images CMOS est très visible sur l'image originale et perturbe l'étape d'appariements. b) et d) : le même couple d'images lorsqu'un filtre gaussien a été appliqué préalablement.

de l'image est alors distordu. Considérer que le capteur utilisé est réellement assimilable à un modèle sténopé implique une compensation de cette distorsion. Cependant, cette compensation induit un flou sur une partie de l'image et peut être imprécise, car les paramètres de distorsion sont difficiles à estimer très précisément.

La résolution est ensuite définie de façon à respecter une contrainte de précision ou de portée. Connaissant alors la taille d'un pixel et la longueur focale, les paramètres intrinsèques α_u et α_v sont fixés. Ce sont ces paramètres qui fixent la précision de la reconstruction 3D possible en aval. Le tableau A.1 récapitule les paramètres des trois types de caméras que nous avons utilisé dans le cadre de cette thèse. La caméra SMaL a été désignée pour les projets DO30 et LOVE en raison de sa grande dynamique et de l'hypothèse réaliste qu'elle représente par rapport à de futures caméras qui pourraient être intégrées dans des véhicules commercialisés.

A.3 La base stéréoscopique

Après cet aperçu rapide des étapes de la spécification du choix d'un capteur vidéo, voyons l'influence d'un paramètre essentiel du capteur stéréoscopique : la base. Si l'on considère les équations fondamentales du capteur stéréoscopique rectifié (équation 2.16), un lien de proportionnalité apparaît entre la disparité d'un point donné et la base stéréoscopique. L'augmentation

A.3. La base stéréoscopique

marque	résolution	technologie	matrice	focale	α_u	α_v	fréquence
SONY	768*576	CCD	1/2"	8,5 mm	1100	1100	12,5 Hz
SONY	384*288	CCD	1/2"	8,5 mm	550	550	25 Hz
SMaL	640*480	CMOS	1/3"	6 mm	1100	1100	30 Hz
Jai Pulnix	640*480	CCD	1/3"	12 mm	1100	1100	200 Hz
Jai Pulnix	320*120	CCD	1/3"	12 mm	1100	1100	514 Hz

TAB. A.1 – *Caractéristiques des différentes caméras utilisées dans cette thèse.*

de celle-ci implique donc l'augmentation de la plage de disparités explorées et en conséquence de la dynamique du capteur pour la mesure de distance. La précision et la portée du capteur en bénéficient directement.

En revanche, l'augmentation de cette plage de disparité amène son lot de complications lors du calcul de la carte de disparité. Cette augmentation du nombre d'hypothèses d'appariements conduit inévitablement à une augmentation du temps de calcul et du risque d'erreurs d'appariements. Par ailleurs, plus la différence de point de vue est grande entre les deux caméras, plus les objets non fronto-parallèles sont dissemblants entre les deux images.

Nous avons également vu au chapitre 2 que l'on peut tirer avantage de cette différence de point de vue pour la discrimination des points "route" et "obstacles" par l'utilisation d'une fenêtre de corrélation adaptée. Une base de faible dimension réduit l'intérêt de cette approche. Ce phénomène s'observe lorsque la scène est représentée dans le plan v -disparité (figure A.3). La pente du profil de la route est moins marquée avec une grande base, ce qui la distingue d'autant plus des plans verticaux. En revanche, avec une base courte, la séparation est d'autant plus ambiguë.

En dehors de ces considérations techniques, si l'on se place du point de vue du constructeur ou de l'intégrateur du système, une petite base est préférable. Un lien rigide étant indispensable entre les caméras pour assurer qu'elles soient constamment alignées, il est difficilement envisageable d'installer une barre en métal d'un mètre de long derrière le pare-brise de véhicules grand public. Pour de petits systèmes, certains industriels proposent en revanche de couler les caméras dans une résine qui assure une rigidité parfaite et constante au système stéréoscopique.

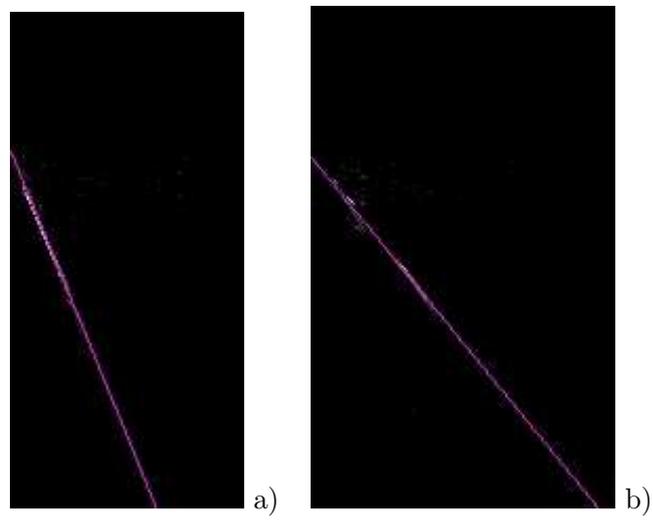


FIG. A.3 – Influence de la base stéréoscopique sur la représentation v -disparité : a) base de 65 cm
b) base de 103 cm.

INFLUENCE DE LA TAILLE DE LA FENÊTRE DE CORRÉLATION

La forme de la fenêtre utilisée pour les méthodes d'appariements par corrélation a une grande influence sur les résultats obtenus. En particulier, ses dimensions jouent un rôle important sur la qualité des cartes de disparité obtenues et le temps de calcul nécessaire à leur obtention. Cela est lié à l'hypothèse selon laquelle une fenêtre de corrélation ne contient qu'un seul objet, fronto-parallèle. Cette contrainte plaide donc pour l'usage de fenêtres de petites dimensions, permettant de correctement appairier les pixels situés sur des objets lointains ou de petite taille. Mais à l'inverse, le critère de corrélation est d'autant plus discriminant que la fenêtre est grande, car elle contient ainsi plus d'information. Augmenter la taille de cette fenêtre permet alors une diminution du nombre d'appariements faux ou ambiguës, et une plus grande robustesse au bruit présent sur les images stéréoscopiques.

Afin d'illustrer ce phénomène, observons les cartes de disparité obtenues pour la paire d'images représentées sur la figure B.1. Celles-ci ont été acquises avec le système stéréoscopique installé dans le véhicule CARLLA (présenté au chapitre 5). Elles comportent plusieurs particularités influant sur la qualité des appariements. Un bruit important, caractéristique des capteurs CMOS, est présent. La distorsion optique et le défaut de rectification épipolaire ne sont pas complètement compensés, ce qui est visible, par exemple, sur le plot situé en bas à droite de l'image de gauche. Enfin, la large base stéréoscopique utilisée impose une forte distorsion perspective sur la surface de la route et une différence de point de vue importante modifiant l'apparence des objets entre les deux images, au niveau des contours d'occultation (la tête du piéton de droite est vue soit devant un buisson, soit devant une cible de calibrage).

Les images de la figure B.2 représentent des cartes de disparité calculées à partir de cette paire d'images, pour différentes tailles de fenêtre. Le critère utilisé est ZSSD, la décision étant prise suivant la stratégie WTA.

L'image (a) illustre l'utilisation d'une fenêtre de petite taille ($3 * 3$ pixels). La carte de disparité résultante est très bruitée.

Si l'on décide d'augmenter fortement la dimension horizontale de la fenêtre ($31 * 3$ pixels), un grand nombre de pixels posant précédemment problème donnent finalement lieu à des appariements corrects (image (b)). En particulier, la différence est très importante sur les arbres, qui contiennent une texture complexe, et sur le marquage situé au centre de l'image. En revanche, la largeur de la fenêtre est trop importante comparée à la dimension de certains obstacles, causant des erreurs aux bords de ceux-ci. La tête du piéton de droite n'est donc plus correctement appariée.

