



HAL
open science

Localisation robuste d'un véhicule en environnement urbain à partir d'un système de stéréo-vision

Nicolas Simond

► **To cite this version:**

Nicolas Simond. Localisation robuste d'un véhicule en environnement urbain à partir d'un système de stéréo-vision. Automatique / Robotique. Université Nice Sophia Antipolis, 2005. Français. NNT : . tel-00009694

HAL Id: tel-00009694

<https://theses.hal.science/tel-00009694>

Submitted on 6 Jul 2005

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.

UNIVERSITÉ DE NICE - SOPHIA ANTIPOLIS UFR Sciences

École Doctorale STIC

Thèse

pour obtenir le titre de

Docteur en Sciences

de l'Université de Nice - Sophia Antipolis

Spécialité : AUTOMATIQUE TRAITEMENT DU SIGNAL ET DES IMAGES

présentée et soutenue par

Nicolas SIMOND

Localisation robuste d'un véhicule en environnement urbain à l'aide d'un système de stéréo-vision

Thèse dirigée par Patrick RIVES

Équipe d'accueil : ICARE – INRIA Sophia-Antipolis

Soutenue publiquement le 1^{er} juillet 2005

devant le jury composé de :

M. Tarek	HAMEL	Professeur à l'Université de Nice/Sophia-Antipolis	Président
M. Didier	AUBERT	Chargé de Recherche LIVIC, Versailles	Rapporteur
M. Eric	MARCHAND	Chargé de Recherche IRISA, Rennes	Rapporteur
M. Michel	PARENT	Responsable du projet IMARA, INRIA Rocquencourt	Examineur
M. François	PEYRET	Directeur de Recherche LCPC, Nantes	Examineur
M. Patrick	RIVES	Directeur de Recherche INRIA Sophia-Antipolis	Directeur

Remerciements

Je remercie les membres du jury :

- Tarek Hamel, professeur de l'Université de Nice/Sophia-Antipolis, de le présider,
- Didier Aubert et Eric Marchand, respectivement chargé de recherche au LIVIC et à l'IRISA, de rapporter cette thèse,
- François Peyret, directeur de recherche au LCPC, et Michel Parent, responsable du projet IMARA de l'INRIA Rocquencourt, d'avoir accepté d'être examinateurs.

J'exprime mes plus vifs remerciements à Claude Samson, responsable du projet ICARE de l'INRIA Sophia-Antipolis, de m'avoir accueilli au sein de son équipe ainsi qu'à Patrick Rives, mon directeur de thèse, d'une part pour sa patience et d'autre part pour le fait d'être parvenu à boucler un dossier de thèse en moins d'un mois durant une période estivale alors qu'aucune bourse n'était originellement prévue.

Bien qu'une thèse soit un travail personnel, son aboutissement est étroitement lié à l'ambiance et aux contacts que l'on peut nouer dans l'équipe d'accueil. Je tiens donc à exprimer ma reconnaissance aux discussions quotidiennes enflammées d'Ezio et Selim, à la disponibilité et la pédagogie de Guillaume lorsque le monde informatique s'évertuait à me créer des misères, aux franches rigolades avec Matthieu, Alexandre, FX et l'impassibilité de Pascal qui m'inspire.

Je salue la bravoure d'Alessandro, Jaime, Nicolás, Christopher et Geraldo qui, du simple fait de partager un bureau, ont du subir pendant plusieurs mois mes frasques de sommeil et d'allergies en tous genres. Qu'ils en soient tous chaleureusement remerciés car cela a été un réel plaisir de passer tout ce temps en leur compagnie. Ce tour d'horizon ne saurait être complet sans citer Agnès, Patricia, Georges, Christophe, Baptiste, Pierre et Jean-Jacques avec qui j'aurai apprécié apprendre encore.

Travailler dans les conditions qu'offre l'INRIA Sophia-Antipolis est un réel luxe pour le cadre et les moyens techniques mis à notre disposition. Mais le prestige de cet établissement tient essentiellement aux rencontres que l'on peut y faire. Personnellement, je n'ai jamais autant voyagé qu'en discutant avec Manuel, Mauro, Doo-Jin, Hahnsang et ceux qui sont devenus depuis mes amis : Albi, Dami, Ben, Tony, Marianne, MC, Alain, Alex, Edi sans oublier les petits français Maga, Pierre, Fred, Caro et Céline.

Je n'oublie pas ceux qui ont participé de manière directe à la rédaction de cette thèse : merci à mes lecteurs occasionnels Sandor, Henri et assidue Virginie, qui ne cesse de m'épater.

Resumé

Localisation robuste d'un véhicule en environnement urbain à partir d'un système de stéréo-vision

Cette thèse s'inscrit dans le contexte de la robotique mobile et plus précisément dans le cadre de la localisation de véhicules en environnement urbain. Notre tâche a consisté à extraire le maximum d'informations sur le déplacement d'un système de stéréo-vision non-calibré embarqué à bord d'un véhicule en s'appuyant sur l'environnement structuré que forme les rues. L'objectif de cette étude est d'estimer le déplacement d'un véhicule lorsque la configuration de l'environnement s'oppose à sa localisation par un système de géo-référencement satellitaire.

Une modélisation polyédrique des rues qui forment l'environnement urbain de cette étude permet de segmenter trois plans principaux : la route et les façades verticales des bâtiments qui délimitent l'espace de navigation. La segmentation du plan de la route est facilitée par la matérialisation des voies de navigation et autres trottoirs ou terre-pleins, parallèles entre eux au premier plan, qui sont à l'origine dans les images de contours qui convergent vers un point de fuite dominant.

Les conditions d'illumination et la présence d'obstacles dynamiques qui obstruent le champ de vision compliquent la tâche de suivi des plans segmentés. L'emploi d'une paire stéréo permet à chaque nouvelle pose de déterminer quelles sont les régions des images qui correspondent aux plans segmentés en calculant l'homographie induite par chaque plan entre les deux images de la paire. Le déplacement du véhicule étant supposé plan, ses mouvements d'inclinaison et de lacet influent peu sur chacune des homographies estimées entre les images de la paire stéréo. Nous possédons ainsi un outil fiable de segmentation des plans principaux qui forment l'environnement.

L'estimation du déplacement entre deux poses de chaque caméra consiste à estimer l'homographie induite par le mouvement relatif du plan considéré. Le calcul est entrepris grâce à la mise en correspondance des primitives (points, droites) détectées dans les régions planaires segmentées. La précision et la fiabilité des résultats dépendent essentiellement de la répartition spatiale des primitives mises en correspondance.

L'introduction de la notion de super-homographie permet de robustifier les estimations des homographies en calculant de manière simultanée toutes les homographies induites par un même plan, observé dans plusieurs images. La redondance des contraintes imposées par le mouvement relatif des primitives coplanaires mises en correspondance améliore l'estimation du mouvement de la projection du plan dans les différentes images. L'apport de la super-homographie est double : elle permet de détecter d'éventuelles erreurs de mise en correspondance et fournit une estimée des coordonnées des primitives coplanaires avec une précision sub-pixelique.

Mots-clés : stéréo-vision, robotique mobile, environnement urbain, homographie, super-homographie, trajectographie

Abstract

Robust localization of a vehicle in urban environments with a stereo-vision system

The thesis deals with mobile robots in the specific context of urban environment. We want to extract the maximum of data on the motion of an uncalibrated stereo-rig embedded in a vehicle with in a structured environment. The aim of this study is to estimate the motion of a vehicle when the localization is impossible with a satellite position system due to the occlusion of the free space.

A polyhedral model of the streets, which is the considered environment in this study, allows the segmentation of three main planes : the road and the vertical façades which represent the frontier of the navigation area. We assume that the white lanes which split the different ways and the kerbs which delimit the road are all parallel in the scene, at least on the foreground. Therefore the detection of these induced edges in image is easy because they all converge in a dominant vanishing point.

The lighting conditions and the dynamical obstacles which occlude the vision field make the tracking task of the segmented planes difficult. The use of a stereo-vision system allows the segmentation in image of the coplanar area of the scene assuming that each plane induces a homography between both images of the stereo-rig. The vehicle motion is assumed to be plane, the pan and tilt rotation motions have low contributions on the computed homographies. We hence possess a reliable tool to segment the main planes which structure the environment.

The motion of the cameras between two poses are obtained with the homography induced by the relative motion of one plane. The homography computation requires the matching of features (lines, points) detected in the coplanar regions of images. The precision and the reliability of the result highly depends on the spatial distribution of the matched features.

The introduction of the notion of super-homography improves the robustness of the homography computation by computing in one shot all the homographies induced by the same plane in several images. The redundancy of the constraints imposed by the relative motion of the matched coplanar features increases the constraints of the plane projection in each image. The super-homography has two advantages : it allows the detection of mismatched features and supplies coordinates with a sub-pixellic precision. When the calibration of the stereo-rig is available, the extraction of the homography induced by the relative motion of a camera from the super-homography supplies the trajectory of the vehicle.

Key-words : stereo-vision, mobile robotics, urban environment, homography, super-homography, trajectory

Notations et acronymes

Géométrie projective

\mathbb{R}^n	:	espace euclidien de dimension n
\mathcal{P}	:	espace projectif de dimension n
\mathbf{P}	:	point de l'espace euclidien
\mathbf{L}	:	droite de l'espace euclidien
\mathbf{p}	:	point du plan projectif
\mathbf{l}	:	droite du plan projectif
$\bar{\mathbf{p}}$:	coordonnées non-homogènes du point du plan projectif
Π	:	plan de la scène
π	:	projection du plan de la scène dans l'image
\mathbf{X}	:	point du plan à l'infini Π_∞
\mathbf{x}	:	point de fuite
	:	
\mathbf{I}	:	matrice image $[l, h]$ de 256 niveaux de gris
u	:	abscisse de l'image
v	:	ordonnée de l'image
f	:	focale de la caméra
u_0	:	abscisse du point central
v_0	:	ordonnée du point central
	:	
\mathbf{C}	:	coordonnées du repère caméra dans le repère global
\mathbf{K}	:	matrice de paramètres intrinsèques d'une caméra
\mathbf{E}	:	matrice essentielle
\mathbf{F}	:	matrice fondamentale
\mathbf{H}	:	matrice d'homographie
\mathcal{H}	:	matrice de super-homographie
	:	
\mathbf{l}_3	:	matrice identité de dimension 3
\mathbf{R}	:	matrice de rotation
\mathbf{T}	:	vecteur de translation
\mathbf{N}	:	normale au plan π
d	:	distance entre le plan π et le repère caméra
	:	
$\mathbf{p}_1 \wedge \mathbf{p}_2$:	produit vectoriel entre les vecteurs $bf p_1$ et \mathbf{p}_2
$\mathbf{p}_1 \simeq \mathbf{p}_2$:	égalité projective entre les vecteurs $bf p_1$ et \mathbf{p}_2

Algorithmie

Γ	: vecteur d'état
$\tilde{\Gamma}$: prédiction du vecteur d'état
n	: numéro de l'image depuis le début de la séquence
r	: numéro de l'itération dans le cas de processus itératif
$dist_max$: distance de référence (10 <i>pixels</i>)
$seuil_theta$: écart angulaire de référence (10°)
Q	: fenêtre carrée centrée sur les coordonnées d'un point d'intérêt
q	: demi largeur de la fenêtre Q (10 <i>pixels</i>)
$seuil_secu$: distance minimale autour du point de fuite dominant (50 <i>pixels</i>)
$seuil_corr$: seuil du coefficient de corrélation (0,5)

Acronymes

FP	: point d'intérêt (<i>Feature Point</i>)
VL	: ligne de fuite (<i>Vanishing Line</i>)
DVP	: point de fuite dominant (<i>Dominant Vanishing Point</i>)
FOE	: foyer d'expansion (<i>Focus Of Expansion</i>)
SFP	: super-point d'intérêt (<i>Super Feature Point</i>)
SVL	: super-ligne de fuite (<i>Super Vanishing Line</i>)
VFP	: point d'intérêt virtuel (<i>Virtual Feature Point</i>)

Table des matières

1	Introduction générale	3
1.1	Le cadre	3
1.1.1	Les Systèmes de Transports Intelligents	3
1.1.2	Le géo-référencement satellitaire	4
1.1.3	La localisation d'un robot mobile en milieu urbain	4
1.1.4	Vers un couplage géo-référencement-vision?	5
1.2	Notre approche	6
1.2.1	Présentation des objectifs	6
1.2.2	Modélisation	6
1.2.3	Approche proposée	6
2	Etat de l'art : la vision appliquée aux robots mobiles	9
2.1	Le contexte	9
2.1.1	Les tâches de localisation et de navigation	9
2.1.1.1	Localisation à l'estime ou par intégration de la vitesse	9
2.1.1.2	Localisation par mesure directe de la pose	10
2.1.2	L'environnement urbain	11
2.2	La localisation en robotique mobile d'intérieur	12
2.2.1	L'emploi de primitives	13
2.2.2	L'emploi d'amers	14
2.2.2.1	Scènes non-instrumentées	14
2.2.2.2	Scènes instrumentalisées	15
2.3	La localisation en robotique mobile d'extérieur	16
2.3.1	Evolution au cours du temps	16
2.3.2	La localisation de robots mobiles sur route	17
2.3.2.1	Etat de l'art en 1990	18
2.3.2.2	La détection de la route	18
2.3.3	Applications	20
2.3.3.1	Contrôle latéral du véhicule	21
2.3.3.2	Estimation du déplacement du véhicule	21
2.3.3.3	Détection, caractérisation et évitement d'obstacles	22
2.4	Le cadre applicatif de la thèse	25
2.4.1	La structuration de l'environnement urbain	26
2.4.2	La localisation et la navigation en environnement urbain	27
2.5	Conclusion	28

3	Outils mathématiques de la vision	31
3.1	Géométrie de la vision	31
3.1.1	L'espace projectif \mathcal{P}^n	31
3.1.2	Les transformations dans \mathcal{P}^n	32
3.1.2.1	Les transformations projectives	32
3.1.2.2	Les transformations affines ou affinités	32
3.1.2.3	Les similitudes	33
3.1.2.4	Les isométries	33
3.1.3	Particularités des espaces projectifs	33
3.1.3.1	Le plan à l'infini	34
3.1.3.2	La conique absolue	34
3.1.3.3	Le point de fuite et les lignes de fuite	35
3.1.3.4	Le bi-rapport	36
3.2	Les modèles géométriques de caméras	37
3.2.1	Le modèle sténopé	37
3.2.2	Paramètres intrinsèques et extrinsèques	38
3.2.2.1	Le changement de repère	38
3.2.2.2	La projection perspective	39
3.2.2.3	La tranformation caméra-image	40
3.2.2.4	La transformation perspective	40
3.2.3	Géométrie d'une paire stéréo	41
3.2.4	La géométrie épipolaire	42
3.2.4.1	La matrice fondamentale F	42
3.2.4.2	La matrice essentielle E	44
3.3	L'homographie planaire	45
3.3.1	Description	45
3.3.1.1	Projection d'un plan sur une image	45
3.3.1.2	Transformations entre images de scène planaire	45
3.3.1.3	Cas particulier des caméras calibrées	46
3.3.1.4	Relations entre géométrie épipolaire et homographie planaire	47
3.3.1.5	Homographie du plan à l'infini Π_∞	48
3.3.1.6	Relations entre deux homographies planaires	49
3.3.2	L'estimation de l'homographie	50
3.3.2.1	Algorithme de transformation linéaire direct (DLT)	50
3.3.2.2	Modèle de bruit	51
3.3.2.3	Configurations dégénérées	52
3.3.2.4	Normalisation du système	53
3.3.2.5	Estimation itérative de la matrice d'homographie	55
3.4	Les déplacements rigides	56
3.4.1	Base projective d'une paire stéréo à partir d'une homographie planaire	56
3.4.2	Dualité entre les déplacements de caméras et d'objets	56
3.4.3	Analyse algébrique du mouvement rigide	57
3.5	Conclusion	57