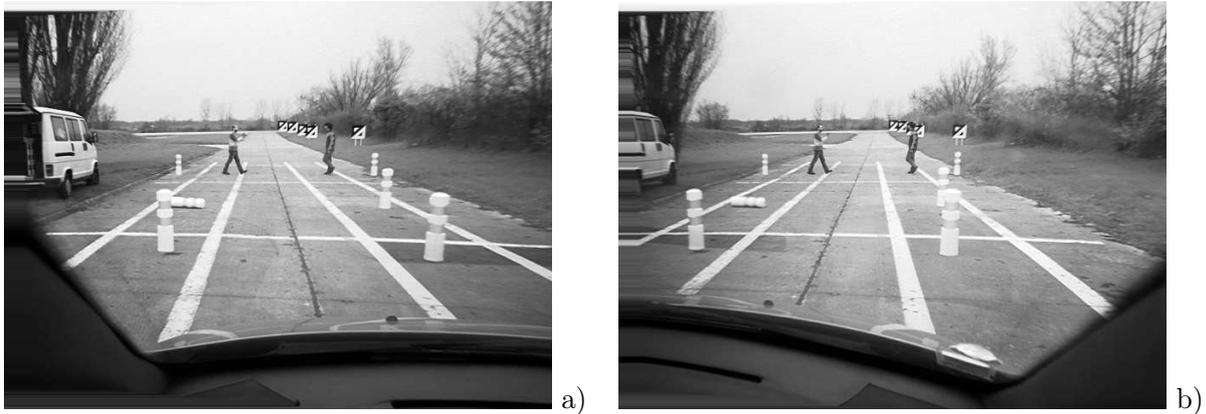


FIG. B.1 – Paire d'image utilisée pour illustrer l'influence de la taille de la fenêtre de corrélation sur la carte de disparité.

Si à l'inverse, la fenêtre de corrélation est agrandie suivant sa direction verticale, comme sur l'image (c) ($3 * 31$ pixels), l'appariement sur les obstacles verticaux est nettement facilité. Ceci est clairement visible sur le piéton de droite, le plot de gauche ou la camionnette. Dès lors, c'est la surface de la route qui pose problème, et le marquage central donne lieu à des valeurs de disparité abhéroentes (il ne présente plus le dégradé de disparité visible sur l'image (b)). Ceci est du au non respect de la contrainte fronto-parallèle. Sur la surface de la route, les pixels contenus dans une fenêtre de corrélation de $3 * 31$ pixels présentent en effet une vingtaine de valeur de disparité différentes, ce qui ne permet pas de trouver une valeur de disparité optimale. La fenêtre mesurant $31 * 3$ pixels est insensible à ce phénomène car, en l'absence de dévers ou de roulis, chaque ligne à la surface de la route a une disparité unique.

Enfin, l'utilisation d'une fenêtre carrée de taille intermédiaire ($15 * 15$ pixels) ne permet pas de compenser les inconvénients des deux autres types de fenêtre (image (d)). Ainsi, si l'appariement est meilleur pour le piéton ou la camionnette, la fenêtre est trop haute pour permettre d'obtenir une estimation correcte de la disparité sur le marquage central, et trop large pour donner de bons résultats sur le plot de gauche.

L'approche que nous avons proposée au chapitre 2 exploite deux types de fenêtres de corrélation, rectangulaire pour les objets frontaux et en parallélogramme pour la surface de la route. Elle permet de résoudre le compromis nécessaire sur les dimensions de la fenêtre de corrélation, imposé par la présence de deux types de surface. La figure B.3 illustre le résultat obtenu avec l'utilisation d'une telle fenêtre de dimension $15 * 15$ pixels. Il est notable que cette stratégie permet d'obtenir des estimations correctes sur la surface de la route comme sur les objets présents dans la scène.

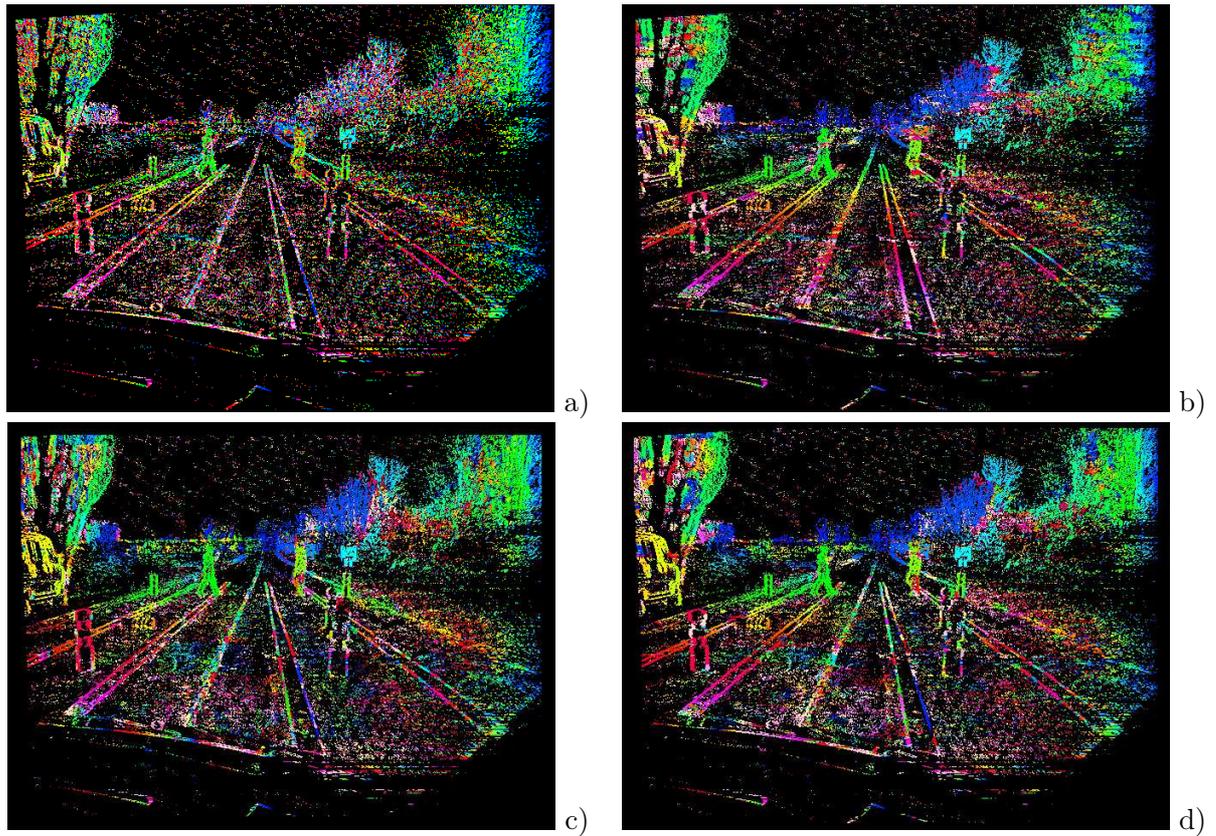


FIG. B.2 – Influence des dimensions de la fenêtre de corrélation sur une image de disparité. a) fenêtre de $3 * 3$ pixels. b) fenêtre de $31 * 3$ pixels. c) fenêtre de $3 * 31$ pixels. d) fenêtre de $15 * 15$ pixels.

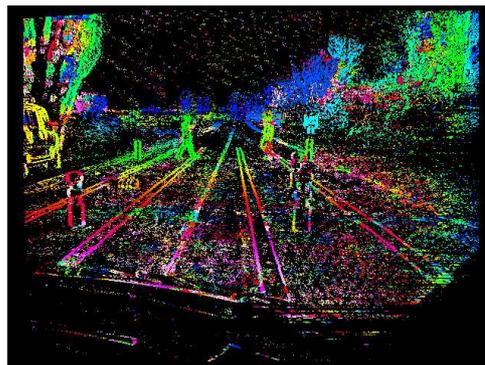


FIG. B.3 – Carte de disparité obtenue avec une fenêtre de corrélation inclinable de taille $15 * 15$ pixels.

Afin d'évaluer de façon quantitative l'influence de la taille de la fenêtre de corrélation sur l'estimation de la disparité, nous avons exploité des images issues du simulateur SiVIC (figure B.4). Celles-ci permettent de disposer d'une carte de disparité de référence obtenue à partir du

z-buffer associé à la représentation 3D de la scène. Les cartes de disparité calculées sont donc comparées à cette référence de façon à calculer un pourcentage de bons appariements, défini comme le nombre de pixels dont la disparité diffère de moins de 1 par rapport à la référence, rapporté au nombre de pixels appariés [SS02a]. Les zones d'occultations partielles sont déterminées par comparaison des z-buffers des caméras droite et gauche, et ne sont pas prises en compte pour l'évaluation de ce critère.



FIG. B.4 – Images issues de la première version du simulateur de capteurs SiVIC, utilisées pour comparer les performances de différentes tailles de fenêtres de corrélation.

a) image de gauche. b) image de droite. c) image de référence de disparité.

Le système stéréoscopique simulé a des caractéristiques équivalentes au système embarqué dans le véhicule VIPER (chapitre 5), pour une résolution 1/4 PAL. Compte tenu de l'absence de bruit sur les images fournies, un bruit gaussien de variance 6 a été ajouté pour rendre l'expérience plus réaliste et éviter que l'appariement soit artificiellement trop simplifié. La figure B.5 illustre les résultats obtenus en faisant varier les dimensions de la fenêtre de corrélation. Le taux de bons appariement reste assez faible car les images utilisées contiennent un grand nombre de zones uniformes qui, une fois bruitées, induisent un grand nombre d'erreurs. Néanmoins, il est notable que l'augmentation de la largeur de la fenêtre de corrélation conduit à une amélioration des résultats, contrairement à l'augmentation de sa hauteur qui, au delà de 5 pixels, diminue la qualité globale de l'image obtenue.

Lorsque l'on utilise la stratégie de la fenêtre "inclinable", l'augmentation des dimensions de la fenêtre dans les deux directions améliore de façon égale la qualité du résultat. Cela illustre l'intérêt de l'approche proposée pour le traitement des scènes routières acquises avec un système à large base stéréoscopique, ou situé à faible hauteur, qui induit une importante distorsion perspective à la surface de la route. En revanche, la courbe tendant vers un palier asymptotique, il semble peu utile d'utiliser un voisinage de corrélation de grande taille pour le traitement de scènes simples. Ceci est important car les dimensions de ce voisinage influent fortement sur le temps de calcul nécessaire au traitement d'une image de disparité. Comme l'illustre la figure B.6, celui-ci augmente linéairement avec l'augmentation de la taille de la fenêtre dans une direction comme une autre.

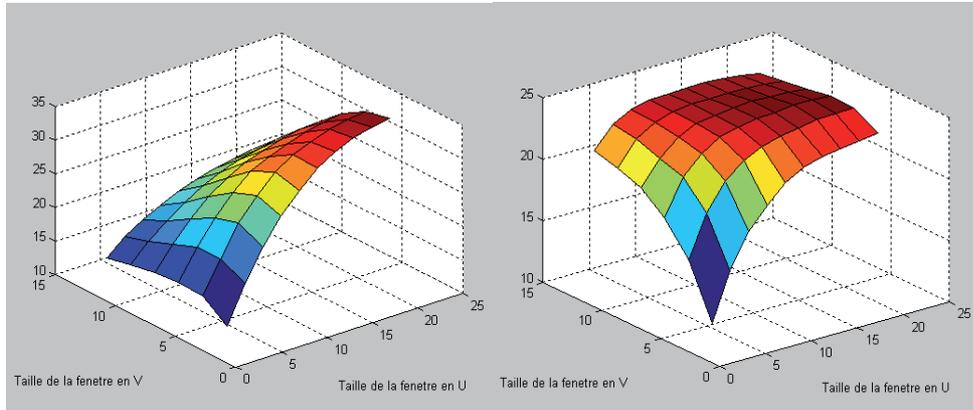


FIG. B.5 – *Evaluation du taux de bons appariements (en %) sur une scène synthétique en fonction de la taille de la fenetre : à gauche avec une fenetre classique, à droite avec une fenetre inclinable.*

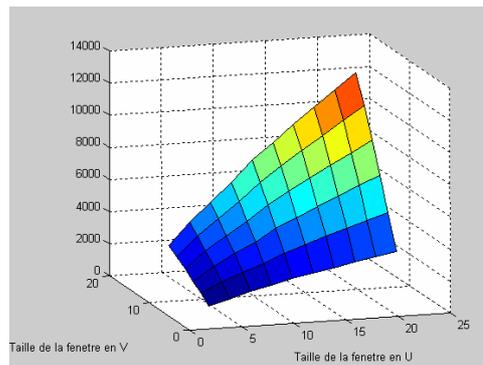


FIG. B.6 – *Temps de calcul d'une image de disparité (en ms) en fonction de la taille de la fenetre de corrélation utilisée.*

Résultats obtenus sur un Pentium 4-C, 2,4 GHz.

TRANSFORMÉE DE HOUGH POUR LA DÉTECTION DE PLANS DANS L'ESPACE DE DISPARITÉ

C.1 Transformée de Hough pour la détection de droites dans l'image

Adaptée au traitement d'images à partir d'un brevet déposé par Hough [Hou62], la transformée de Hough permet de détecter de façon robuste diverses formes dans une image. Pour cela, l'image est projetée dans un espace de paramètres bien adapté, dans lequel la détection se résume à une détection de pics. L'application la plus courante de cette transformation est la détection d'alignement de pixels dans une image binaire.

Pour réaliser cette transformation, il faut considérer la paramétrisation normale des droites présentés sur la figure C.1. Un couple (ρ_h, Θ_h) correspond alors à une droite de type $y = a_1 \cdot x + a_0$ avec :

$$\begin{cases} a_1 &= -\frac{1}{\tan \theta_h} \\ a_0 &= \frac{\rho_h}{\sin \Theta_h} \end{cases} \quad (\text{C.1})$$

Ainsi, si l'on considère un plan repéré par les variables (ρ_h, Θ_h) , chaque point de ce plan correspond à une droite du plan cartésien classique.

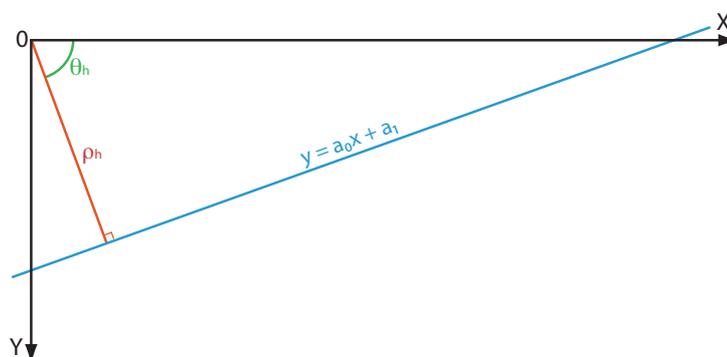


FIG. C.1 – Paramétrisation d'une droite dans l'espace de Hough.

A l'inverse, un point (x, y) du plan cartésien passe une infinité de droites. Ces droites appar-

tiennent à une même famille et respectent toutes l'équation :

$$\rho_h = x * \cos(\Theta_h) + y * \sin(\Theta_h) \quad (\text{C.2})$$

Ainsi, la projection d'un point (x, y) dans l'espace paramétrique de Hough est la sinusoïde définie par l'équation C.2. En conséquence, lorsque plusieurs points sont alignés dans l'image initiale, les sinusoïdes de leurs projections se coupent en un même point de l'espace de Hough qui définit justement la droite formée par cet alignement. Une droite apparaît donc comme un maximum dans le plan de Hough.