4	Caractérisation des plans principaux par stéréovision	59
4.1	Segmentation des plans principaux	59
4.1.1	Segmentation de la région apparentée à la route	60
4.1.1.1	Extraction de contours	60
4.1.1.2	Le point de fuite dominant	65
4.1.1.3	Extraction de droites-supports	66
4.1.1.4	Les lignes de fuite	72
4.1.1.5	Mise en correspondance de lignes de fuite	77
4.1.1.6	Détection des bords de la route : la région d'intérêt	82
4.2	Extraction de points d'intérêt appartenant à la route	83
4.2.1	Extraction de points d'intérêt	83
4.2.2	Le détecteur de Harris-Stephens	84
4.2.2.1	Rappel	84
4.2.2.2	Implémentation	86
4.2.3	Exploitations des contraintes de rigidité et de coplanarité	89
4.2.4	Mise en correspondance de points d'intérêt	91
4.2.4.1	Les points d'intérêt détectés sur la route	91
4.2.4.2	Mise en correspondance de points d'intérêt par la méthode de Pilu	93
4.2.4.3	Adaptation de l'algorithme de Pilu à notre cas	93
4.2.4.4	Récapitulatif de la méthode employée	99
4.3	Estimation de l'homographie	101
4.3.1	Difficulté de mise en oeuvre	101
4.3.2	Résolution du système linéaire	101
4.3.3	Initialisation de l'algorithme en début de séquence	102
4.3.4	Décomposition de l'homographie	103
4.4	Résultats : évolution de la géométrie du banc stéréo au cours du temps	103
4.4.1	La séquence du port d'Antibes	103
4.4.2	Décomposition de l'homographie stéréo	104
4.5	Conclusion	106
5	Suivi des plans principaux par vision dynamique	109
5.1	Segmentation du plan de la route par filtrage temporel	110
5.1.1	Filtrage de Kalman	110
5.1.1.1	Rappel	110
5.1.1.2	Implémentation	112
5.1.2	Estimation du point de fuite dominant	113
5.1.2.1	Particularités de la primitive point de fuite dominant	114
5.1.2.2	Filtrage temporel des coordonnées du point de fuite dominant	114
5.1.3	Identification et suivi des bandes principales	119
5.1.3.1	Identification des bandes principales	119
5.1.3.2	Mise en correspondance des mesures et des prédictions des bandes principales	122
5.1.3.3	Récapitulatif de l'étape de suivi des bandes principales	123
5.1.3.4	Résultats du suivi des bandes principales sur la séquence du port d'Antibes	124
5.2	Suivi des points d'intérêt coplanaires	129

5.2.1	Mouvement relatif des points d'intérêt coplanaires entre deux poses de la caméra	131
5.2.1.1	Observations	131
5.2.1.2	Linéarisation des mouvements de rotation	133
5.2.2	Recherche d'une prédiction de l'homographie	134
5.2.2.1	Connaissance a priori du mouvement apparent des points d'intérêt	134
5.2.2.2	Sélection des FPs potentiellement en correspondance	135
5.2.2.3	Hypothèse d'un déplacement du véhicule à vitesse constante	138
5.2.2.4	Hypothèse de mouvement relatif des points d'intérêt à vitesse constante	140
5.2.3	Suivi de points d'intérêt mal caractérisés	142
5.2.4	Sélection des points d'intérêt coplanaires	145
5.2.5	Résultats	147
5.3	Discussion et conclusion	149
6	Trajectographie du véhicule	151
6.1	Position du problème	151
6.2	Décomposition des homographies	153
6.2.1	Résolution	153
6.2.1.1	Rappel	153
6.2.1.2	Décomposition du mouvement selon l'ordre de multiplicité	154
6.2.1.3	La prise en compte de la contrainte d'observabilité	155
6.2.2	Résultats	156
6.3	Définition et propriété de la super-homographie	159
6.3.1	Propriétés de la matrice de super-homographie	159
6.3.1.1	Rappel	159
6.3.1.2	Super-primitives	160
6.3.1.3	Décomposition de la matrice de super-homographie	161
6.3.1.4	Estimation itérative de la super-homographie	162
6.3.2	Implémentation de la méthode	163
6.3.2.1	Adaptation à notre configuration	164
6.3.2.2	Prise en compte des lignes de fuite	164
6.3.2.3	Initialisation de la super-homographie	169
6.3.2.4	Estimation de la super-homographie	172
6.3.2.5	Récapitulatif	178
6.3.3	Reconstruction de la route au long d'une séquence	178
6.4	Conclusion	182
7	Conclusion et perspectives	185
A	Détail des séquences utilisées	189
A.1	Séquence du port d'Antibes	189
A.2	Séquence descente enregistrée à Versailles	190
B	Algorithme de RANSAC	191
B.1	Principe	191
B.2	Implémentation	192

Table des figures

1.1	Exemples de scènes urbaines.	7
1.2	Homographies induites par un plan observé par une paire stéréo en déplacement.	7
2.1	Modélisation des frontières considérées parallèles d'une route.	20
3.1	Toutes les droites ayant la même orientation D convergent vers le même point X_D du plan à l'infini Π_∞	35
3.2	La ligne de fuite correspond à la projection sur le plan image de la droite d'interception de plusieurs plans parallèles avec le plan à l'infini Π_∞	36
3.3	Bi-rapport d'un faisceau de quatre droites	37
3.4	Projection m du point M sur le plan image selon un modèle de caméra sténopé.	38
3.5	Géométrie épipolaire de deux caméras.	43
3.6	Composition d'homographies.	46
3.7	Projection d'un point dans deux images.	48
3.8	Projections de points appartenant à des plans différents.	49
3.9	Les différents types d'erreurs de projection.	52
3.10	Equivalences géométriques entre FPs et faisceau de VLs.	53
3.11	Le déplacement rigide d'un objet devant une caméra fixe est équivalent au déplacement d'une caméra mobile dans un environnement statique.	57
4.1	Comparaison des résultats de différents extracteurs de contours sur différentes images de scènes urbaines.	62
4.2	Détections de contours par filtrage de Canny-Deriche.	64
4.3	Exemples de rares scènes où des contours orientés selon la direction perpendiculaire à la direction de la route peuvent être détectés.	65
4.4	Contours détectés par le filtrage de Canny sur une paire d'images stéréo.	67
4.5	Regroupement des segments représentant un même contour.	69
4.6	Détail des différentes opérations nécessaires pour extraire le faisceau de droites-supports à partir des contours détectés par le filtre de Canny.	73
4.7	Représentation à l'aide des coordonnées théoriques des droites-supports qui convergent vers le DVP.	74
4.8	Résultat de la détection de contours autour d'une longue bande rectiligne.	76
4.9	Bandes principales extraites pour calculer le bi-rapport.	79
4.10	Projection des bandes extraites des faisceaux de droites grâce à une prédiction et une estimation de l'homographie stéréo.	81
4.11	Superposition des bandes extraites des deux faisceaux de droites et des prédictions des bandes principales.	82

4.12	Création des ROIs suite à l'identification des bandes principales.	83
4.13	Résultats de l'extraction de points d'intérêt avec un filtre de Harris et le score de Benhimane.	87
4.14	La résolution des pixels dépend de la distance qui sépare la scène observée du foyer de la caméra.	88
4.15	Caractérisation des points d'intérêt détectés au premier plan.	88
4.16	Evolution de l'homographie stéro entre deux poses successives.	90
4.17	Evolution de l'angle d'inclinaison α d'une caméra entre deux positions successives.	90
4.18	Représentation des déplacements des FPs coplanaires mis en correspondance entre deux images de la paire stéréo.	92
4.19	Représentation des déplacements entre FPs en correspondance lorsque les axes optiques des caméras du banc stéréo sont alignés.	92
4.20	Représentation des FPs détectés dans les ROIs.	94
4.21	Détail de la mise en correspondance des FPs en utilisant par l'algorithme de Pilu.	96
4.22	Mise en oeuvre de la pondération sur le critère de distance appliqué aux matrices de proximité lors de la présélection.	97
4.23	Comparaison entre les résultats fournis par les deux méthodes de mises en correspondance.	98
4.24	Recherche de couples de FPs en correspondance lorsque la prédiction de l'homographie stéréo est perfectible.	100
4.25	Décomposition de l'homographie stéréo au cours d'une séquence.	104
4.26	Cas dégénérés où la répartition des FPs en correspondance n'est pas suffisamment uniforme pour contraindre correctement l'homographie stéréo induite par le plan de la route.	105
5.1	Chronogrammes des coordonnées prédites, mesurées et estimées du DVP fournies par le filtre prédictif de Kalman.	115
5.2	Droites-supports extraites lors des brusques mouvements d'inclinaison du véhicule.	116
5.3	Ecart entre les coordonnées mesurées et prédites des DVPs, fournies par le filtre de Kalman.	117
5.4	Cas où l'ordonnée de la prédiction du DVP est éloignée de la valeur réelle.	118
5.5	Evolution du faisceau de bandes principales en fonction des mouvement de rotation de la caméra.	120
5.6	Code de couleurs employé pour identifier les caractéristiques de trois bandes principales.	121
5.7	Evolution des caractéristiques des bandes principales au cours de la séquence du port d'Antibes.	125
5.8	Evolution de la largeur des bandes mesurée en fonction de l'angle de lacet de la caméra par rapport à celui de la route.	126
5.9	Relevé des caractéristiques des bandes principales pour la séquence Descente.	127
5.10	Dénombrement des cas où l'identification des projections des trois bandes principales a posé problèmes.	128
5.11	Détail de détection de contours qui posent problèmes.	129
5.12	Identification des bandes principales dès leur réapparition dans le couple d'images.	130
5.13	Résultats de la mise en correspondance de FPs suite à un mouvement de translation pure et un mouvement de rotation dominant.	132

5.14	Champ de composition des vitesses de FPs coplanaires entre deux poses de la caméra.	133
5.15	Chronogramme de l'estimation du mouvement vertical maximal entre deux images consécutives.	135
5.16	Présélection des FPs de l'image courante grâce à la seule connaissance a priori du mouvement entre deux images successives.	136
5.17	Cas particulier où le fort mouvement d'inclinaison du véhicule entre deux poses de la caméra induit un mouvement relatif vers le haut de certains FPs coplanaires dans les deux images du couple stéréo.	137
5.18	La présélection de FPs en correspondance est complexifiée lorsque certains FPs sont détectés le long d'un motif répétitif.	139
5.19	Confrontation du choix de la prédiction d'une homographie lorsque la distribution des primitives n'est pas uniforme dans la ROI. A gauche, la prédiction de l'homographie correspond à l'estimée obtenue à l'itération précédente.	143
5.20	La détection de points d'intérêt selon l'angle de vue et la proximité de la caméra.	144
5.21	Sectorisation et découpage de la ROI.	145
5.22	Détection de FPs coplanaires en présence d'obstacles.	146
5.23	Suivi dans les images droites des FPs qui délimitent les extrémités de la flèche peinte située devant le véhicule en stationnement.	148
6.1	Détail des homographies induites entre 2 couples de paires d'images.	152
6.2	Estimation des mouvements de translation et rotation de chacune des caméras entre chaque pose.	157
6.3	Les FPs détectés aux coins saillants des parallélogrammes formés par le marquage horizontal sont différents suivant la position relative de la caméra par rapport à la bande principale centrale.	158
6.4	Comparaison de la qualité de l'homographie stéréo H_{st}^n extraite de \mathcal{H}^n et celle directement estimée à partir des FPs mis en correspondance.	166
6.5	Cas où la distribution spatiale des SFPs n'est pas suffisamment uniforme pour contraindre l'intégralité de la projection du plan de la route.	168
6.6	Synthèse des couples de primitives mis en correspondance grâce aux estimations des homographies courantes.	170
6.7	Caractérisation des super-primitives qui ont permis de calcul de la super-homographie.	173
6.8	Représentation de la composition du bloc triangulaire inférieur des matrices \mathcal{H}^{n-1} et \mathcal{H}^n à partir des homographies qui les constituent.	174
6.9	Evolution du nombre de points d'intérêt au cours de la séquence "descente".	175
6.10	Choix de la première image pour reconstituer le plan de la route au long de la séquence.	179
6.11	Reconstruction du plan de la route sur la séquence du port d'Antibes.	180
6.12	Décomposition des mouvements de translation et rotation de la caméra droite au cours de la séquence "descente".	181
6.13	Reconstitution du plan de la route à partir de deux instants différents.	183
A.1	Echantillons de la séquence du port d'Antibes.	189
A.2	Echantillons de la séquence descente enregistrée à Versailles.	190

Liste des tableaux

4.1	Disparité relative des seuils utilisés pour chacune des images-tests en fonction du filtre employé.	63
4.2	Caractéristiques des droites supports puis des bandes extraites de l'image D0772.	74
4.3	Caractéristiques des segments détectés autour de la bande principale.	76
4.4	Bi-rapport des bandes principales.	79
4.5	Caractéristiques des FPs mis en correspondance grâce à l'algorithme de Pilu.	96
5.1	Orientations et abscisses des bandes coplanaires en correspondance dans le couple d'images stéréo.	121
6.1	Exemples de liaisons établies suite aux estimations des homographies courantes à partir des 2 couples d'images précédents.	170

Chapitre 1

Introduction générale

1.1 Le cadre

La question des transports dans les états membres de l'Union Européenne est désormais critique. L'impact économique de la congestion des voies de transports routiers a été évaluée par la Commission Européenne à une perte de 40 milliards d'euros/an. D'autre part, il a été démontré ces dernières années que le rejet de gaz nocifs par les véhicules en circulation étaient à l'origine d'une pollution sonore mais surtout atmosphérique de l'environnement, elle-même responsable de la dégradation de la santé des citoyens. Enfin, le transport routier est la principale cause de mortalité dans les différents états de l'Union : 40 000 décès/an. La Commission Européenne s'est donc engagée à améliorer les modes de transports dans *Le livre blanc sur la politique européenne des transports à l'horizon 2010 : l'heure des choix* qui préconise le développement massif des Systèmes de Transports Intelligents.

1.1.1 Les Systèmes de Transports Intelligents

”Les systèmes de transports intelligents (*ITS*) recourent aux technologies de l'information et de la communication pour faciliter le transport sans rupture de charges des voyageurs et des marchandises”¹. Fort du constat que le seul élément de régulation des transports routiers était le feu de signalisation (inventé en 1923), la Commission Européenne a donc décidé de favoriser le développement de réponses technologiques aux questions posées actuellement par les transports routiers que sont la congestion, la pollution, la santé et la sécurité.

La Commission Européenne finance ainsi le développement de véhicules automatiques dans le cadre spécifique des grandes agglomérations où le nombre de trajets quotidiens ne cessent d'augmenter ce qui réduit d'autant plus la fluidité des trajets. Le projet Cybermove² a pour objectif de promouvoir l'intérêt de tels véhicules auprès des instances politiques et décisionnelles ainsi que de créer des normes et textes de lois qui réglementent l'utilisation de ces nouveaux véhicules. Le projet Cybercars³ est, quant à lui, un consortium technologique qui a pour but de développer de nouvelles solutions issues de la recherche dans le cadre de la robotique mobile en milieu urbain et financer des travaux dans de nouveaux axes de recherches. Ce travail de thèse s'inscrit dans le contexte du projet Cybercars.

¹http://europa.eu.int/comm/transport/themes/network/english/its/pdf/its_brochure_2003_fr.pdf

²<http://www.cybermove.org>

³<http://www.cybercars.org>

A terme, la Commission Européenne souhaite que les citoyens utilisent d'autres types de transports alternatifs à leurs véhicules particuliers représentant 80% des déplacements en milieu urbain. L'idée est que les grandes agglomérations s'équipent de flottes de véhicules automatiques en libre service qu'elles mettront à disposition des usagers qui passeront commande de véhicules pour des trajets en fonction de leurs besoins. Ces véhicules exploiteront les nouvelles possibilités fournies par les STIC et pourront être dotés de plus ou moins grande autonomie. Outre la résolution des problèmes de stationnement, ce système aura l'intérêt de fluidifier les temps de parcours en limitant le nombre de véhicules et limiter les problèmes d'embouteillages par une gestion prédictive de flots de véhicules, localisés par exemple GSM et/ou Galileo.

1.1.2 Le géo-référencement satellitaire

Les applications de ce type de système de positionnement sont désormais connues du grand public : du récepteur de poche pour randonneur, aux systèmes d'aide à la navigation des véhicules en passant par les récepteurs employés dans les domaines spécifiques que sont la navigation maritime et aérienne. Les sociétés qui gèrent des flottes de véhicules (voiries, transports urbains, sociétés de transports) ont aussi trouvé en ce système un moyen d'améliorer leur productivité en connaissant en temps réel la position de chacun de leurs véhicules.

Le projet de Système de Navigation Globale par Satellites (GNSS) Galileo⁴ a été lancé en 1994 par la Commission Européenne pour faire face aux "problèmes de mobilité et de transport". La constellation de 30 satellites (27 + 3 de secours) sera mise en orbite en 2006 et devra être opérationnelle en 2008.

Répondant à une volonté politique de la Commission Européenne de ne plus dépendre de systèmes militaires que sont le GPS américain et son clone russe Glonass, Galileo "constitue un véritable service public", néanmoins compatible avec son concurrent américain car "l'utilisation harmonieuse des deux infrastructures apporte un réel avantage en termes de précision et en termes de sécurité en cas de défaillances de l'un des deux systèmes".

Les avantages de Galileo sont de trois ordres :

- ses performances technologiques : une couverture mondiale en zones urbaines et même aux latitudes extrêmes, un "message d'intégrité" informant l'utilisateur d'erreurs,
- son financement civil garantit juridiquement son fonctionnement "en matière de responsabilité contractuelle",
- son faible coût financier : "Galileo n'est pas cher (3,2 milliards d'euros) soit l'équivalent de 150 km d'autoroutes semi-urbaines".

Ce système de positionnement a été conçu dans la perspective de faciliter les déplacements en milieux urbains : "comparé au système américain existant GPS, Galileo sera caractérisé par une précision plus élevée, une plus grande fiabilité et une couverture plus homogène (95% de l'ensemble des zones urbaines couvertes sans interruption, contre 50% aujourd'hui)". A ce propos, [Chen, 1999] confirmait que le géo-référencement satellitaire était impossible sur 95% de la surface de la métropole de Tokyo. Nous sommes alors en droit de nous poser la question : quelle sera la couverture de Galileo dans un tel cas ?