C.2 Adaptation à la détection d'un plan : transformée en niveaux de gris

L'approche v-disparité [Lab03] suppose que, pour détecter un plan dans l'espace de disparité E_Δ , il est préférable de projeter avec accumulation la totalité des points de cet espace suivant une direction connue du plan à détecter. Ce procédé forme une image d'accumulation dans laquelle le niveau de gris d'un pixel correspond au nombre de points de E_Δ projetés sur celui-ci. Le plan à détecter apparaît alors comme une alignement de points. La transformée de Hough semble être alors un outil très adapté à la détection de cet alignement.

Cependant, la version classique de la transformée de Hough pour détecter des droites suppose la binarisation préalable de l'image à l'aide d'un seuillage. Mis à part le fait que cette méthode suppose l'emploi d'un seuil à ajuster en fonction du bruit présent dans l'image, cette approche supprime partiellement l'aspect "global" de la détection de la route dans l'image v-disparité, puisqu'elle supprime l'information de cumul des pixels pour chaque ligne de l'image v-disparité. Si, au contraire, l'accumulation dans l'espace de paramètres de Hough tient compte du niveau de gris de chacun des pixels de l'image v-disparité, la valeur d'accumulation de la droite détectée représente bien la détection d'un plan, en révélant le nombre total de points présents sur ce plan, et non pas simplement le nombre de points présents sur la droite qui le caractérise dans l'image v-disparité.

La transformée de Hough en niveaux de gris peut néanmoins poser problème lorsqu'un pixel de l'image v-disparité a une valeur très élevée. Cela peut être le cas lorsqu'un obstacle contient un grand nombre d'indices visuels sur une ligne donnée. Le point ayant un niveau très élevé peut alors avoir un effet de "point attracteur", car tous les alignements de pixels passant par ce point ont une valeur de cumul très importante dans l'espace de Hough. Cette valeur peut alors être supérieure au cumul formé par des alignements réels de pixels de faible niveau de gris. La figure C.3 illustre un tel cas, mettant en défaut la transformée en niveaux de gris.

Lorsque l'approche v-disparité est utilisée pour l'estimation de la surface de la route, le phénomène de "point attracteur" peut donc fausser cette estimation, comme cela a été présenté dans [SPLA07]. Nous avons proposé au chapitre 2 une méthode permettant de supprimer ce problème en ne projetant sur l'image v-disparité que les pixels appartenant vraisemblablement à la surface de la route.

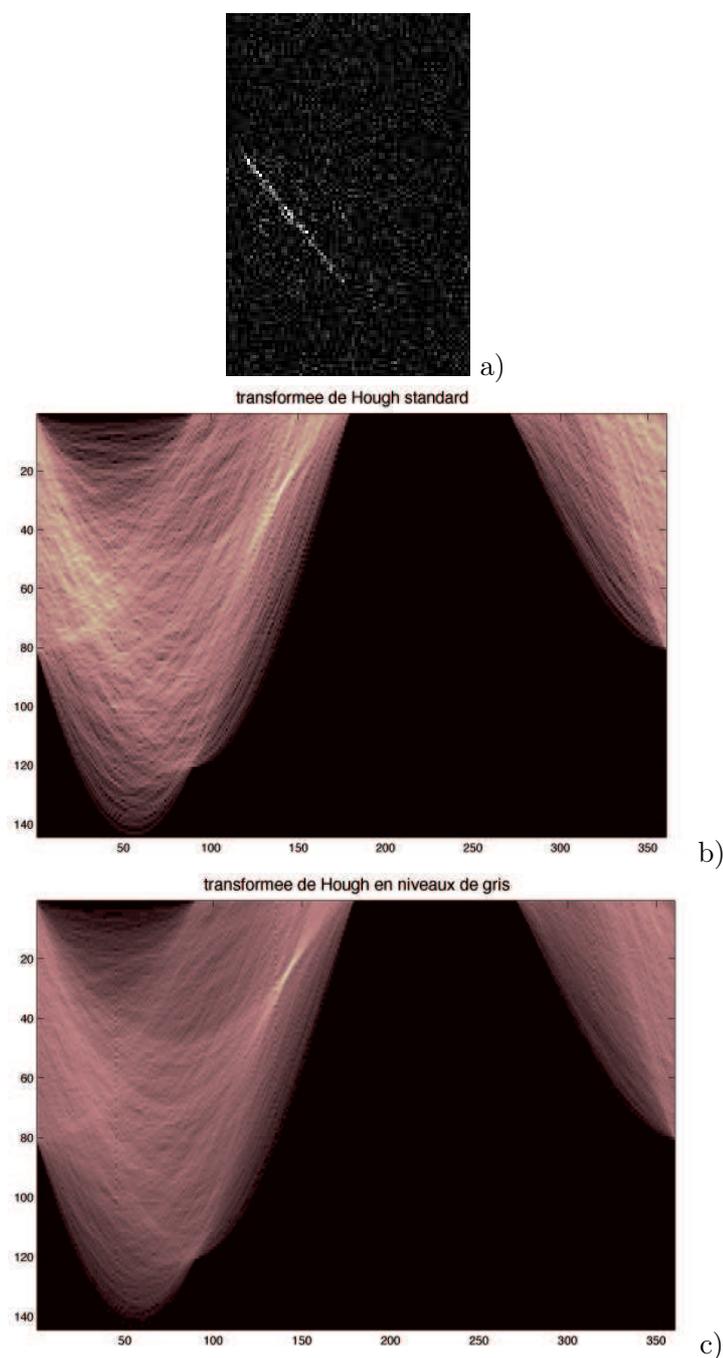


FIG. C.2 – Les deux approches de la transformée de Hough appliquées à une image v -disparité bruitée (a) pour trouver le profil global de la route. .

b) Pour la transformée classique, l'image v -disparité est seuillée avant projection dans l'espace de paramètres (le seuil utilisé est 10). c) La transformée en niveaux de gris tient compte de la valeur d'accumulation de chacun des pixels de l'image v -disparité.

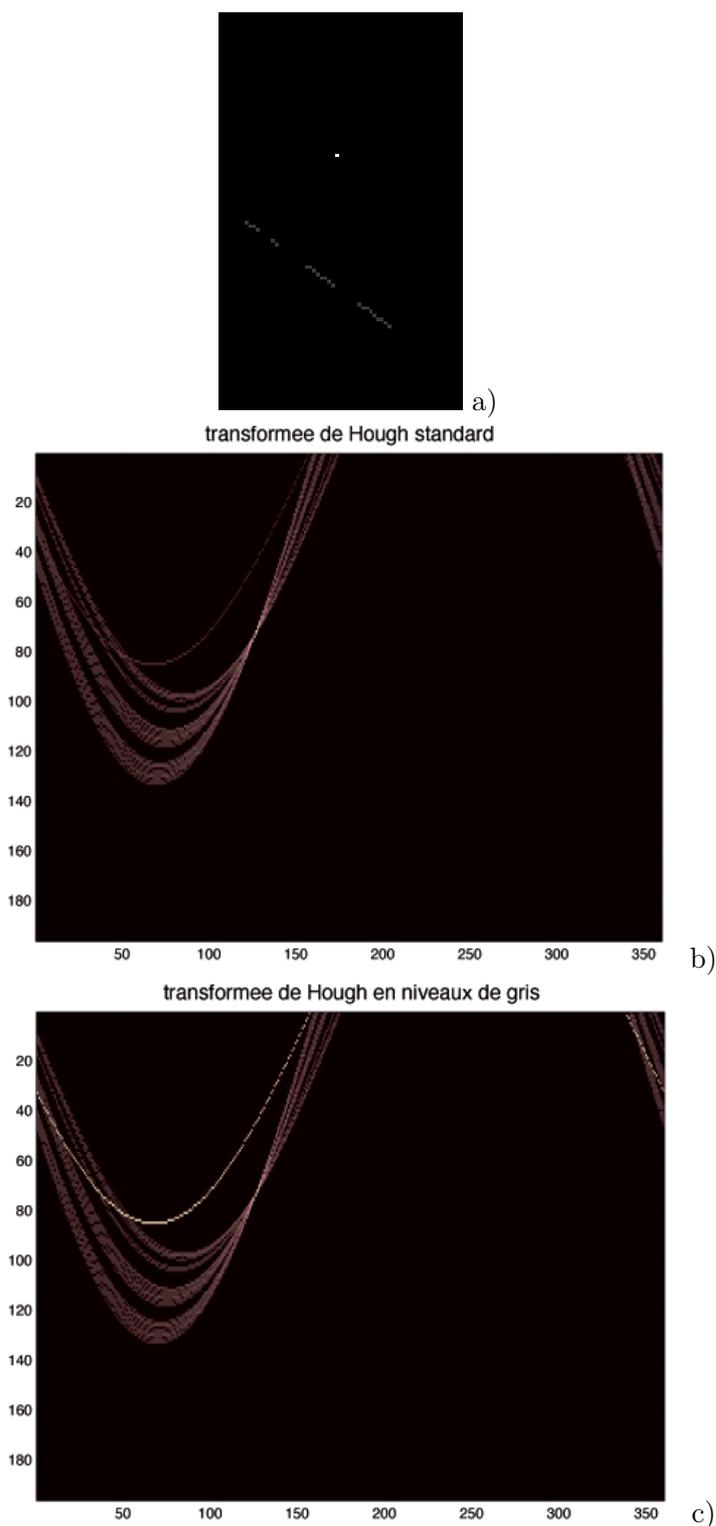


FIG. C.3 – Limites de la transformée de Hough en niveaux de gris.

a) image contenant un alignement et un point attracteur. b) Dans la transformée de Hough classique, l'alignement est visible comme un maximum. c) Dans la transformée en niveaux de gris, le point attracteur crée une sinusoïde de forte intensité qui perturbe la détection.

CALIBRAGE ET CONFIGURATION DES CAPTEURS

Afin de permettre une coopération entre les trois capteur que nous avons utilisés, il est indispensable que leur position respective par rapport au référentiel R_a soit connue. Pour cela, une étape préalable de calibrage et de configuration est nécessaire. Celle-ci est réalisée manuellement sur le banc de calibrage des pistes d'essai du site de Satory, représenté sur la figure D.1.



FIG. D.1 – Banc de calibrage du site de Satory utilisé pour réaliser la configuration des différents capteurs.

Les marquages au sol permettent de positionner les caméras. Les objets identifiés sont placés dans des positions connues pour finaliser la configuration du lidar et de l'identification optique.

Dans un premier temps, les paramètres intrinsèques de toutes les caméras utilisées (stéréovision et identification optique) doivent être correctement estimés. A cette fin, la boîte à outil matlab proposée par Bouguet⁴³ est très efficace.

Concernant le capteur stéréoscopique, la configuration est réalisée grâce aux marquages du banc de calibrage, en suivant la méthodologie proposée en annexe de la thèse de Labayrade [Lab03]. Celle-ci consiste à vérifier que les plans image sont alignés, que la géométrie épipolaire est correctement rectifiée, et que les pixels situés à l'infini ont une disparité nulle.

Tout d'abord, les marquages transverses doivent apparaître horizontaux dans les images, ce qui signifie que le roulis des caméras est nul. Ceux-ci doivent également être alignés entre les deux

⁴³http://www.vision.caltech.edu/bouguetj/calib_doc/

images d'une paire stéréoscopique, pour s'assurer que le tangage des deux caméras est identique. Enfin, l'intersection des marquages longitudinaux et parallèles du banc de calibrage doit avoir u_o pour abscisse dans les deux images, certifiant alors que l'angle de lacet des deux caméras est nul.

Cette configuration est ensuite affinée par une rectification logicielle des images stéréoscopiques, en exploitant la méthode décrite dans [LZ99]. L'étape préalable est néanmoins indispensable, car la configuration initiale du système doit être proche de la solution recherchée pour que la méthode logicielle fonctionne correctement.

Les configurations du lidar et de l'identification optique sont ensuite réalisées en positionnant des objets identifiés (lampes émettrices) à des positions connues du banc de calibrage et en réglant manuellement le système pour que la reconstruction soit correcte.

Un objet positionné le long de l'axe $X_a = X_l^0$ permet d'obtenir une mesure de l'angle de lacet du lidar. Ce dernier fournit en effet une mesure polaire (ρ_l, ϕ_l) pour l'impact situé sur le support vertical de la lampe. Théoriquement, $\phi_l = 0$. En pratique, ϕ_l est égal à l'opposé de l'angle de lacet.

Une fois corrigé le lacet du télémètre, son tangage est estimé en mesurant la hauteur Y_a^0 à laquelle le plan de balayage coupe la droite définie par $(X_a = X_l^0, Z_a = Z_a^0)$. La hauteur Y_l^0 du lidar étant mesurée par ailleurs, le tangage est donné par : $\theta_l = atan(\frac{Y_a^0 - Y_l^0}{Z_a^0})$.

La mesure de position d'un autre objet, positionné différemment, permet de valider le résultat de cette étape de calibrage.

Pour la configuration du capteur d'identification optique, la caméra est en premier lieu positionnée à l'aide des marquages au sol, de façon à annuler ses angles de roulis et de lacet. Le tangage est ensuite estimé par une procédure similaire à celle exploitée pour le lidar, en comparant la mesure de position fournie par le capteur à la position connue de la lampe émettrice. Compte tenu de la précision assez faible du capteur d'identification, cette procédure est effectuée pour diverses positions de la lampe, de façon à obtenir une valeur moyenne des paramètres.

Enfin, précisons que pour les deux systèmes exploitant des caméras, les étapes de calibrage et de configuration sont réalisées en utilisant la plus haute résolution d'images que le capteur est susceptible de fournir, afin de profiter d'un maximum de précision.

DIMENSIONNEMENT DU CAPTEUR D'IDENTIFICATION OPTIQUE

Comme cela a été présenté au chapitre 3, le capteur utilisé pour l'identification optique est une caméra sensible au rayonnement infrarouge proche et équipée d'un filtre passe-bande. Si la plupart des caméras pourrait convenir, il est préférable d'utiliser une caméra qui ait une fréquence d'acquisition élevée. Pour s'en convaincre, définissons F_{cam} et F_{em} , respectivement les fréquences d'acquisition de la caméra et d'émission de la lampe. Le théorème d'échantillonnage de Shanon impose que la fréquence d'échantillonnage d'un signal soit supérieure ou égale au double de sa fréquence la plus élevée. Afin de respecter cette condition, définissons un taux de sur-échantillonnage N_{ipb} comme étant le nombre d'images acquises par la caméra par bit émis par la lampe :

$$N_{ipb} = \frac{F_{cam}}{F_{em}} \quad (E.1)$$

Le temps de décodage d'une trame de longueur N_{trame} est alors :

$$T_{reception} = \frac{N_{trame}}{F_{em}} = \frac{N_{trame} * N_{ipb}}{F_{cam}} \quad (E.2)$$

A titre d'exemple, prenons une trame courte de longueur $N_{trame} = 16$ bits ($N_{start} = 4$ bits (1101), $N_{ID} = 10$ bits). Un tel format de trame permet de coder moins de 1024 identifiants différents. Si l'on pose $N_{ipb} = 2,5$ pour assurer un peu de marge, on obtient pour une fréquence vidéo classique ($F_{cam} = 25$ Hz) : $T_{reception} = 1,6$ s.

Ce chiffre est trop élevé pour un système qui se veut réactif et met en évidence le besoin d'un récepteur à haute fréquence surpassant largement la cadence vidéo. Heureusement, l'évolution des systèmes d'acquisition et de l'électronique associée a récemment augmenté l'offre en caméras haute fréquence. En particulier, certains modèles de caméra permettent d'acquérir alternativement les lignes adjacentes pour augmenter la fréquence d'acquisition au détriment de la résolution. Nous avons exploité pour cela une caméra PULNIX TM-6740CL, permettant d'atteindre une fréquence d'acquisition de 514 Hz, pour une résolution de 320*120 pixels. La figure E.1-bas illustre l'effet produit par un tel mode d'acquisition.