1.1.3 La localisation d'un robot mobile en milieu urbain

Le développement des systèmes de géo-référencement a permis lors de cette dernière décennie de transposer la plupart des techniques de localisation des robots mobiles des environnements

⁴http://europa.eu.int/comm/dgs/energy_transport/galileo

intérieurs vers l'extérieur. La mise à disposition d'un capteur de position à très grande échelle relègue alors la tâche de localisation à une tâche secondaire, moins critique que celle de navigation.

Cependant, dans le cadre spécifique des milieux urbains, la précision ainsi que la continuité des informations fournies par les systèmes de géo-référencement actuels ont été un réel frein au développement de la robotique mobile. En effet, la localisation d'un récepteur exige la réception d'émissions d'au moins quatre satellites et la fiabilité de la mesure est proportionnelle à la durée pendant laquelle les satellites sont détectés en visée directe. Le nombre de satellites varie généralement de 8 en pleine mer à aucun dans des configurations de type vallées étroites. Les satellites étant sur des orbites basses, leurs positions évoluent rapidement au cours du temps et un horizon peu dégagé est à l'origine de pertes de réception fréquentes.

Dès lors, la tâche de localisation de robots mobiles évoluant en extérieur est dans la majorité des cas réalisée par le couplage d'un système de géo-référencement avec des capteurs de localisation traditionnels (odomètres, centrales inertielles, etc.). Ces derniers très précis sur de "courtes distances" subissent généralement l'effet de dérive au cours du temps. La complémentarité de tels capteurs avec un système de géo-référencement est alors évidente. La mise en oeuvre d'un système de localisation d'un robot mobile évoluant en extérieur consiste alors à identifier la configuration optimale de capteurs nécessaires, capables de fournir une estimée de la pose du robot à tout instant. Les solutions proposées varient suivant l'application, la précision souhaitée et la structuration de l'environnement du robot mobile.

Dans le cas de la localisation d'un robot mobile en milieu urbain, l'étroitesse des rues s'oppose à la mise en oeuvre d'un système de localisation principalement fondé sur le système GPS. Or, parallèlement au développement des systèmes de géo-référencement, les progrès dans le domaine de la vision par ordinateur ont été tels ces vingt dernières années que la robotique mobile emploie désormais largement ce type de capteur qui permet d'appréhender en une mesure la complexité d'une scène. Il devient alors envisageable d'essayer d'utiliser des informations visuelles pour réaliser une tâche de localisation en temps réel.

1.1.4 Vers un couplage géo-référencement-vision ?

L'intégration de système de géo-référencement sur les véhicules a apporté des réponses satisfaisantes dans des configurations de visibilité appropriées (autoroute, inter-urbain) mais sans se révéler fiable dans les situations qui nous intéressent.

Notre idée est relativement simple : nous souhaitons estimer le déplacement d'un véhicule à partir d'un système de vision lorsque le système de géo-référencement n'est plus capable de fournir une localisation acceptable. Vu autrement, ce travail consiste à localiser un véhicule à partir de capteurs de vision embarqués en supposant qu'un recalage soit possible par géo-référencement ou par fusion de données avec un Système d'Information Géographique (SIG) intégrant une carte cadastrale précise et un Modèle Numérique de Terrain (MNT).

La navigation d'un robot mobile exige la satisfaction simultanée de tâches qui peuvent toutes être réalisées à partir d'un système de vision :

- l'estimation du déplacement soit par flot optique soit par estimation d'homographie(s), ces méthodes sont généralement regroupées sous l'appellation d'odométrie optique,
- la détection d'obstacles par segmentation de régions de l'image qui vérifient les mêmes propriétés de colorimétrie, texture ou qui subissent le même déplacement,
- la localisation par la recherche d'amers naturels ou d'autres balises qui structurent l'environnement du véhicule,

- la reconstruction de l’environnement, si besoin est, dépend du niveau de complexité du modèle dont on dispose,
- la navigation ou planification de trajectoires en fusionnant les informations issues de l’ensemble des tâches précédentes.

L’objectif de cette thèse est de déterminer la portée et les limitations de l’introduction de capteurs de vision pour localiser un véhicule en milieu urbain et de définir dans quelles proportions ce seul capteur permet de compléter/remplacer le positionnement par GPS.

1.2 Notre approche

1.2.1 Présentation des objectifs

Nous souhaitons réaliser la construction d’un modèle 2D augmenté de l’environnement d’un véhicule en déplacement en milieu urbain à partir d’un système de vision embarqué. La segmentation des principaux éléments structurants de la scène permet conjointement d’estimer la trajectographie du véhicule en mesurant le déplacement relatif par rapport à chacun des éléments de la scène segmentés.

1.2.2 Modélisation

L’environnement urbain tel que nous le concevons dans cette thèse correspond aux parties les plus urbanisées, là où la probabilité d’obtenir une réception continue d’un système de positionnement satellitaire est la plus réduite. Notre priorité est donc ce que nous pourrions définir comme des canyons urbains, structurés par les constructions qui délimitent l’espace de navigation du véhicule qu’est la route.

L’environnement urbain a la réputation d’être parmi les environnements les plus complexes qu’il soit du fait de son fort caractère dynamique. Les différents ”obstacles” (piétons, véhicules) qui animent la scène suivent des trajectoires complexes à des vitesses dont la disparité est importante. L’emploi d’un système de stéréo-vision embarqué nous permet de discriminer à tout instant la cohérence des éléments segmentés en autorisant une mesure de distance, ne serait-ce qu’en fonction d’un facteur d’échelle.

Nous présentons dans la Fig. 1.1 des exemples de séquences d’images sur lesquelles nous avons travaillé. Toutes correspondent à des situations types rencontrées en environnement urbain.

1.2.3 Approche proposée

L’approche que nous proposons consiste à extraire un maximum d’informations à partir de la complémentarité de la stéréo-vision et de la vision dynamique. La stéréo-vision est particulièrement bien adaptée à la segmentation de plans statiques de la scène. La vision dynamique permet, elle, d’estimer le mouvement du véhicule dans son environnement et d’appréhender le caractère dynamique de la scène (voir Fig. 1.2). La conjonction des deux approches nous permet de simplifier le problème de modélisation de l’environnement et d’estimation simultanée de la trajectographie du véhicule en imposant des contraintes et de la redondance dans les observations.

Le chapitre 2 propose un état de l’art sur les méthodes de localisation de robots mobiles basées sur la vision. Les méthodes présentées sont distinguées selon l’environnement dans lequel évolue les robots : en intérieur, en extérieur puis dans le cadre spécifique de l’environnement urbain.



FIG. 1.1 – Exemples de scènes urbaines.

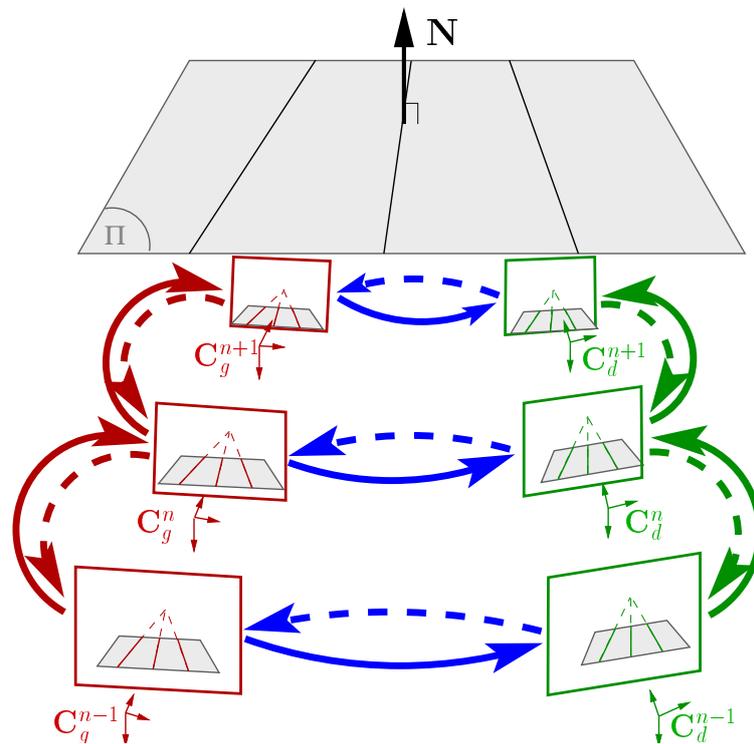


FIG. 1.2 – Homographies induites par un plan observé par une paire stéréo en déplacement. A chaque pose de la paire stéréo, trois homographies seront estimées : une homographie entre les images du couple stéréo pour segmenter les plans principaux (en bleu) et celles relatives au déplacement de la caméra gauche (en rouge) et droite (en vert).

Le chapitre 3 est un rappel des principaux outils utilisés en géométrie projective. Nous présentons les espaces projectifs en nous focalisant sur celui de dimension 2 en détaillant les transformations ainsi que les particularités liées à cet espace. Nous présentons succinctement ensuite le modèle de caméra sténopé pour introduire la géométrie épipolaire, la vision stéréographique et l'homographie planaire. Nous terminons cette partie en analysant les déplacements rigides observés par une caméra.

Le chapitre 4 correspond à la segmentation des plans principaux de la scène par stéréovision. Les différents plans sont représentés par des primitives (droites, points) dont les coordonnées vérifient une homographie commune. La calibration de la paire stéréo pourra être entreprise dès que la répartition spatiale des primitives détectées sera suffisamment contraignante sur au moins deux plans.

Le chapitre 5 correspond au suivi par vision dynamique des plans principaux détectés par stéréovision. Les primitives coplanaires détectées aux deux poses de chacune des caméras sont mises en correspondance en recherchant l'homographie induite par le déplacement de la caméra.

Le chapitre 6 est dédié à la trajectographie du véhicule qui est calculée à partir de la décomposition des homographies estimées entre chaque pose des deux caméras. Les résultats étant peu fiables, l'introduction de la super-homographie permet de calculer de manière croisée les homographies qui lient plusieurs vues d'une même scène planaire. La prise en compte des projections des primitives coplanaires observées dans plusieurs vues permet d'introduire de manière implicite les contraintes de géométrie épipolaire pour chaque couple d'images. Les homographies extraites de la super-homographie étant plus contraintes que celles obtenues initialement s'avèrent de meilleure qualité.

En conclusion, nous mettons en évidence les contributions principales de cette étude et proposons quelques perspectives de recherches.

Chapitre 2

Etat de l'art : la vision appliquée aux robots mobiles

2.1 Le contexte

2.1.1 Les tâches de localisation et de navigation

La localisation et la navigation d'un robot mobile sont deux tâches étroitement liées qu'il est parfois périlleux de vouloir distinguer. En effet, d'après le dictionnaire de l'Académie Française, la localisation se définit comme "l'action de déterminer précisément la position d'une personne ou d'une chose". Localiser un robot mobile consiste donc à le situer par rapport aux éléments structurants de son environnement, c-à-d en faisant abstraction d'éléments mobiles. Cette tâche nécessite l'usage de capteurs qui permettent aux robots mobiles d'appréhender l'environnement dans lequel ils évoluent.

La navigation, quant à elle, est "le fait de naviguer, voyager sur la mer, les rivières", cette définition implique un déplacement en milieux dynamiques et suppose l'exploitation d'informations fournies par la localisation. La principale tâche de la navigation consiste par conséquent à intégrer le caractère dynamique de la scène de façon à mettre en oeuvre une commande adaptée.

2.1.1.1 Localisation à l'estime ou par intégration de la vitesse

La localisation à l'estime est calculée par la mesure de déplacements, de vitesses ou d'accélération du robot. Suivant le cadre applicatif, le déplacement du robot est mesuré :

- directement par odométrie. L'odomètre est un codeur optique monté sur l'axe de roues non-motrices, quand cela est possible, qui équipe la plupart des robots mobiles qui se déplacent en terrain non-accidenté. Il est cependant peu précis à faible vitesse lorsque le déplacement est inférieur à un secteur du codeur. L'odométrie peut fournir des estimations de déplacements 3D lorsque ses données sont fusionnées avec celles d'un inclinomètre.
- indirectement par intégration de la vitesse ou de l'accélération. Ces mesures sont obtenues à partir de capteurs inertiels (accéléromètres, gyromètres), employés de préférence en environnement extérieur. Ces capteurs sont connus pour leur grande précision mais ont l'inconvénient d'être onéreux à l'achat comme à l'utilisation puisqu'ils nécessitent une étape de calibration et un recalage périodique. Ils présentent de surcroît l'inconvénient de nécessiter une à deux intégrations pour fournir une estimée du déplacement et sont par conséquent très sensibles aux bruits de mesure.

La localisation à l'estime est largement employée car les estimations de déplacement fournies sont rapidement disponibles et peuvent être assimilées à des estimations fiables tant que l'incertitude sur les erreurs de mesure demeure bornée. Cette technique est en effet soumise à la dérive : un système de localisation fiable nécessite une étape de recalage.

2.1.1.2 Localisation par mesure directe de la pose

La localisation par mesure directe consiste à estimer la position d'un robot mobile à partir d'observations de l'environnement statique du robot. L'interprétation de l'intégralité de l'environnement n'est pas nécessaire à la tâche de localisation. L'estimation de la pose d'un robot s'effectue en mesurant la distance ou le gisement, écart angulaire entre la direction de déplacement de la position observée de l'amer.

D'après le dictionnaire de l'Académie Française, le mot amer est dérivé du normand "merc" du *XVII^{ème}* s. qui évoque une borne de séparation. Aujourd'hui, l'amer désigne "un point fixe sur la côte ou en mer, visible de loin et servant de repère aux navigateurs". Le fait d'utiliser ce terme d'origine maritime plutôt que balises ou repères renforce la filiation qu'assume les roboticiens avec les grands navigateurs de par la similitude des problèmes rencontrés.

Les amers sont des éléments statiques de la scène que le robot pourra facilement identifier. Suivant le type d'applications, les amers utilisés seront naturels (éléments caractéristiques de la scène) ou artificiels, on parle alors d'une instrumentation de l'environnement à partir de cibles.

La précision des mesures dépend essentiellement du type de capteur employé et de la configuration géométriques des éléments observés. Cette technique de localisation est complémentaire de celle à l'estime car elle permet par exemple de compenser les effets de dérive et fournir une estimée de la localisation dont l'incertitude est bornée. Deux critères doivent alors pris en compte : son niveau de structuration ainsi que la connaissance a priori de l'environnement et si tel est le cas son éventuelle instrumentation.

En environnement d'intérieur, les éléments structurants (murs, sol et plafond) sont formés de surfaces majoritairement planaires facilement identifiables à partir de mesures télémétriques (ultrasons, lasers, vision). Les dimensions réduites de l'environnement ainsi que la détection d'arêtes selon des directions privilégiées (encadrements de portes et intersections de couloirs) autorisent une instrumentation relativement réduite.

Les robots mobiles évoluant en terrain naturel inconnu utilisent des capteurs de distance (télémètres ou stéréoscopie) pour détecter des amers qu'ils vont suivre au cours de leur déplacement. Deux possibilités sont alors envisageables : soit le déplacement est estimé à partir du déplacement relatif des éléments détectés, soit leur position est mémorisée de façon à construire de manière incrémentale une carte, qualifiée de globale. La localisation du robot est alors obtenue suite à une étape de mise en correspondance entre la carte locale (instantanée) et la carte globale. En terrain naturel connu, la localisation est estimée par trilatération (calcul de distance) ou triangulation (intersection de trois angles de relèvement) des balises géo-référencées. Deux types de balises se distinguent :

- les artificielles passives : catadioptrique avec le système Lasernet ([Borenstein et al., 1996]), à base de trièdres en métal ([Bechetoille, 1994]),
- les artificielles actives : infra-rouges, laser ([Victorino, 2002]), lumineuses de type SIREM ([Bonnifait, 1997]), hyperfréquences radars ou GPS ([Botton et al., 1997]).

L'environnement urbain est par définition un environnement d'extérieur qui peut présenter des similitudes importantes avec les environnements d'intérieur. Les méthodes de localisation dans cet environnement particulier peuvent donc s'inspirer largement des deux techniques

décrites précédentes.

2.1.2 L'environnement urbain

La localisation et la navigation d'un robot mobile en environnement urbain exige l'élaboration d'un modèle de l'environnement qui soit à la fois réaliste et transposable dans n'importe quelle ville. Ce modèle facilitera la représentation des éléments structurants de l'environnement. Or, la diversité des scènes susceptibles d'être rencontrées dans une métropole est telle que l'énumération des variables à prendre en compte paraît déjà fastidieuse :

- a-t-on affaire à une ruelle, une rue, une avenue, un boulevard ?
- quelle est le type de voies de circulation ? Une voie à sens unique, une route à double sens, une route à plusieurs voies ?
- quel est le type d'architecture environnant ? La scène correspond-elle à une allée d'arbres qui longe des voitures en stationnement de part et d'autre de la route, des pavillons entourés de leur jardin, une juxtaposition de maisons ou immeubles qui forment une rue comme nous en trouvons dans les centre des villes européennes ou une succession de gratte-ciels entrecoupées de temps à autres par une esplanade ?

dans le cas où un système de vision est employé :

- quel est le champ de vision optimal ? Peut-on se contenter d'observer la scène devant le véhicule ou doit-on tenir compte d'un champ de vision plus latéral ?
- la signalisation horizontale est-elle omni-présente ou apparaît-elle de manière discontinue ?
- les conditions de luminosité sont-elles optimales, tolérables ou minimales ?