Dès lors, en configurant la lampe émettrice de façon à émettre 210 bits par seconde (soit $N_{ipb} = 2,36$), un message de longueur $N_{trame} = 14$ bits est théoriquement décodé en : $T_{reception} = 66$ ms.

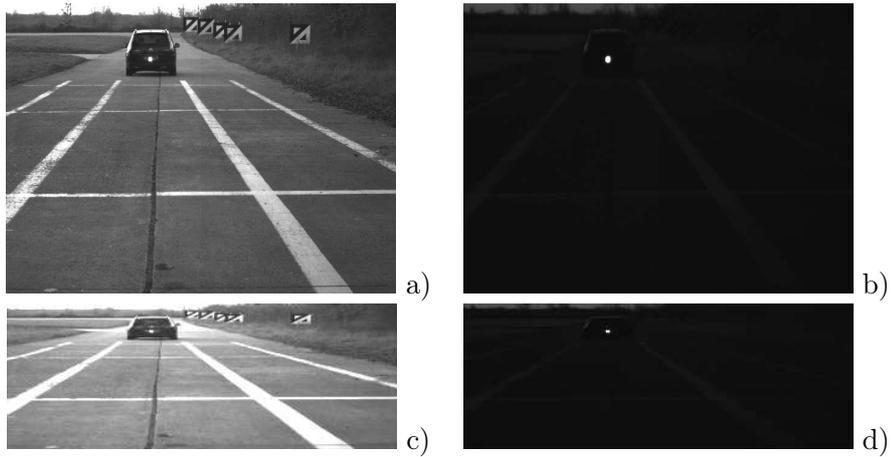


FIG. E.1 – Image d’une scène routière acquise avec la caméra utilisée comme récepteur sans (a) et avec le filtre passe bande 880 nm (b). La même scène acquise en résolution réduite à une fréquence de 514 images par seconde (c et d).

Afin de disposer d’un nombre suffisant d’identifiants, nous avons choisi d’utiliser des messages de longueur $N_{trame} = 16$ bits. Le temps théorique de confirmation est alors 73 ms. En pratique, celui-ci est en moyenne de 100 ms, car la réception d’un message ne commence pas nécessairement par le code de début de celui-ci.

PRODUCTION SCIENTIFIQUE AU COURS DE LA THÈSE

F.1 Publications

Communications dans des congrès internationaux

- M. Perrollaz, R. Labayrade, C. Royère, N. Hautière & D. Aubert : Long Range Obstacle Detection Using Laser Scanner and Stereovision (IEEE Intelligent Vehicles Symposium, Tokyo, 2006).
- N. Hautière, R. Labayrade, M. Perrollaz, & D. Aubert : Road Scene Analysis by Stereovision : a Robust and Quasi-Dense Approach (International Conference on Control, Automation, Robotics and Vision, Singapour, 2006).
- A. von Arnim, M. Perrollaz, A. Bertrand & J. Ehrlich : Vehicle Identification Using Infrared Vision and Applications to Cooperative Perception (IEEE Intelligent Vehicles Symposium, Istanbul, 2007).
- M. Perrollaz, R. Labayrade, R. Gallen & D. Aubert : A Three Resolution Framework for Reliable Road Obstacle Detection Using Stereovision (IAPR International Conference on Machine Vision and Applications, Tokyo, 2007).
- N. Soquet, M. Perrollaz, R. Labayrade & D. Aubert : Free Space Estimation for Autonomous Navigation (IAPR International Conference on Vision Systems, Bielefeld, 2007).

Communications avec publication dans des actes

- M. Perrollaz : Détection d'obstacles longue portée par télémétrie laser et stéréovision (Acte INRETS N°108).

Chapitre d'ouvrage

- R. Labayrade, D. Gruyer, C. Royère & M. Perrollaz : Obstacle Detection in Outdoor Environments Based on Fusion Between Stereovision and Laser Scanner. (Chapitre 7 de "Moving Robots, Moving Intelligence", Novembre 2006. Coordonné par l'éditeur en chef de : International Journal of Advanced Robotic Systems).

F.2 Brevet

- A. von Arnim, M. Perrollaz & J.M. Chevreau : «Système et procédé de communication et d'identification intervehiculaire», Brevet Français déposé le 15 Janvier 2007 sous le numéro

07 52663.

F.3 Participation à des projets de recherche

- DO30 : Detection d'obstacles à 30 m. (ANR)
- LOVE : Logiciel d'Observation de Vulnérables. (ANR)
- TRACKSS : Technologies for Road Advanced Cooperation Knowledge Sharing Sensors. (EUROPE)
- Collaboration avec ATMEL : Conseil technique pour un projet de développement d'un corrélateur matériel pour le calcul de carte de disparité à base d'AVR32.

F.4 Séminaires et Démonstrations

- Démonstrations lors des journées portes-ouvertes du LIVIC, 2005.
- Démonstrations au comité d'évaluation du LIVIC, 2006.
- Démonstrations lors congrès mondial de la route, 2007.
- Démonstrations lors du carrefour PREDIT, 2008.
- Démonstrations lors d'un symposium franco-japonais à l'INRIA, 2008.
- Diverses démonstrations au LIVIC.

F.4.1 Rapport d'articles soumis à la publication

- Revue d'un article pour IEEE International Conference on Intelligent Transportation Systems, 2007.
- Revue d'un article pour IEEE International Conference on Intelligent Transportation Systems, 2008.
- Revue d'un article pour IEEE Transactions on Vehicular Technologies, 2008.

F.5 Vulgarisation scientifique

- Animation scientifique sur le site de l'INRETS lors pour la "Fête de la Science" en 2004 et 2006 (Jardin du Luxembourg, Paris).
- Démonstrations et interviews télévisées pour les chaînes TF1 et i-TV, lors du carrefour PREDIT, 2008.

Bibliographie

- [AMAN04] A. Albiol, J.M. Mossi, A. Albiol, and V. Naranjo. Automatic license plate reading using mathematical morphology. In *Proceedings of the VIIP Conference*, Marbella, Spain, 2004.
- [AMC⁺00] R. Aufrère, F. Marmoiton, R. Chapuis, F. Collange, and J.P. Dérutin. Detection de route et suivi de véhicules par vision pour l'acc. *Traitement du Signal*, 17(3) :233–248, 2000.
- [Amm07] S. Ammoun. *Contribution des communications intervéhiculaires pour la conception de systèmes avancés d'aide à la conduite*. PhD thesis, Ecole des mines de Paris, Decembre 2007.
- [BB98] M. Bertozzi and A. Broggi. Gold : A parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Transactions on Image Processing*, 7(1), January 1998.
- [BBFN00] M. Bertozzi, A. Broggi, A. Fascioli, and S. Nichele. Stereo vision-based vehicle detection. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Detroit, USA, October 2000.
- [BBLR05] M. Bertozzi, A. Broggi, A. Lasagni, and M. Del Rose. Infrared stereo vision-based pedestrian detection. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Las Vegas, USA, 2005.
- [BBM⁺06] M. Bertozzi, A. Broggi, P. Medici, P.P.Porta, and A. Sjögren. Stereo vision-based start-inhibit for heavy goods vehicles. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, 2006.
- [BCPZ06a] A. Broggi, C. Caraffi, P.P. Porta, and P. Zani. The single frame stereo vision system for reliable obstacle detection used during the 2005 darpa grand challenge on terramax. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada, September 2006.
- [BCPZ06b] A. Broggi, S. Cattani, P.P. Porta, and P. Zani. A laserscanner-vision fusion system implemented on the terramax autonomous vehicle. In *Proceedings of the International Conference on Intelligent Robots and Systems*, Beijing, China, October 2006.
- [BFT⁺06] A. Broggi, R. I. Fedriga, A. Tagliati, T. Graf, and M. Meinecke. Pedestrian detection on a moving vehicle : an investigation about near infra-red images. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, June 2006.
- [BHD97] M. Betke, E. Haritaogiu, and L.S. Davis. Highway scene analysis in hard real-time. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, Boston, USA, 1997.
- [BN98] M. Betke and H. Nguyen. Highway scene analysis form a moving vehicle under reduced visibility conditions. In *Proceedings of the IEEE International Conference on Intelligent Vehicles*, Stuttgart, Germany, October 1998.