Enfin, il nous faut établir le type d'obstacles devant être pris en compte. Si l'on se réfère à un horizon temporel, on peut tenter de proposer des scénarios suivants en fonction de la complexité croissante des problèmes posés par un robot mobile en milieu urbain. Cette projection n'engage que son auteur :

- à court et moyen terme, le robot mobile n'est pas encore autonome. L'utilisateur bénéficie d'un véhicule équipé de capteurs et d'un système informatique capables :
 - de donner la position exacte du véhicule à tout moment sur un plan de la ville,
 - de fournir une aide à la conduite active par des indications de changements de direction ou de dépassement de vitesse autorisée,
 - d'indiquer les possibilités de stationnement et d'activités culturelles et autres au cours du déplacement,
 - d'attirer l'attention sur de possibles obstacles ou manoeuvres dangereuses.

Le système d'assistance au conducteur doit alors être susceptible :

- d'appréhender son environnement,
- de se localiser quel que soit le type d'obstruction,
- de détecter tous les obstacles,
- de classer leur dangerosité en fonction de leur déplacement relatif par rapport à la trajectoire du véhicule.
- à long terme, le robot mobile se déplace de manière autonome sur des voies semi-protégées, réservées à son passage. En cas de croisement, l'interaction avec d'autres véhicules non-autonomes est réglementée et codifiée de façon à ce qu'aucun litige ne soit possible. Le robot mobile devra alors être capable d'interagir avec l'environnement et de détecter et suivre les indicateurs qui réglementent le code de la route.
- dans un futur lointain, le robot mobile se déplace en milieu urbain en étant inséré dans le trafic des véhicules non-autonomes. Le robot mobile devra alors aussi interagir avec les

véhicules environnants afin de se déplacer de façon sécurisée sans perturber la fluidité du trafic.

Ainsi, nous pouvons vérifier que la notion d'obstacle évolue au fur et à mesure que les tâches à remplir par le robot mobile se complexifient. Dans un premier temps, l'obstacle s'avère être tout objet immobile ou non qui délimite le libre espace formé par la route. Les éléments structurants de la scène statique sont principalement les trottoirs, les remblais de séparation entre voies, le mobilier urbain et les façades des bâtiments. Les éléments de la scène dynamique à considérer en priorité sont les véhicules et les piétons. Lorsque le robot mobile sera inséré dans un flux de circulation, seuls les éléments mobiles et statiques dont la trajectoire coupe celle planifiée par le robot mobile seront considérés comme obstacles.

Dès lors, la mise en oeuvre d'un modèle de rue générique sur lequel nous pourrions nous appuyer pour réaliser notre tâche de localisation paraît difficile voire impossible tant les possibilités de configurations sont nombreuses. Il nous faudra donc trouver un moyen de nous en affranchir. Nous proposons dans les deux parties qui suivent des méthodes de localisation et navigation de robots mobiles fondées sur la vision et spécifiques aux environnements d'intérieur puis d'extérieur. Dans une troisième partie, nous nous focaliserons plus précisément sur le cadre spécifique de cette étude que sont les environnements urbains.

2.2 La localisation en robotique mobile d'intérieur

Les techniques de localisation et de navigation traditionnelles développées pour la robotique mobile d'intérieur ont d'abord employé des combinaisons de capteurs des deux types. Le premier mesure des variations de paramètres internes au robot, il est alors qualifié de proprioceptif, à l'opposé des extéroceptifs qui mesurent, eux, des informations relatives à des éléments de l'environnement. L'emploi de capteurs proprioceptifs tels les odomètres et les capteurs inertiels se justifie par les faibles distances parcourues et la vitesse réduite du robot. L'espace navigable est généralement contraint par des surfaces planaires où des capteurs de type sonars, radars, lidars ou télémètres lasers suffisent à appréhender l'environnement du robot.

Au début des années 1990, les caméras ont été introduites pour mieux appréhender l'espace libre devant le robot et réduire le nombre de capteurs embarqués. L'exploration de milieux inconnus par des robots mobiles autonomes a suscité le développement d'applications exigeant une observation latérale : le capteur de vision est alors monté sur tourelle ou bras articulés. A l'heure actuelle, la caméra omni-directionnelle est le capteur qui connaît le plus d'intérêt de la part de la communauté roboticienne du fait de sa propriété de pouvoir saisir une vue panoramique de l'environnement en une image.

Vues d'un véhicule, les rues de nos métropoles peuvent facilement être assimilées aux couloirs observés par un robot mobile d'intérieur. Les façades des bâtiments sont généralement alignées avec la route qui représente l'espace de champ libre. La route est d'ailleurs divisée en "couloirs de circulation" matérialisés par la signalisation horizontale. Dès lors, l'idée de transposer les méthodes visuelles éprouvées de robotique d'intérieur à la localisation et la navigation d'un robot mobile en milieu urbain s'impose d'elle-même.

Cependant, l'assimilation de l'environnement urbain à un simple couloir est audacieuse. La première référence qui traite de la spécificité de la localisation en milieu urbain que nous avons trouvé est en date de 1997 [Snaith et al., 1997]. La variabilité et le caractère dynamique des scènes rencontrées ont été jusqu'alors un réel frein au développement de méthodes visuelles spécifiques à la localisation d'un robot mobile en ville.

2.2.1 L'emploi de primitives

Nous appellerons par la suite primitives toutes les caractéristiques retenues suite à un filtrage de l'image pour détecter des points d'intérêt ou des contours. Ces caractéristiques sont extraites grâce aux variations de photométrie (niveaux de gris, couleurs) et de textures des éléments constituant la scène projetée dans l'image. Les auteurs conjecturent que la majorité des primitives détectées permettent de segmenter dans l'image les contours ou coins des éléments structurants de la scène. Néanmoins, des variations d'éclairage tels les reflets et les ombres portées sont à l'origine de mesures erronées qui détériorent la qualité du rapport signal sur bruit.

Les environnements d'intérieurs sont essentiellement caractérisés par la géométrie des lieux dont l'espace libre est limité par des surfaces généralement planaires. Le sol étant considéré horizontal et plan, l'environnement est modélisé par des surfaces verticales. Le mouvement du robot mobile est de surcroît plan ce qui a l'intérêt de réduire le nombre de paramètres de déplacement de six en espace libre à trois : une rotation autour de l'axe vertical et deux translations, une selon l'axe longitudinal du robot et une seconde selon le latéral. Cette hypothèse de sol plan est d'ailleurs communément admise par l'ensemble de la communauté dans le cadre de déplacements en environnements d'intérieur.

L'extraction de points d'intérêt permet de détecter la présence d'obstacles à partir d'une caméra non-calibrée. [Pears and Bojian, 2001] suppose que les points d'intérêt coplanaires vérifient la même homographie entre deux images prises de points de vues différents. La mise en correspondance est effectuée à partir d'une prédiction de l'homographie, fournie par un filtre de Kalman considérant un déplacement de la caméra à vitesse constante. [Branca et al., 1997] compare le mouvement apparent des points d'intérêt entre deux images à la base des mouvements rigides associés aux translations et rotations possibles, exprimée dans le plan image. Les expérimentations montrent que l'erreur d'approximation du champ de vecteur est plus importante lorsque l'on modélise le déplacement par le mouvement de rotation que par celui d'une translation latérale. Il est donc plus avantageux de ne considérer que le mouvement de translation selon la direction du foyer d'expansion pour estimer le mouvement plan d'un robot mobile.

La structuration des environnements d'intérieur, formés de surfaces essentiellement planaires, est à l'origine de nombreuses méthodes basées sur la détection de contours. La projection des arêtes parallèles de la scène dans l'image est à l'origine de contours qui convergent vers des points de fuite communs. L'extraction des points de fuite permet de faciliter la mise en correspondance des contours ([Lebegue and Aggarwal, 1994; Guerrero and Sagues, 2001]) qui sont ensuite utilisés dans des tâches de navigation à partir du suivi de contours verticaux.

Certains auteurs supposent que de nombreux contours sont alignés selon les trois directions principales que sont l'axe du couloir, sa transversale et la verticale. L'extraction précise des trois points de fuite correspondants facilite une reconstruction de la scène et robuste les estimations de déplacements car le mouvement apparent des points de fuite ne dépend que des mouvements de rotation de la caméra entre deux poses. Dans [Rother, 2000], le plan image illimité forme l'espace d'accumulation de l'intersection des contours, [Leung and McLean, 1996; Lobo and Dias, 2003] utilisent la sphère gaussienne et [Kosecka and Zhang, 2002] emploie un algorithme basé sur le maximum de vraisemblance pour déterminer si un contour est effectivement orienté vers un point de fuite. La dernière méthode, plus récente, semble avoir des performances supérieures aux deux premières car elle s'affranchit de la méthode d'espace d'accumulation.

Le domaine de la réalité augmentée ([Yoon et al., 2002; Kim et al., 2002; Gibson, 1950]) s'est d'ailleurs largement inspiré de ces travaux pour développer des méthodes de reconstruc-

tion semi-automatique d'environnements structurés à partir de contraintes de parallélisme et d'orthogonalité ou des méthodes de calibration de caméra à partir d'une unique image d'une boîte cubique.

Enfin, nous ne pouvons passer sous silence les travaux du projet MOVI¹, dirigé par R. Horaud, sur la modélisation, la reconstruction grâce au déplacement d'une caméra (*Structure From Motion*) et la calibration de caméra à partir de scènes planaires sont en tout point remarquable. Les méthodes développées étant innovantes et performantes, il peut être intéressant de chercher à transposer certaines d'entre elles dans des cadres applicatifs moins théoriques que celles dans lequel elles sont généralement présentées.

2.2.2 L'emploi d'amers

La localisation à partir d'amers nécessite l'emploi d'une liste ou carte où sont répertoriées les coordonnées mais aussi les signes distinctifs des amers dans le cadre de la navigation. Le relevé de l'angle de gisement permet par triangulation d'estimer la position de l'observateur. Lorsque les amers sont observés à partir d'un capteur qui fournit de surcroît une estimation de la distance (télémètre laser et balises réfléchissantes, par exemple), la détection de seulement deux amers est nécessaire pour faire le point. Néanmoins, nous nous restreignons dans ce qui suit à l'utilisation d'amers visuels.

La connaissance a priori de l'environnement est nécessaire aux localisations dites absolues à partir d'amers détectés à partir de capteurs extéroceptifs. Une variante de cette méthode consiste à s'affranchir de la connaissance a priori de la position des amers en estimant et affinant leurs positions respectives au cours d'une tâche d'exploration. Lorsque l'environnement contient peu d'éléments susceptibles de jouer le rôle d'amer, l'environnement a été instrumentalisé en y insérant des amers artificiels ou cibles. L'instrumentation d'un environnement suppose que ce dernier soit de dimensions finies.

Deux cas sont envisageables : la tâche de navigation peut être remplie en observant la position relative d'amers non-référencés au cours d'un déplacement. Les amers sont principalement utilisés en robotique mobile dans des tâches de recalage. La détection et le suivi d'amers non-référencés permettent généralement d'améliorer la robustesse de l'étape de localisation. Lorsque la détection des amers est opérée à partir d'un capteur de vision, on lui adjoint généralement au minimum un odomètre et un capteur extéroceptif qui permettent d'estimer le déplacement du robot et mesurer la distance auxquelles sont détectées les amers.

2.2.2.1 Scènes non-instrumentées

Les méthodes directes qui estiment les mouvements des éléments dans l'image sans connaissance a priori sur le déplacement 3D de la caméra et/ou des objets sont largement employées en robotique mobile d'intérieur car elles sont adaptées aux faibles déplacements dans des environnements confinés. Deux méthodes sont couramment utilisées pour détecter des amers naturels. La "*Normalized Cross Correlation*" (NCC), est généralement employée pour s'affranchir de variations d'illuminations importantes et fournit des amers invariants aux transformations d'échelles et aux déplacements. La "*Sum of Absolute Difference*" (SAD) semble plus appropriée aux environnements d'intérieurs, plutôt statiques, elle permet de détecter très rapidement des balises stables.

¹<http://www.inrialpes.fr/movi/>

[Werman et al., 1999] cherche à auto-calibrer une caméra en estimant les paramètres intrinsèques à partir de combinaisons linéaires des positions 3D connues d'amers et de leurs représentations à l'image. Différents cas sont abordés : mouvement connu, mouvement inconnu, mouvement plan sans puis avec rotation. Les résultats sont corrects mais cet article demeure trop proche de la théorie pour être réellement exploitable dans des applications d'extérieur.

Ainsi, les quatre améliorations que proposent [Watman et al., 2004] d'un algorithme de corrélation fondé sur la SAD permettent une détection d'amers en environnement d'intérieur en réduisant le temps de calcul d'un facteur pouvant atteindre 1000 dans les meilleures conditions. Lorsque le coefficient de corrélation entre deux régions est maximal, le résultat de la SAD présente un minimum au centre de la fenêtre de recherche. La détection de contours de l'image des disparités se fait en deux passes : la première sur toute l'image à l'aide d'une fenêtre glissante 3x3 pour segmenter les dix meilleurs scores de la corrélation par SAD, la seconde pour affiner le résultat en quadruplant la taille du masque. Les régions uniformes de chaque image sont ensuite éliminées en comparant leurs couleurs à celles de leurs voisines. La détermination du déplacement optimal de chaque région entre deux images est alors estimée en utilisant un seuil adaptatif en parcourant la région considérée à l'aide d'une trajectoire en spirale "Turn Back and Look".

2.2.2.2 Scènes instrumentalisées

Les méthodes fondées sur l'indexation d'images se distinguent en deux familles selon le type d'amers artificiels à détecter. La première utilise une collection d'images des différents amers à détecter au cours d'une tâche de navigation tandis que la seconde suppose que les amers sont représentés par une cible de même motif. Une carte des amers référencés dans le repère de l'environnement est alors utilisée.

[Dudeck and Jugessur, 2000] a étudié l'intérêt d'intégrer une analyse en composantes principales pour détecter des amers artificiels dans des images en couleur. L'objectif est de faciliter la navigation à l'estime d'un robot mobile en employant des imagerie de balises uniques. Plusieurs avantages de la méthode peuvent être dégagés : la détection des objets est invariante par rotation si l'on effectue une transformée de Fourier sur l'image de référence lorsqu'elle est exprimée avec une représentation polaire et l'insensibilité aux occultations partielles. L'analyse en composante principale se résume à une compression des informations contenues dans une imagerie de référence. Un opérateur qui délimite de manière automatique les régions d'intérêt des images à tester est utilisé pour estimer le centre de chaque fenêtre sur laquelle se fera l'analyse. Les auteurs demeurent cependant imprécis quant au type d'opérateur qu'ils emploient.

Le suivi d'une cible plane texturée permet le calcul de l'homographie relative au mouvement apparent de cette dernière dans l'image. [Jurie and Dhome, 2001] propose de minimiser la différence de niveaux de gris de points équi-répartis en s'appuyant sur la méthode des hyperplans. Malgré un coût de calcul réduit, la méthode nécessite une phase d'apprentissage pour chaque région cible. Inspiré par la problématique connue du train de véhicules, [Benhimane and Malis, 2005] propose de réaliser le lien virtuel qui lie un véhicule maître à un véhicule esclave muni d'une caméra en calculant de manière robuste et rapide l'homographie d'une cible plane texturée, disposée sur le véhicule maître.

La structuration des environnements de type couloirs suppose qu'il est possible d'observer des amers plans de forme quadrangulaire et circulaires. Ces amers présentent des contraintes géométriques fortes qui permettent l'estimation de l'orientation de la caméra qui les observe. Ainsi [Mata et al., 2001] utilise un algorithme génétique pour détecter dans chaque scène

des régions d'intérêt autour d'amers préalablement définis. Les régions d'intérêt sont des parallélogrammes définis par les coordonnées du coin supérieur gauche, les dimensions de largeur et de hauteur ainsi que d'un facteur d'obliquité, représentatif de la distorsion projective. L'identification de l'amer est réalisé par corrélation de quatre fenêtres rectangulaires situées aux extrémités de la région d'intérêt. Une décomposition des couleurs selon une représentation cylindrique HSV (teinte, saturation, luminance) est utilisée en prenant garde de rejeter des amers dont la valeur de saturation est faible car peu discriminante.

[Hayet, 2003] fusionne les données télémétriques d'un capteur laser avec l'indexation d'amers visuels texturés, observés par une caméra monochrome. Leur identification et leur suivi font appel aux notions de saillance (seuillage de la covariance et le nombre de points d'intérêt qui forment un amer) et de visibilité afin d'augmenter le pouvoir de discrimination. Un recalage par photogrammétrie (applications de la stéréoscopie aux levés topographiques) est réalisé à partir du mouvement de points d'intérêt détectés dans les amers en considérant que le déplacement de la caméra vérifie une homographie. Les données du télémètre laser sont, elles, utilisées pour entreprendre la segmentation robuste des murs qui forment le couloir et réaliser la calibration de la caméra en fusionnant le résultat d'une segmentation de contours, facilitée par la détection du point de fuite dominant.

L'intérêt de telles méthodes est évident lors de la transposition à notre problème puisqu'il serait possible soit de reconnaître des intersections, soit de détecter des amers artificiels (panneaux de signalisation, feux rouges, flèches indicatrices sur la route) à partir d'une simple image.

Cependant, la contrainte de faible mobilité du robot d'intérieur pose un certain nombre de problèmes lorsque l'on souhaite appliquer directement les méthodes dédiées à la robotique mobile au milieu dynamique structuré qu'est le milieu urbain. En effet, l'ordre de grandeur des vitesses de déplacement est dans le meilleur des cas décuplé et la variabilité de la scène au cours du temps est difficilement paramétrable. C'est pourquoi, il nous semble opportun de nous inspirer des méthodes visuelles de localisation et navigation développées en environnement extérieur non-urbain pour les transposer en milieu urbain par la suite.

2.3 La localisation en robotique mobile d'extérieur

2.3.1 Evolution au cours du temps

A la fin des années 1980, une réelle volonté politique accompagnée de crédits importants a permis l'accélération du développement des robots mobiles d'extérieur. Le laboratoire américain DARPA a alors obtenu des subventions pour qu'un robot suive une route de manière autonome (projet "Autonomous Land Vehicle") et explore un terrain inconnu non-structuré (projet "Unmanned Ground Vehicle"). L'émulation aidant, les autres laboratoires de robotique se sont alors rapidement intéressés à ces nouvelles problématiques.

L'abaissement des coûts et les progrès technologiques des caméras ont rapidement favorisé l'intégration de ce capteur supplémentaire aux configurations existantes embarquées sur les robots mobiles. L'apport de la vision a permis une modélisation 3D de l'environnement plus complète où la notion d'espace libre a fait place à la distinction entre environnement statique et dynamique. De manière plus anecdotique, les projets d'exploration planétaire à partir de robots autonomes n'auraient pas vu le jour sans l'apport de la vision.

Conjointement, l'introduction des systèmes de géo-référencement a permis la transposition de la plupart des techniques de localisation et de navigation fondées sur des connaissances a priori employées en robotique mobile d'intérieur aux environnements ouverts. Les systèmes de

géo-référencement sont à l'origine d'une véritable explosion de l'échelle des distances des champs d'action de la robotique mobile en environnement extérieur.

Il ne faut pas pour autant perdre de vue que la résolution du système GPS était de quelques dizaines de mètres jusqu'en 2000, date à laquelle le code SA de brouillage a été désactivé. Elle est depuis inférieure à 10 m. En mode différentiel (DGPS), la résolution est comprise entre 1 à 10 m. A l'horizon 2008, avec la mise en exploitation des systèmes Galileo et Egnos, la résolution des systèmes de géo-référencement est annoncée à moins de 5m en mode de base.

La précision de la localisation de récepteurs fixes peut être augmentée d'un facteur 100. L'information de position est, certes, précise en comparaison de la couverture quasi-mondiale mais ne peut être utilisée telle quelle pour localiser avec précision un robot mobile. L'intégration de récepteurs GPS a donc favorisé le développement de robots mobiles d'extérieur capables soit de parcourir plusieurs centaines de kilomètres sur autoroutes, soit d'entreprendre des tâches de navigations en milieu ouverts non-structuré comme des champs.

Les champs d'applications de la robotique mobile en milieu naturels non-structurés sont nombreux et variés : de l'observation sous-marine ([Espiau, 2002]) aux robots explorateurs sur Mars ([Mallet, 2001]) en passant par l'automatisation des engins agricoles ([Gaudin, 2000]) ou la cartographie de paysages par des drones ou dirigeables ([Silveira et al., 2003]). Ils représentent une part importante en volume et investissements de la robotique mobile d'extérieure. Malgré la diversité apparente, ces quelques exemples ont tous un point commun : la vision. L'introduction des caméras a en effet permis soit de rendre les tâches de navigation plus robustes en fusionnant les données visuelles à celles des capteurs de navigation, soit de rendre autonome des robots qui étaient jusqu'alors télé-opérés.

2.3.2 La localisation de robots mobiles sur route

Nous nous limiterons dans cette partie, à de rares exceptions près, aux applications concernant la robotique mobile sur route qui correspond à la thématique de cette thèse. Une route peut-elle être considérée comme un élément structurant de la scène ? Vue du ciel, certainement mais observée d'un véhicule, elle apparaîtra comme une région uniforme plutôt foncée et homogène, éventuellement délimitée par un marquage au sol. Cependant dans le cadre spécifique de véhicules autonomes équipés de caméra(s), la route s'avère être le seul lien commun aux images issues de séquences vidéo enregistrées en campagne, sur autoroutes ou en ville.

La caractérisation de la route est donc prépondérante pour notre travail mais elle n'est pas pour autant toujours facile. La plupart des expérimentations de navigation autonome réalisées jusqu'alors ont été effectuées sur circuit ou sur autoroute. Les méthodologies sont maintenant éprouvées mais les configurations de test imposent toujours d'avoir un champ de vision dégagé permettant une segmentation aisée de la route dans les images. Ainsi, une distance relativement importante (plusieurs dizaines de mètres) est maintenue sur la plupart des applications présentées de manière à ne pas compliquer la tâche de navigation par une gestion d'obstacle.

La majorité des travaux suppose des hypothèses communes concernant la modélisation des routes. Elles sont considérées :

- planes ou localement planes car cela facilite l'estimation de déplacement,
- à bords parallèles, ce qui demeure vrai même en cas de virage, tout du moins pour une approximation linéaire locale de ses limites latérales observés au premier plan,
- matérialisée : l'asphalte et le béton forment à l'image des régions homogènes dont les limites ainsi que les voies de circulation sont mises en évidence par des bandes blanches (ou jaunes selon les pays) de la signalisation horizontale. Cette dernière contrainte restrictive

n'est cependant pas supposée par tous. Quelques auteurs ([Turchetto and Manduchi, 2003; Okutomi et al., 2002]) parviennent à segmenter la région de la route dans les images en supposant généralement une faible variation de sa projection entre deux acquisitions.

2.3.2.1 Etat de l'art en 1990

L'article de [Aubert et al., 1990] est, à ce propos, des plus intéressants. En effet, un état de l'art était mené sur les différentes pistes qui étaient étudiées par les laboratoires les plus avancés pour résoudre les problèmes de robotique mobile à partir de systèmes de vision. Une étude comparative était ainsi menée sur 11 méthodes différentes de navigation sur routes structurées. La comparaison détaillait le type de segmentation, le modèle de route retenu et relevait les hypothèses simplificatrices ainsi que les connaissances a priori de l'environnement utilisées, la prise en compte ou non des problèmes liés aux intersections.

À l'époque déjà, les problèmes étaient clairement identifiés : le suivi de route de manière automatique exige :

1. la segmentation de la route dans l'image pour en extraire des primitives ou caractéristiques,
2. la modélisation locale de la géométrie de la route en vue d'une phase de localisation et l'estimation de la position du véhicule par rapport à cette dernière,
3. la détermination de la trajectoire à suivre et la gestion des problèmes de navigation dans les intersections.

De nombreux moyens de segmenter la route dans les images avaient été expérimentés : de la classification par colorimétrie à la propagation des caractéristiques de la signalisation horizontale détectée par transformation de Hough puis filtrée par un réseau de neurones ou filtrage de Kalman, en passant par l'extraction de point de fuite et l'exploitation de la connaissance des paramètres de la route comme la largeur des voies, les caractéristiques des bandes qui les délimitent.

Le laboratoire de robotique de CMU (Carnegie Mellon University) qui ne manquait pas d'humour avait d'ailleurs baptisé son robot YARF pour Yet Another Road Follower ! Différents modèles de courbures de la route (linéaire, parabolique, spine curve) étaient éprouvés, des méthodes de corrélation à partir de masques orientés étaient employées de façon à identifier les bandes de la signalisation horizontale.

La bibliographie importante sur le sujet de la segmentation de la route encore aujourd'hui atteste que si les problèmes ont été clairement établis en 1990, la plupart d'entre eux n'ont toujours pas été résolus depuis.

2.3.2.2 La détection de la route

Le cas des routes peu ou pas matérialisées par un marquage au sol correspondent à des applications spécifiques de robotique mobile : suivi de pistes de désert ([Rasmussen, 2004]) à partir de textures parallèles au tracé ou en terrains accidentés pour applications militaires ([Kluge and Lakshmanan, 1995; Hong et al., 2002]). Le premier utilise une approche de type contours actifs en cas de route très ombragée tandis que le second fusionne les informations fournies par un capteur d'élévation pour segmenter la région de l'image correspondant au chemin à suivre. Plus proche de notre problématique, les conditions d'illuminations extrêmes en tenant compte d'ombres portées ont été étudiées par [Turchetto and Manduchi, 2003]. Tous supposent néanmoins que la route est délimitée par des bords parallèles ou des contours linéaires par morceaux, plus ou moins faciles à détecter suivant les cas. [Wang et al., 2002] par exemple

est d'avis que les contours naturels des routes chinoises sont plus fiables que la signalisation horizontale!

Le suivi des bords de la route La représentation de la route par une surface dont les frontières peuvent être matérialisées est couramment employée par la majorité des auteurs. En effet, les bords de la route sont généralement des régions de l'image à fort contraste, continues, facilement détectables par une segmentation de contours.

La représentation de la surface par ses deux frontières principales est complétée par l'hypothèse que les paramètres liés au modèle évoluent peu entre deux images. Ainsi, il est envisageable d'utiliser une approche de type filtrage de Kalman sur les paramètres du modèle de façon à adapter le modèle à la qualité des contours détectés. L'objectif est alors de s'affranchir d'évolutions lentes des valeurs des paramètres obtenus lors de la phase d'initialisation ou d'apprentissage selon les cas.

Le cas de modèle de route linéaire est largement employé du fait de sa simplicité et de la pertinence d'une telle approximation sur la plupart des routes à grands rayons de courbure qui représentaient l'essentiel des cas étudiés jusqu'à peu. La transformée de Hough est alors largement répandue ([Kasprzak et al., 1994], [Taylor et al., 1996]) comme outil de discrimination pour détecter les contours qui représentent les délimitations de la route. D'autres comme [Lu et al., 2002] exploitent directement l'évolution du point de fuite dominant pour identifier l'évolution du modèle.

Cette méthode communément employée a trois principaux avantages :

1. elle répond aux contraintes de continuité imposées par la scène,
2. elle diminue le temps de calcul nécessaire à sa détection,
3. elle permet de s'affranchir d'éventuelles occlusions d'obstacles qui pourraient perturber un algorithme simplement basé sur la détection de contours.

Dans le cas de modèles plus élaborés, l'introduction d'un capteur supplémentaire de position, vitesse ou accélération permet de corroborer les observations faites par le capteur de vision ([Hayet, 2003; Kluge and Lakshmanan, 1995]). L'apport principal de la fusion de données est de pouvoir discriminer parmi les solutions plausibles celles qui pourraient être obtenues à partir d'artefacts.

D'autres types de modèles sont couramment employés : polynomiaux ([Aufrère et al., 2001] voir Fig. 2.1) mais aussi des clothoïdes ([Hu and Uchimura, 2002]) et même des hyperboloïdes ([Kluge and Lakshmanan, 1995]). Le degré de complexité associé aux modèles dépend essentiellement de la profondeur de champ que le système de vision est censé observer bien que cet argument apparaisse rarement dans les justifications invoquées. Les auteurs qui exploitent des scènes d'autoroutes utilisent généralement les modèles les plus complexes qui tiennent compte de courbures variables de la route. L'objectif est alors de déterminer le meilleur jeu de paramètres qui représente la courbure de la route dans sa globalité. A l'opposé, les auteurs travaillant sur des scènes urbaines se contentent de représentation linéaire. Le trafic est tel qu'espérer observer la route sur plusieurs dizaines de mètres est une gageure.

Nous nous sommes délibérément limités à quelques exemples parmi la nombreuse bibliographie qui existe dans le domaine de la robotique mobile sur route. Cette limitation trouve sa justification dans le fait que les environnements urbains présentent parfois des scènes où la présence de marquage au sol n'est pas systématique. Dès lors, la mise en oeuvre d'une méthode basée sur ce type d'amers peut restreindre le champ d'application. Il apparaît néanmoins évident que le modèle de bords de voie le plus adapté à l'environnement urbain est le linéaire.

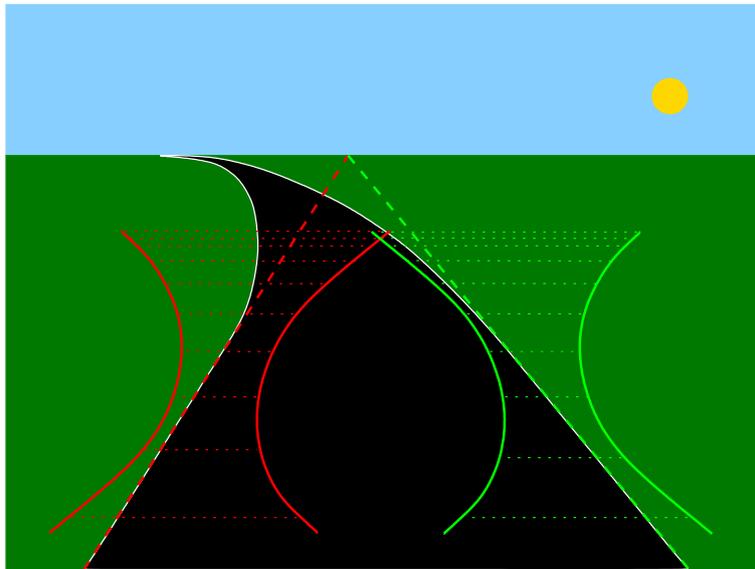


FIG. 2.1 – Modélisation des frontières considérées parallèles d’une route. Un modèle linéaire représenté en pointillés est généralement acceptable au premier plan. Le modèle employé par Aufrère est polynomial : chaque contour est estimé en 10 ordonnées test où l’intervalle de confiance est symbolisé par les pointillés. Lorsque le modèle s’affine, les arcs représentant les extrémités des intervalles de confiance sont de plus en plus proches de la frontière de la route.

La continuité des caractéristiques Outre le modèle de contours qui délimitent la région de l’image, il est aussi courant de supposer que les caractéristiques liées à la région de la route dans l’image varient peu entre deux images enregistrées à la cadence vidéo. Ainsi, des algorithmes traquent des primitives (points, droites et combinaisons des deux), recherchent des régions dont la colorimétrie est identique à celles obtenues à l’itération précédente : décomposition RGB pour [Heisele, 2000], HSV pour [Sotelo et al., 2002]. Ce genre de méthodes restreint alors la région de recherche grâce à une connaissance a priori de l’environnement : la région d’intérêt se situe dans la partie basse de l’image sous la ligne d’horizon, limitée par les frontières de route.

A partir d’une paire stéréo calibrée, [Okutomi et al., 2002] segmente la région apparentée à la route en tant que région de l’image qui vérifie l’homographie induite par le plan de la route entre les deux images du couple. La méthode s’affranchit de la présence de marquage au sol ou de la détection de délimitations puisque l’estimation de la matrice d’homographie est obtenue par minimisation de la différence des niveaux de gris entre les deux images.

2.3.3 Applications

Contrairement aux autres capteurs qui mesurent soit un état, soit une variable physique, la caméra enregistre une image instantanée d’une scène. Ainsi, l’introduction d’un système de vision sur un robot mobile ne répond pas à la nécessité de mesurer une variable donnée mais plutôt d’interpréter la scène dans son intégralité. Les applications de l’introduction d’un système de vision sur un robot mobile que nous présentons dans cette partie ne dépendent pas du type d’environnement dans lequel évolue le robot.

L’ordre de grandeur des échelles de distance, de vitesse ainsi que le caractère dynamique des scènes extérieures est plus contraignant que ceux qui caractérisent les applications en en-

vironnement d'intérieur. Nous restreignons volontairement dans ce qui suit des applications de l'introduction de système de vision embarqué à bord de véhicules sur routes. La détection d'une route dans une scène par l'intermédiaire d'un système de vision est en effet étroitement liée au fait de situer le véhicule sur la chaussée, d'éventuellement estimer son déplacement entre deux acquisitions et de distinguer d'éventuels obstacles.

2.3.3.1 Contrôle latéral du véhicule

La bande de roulement virtuelle² mise au point par le Centre de Robotique de l'Ecole des Mines de Paris, exploite le résultat de la détection de contours opéré sur les bandes blanches qui bordent la voie de circulation empruntée pour estimer la position latérale et l'orientation du véhicule. L'information de vitesse du véhicule permet d'estimer le temps nécessaire pour franchir une des deux bandes et éventuellement déclencher une alerte.