- [BP99] S. Blackman and R. Popoli. *Design and analysis of modern tracking systems*. Artech House Publishers, 1999.
- [BT98] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4) :401–406, April 98.
- [BTG06] H. Bay, T. Tuytelaars, and L. Van Gool. Surf : Speeded up robust features. In *Proceedings of the 9th European Conference on Computer Vision*, Graz, Austria, 2006.
- [Cha05] S. Chambon. *Mise en correspondance stéréoscopique d'images couleur en présence d'occultations*. PhD thesis, Université Paul Sabatier - Toulouse III, 2005.
- [col05] Ouvrage collectif. Sur la route... la sécurité. In *Les collections de l'INRETS*. INRETS, 2005.
- [Dev97] F. Devernay. *Vision stéréoscopique et propriétés différentielles des surfaces*. PhD thesis, Ecole Polytechnique, 1997.
- [DLLC05] J. Douret, R. Labayrade, J. Laneurit, and R. Chapuis. A reliable and robust lane detection system based on the parallel use of three algorithms for driving safety assistance. In *IAPR Conference on Machine Vision and Applications*, Tsukuba, Japan, May 2005.
- [EW02] G. Egnal and Richard P. Wildes. Detecting binocular half-occlusions : Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8) :1127–1133, 2002.
- [Fab04] J. Fabrizio. *Localisation d'obstacles coopératifs par systèmes de vision classiques et panoramiques*. PhD thesis, Université Pierre et Marie Curie - Paris 6, 2004.
- [FDW⁺06] B. Fardi, J. Dousa, G. Wanielik, B. Elias, and A. Barke. Obstacle detection and pedestrian recognition using a 3d pmd camera. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, June 2006.
- [FJ00] U. Franke and A. Joos. Real-time stereo vision for urban traffic scene understanding. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Dearborn, USA, October 2000.
- [FMDH00] B. Fremont, A. Menhaj, P. Deloof, and M. Heddebaut. A cooperative collision avoidance and communication system for railway transports. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Dearborn, USA, Octobre 2000.
- [FRT97] A. Fusiello, V. Roberto, and E. Trucco. Efficient stereo with multiple windowing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 858–863, Puerto-Rico, June 1997.
- [FTV00] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1) :16–22, 2000.
- [GE96] V. Graefe and W. Efenberger. A novel approach for the detection of vehicles on free-ways by real-time vision. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, September 1996.
- [GLM⁺01] P. Griffiths, D. Langer, J. A. Misener, M. Siegel, and C. Thorpe. Sensor-friendly vehicle and roadway systems. In *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hongrie, 2001.

- [GP99] D.M. Gavrila and V. Philomin. Real-time objet detection for "smart" vehicles. In *Proceedings of the IEEE International Conference on Computer Vision*, pages p.87–93, Kerkyra, 1999.
- [Gru99] D. Gruyer. *Etude du Traitement de données imparfaites pour le suivi multi-objets : Application aux situations routières*. PhD thesis, Université de Technologie de Compiègne, Decembre 1999.
- [GSLP07] D. Geronimi, A. D. Sappa, A. Lopez, and D. Ponsa. Adaptive image sampling and windows classification for on-board pedestrian detection. In *Proceedings of the 5th International Conference on Computer Vision Systems*, Bielefeld, Germany, April 2007.
- [GTAN02] E. Gerbenne, E. Tournie, R. Artur, and C. Narduzzi. Arcos 2004 - analyse fonctionnelle. Technical report, ARCOS, 2002.
- [Han99] J. A. Hancock. *Laser Intensity-Based Obstacle Detection and Tracking*. PhD thesis, Carnegie Mellon University, 1999.
- [HIG02] H. Hirschmüller, P. R. Innocent, and J. M. Garibaldi. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, 47(1-3) :229–246, 2002.
- [HLPA06] N. Hautière, R. Labayrade, M. Perrollaz, and D. Aubert. Road scene analysis by stereovision : a robust and quasi-dense approach. In *Proceedings of the 9th IEEE International Conference on Control, Automation, Robotics and Vision*, Singapore, 2006.
- [HM95] R. Horaud and O. Monga. *Vision par ordinateur : Outils fondamentaux*. Hermes, 1995.
- [Hou62] P.V.C. Hough. A method and means for recognizing complex patterns. U.S. Patent No. 3,069,654, 1962.
- [HZ03] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision (second edition)*. Cambridge University Press, 2003.
- [IKN98] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11) :1254–1259, 1998.
- [Jau02] Cahier Technique Série Jaune. Régulation auto-adaptative de la vitesse de roulage acc. Technical report, Bosch, 2002.
- [KBD05] N. Kaempchen, M. Buehler, and K. Dietmayer. Feature-level fusion for free-form object tracking using laserscanner and video. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Las Vegas, USA, 2005.
- [KKO⁺06] Y. Kimura, T. Kato, M. Ohta, Y. Ninomiya, Y. Takagi, M. Usami, and S. Tokoro. Stereovision for obstacle detection. In *Proceedings of the 13th World Congress on Intelligent Transportation Systems*, London, UK, 2006.
- [KO94] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptative window : theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9) :920–932, September 1994.
- [Kon97] K. Konolige. Small vision system : Hardware and implementation. In *Proceedings of the 8th International Symposium on Robotics Research*, October 1997.

- [LA03a] R. Labayrade and D. Aubert. In-vehicle obstacle detection and characterization by stereovision. In *Proceedings of the 1st International Workshop on In-Vehicle Cognitive Computer Vision Systems*, Graz, Austria, November 2003.
- [LA03b] R. Labayrade and D. Aubert. A single framework for vehicle roll, pitch, yaw estimation and obstacles detection by stereovision. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Columbus, USA, June 2003.
- [Lab03] R. Labayrade. *Détection Générique, Robuste et Rapide d'obstacles Routiers par Stéréovision Embarquée*. PhD thesis, Université Paris 6, 2003.
- [LAC06] S. Lefevbre, S. Ambellouis, and F. Cabestaing. Obstacle detection on a road by dense stereovision with 1d correlation windows and fuzzy filtering. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada, 2006.
- [Lem05a] V. Lemonde. Détection d'obstacles par stéréovision sur véhicules intelligents. In *Proceedings des 9èmes journées Jeunes Chercheurs en Vision par Ordinateur (ORASIS 05)*, 2005.
- [Lem05b] V. Lemonde. *Stéréovision Embarquée sur véhicule : de l'autocalibrage à la détection d'obstacles*. PhD thesis, Institut National des Sciences Appliquées de Toulouse, 2005.
- [Low04] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2) :91–110, 2004.
- [LP03] H. S. Liu and G. Pang. Positioning beacon system using digital camera and leds. *IEEE Transactions on Vehicular Technology*, 52(2), Mars 2003.
- [LQ02] 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.
- [LRA05] R. Labayrade, C. Royère, and D. Aubert. A collision mitigation system using laser scanner and stereovision fusion and its assesment. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 440–446, Las Vegas, USA, 2005.
- [LRGA05] R. Labayrade, C. Royere, D. Gruyer, and D. Aubert. Cooperative fusion for multi-obstacles detection with the use of stereovision and laser scanner. *Autonomous Robots*, 19(2) :117–140, 2005.
- [LZ99] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 1999.
- [MBN04] A. Mendes, L. Conde Bento, and U. Nunes. Multi-target detection and tracking with a laserscanner. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, University of Parma, Italy, 2004.
- [MGL06] B. Mourllion, D. Gruyer, and A. Lambert. A study on the safety-capacity tradeoff improvement by warning communications. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada, 2006.
- [MGS⁺07] T. Michalke, A. Gepperth, M. Schneider, J. Fritsch, and C. Goerick. Towards a human-like vision system for resource-constrained intelligent cars. In *Proceedings of the 5th International Conference on Computer Vision Systems*, 2007.
- [Mor77] H. Moravec. Toward automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pages 584–590, 1977.
- [Mou06] B. Mourllion. *Extension d'un système de perception embarqué par communication. Application à la diminution du risque routier*. PhD thesis, Université Paris XI, Decembre 2006.