Sur le même principe, [Takahashi et al., 2002] propose, d'utiliser une caméra grand angle située à l'arrière du véhicule, inclinée vers le sol en direction opposée au déplacement. Cette caméra est généralement utilisée pour l'aide au stationnement sur certains véhicules. La détection du franchissement nécessite d'abord de redresser l'image de façon à linéariser les bandes observées. Ensuite, une transformée de Hough étendue est appliquée de façon à extraire les paramètres de la voie, limitée par deux bandes. La transformation du plan image en vue de dessus ainsi que la table de conversion qui permet la transformation de l'espace image à l'espace des paramètres requièrent une étape d'initialisation.

2.3.3.2 Estimation du déplacement du véhicule

Comme nous l'avons vu jusqu'alors, différentes méthodes permettent de segmenter la route et l'environnement statique d'une scène. Cependant estimer le déplacement du véhicule grâce à la vision demeure encore peu étudié. Cette tâche réclame de pouvoir extraire, caractériser et suivre des amers statiques de l'environnement. La fusion de données à l'aide de capteurs proprioceptifs ou extéroceptifs robuste le processus de vision. Nous souhaitons décrire quelques méthodes fondées sur l'exploitation des seules données visuelles pour estimer le déplacement entre deux images.

La méthode couramment utilisée dans de tels cas est une estimation du flot optique calculé à partir de primitives ponctuelles et de considérations sur la scène observée. Le premier plan est en effet généralement dépourvu d'obstacles et le nombre de primitives pouvant être extraits est conséquent ([Adiv, 1985; Horn and Weldon, 1988]) ont été parmi les précurseurs dans ce domaine en environnement d'intérieur. Lorsque l'environnement est inconnu, l'estimation du déplacement du robot à partir d'un système de vision sous-entend de pouvoir bénéficier de la calibration du système de vision de manière à obtenir une information métrique. Le problème de la qualité de la calibration pose des difficultés importantes dans le cas de caméras embarquées. Des méthodes d'auto-calibration en ligne sont mises en oeuvre afin de permettre une compensation des dérives dues aux conditions d'utilisation (vibrations, accélérations, ...).

[Stein et al., 2000] parvient à estimer le mouvement d'un véhicule à partir d'une unique caméra calibrée en observant une route non matérialisée supposée plane. La méthode est fondée sur l'hypothèse que l'image des disparités entre deux images consécutives obéit à une fonction de vraisemblance. L'auteur suppose que le déplacement observé entre deux patches d'images consécutives suit une loi bayésienne. Il cherche donc à obtenir la meilleure estimation du

²<http://caor.ensmp.fr/french/recherche/sti/BandeRoulementVirtuelle.php>

déplacement a priori qui permet de minimiser la différence entre le warping du patch de l'image courante vers le patch de l'image précédente en utilisant comme prédiction du déplacement celui estimé à l'image précédente. L'estimation du mouvement de la caméra est alors calculée en sommant les densités de probabilités des mouvements qui minimisent la différence entre patches en tenant compte de leur appartenance à la route et d'une information relative à leur gradient. Les résultats obtenus sont remarquables : l'erreur sur l'estimation du mouvement de rotation, suite à un tour de rond-point, n'est que de 6° et la méthode fonctionne aussi de nuit.

[Ke and Kanade, 2003] propose une amélioration de la méthode précédente en exploitant le résultat de l'observation d'une caméra virtuelle orientée selon la verticale vers le sol. L'apport de cet artifice de calcul est majeur : le mouvement apparent entre les deux patches n'est alors plus soumis à l'ambiguïté des contributions des mouvements de rotation et de translation. De plus, en positionnant le centre de la caméra virtuelle au centre du patch, le conditionnement du Hessien est clairement amélioré, ce qui robuste l'estimation de l'homographie induite entre les deux patches. Cette fois encore, les résultats calculés à partir de patches d'une centaine de pixels de coté d'une région de la route sans la moindre structure sont d'une grande précision. La calibration de la caméra peut s'avérer un réel frein à l'emploi d'une telle méthode qui propose néanmoins une solution robuste au problème de l'estimation du déplacement d'une caméra.

La route étant une surface peu texturée, la plupart des auteurs (entre autres [Mandelbaum et al., 2000; Horn and Weldon, 1988]) s'accordent sur l'emploi des méthodes directes pour estimer le déplacement d'une caméra embarquée au profit de méthodes basées sur le flot optique qui nécessitent, elles, des surfaces très texturées.

2.3.3.3 Détection, caractérisation et évitement d'obstacles

En général La tâche de détection d'obstacles est l'apanage de méthodes fondées sur l'emploi de capteurs de type télémètres qui fournissent une information 3D (gisement et distance) de la position d'un éventuel obstacle. L'inconvénient majeur de ce type de capteur est de ne fournir que des coupes de l'environnement puisque les tirs se font pour des élévations et des angles de gisement prédéterminés. L'introduction de la vision a permis de robustifier les méthodes développées jusqu'alors en permettant de mieux saisir la complexité des scènes selon l'axe verticale et d'offrir une perception plus homogène des éléments qui structurent la scène.

La fusion de données télémètre-vision est effectuée dans le cadre d'applications où les éléments qui composent la scène ont des caractéristiques morphologiques, colorimétriques ou de texture qui leurs soient propres. Les informations extraites par les deux types de capteurs sont complémentaires et permettent d'effectuer une reconstruction en ligne de l'environnement. De plus, la vision est particulièrement bien adaptée à la reconnaissance d'objets (amers, obstacles) du fait de son pouvoir de segmentation et de la connaissance a priori sur l'environnement qui permet d'introduire des contraintes spatio-temporelles sur les mouvements apparents dans les images.

Comme nous l'avons expliqué dans l'introduction de ce chapitre, la notion d'obstacle est relative au type d'application que doit remplir le robot mobile. Dans le cas du déplacement d'un véhicule sur route, la détection d'obstacles consiste à estimer l'espace libre devant le véhicule. Actuellement, l'accent est essentiellement mis sur la détection des piétons de façon à fournir un système d'aide à la conduite qui permette de diminuer le nombre d'accidents graves dus à leur renversement. A terme, l'autonomie des robots mobiles nécessite qu'il soient capables de non seulement distinguer la scène statique de la dynamique mais surtout d'adapter leur commande en fonction du type d'obstacle (piétons/véhicules) et du risque de croisement avec sa trajectoire.

La tâche de détection d'obstacle à partir de la vision est fondée sur le principe qu'un obstacle ne satisfait pas le modèle retenu pour la route. Les quatre méthodes suivantes supposent donc qu'une rupture de la représentation de la route dans l'image est synonyme de la présence d'obstacle(s) :

- le flot optique est la méthode employée par la majorité des auteurs. A partir de suppositions sur le mouvement de la caméra, l'obstacle dynamique présente une discontinuité dans le flot optique dont une composante est due au mouvement relatif entre la caméra et ce dernier. [Talukder et al., 2003] propose de calculer la corrélation entre deux images à partir d'une fenêtre glissante,
- la continuité des propriétés chromatiques : en supposant que la représentation de la route est continue et homogène, la segmentation de la scène en régions qui présentent des propriétés communes. Différentes propriétés sont étudiées : [Heisele, 2000] agrège les régions suivant une décomposition de la colorimétrie RGB en utilisant un algorithme dédié du k^{eme} plus-proche-voisin ("*k-means*"),
- l'extraction et le suivi de points d'intérêt robustes de type KLT ([Rabie et al., 2001]) pour l'analyse de scène à partir d'une caméra embarquée, montée sur une tourelle *pan-and-tilt*,
- la disparité entre deux images d'une paire stéréo calibrée : [Yu et al., 2003] compare la normale extraite du plan de la route avec celles calculées à partir de trois points tirés aléatoirement, [Okutomi et al., 2002] calcule l'homographie induite par le plan de la route entre les deux images du couple stéréo,

Il est à noter que seules les méthodes qui exploitent la disparité entre les deux images sont capables de détecter des obstacles statiques mais nécessite l'emploi d'une paire stéréo. Toutes les méthodes ont dans le pire des cas des performances temporelles proches des contraintes Temps-Réel. Les performances de détection des méthodes présentées sont relatives : sur les exemples proposés, les obstacles se situent rarement avant le second plan, la proportion de la scène statique à l'image étant en effet largement prépondérante.

La détection de véhicules Les méthodes suivantes ont été développées dans l'optique de fonctions d'aide à la conduite pour les véhicules actuels. Elles se distinguent des précédentes en fournissant une donnée métrique de la distance qui sépare l'obstacle détecté du véhicule. Deux méthodes sont actuellement mis en oeuvre pour recouvrer cette information de qualité : la fusion de données entre une caméra et un capteur de type télémètre, Lidar, Ladar, Laser ou l'emploi d'une paire stéréo calibrée voire rectifiée. De plus, la recherche sélective d'obstacles de type véhicule a nécessité pour chacune des méthodes présentées la mise en oeuvre de modèles plus ou moins évolués.

La v-disparité développée par [Labayrade and Aubert, 2003] exploite les informations contenues dans l'image de v-disparité pour estimer le profil longitudinal de la route ainsi que la distance qui sépare le système de vision de l'intersection des véhicules avec la route sur laquelle ils circulent. La v-disparité (v axe vertical des images) permet d'identifier dans les plans images d'une paire stéréo calibrée et rectifiée les points de la scène qui présentent la même disparité (mouvement relatif en abscisse). Le calcul de la v-disparité revient à accumuler les *pixels* d'une même ligne qui présentent une disparité commune. L'image de v-disparité présente pour chaque ligne de l'image la fonction de répartition en niveaux de gris des disparités mesurées, reportées en abscisse. La v-disparité permet de segmenter de manière fiable et rapide des surfaces planaires observées dans la scène. C'est à notre connaissance une des rares méthodes capable de fournir des résultats de segmentation en tenant compte d'un profil de route planaire par morceaux (jusqu'à 15 pour un même profil). La distinction entre véhicules particuliers et poids-lourds est

effectuée en supposant que les premiers ne dépassent pas 2m de hauteur, contrairement aux seconds compris entre 2 et 5m.

[Bensrhair et al., 2002] a développé un algorithme qui, à partir d'une paire stéréo calibrée et rectifiée, détecte en Temps Réel des véhicules en considérant la symétrie de leurs contours verticaux. Les extrémités des segments qui forment les contours sont obtenus en vérifiant leur déclinitivité. La déclinitivité est la propriété qu'ont les pixels qui forment un contour d'avoir un niveau de gris qui évolue linéairement. Une phase de mise en correspondance est alors effectuée avec les contours sélectionnés en utilisant l'information 3D de grilles d'occupation.

[Hilario et al., 2004] modélise, quant à lui, les véhicules à partir de 7 paramètres : leurs coordonnées dans l'image, leur hauteur, leur inclinaison de pare-brise par rapport à la verticale vue de derrière, leur proportion de surface vitrée et de caisse par rapport à la hauteur globale. A partir de ce modèle est calculée la distribution de probabilité $P(I/Z)$ qu'une déformation du modèle en fonction des paramètres Z ait lieu dans l'image I . Dès lors, la détection d'obstacles revient à chercher le maximum a posteriori de la fonction de probabilité $P(Z/I)$. Une fonction énergie est introduite pour simultanément tenir compte des facteurs symétrie, forme et ombre générée par l'obstacle. L'obstacle est correctement détecté par l'algorithme génétique lorsque la fonction énergie est minimale.

Le système Fonction d'Analyse et de Détection de l'Environnement (FADE³) du Centre de Robotique de l'Ecole des Mines de Paris parvient à élargir le champ de détection des véhicules roulant sur les voies latérales de manière à s'affranchir des erreurs de détection du radar anti-collision qui équipe les véhicules possédant un régulateur de vitesse adaptatif (*Auto Cruise Control*). L'algorithme réalise en Temps-Réel la fusion de plusieurs techniques qui travaillent en parallèle. Le modèle de véhicule comprend outre des détecteurs de gabarit vus précédemment (détecteur de symétrie, de lignes verticales et horizontale), un détecteur d'ombre formée entre les roues arrières d'un véhicule suivi et un détecteur de feux arrières rouge (à partir d'une caméra monochrome). Les résultats obtenus sont présentés sur une séquence tournée en trafic dense à vive allure, les véhicules sont détectés dès qu'ils apparaissent entièrement à l'image.

Enfin, l'originalité de la méthode développée par [Okada et al., 2003] mérite de figurer dans cette liste bien que nous émettions quelques restrictions quand à la reproductivité des résultats. En effet, les véhicules détectés sont supposés proches et vus de derrière. La méthode s'affranchit du calcul du mouvement apparent de la route car elle emploie l'invariance du bi-rapport de quatre droites. Le bi-rapport de référence est dans un premier temps calculé dans une portion de l'image, libre de tout obstacle. Une contrainte de déplacement (pour une vitesse donnée) en substituant la 4^{eme} ordonnée par la ligne de fuite $y = 0$. En présence de véhicule, l'opération est répétée en substituant cette fois-ci la 4^{eme} ordonnée par la ligne de fuite $y = \infty$. A l'image suivante, si les ordonnées du groupe d'horizontales sont plus proches des ordonnées prédites par la contrainte de déplacement de l'obstacle que celle de la route alors le jeu d'horizontales est effectivement un obstacle.

La solution FADE mise à part du fait qu'elle combine les performances de plusieurs algorithmes de suivi et de détection, la méthode de la v-disparité apparaît de loin celle qui fournit les meilleurs résultats pour une implémentation aisée qui ne nécessite pas à proprement parler de modèle pour détecter des obstacles. Cette méthode sera néanmoins difficilement à mettre en oeuvre en milieu urbain car la densité du trafic peut s'opposer à une identification correcte du profil de la route.

³<http://caor.ensmp.fr/french/recherche/sti/FusionRadarVisionFade.php>

La détection de piétons Parmi les obstacles qu'un robot mobile peut rencontrer figurent les personnes. Or, la complexité de l'élaboration d'un modèle de piétons rend la tâche quasiment impossible : le contour de leur apparence à l'image est variable et de taille plus réduite que les voitures, il n'est pas possible de les représenter par une surface plane à moins d'une certaine distance. De plus, leurs silhouettes respectent rarement une uniformité de colorimétrie et les trajectoires suivies ne peuvent être considérées comme contraintes.

Différentes voies sont testées de manière à caractériser les piétons malgré la difficulté de les modéliser : [Shashua et al., 2004] propose de segmenter la silhouette des piétons observés à partir d'une caméra en un nombre de régions différents suivant que le piéton est vu de face ou de profil. Les régions représentant le corps et les membres sont obtenues à partir d'une caméra calibrée. D'autres comme [Broggi et al., 2000] exploitent l'amplitude du gradient de l'image de disparité entre deux poses et des contraintes de symétrie. Certains comme [Tian et al., 2004] utilisent des capteurs travaillant dans le domaine de l'infra-rouge pour détecter des piétons de nuit comme de jour. Enfin, la fusion des données de capteur de vision et radar est en voie d'expérimentation chez les constructeurs automobiles : un exemple des récentes avancées est proposé dans [Gavrila et al., 2004].

De nombreuses techniques de détection de piétons ont été développées ces dernières années dans le cadre de la vidéo-surveillance. Les résultats obtenus par [Zhao and Thorpe, 2000] en utilisant une tête stéréo embarquée sont de qualité comparables à ceux fournis par les méthodes dédiées à la vidéo-surveillance : deux piétons marchant l'un à côté de l'autre sont correctement distingués. La méthode exploite l'image de disparité calculée à partir des images d'une paire stéréo pour segmenter des régions de l'image qui présentent des discontinuités. Seules les régions dont les dimensions et les tailles correspondent à celles de piétons sont conservées. Un réseau de neurones détecte alors les piétons à partir de critères selon les gradients

De manière plus anecdotique, [Reisman et al., 2004] parvient à localiser des mouvements de foule en Temps Réel en analysant le contenu spatio-temporel de 5 lignes horizontales équidistantes et fixes dans l'image. Les conditions de détection sont assez restrictives : les piétons traversent la route sur un passage réservé. L'idée est de détecter dans l'image des régions qui ne vérifient pas le flot optique induit par le déplacement de la caméra. Contrairement à ceux des obstacles mobiles, les mouvements de foule sont caractérisés par leurs apparences erratiques car une partie des piétons se croisent ou ne marchent pas à la même allure. L'auteur exploite ainsi ces particularités pour tracer les histogrammes de mouvements apparents de chacune des 5 lignes test. Pour cela des tranches d'images sont créées en superposant les niveaux de gris de chaque ligne au cours des 16 dernières images. Une transformée de Hough est alors appliquée à ces tranches d'images pour détecter les abscisses des points qui présentent les plus forts déplacements latéraux.

Enfin, nous relevons que très peu de méthode ([Labayrade and Aubert, 2003; Heisele, 2000] toutes deux à partir de stéréo-vision) permettent la détection de deux roues. A mi-chemin entre les caractéristiques des véhicules et des piétons, ces véhicules circulent sur la route mais ont des dimensions relativement réduites. Or, l'essor qu'a eu ces dernières années ce type de transport mérite certainement que la communauté s'intéresse de près à la détection de ces petits véhicules qui parviennent à se faufiler entre deux voies de circulation.

2.4 Le cadre applicatif de la thèse

Comparées aux tâches précédemment détaillées, la localisation et la navigation d'un robot mobile dans le contexte délicat de l'environnement urbain sont des tâches beaucoup plus difficiles qui réclament des hypothèses simplificatrices. Des conjectures sur la scène sont effectuées

en supposant par exemple la présence d'éléments structurants (éventuellement partiellement cachés) ou encore que le trajet ait été enregistré lors d'une phase d'apprentissage, et que des amers aient été détectés puis référencés. Le géo-référencement n'étant pas exploitable par l'emploi de systèmes satellitaires, les auteurs s'accordent sur l'emploi d'un MNT ou d'un Système d'Information Géographique (SIG) afin de faciliter la tâche de navigation.

2.4.1 La structuration de l'environnement urbain

Les articles de [Franke et al., 1999] et [Bertozzi et al., 2002], respectivement de 1999 et 2002, détaillent les prémisses des futurs systèmes visuels ADAS (Advanced Driving Autonomous System) qui commencent à équiper nos véhicules actuellement. Ils se placent tous deux dans le contexte particulier des trajets urbains réputés pour leur difficulté. Le premier énumère les avancées du laboratoire de Mercedes-Benz pour les applications typiquement urbaines comme le suivi de route, la détection d'obstacles, de piétons, la reconnaissance de flèches de circulation peintes sur la route, la détection et l'interprétation des panneaux de signalisation et des feux rouges, le tout modélisé en réalité virtuelle. Le second compare les méthodes, hypothèses et résultats obtenus pour satisfaire les tâches caractéristiques de la conduite en milieu urbain (suivi de route, détection d'obstacles). Un nouveau point est cependant abordé : l'intégration des architectures sur des plate-formes ad-hoc de façon à respecter les contraintes temporelles malgré le volume conséquent des données à traiter.

L'instrumentation des voies de circulation pour faciliter le déplacement de robots mobiles en ville n'est pas envisageable pour des questions d'échelles, matérielles ou plus simplement financières. La complexité des réseaux urbains et la nécessité de s'affranchir de l'introduction de nouveaux équipements favorisent le développement de solutions basées sur des amers naturels, tout du moins pour les applications spécifiquement urbaines.

De par son architecture, l'environnement urbain présente une structuration que les auteurs s'efforcent d'intégrer de façon à discriminer de manière robuste et rapide les parties dynamique et statique de la scène. La structuration de la scène selon l'axe vertical est généralement riche et s'avère de surcroît rarement occultée par la présence d'obstacles de hauteur réduite en comparaison des structures visibles.

Quoique apparemment un peu éloigné dans les objectifs du travail réalisé dans cette thèse, l'article de [Snaith et al., 1997] en est proche par les problèmes posés et les solutions mises en oeuvre. L'objectif est en effet de faciliter le déplacement des personnes mal-voyantes en développant une canne intelligente pour détecter les pas de portes et faciliter leurs franchissements. Une caméra fixée sur la canne constitue l'unique capteur pour appréhender l'environnement urbain, structuré à l'aide de contours essentiellement orientés selon les trois axes principaux. La reconnaissance des pas de portes se fait en deux temps : d'abord une extraction de contours quasi-verticaux puis une recherche localisée de segments quasi-horizontaux qui présentent des intersections avec les premiers. Le franchissement des pas de portes se fait quant à lui par la recherche du point de fuite dominant, l'hypothèse étant faite que les portes sont à l'extrémité de couloirs rectilignes.

La possibilité d'observer, même de manière partielle, des éléments texturés structurant l'environnement urbain a permis la transposition de techniques de localisation à partir de primitives linéaires, développées à l'origine pour les environnements d'intérieur. Les auteurs s'accordent en effet sur la robustesse et la précision des contours extraits en comparaison des diverses méthodes développées pour extraire des points d'intérêt. A partir de considérations sur la géométrie épipolaire et le suivi de contours verticaux [Gotoh et al., 1999] parvient à recons-

tituer un modèle de rue à partir de façades de bâtiments observées avec une caméra embarquée, non calibrée.

Dans le même esprit, [Se and Brady, 2003] s'intéresse à la détection d'éléments de scènes urbaines riches en caractéristiques géométriques comme les escaliers et les passages piétons pour les utiliser comme amers naturels lors d'une tâche de navigation avec un système de vision. Ces éléments, constitués de rectangles ou de parallépipèdes équidistants, induisent des contraintes géométriques fortes (parallélisme, orthogonalité et coplanarité) dans les images. Cet article élargit le spectre des éléments susceptibles de faciliter une reconstruction sommaire de l'environnement en améliorant l'estimation des points de fuite des contours détectés. Les différentes méthodes supposaient jusqu'alors de segmenter les contours alignés selon les trois directions principales (axe du couloir, sa transversale et la verticale).

Dans le cadre d'application d'insertion d'objets mobiles en réalité augmentée en milieu structuré, [P.-L. Bazin, 2001] réalise l'insertion de prototypes de véhicules en déplacement dans des environnements urbains. La méthode propose d'adapter la modélisation de l'environnement ainsi que celle de la trajectoire de la caméra en fonction de la structure de la scène observée et de l'application. Les relations entre objets sont classées selon trois familles :

- angulaire comme le parallélisme, l'orthogonalité et l'égalité angulaire,
- dimension soit la colinéarité et la coplanarité,
- métrique pour les distances et longueurs.

Les propriétés paramétriques telles que les angles et distances sont difficilement exploitables car leurs valeurs numériques sont inconnues à l'origine. C'est pourquoi les propriétés retenues préférentiellement dans l'algorithme sont celles qui ne font pas référence à des mesures, soit parallélisme, coplanarité, égalité de longueurs, intersection. A partir de l'indexation réalisée par un opérateur des différents éléments remarquables qui structurent la scène ainsi que de leurs relations géométriques, l'algorithme fournit un jeu de contraintes minimales en optimisant des graphes de relations entre les différents objets.

2.4.2 La localisation et la navigation en environnement urbain

Nous présentons dans cette dernière partie des travaux qui se placent dans une problématique comparable à la notre. L'objectif est de localiser un véhicule en ville en tentant de s'affranchir de la disparités des scènes et de leur caractère dynamique. Les méthodes peuvent être distinguées en deux familles : celles qui supposent une connaissance a priori de l'environnement et celles qui exploitent la fusion de données de capteurs de (géo-)référencement.

La connaissance a priori de l'environnement est fournie par un SIG sur lequel sont référencés les éléments structurants de la scène détectés par une caméra. Ainsi, [Chen, 1999] propose d'estimer les mouvements de rotation et translation d'une caméra dans un milieu urbain dense en réalisant la mise en correspondance de contours. La position initiale étant fournie par la fusion de données de capteurs inertiels, de SIG et de D-GPS lorsqu'elles sont disponibles.

[Kais et al., 2004] s'est, lui, focalisé sur la fusion des primitives verticales enregistrées dans un SIG avec celles issues des images d'une caméra calibrée embarquée. Le principe consiste à générer automatiquement des régions d'intérêt autour de primitives utilisées par le recalage, pour estimer le déplacement et l'orientation d'une caméra virtuelle. Cette tâche est facilitée par la propagation des erreurs de la localisation de chaque primitive en considérant les incertitudes liées aux différentes transformations qui lient les coordonnées 3D à leurs projections dans l'image. L'estimation du déplacement est obtenue lorsque les primitives de la caméra virtuelle sont superposables à celles observées par la caméra.

L'environnement urbain s'oppose généralement à l'utilisation de capteurs de géo-référencement. Le contournement de cet handicap consiste alors à référencer une collection d'images-clés pour lesquelles la pose et l'orientation de la caméra ont été préalablement référencées lors d'une phase d'apprentissage. L'apport de la méthode proposée par [Katsura et al., 2003] consiste à segmenter dans chaque image quatre types de régions : région uniforme et/ou ciel, constructions, arborescente et/ou éventuellement arborescente. L'idée étant de traiter différemment les régions de la scène susceptibles de subir des variations au cours du temps (ciel, feuillage) ou l'obstruction par la présence d'obstacles (régions uniformes).

Toujours à partir d'images-clés d'une séquence d'apprentissage, [Royer et al., 2004] exploite, lui, les projections de primitives dont les coordonnées 3D ont été estimées puis affinées suite à un lourd calcul (quelques heures) d'alignement de faisceaux à partir de triplets d'images. La sélection des images clés retenue est telle qu'il existe un nombre de projections en correspondance minimum (300) entre deux images. Lors de nouveaux déplacements, la position courante de la caméra est calculée en Temps Réel en estimant le déplacement rigide qui permet de transformer l'image courante à l'image de référence la plus proche.

Notons néanmoins que la distinction entre les deux familles de méthodes n'est pas évidente puisque à mi-chemin entre les méthodes de [Chen, 1999] et [Katsura et al., 2003], [Georgiev and Allen, 2004] a développé une architecture multi-capteurs qui permet la navigation d'un robot mobile sur plusieurs centaines de mètres en ville par fusion de données odométriques, inertielles, GPS et visuelles. Dès que la qualité des informations de géo-référencement se dégrade, la caméra motorisée réalise un asservissement visuel entre les façades visibles et une collection d'images géo-référencées au préalable.

2.5 Conclusion

Au cours de ce chapitre, nous avons dressé un état de l'art des systèmes de localisation fondés sur la vision. La localisation en environnement d'intérieur exploite le haut niveau de structuration des environnements de types couloirs et souvent une connaissance a priori de l'environnement pour réaliser une localisation précise et robuste de robots mobiles. L'intégration de capteur de vision a d'une manière générale permis d'alléger la configuration matérielle dédiée à la tâche de localisation et a ouvert de nouvelles perspectives telle que l'accomplissement de tâches.

Les systèmes de localisation en environnement d'extérieur étaient jusqu'à peu limités. L'apparition des systèmes de géo-référencement a suscité un regain d'intérêt en permettant de s'affranchir des problèmes d'échelles. La vision est alors employée comme un capteur susceptible d'appréhender la complexité et le caractère dynamique des scènes.

La localisation d'un véhicule en environnement urbain est un problème encore ouvert. La précision nécessaire à la localisation de robots mobiles autonomes est estimée dans ce cadre d'application particulier à quelques dizaines de centimètres. La transposition des techniques actuelles de localisation et de navigation des véhicules sur routes péri-urbaines et autoroutes exploitent les informations de systèmes de géo-référencement par satellites qui ne sont pas accessibles de manière continue en milieu urbain.

La plupart des auteurs qui exploitent les données fournies par un système de géo-référencement ne précisent pas quelles sont la couverture, la précision et la continuité des informations de position enregistrées durant leurs expérimentations. L'annonce de Chen qui affirme que 95% de la surface de Tokyo ne peut être localisée par le système GPS révèle une troublante contradiction. La plupart des méthodes de localisation en environnement urbain exploitent les informations

fournies par un système de géo-référencement alors que le taux de rafraîchissement et la précision des informations réceptionnées dépendent de la structuration de la scène urbaine.

Ce constat est à l'origine de ce travail de thèse. Notre étude se place dans un contexte différent des méthodes que nous venons de présenter puisque notre objectif est de localiser un véhicule qui se déplace en milieu urbain à partir des seules informations extraites d'un couple stéréo embarqué. A partir d'hypothèses simplificatrices sur l'environnement et le mouvement du véhicule, nous montrerons qu'il est possible d'estimer le déplacement du véhicule entre deux acquisitions en s'appuyant sur les différentes surfaces planaires susceptibles d'être détectées.

Chapitre 3

Outils mathématiques de la vision

Ce chapitre de rappels est largement inspiré des trois références suivantes [Faugeras, 1993; Hartley and Zisserman, 2003; Horaud, 1998]. Nous présentons dans ce chapitre les fondements de la géométrie projective qui sont indispensables à la compréhension des méthodes mises en oeuvre au cours de cette étude.

3.1 Géométrie de la vision

3.1.1 L'espace projectif \mathcal{P}^n

L'espace projectif \mathcal{P}^n est constitué de l'espace \mathbb{R}^{n+1} privé du vecteur $[0, \dots, 0]^t$. Deux points \mathbf{m} et \mathbf{m}' sont liés par une relation d'équivalence si il existe $\lambda \neq 0$ tel que :

$$[m_1, \dots, m_{n+1}] = \lambda \cdot [m'_1, \dots, m'_{n+1}]$$

Les deux points représentent alors le même point de l'espace \mathcal{P}^n .

Les coordonnées projectives (homogènes) d'un point \mathbf{m} de l'espace projectif \mathcal{P}^n s'expriment à partir d'une combinaison linéaire de $(n+1)$ des $(n+2)$ vecteurs \mathbf{e}_k qui forment la base canonique \mathcal{B}_0 . Ces vecteurs sont tels qu'aucun d'entre eux n'appartient à un hyperplan de \mathcal{P}^n :

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{e}_{n+1} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad \mathbf{e}_{n+2} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\mathbf{m} = \sum_{k=1}^{k=n+1} m_k \cdot \mathbf{e}_k$$

Une nouvelle base \mathcal{B} de \mathcal{P}^n sera constituée de $(n+2)$ points \mathbf{m}_k avec $k \in [1, n+2]$ tels qu'aucun sous-ensemble de $(n+1)$ points n'appartient à un hyperplan. Il existe alors une transformation linéaire \mathbf{H} de \mathcal{P}^n dans \mathcal{P}^n , appelée homographie ou collinéation, qui lie les vecteurs de la base \mathcal{B} à ceux de la base canonique \mathcal{B}_0 :

$$\mathbf{m}_k = \lambda_k \mathbf{H} \cdot \mathbf{e}_k \quad \text{avec} \quad \lambda_k \neq 0$$

L'homographie est représentée par un ensemble de matrices $\mathbf{H}[n+1, n+1]$ régulières (inversibles) qui se distinguent par un facteur d'échelle (λ). Chacune des $(n+1)$ colonnes de la matrice

H représentent les coordonnées des vecteurs de \mathcal{B} exprimés dans \mathcal{B}_0 . Le vecteur nul étant proscrié de \mathcal{P}^n , chaque colonne est définie à un coefficient près, ce qui induit n contraintes linéaires indépendantes. La matrice d'homographie comprend donc $(n+1)^2 - 1$ degrés de libertés. Soit, il est nécessaire de connaître l'égalité projective entre $(n+2)$ points pour déterminer entièrement H .

3.1.2 Les transformations dans \mathcal{P}^n

Les transformations projectives de \mathbb{R}^n dans \mathbb{R}^n associées au produit matriciel forme le groupe projectif (GL^n). Il est possible de rassembler en sous-groupes des transformations ayant des propriétés identiques. Ainsi, les transformations affines forment le sous-groupe des homographies particulières dont la dernière ligne est de la forme $[0, \dots, 0, 1]$. Les transformations euclidiennes représentent un sous groupe des applications affines dont la sous-matrice supérieure gauche $[n, n]$ est orthogonale, multipliée par un facteur d'échelle s .

Il est d'usage de distinguer algébriquement les divers sous-groupes des transformations en identifiant leurs invariants. Ce sont alors soit des éléments (points, droites, plans), soit des quantités (longueurs, angles, aires, volumes). Nous nous plaçons désormais dans le cas particulier qui nous concerne plus précisément où $n \leq 3$.

3.1.2.1 Les transformations projectives

Les transformations projectives sont des applications linéaires de \mathbb{R}^n dans \mathbb{R}^n qui à tout point de l'espace projectif \mathbf{M} associe le point \mathbf{m} selon l'équation suivante et où \simeq représente une égalité projective, c-à-d à un facteur près :

$$\mathbf{m} \simeq P_{\text{proj}} \cdot \mathbf{M} \quad (3.1)$$

Ainsi, seuls les rapports entre les éléments qui forment la matrice P_{proj} sont significatifs. Les invariants de telles transformations sont les positions relatives des objets géométriques les uns par rapport aux autres ainsi que le bi-rapport de quatre points collinéaires (développé en 3.1.3.4).

3.1.2.2 Les transformations affines ou affinités

Les transformations affines P_{aff} sont des applications linéaires qui font correspondre à tout point \mathbf{M} de \mathcal{P}^n un point \mathbf{m} de \mathcal{P}^n , telles que les coordonnées non-homogènes respectives vérifient l'équation :

$$\bar{\mathbf{m}} = A \cdot \bar{\mathbf{M}} + \mathbf{B} \quad (3.2)$$

où A est une matrice $[n, n]$ de rang n et \mathbf{B} vecteur colonne de dimension n . L'expression précédente se réécrit sous forme plus réduite en introduisant les coordonnées homogènes des deux points :

$$P_{\text{aff}} \simeq \begin{bmatrix} A & \mathbf{B} \\ \mathbf{0}^t & 1 \end{bmatrix} \quad (3.3)$$

où $\mathbf{0}^t$ représente un vecteur ligne nul de n colonnes. Le rang de la matrice ainsi formée est $(n+1)$ car le noyau de l'application est vide par définition.