-
- [MS07] W. Miled-Souid. *Mise en correspondance stéréoscopique par approches variationnelles convexes ; Application à la détection d'obstacles routiers*. PhD thesis, Université Paris-Est, 2007.
- [NDF⁺04] S. Nedevschi, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, T. Graf, and R. Schmidt. High accuracy stereovision approach for obstacle detection on non planar roads. In *Proceedings of the IEEE Intelligent Engineering Systems*, Cluj Napoca, Romania, September 2004.
- [NOD⁺06] S. Nedevschi, F. Oniga, R. Danescu, T. Graf, and R. Schmidt. Increased accuracy stereo approach for 3d lane detection. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, June 2006.
- [Roy02] C. Royère. *Contribution à la résolution du conflit dans le cadre de la théorie de l'évidence. Application à la perception et à la localisation des véhicules intelligents*. PhD thesis, Université de Technologie de Compiègne, Janvier 2002.
- [SBS00] H. W. Schreier, J. R. Braasch, and M. A. Sutton. Systematic errors in digital image correlation caused by intensity interpolation. *Optical Engineering*, 39(11) :2915–2921, 2000.
- [Sha49] C.E. Shannon. Communication in the presence of noise. In *Proceedings of the Institute of Radio Engineers*, volume 37, pages 10–21, January 1949.
- [SLSW02] B. Steux, C. Laurgeau, L. Salesse, and D. Wautier. Fade : A véhicule detection and tracking system featuring monocular color vision and radar data fusion. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Versailles, France, 2002.
- [SMS00] G.P. Stein, O. Mano, and A. Shashua. A robust method for computing vehicle ego-motion. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Dearborn, USA, 2000.
- [SMW03] M. Skuttek, M. Mekhaïel, and G. Wanielik. Precrash system based on radar for automotive applications. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Columbus, USA, 2003.
- [SO01] M. Shimizu and M. Okutomi. Precise sub-pixel estimation on area-based matching. In *Proceedings of the 8th IEEE International Conference on Computer Vision*, volume 1, pages 90–97, Vancouver, Canada, 2001.
- [SPLA07] N. Soquet, M. Perrollaz, R. Labayrade, and D. Aubert. Free space estimation for autonomous navigation. In *Proceedings of the 5th International Conference on Computer Vision Systems*, 2007.
- [SRBB06] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi. Pedestrian detection using infrared images and histograms of oriented gradients. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, June 2006.
- [SS02a] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3) :7–42, 2002.
- [SS02b] R. Szeliski and D. Scharstein. Symmetric sub-pixel stereo matching. In *Proceedings of the 7th European Conference on Computer Vision*, volume 2, pages 525–540, 2002.
- [SSO04] H. Takahashi S. Sugimoto, H. Tateda and M. Okutomi. Obstacle detection using millimeter-wave radar and its visualization on image sequence. In *Proceedings of the International Conference on Pattern Recognition*, Cambridge, UK, 2004.

- [Ste01] B. Steux. *RTMaps, un environnement logiciel dédié à la conception d'applications embarquées temps-réel. Utilisation pour la détection automatique par fusion radar / vision*. PhD thesis, Ecole des mines de Paris, Decembre 2001.
- [TBM⁺06] G. Toulminet, M. Bertozzi, S. Mousset, A. Bensrhair, and 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, August 2006.
- [TFH97] K. Takada, H. Fujii, and O. Hayashi. Multiple vehicle identification in a longitudinal ranging system. In *Proceedings of the World Congress on Intelligent Transportation Systems*, Berlin, Germany, 1997.
- [TMRM02] A. Talukder, R. Manduchi, A. Rankin, and L. Matthies. Fast and reliable obstacle detection and segmentation for cross-country navigation. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Versailles, France, June 2002.
- [vAPBE07] A. von Arnim, M. Perrollaz, A. Bertrand, and J. Ehrlich. Vehicle identification using infrared vision and applications to cooperative perception. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, June 2007.
- [vdMG06] W. van der Mark and D. M. Gavrila. Real-time dense stereo for intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 7(1) :38–50, 2006.
- [Vek02] O. Veksler. Stereo correspondence with compact windows via minimum ratio cycle. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12) :1654–1660, December 2002.
- [Wil98] T. Williamson. *A High-Performance Stereo Vision System for Obstacle Detection*. PhD thesis, Carnegie Mellon University, 1998.
- [YK05] K. Yoon and I. Kweon. Locally adaptative support-weight approach for visual correspondence search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 924–931, 2005.
- [YKN06] K. Yamaguchi, T. Kato, and Y. Ninomiya. Moving obstacle detection using monocular vision. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, June 2006.
- [Zha98] Z. Zhang. Determining the epipolar geometry and its uncertainty : a review. *International Journal of Computer Vision*, March 1998.
- [ZW94] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proceedings of the European Conference on Computer Vision*, pages 151–158, 1994.
- [ZWH94] Z. Zhang, R. Weiss, and R.A. Hanson. Qualitative obstacle detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

DÉTECTION D'OBSTACLES MULTI-CAPTEURS SUPERVISÉE PAR STÉRÉOVISION

Résumé

Parmi les nouvelles technologies envisagées pour le développement d'aides à la conduite innovantes, la détection d'obstacles tient une place importante. Elle permet en effet d'anticiper d'éventuelles collisions, pour un gain réel en sécurité. Cette thèse propose d'aborder le thème de la détection d'obstacles par une approche multi-capteurs qui se veut robuste et générique, grâce au rôle central conféré à la stéréovision. Dans la méthodologie proposée, les différents capteurs (capteur stéréoscopique, télémètre laser, capteur d'identification optique) fournissent des hypothèses de détection sous la forme de volumes d'intérêt dans l'espace de disparité lié aux images stéréoscopiques. Un traitement localisé dans chacune de ces régions permet ensuite de valider et de caractériser ces hypothèses. Nous proposons dans cette thèse la description de cette méthodologie, trois méthodes de création d'hypothèses de détection et des critères pour la validation de celles-ci. Par ailleurs, des aspects pragmatiques liés à la mise en œuvre de cette approche sont abordés, comme les choix algorithmiques permettant l'obtention en temps réel de données exploitables pour la stéréovision et l'évaluation des méthodes proposées. Enfin, nous présentons trois applications fonctionnant dans des véhicules expérimentaux et anticipant sur de futures aides à la conduite.

Mots-clés: systèmes de transports intelligents, ADAS, détection d'obstacles, stéréovision, multi-capteurs, perception coopérative.

MULTI-SENSOR ROAD OBSTACLE DETECTION CONTROLLED BY STEREOVISION

Abstract

Road obstacle detection is a major topic for the development of future Advanced Driver Assistance Systems (ADAS). In this document, we propose a multi-sensor approach to obstacle detection. It is assumed to be robust and generic, thanks to a control task conferred to stereovision. In the proposed approach, various sensors (stereoscopic sensor, laser scanner, optical identification sensor) provide hypothesis of detections, which are represented as volumes of interest in the disparity space associated to the stereoscopic images. These volumes are then processed by a stereovision algorithm to perform validation and characterization of the hypothesis. We propose a description of this methodology, three methods for the creation of hypothesis, and criteria for their validation. We also explore some practical aspects related to the implementation and the evaluation of the proposed strategy. Particularly, we propose to determine the appropriate method to obtain usable data from the stereoscopic images. Finally, three applications of the proposed methodology, fully operational in experimental vehicles, are presented. They figure how obstacle detection and cooperation with inter-vehicle communication could be used for the development of future ADAS.

Keywords: intelligent transportation systems, ADAS, obstacle detection, stereovision, multi-sensor, cooperative perception.