Une transformation affine peut être décomposée à partir de mouvements élémentaires :

$$A = [\mathbf{R}_1]_x \cdot [\mathbf{R}_2]_{\times}^{-1} \cdot \text{Di} \cdot [\mathbf{R}_2]_{\times} \quad (3.4)$$

où $[\mathbf{R}_1]_\times$ et $[\mathbf{R}_2]_\times$ représentent deux rotations selon les axes \mathbf{R}_1 et \mathbf{R}_2 , d'angles la norme des vecteurs ainsi formés alors que \mathbf{D}_i représente une matrice diagonale de facteurs d'échelle non-isotropique. Ainsi, seules les propriétés induites du parallélisme entre droites sont conservées par transformations affines. C'est le cas du rapport entre distances de droites parallèles ainsi que le rapport entre aires, délimitées par des droites parallèles, qui est multiplié par le produit des facteurs de \mathbf{D}_i .

3.1.2.3 Les similitudes

Les similitudes \mathbf{P}_{sim} sont des transformations affines particulières de \mathcal{P}^n . En effet, la sous-matrice supérieure gauche \mathbf{A} correspond à une matrice de rotation orthogonale $\mathbf{R} [n, n]$ dont le déterminant est égal à s . Le vecteur \mathbf{B} est alors assimilé à un vecteur de translation $\overline{\mathbf{T}}$.

$$\mathbf{P}_{\text{sim}} \simeq \begin{bmatrix} s\mathbf{R}_{n,n} & \overline{\mathbf{T}} \\ \mathbf{0}^t & 1 \end{bmatrix} \quad (3.5)$$

Une similitude peut se décomposer en une translation de l'origine du repère suivie d'une rotation du repère de référence couplée à une homothétie de rapport s . Ainsi, les angles et par conséquent les formes sont conservés, par contre les aires et les volumes sont respectivement multipliés par s^2 et s^3 .

3.1.2.4 Les isométries

Les isométries sont des similitudes dont le rapport d'homothétie s est unitaire. Les isométries appartiennent avec les similitudes aux transformations euclidiennes, décomposables en des mouvements composés d'une translation et d'une rotation. De telles transformations sont qualifiées de "métriques". Les angles, distances, surfaces et volumes sont conservés. Il est d'usage d'appeler les homographies obtenues par isométries des déplacements pour les espaces projectifs \mathcal{P}^n lorsque $n \in [2, 3]$:

$$\mathbf{P}_{\text{iso}} = \begin{bmatrix} \mathbf{R}_{n,n} & \overline{\mathbf{T}} \\ \mathbf{0}^t & 1 \end{bmatrix}$$

3.1.3 Particularités des espaces projectifs

Tout point $\mathbf{M} = [X, Y, Z, T]^t$ de l'espace projectif \mathcal{P}^3 appartient au plan Π , défini par sa normale \mathbf{N} , à condition que $\mathbf{M}^t \cdot \mathbf{N} = 0$. Soit Π un plan de l'espace projectif \mathcal{P}^3 . Un point \mathbf{M} appartient à Π à condition que :

$$\Pi^t \cdot \mathbf{M} = 0 \quad (3.6)$$

Considérons le changement de base \mathbf{P} tel que $\mathbf{M}' \simeq \mathbf{P} \cdot \mathbf{M}$. L'expression du plan Π' dans la nouvelle base s'écrit $\Pi'^t \cdot \mathbf{M}' = 0$, nous en déduisons ainsi :

$$\Pi' \simeq \mathbf{P}^{-t} \cdot \Pi \quad (3.7)$$

Par conséquent, si \mathbf{P} représente la transformation entre deux points de l'espace projectif \mathcal{P}^3 , alors la transformation \mathbf{P}^{-t} relie les plans duaux de ces points.

La même démonstration s'applique aux points \mathbf{M} de \mathcal{P}^2 . La transformation \mathbf{P}^{-t} représente la liaison entre les droites \mathbf{q} et \mathbf{q}' .

3.1.3.1 Le plan à l'infini

Le plan à l'infini est le plan d'équation $T = 0$, ses coordonnées, exprimées dans le repère euclidien, sont donc : $\Pi_\infty = [0, 0, 0, 1]^t$. Il s'agit bien des coordonnées du 4^{eme} vecteur de la base canonique de l'espace projectif. Ce plan contient toutes les directions de la forme $\mathbf{D} = [X, Y, Z, 0]^t$ et l'intersection de plans et/ou droites parallèles.

Le plan à l'infini est une représentation géométrique qui a été introduit de façon à exprimer les spécificités des projections affines dans le repère euclidien. Le plan à l'infini est en effet invariant par transformation affine mais pas par transformation projective. Cette différence entre les deux types de projections justifie que le plan à l'infini comprend les 3 degrés de liberté qui complètent les 12 degrés de liberté d'une transformation affine pour retrouver les 15 degrés de liberté d'une transformation exprimée dans l'espace projectif.

Soit \mathbf{H} , une transformation affine. Le plan à l'infini Π_∞ est donc solution de l'équation $\mathbf{H}^{-t} \cdot \mathbf{\Pi} = \mathbf{\Pi}$. La détermination des plans invariants par une transformation \mathbf{H}^{-t} est obtenue en calculant les vecteurs propres associés aux valeurs propres λ de la matrice \mathbf{H}^{-t} .

3.1.3.2 La conique absolue

La conique absolue Ω correspond à l'intersection de la quadrique d'équation $X^2 + Y^2 + Z^2 + T^2 = 0$ avec le plan à l'infini Π_∞ . Les coordonnées de la conique absolue, exprimées dans le repère euclidien, sont solutions du système suivant :

$$\begin{cases} X^2 + Y^2 + Z^2 + T^2 = 0 \\ T^2 = 0 \end{cases} \quad (3.8)$$

La conique absolue Ω est invariante par similitude. Soit \mathbf{H} une transformation affine, la transformée de la conique absolue par une telle transformation est $\mathbf{A}^{-t} \cdot \mathbf{l}_3 \cdot \mathbf{A}^{-1} \simeq \mathbf{l}_3$, ce qui mène à $\mathbf{A} \cdot \mathbf{A}^t \simeq \mathbf{l}_3$. \mathbf{A} doit par conséquent être une matrice orthogonale, soit une matrice de rotation à un facteur près, ce qui prouve l'invariance de la conique absolue par similitude.

La conique absolue Ω est une représentation géométrique des 5 derniers degrés de liberté nécessaires à l'expression des propriétés de l'espace métrique, à partir d'un repère affine. En effet, dès que Ω et son support Π_∞ sont identifiés, il est possible de mesurer des angles et des longueurs relatives.

Deux directions \mathbf{D}_1 et \mathbf{D}_2 se projettent sur le plan à l'infini en \mathbf{d}_1 et \mathbf{d}_2 . L'angle formé par ces deux directions exprimé dans un repère euclidien est :

$$\cos \theta = \frac{\mathbf{d}_1^t \cdot \mathbf{d}_2}{\sqrt{\mathbf{d}_1^t \cdot \mathbf{d}_1 \cdot \mathbf{d}_2^t \cdot \mathbf{d}_2}} = \frac{\mathbf{d}_1^t \cdot \Omega \cdot \mathbf{d}_2}{\sqrt{\mathbf{d}_1^t \cdot \Omega \cdot \mathbf{d}_1 \cdot \mathbf{d}_2^t \cdot \Omega \cdot \mathbf{d}_2}} \quad (3.9)$$

De même, l'orthogonalité de deux directions s'exprime en fonction de la conique absolue :

$$\mathbf{d}_1^t \cdot \mathbf{d}_2 = \mathbf{d}_1^t \cdot \Omega \cdot \mathbf{d}_2 = 0 \quad (3.10)$$

Il est à noter que cette égalité est toujours vérifiée quel que soit le type de repère employé : seule la matrice associée à la conique absolue sera différente de l'identité si le repère n'est pas orthonormé.

3.1.3.3 Le point de fuite et les lignes de fuite

La projection perspective a la propriété d'afficher dans le plan de l'image tous les objets d'une scène sans distinction de la distance qui les sépare du foyer de la caméra. Parmi ceux-là, il est particulièrement intéressant d'identifier à l'image ceux qui s'étendent jusqu'à l'infini dans l'espace projectif. Ainsi, la projection d'une ligne infinie est délimitée dans le plan image par un point de fuite (Vanishing Point) et l'intersection de la ligne avec le plan image.

D'un point de vue de géométrie projective, le plan à l'infini Π_∞ (4^{ème} coordonnée nulle) est le plan des directions. Deux droites qui partagent la même direction $\mathbf{D} = [\overline{\mathbf{D}}^t, 0]^t$ s'y interceptent au même point \mathbf{X}_D . Le VP \mathbf{x}_D apparaît comme la projection du point \mathbf{X}_D selon la transformation projective $P=K[l, \mathbf{0}]$:

$$\mathbf{x}_D = P.\mathbf{X}_D = K[l, \mathbf{0}] \cdot \begin{vmatrix} \overline{\mathbf{D}} \\ 0 \end{vmatrix} = K.\overline{\mathbf{D}} \quad (3.11)$$

Les coordonnées du VP \mathbf{x}_D dépendent uniquement de la direction des droites auxquelles il est associé. Parmi toutes les droites parallèles ayant la même direction \mathbf{D} , le rayon qui passe par le foyer de la caméra et \mathbf{X}_D est remarquable puisque son intersection avec le plan de l'image forme le VP \mathbf{x}_D (Fig. 3.1). Toutes les projections des droites ayant la même direction convergent donc vers ce point. Si la projection perspective de deux droites parallèles est représentée par deux droites parallèles alors les droites appartiennent alors à des plans fronto-parallèle du plan image. Le point d'intersection (VP) est alors repoussé à l'infini.

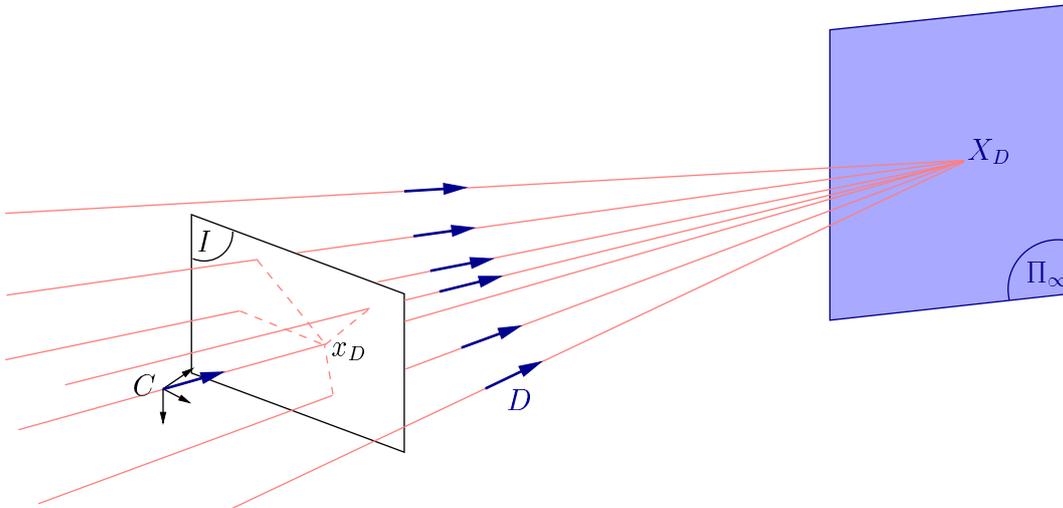


FIG. 3.1 – Toutes les droites ayant la même orientation D convergent vers le même point X_D du plan à l'infini Π_∞ . Parmi l'ensemble de ces droites, celle qui passe par le foyer C de la caméra est remarquable car son intersection du plan de l'image I forme le VP x_D de la direction considérée.

Le VP \mathbf{x}_D connaît aussi une définition algébrique. Une droite de l'espace projectif peut être caractérisée à partir d'un couple formé d'un point \mathbf{A} et d'une direction $\mathbf{D} = [\overline{\mathbf{D}}^t, 0]^t$. Tout point \mathbf{M} de cette droite est alors défini par un paramètre $\delta \in [0, +\infty[$ tel que :

$$\mathbf{M}(\delta) = \mathbf{A} + \delta.\mathbf{D} \quad (3.12)$$

La projection du point \mathbf{M} dans le plan image est alors obtenue grâce à la matrice de projection $P=K[l, \mathbf{0}]$:

$$\mathbf{m}(\delta) = P.\mathbf{A} + \delta.P.\mathbf{D} = \mathbf{a} + \delta.K.\overline{\mathbf{D}} \quad (3.13)$$

avec \mathbf{a} et $\overline{\mathbf{D}}$ projections respectives du point \mathbf{A} et du vecteur \mathbf{D} . Le VP \mathbf{x}_D représente donc la limite à l'infini de l'expression précédente :

$$\mathbf{x}_D = \lim_{\delta \rightarrow +\infty} \mathbf{m}(\delta) = \lim_{\delta \rightarrow +\infty} (\mathbf{a} + \delta \cdot \mathbf{K} \cdot \overline{\mathbf{D}}) = \mathbf{K} \cdot \overline{\mathbf{D}} \quad (3.14)$$

Nous retrouvons les coordonnées d'un rayon et vérifions ainsi que les coordonnées du VP ne dépendent que de la direction \mathbf{D} du jeu de droites parallèles et en aucun cas des coordonnées des points \mathbf{A} qui appartiennent à ces droites.

Les plans parallèles de l'espace projectif s'interceptent sur le plan à l'infini Π_∞ le long d'une ligne commune dont la projection sur le plan image forme une ligne de fuite (*Vanishing Line*). Ainsi par construction, une VL correspond à l'intersection du plan image avec le plan parallèle au plan considéré et passant par le foyer \mathbf{C} de la caméra. La VL dépend seulement de la direction de la normale au plan \mathbf{N} considéré.

L'intersection de plans régulièrement espacés, orthogonaux à un même plan Π forment dans l'espace projectif un jeu de droites parallèles. La projection sur le plan image de ce jeu de droites est caractéristique. En effet, le jeu de droites projetées demeure parallèle mais la distance qui les séparent diminue de façon monotone jusqu'à ce que l'une d'entre elles interceptent le VP, il s'agit de la VL (Fig. 3.2). De manière générale, les VLs sont définies dans le plan image comme les droites passant par deux VPs.

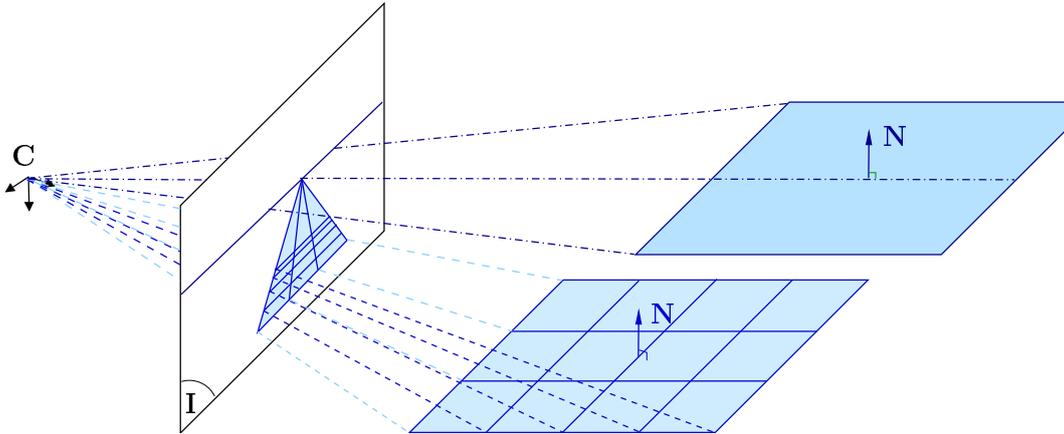


FIG. 3.2 – La ligne de fuite correspond à la projection sur le plan image de la droite d'interception de plusieurs plans parallèles avec le plan à l'infini Π_∞ . La VL est par construction l'interception du plan, perpendiculaire à la normale N considérée, qui passe par le foyer C de la caméra et intercepte le plan de l'image.

3.1.3.4 Le bi-rapport

Le bi-rapport est le second invariant de la transformation projective après l'ordonnement qui signifie que l'ordre d'apparition de primitives est respecté. Il existe plusieurs définitions équivalentes du bi-rapport : elles se distinguent seulement par le type de primitives employées. Pour notre cas, nous exploitons la définition issue de la projection d'un faisceau de droites car il est admis que la précision sur les coordonnées des primitives requise n'est généralement pas fournie par les détecteurs de points d'intérêt.

Le bi-rapport d'un faisceau de quatre droites est calculé grâce aux écarts angulaires signés α_i avec $i \in [1, 2, 3, 4]$ entre les orientations θ_i des droites tel qu'il est défini sur la Fig. 3.3(a) :

$$BR(l_a, l_b, l_c, l_d) = \frac{\sin(\alpha_1)}{\sin(\alpha_2)} / \frac{\sin(\alpha_3)}{\sin(\alpha_4)} \quad (3.15)$$

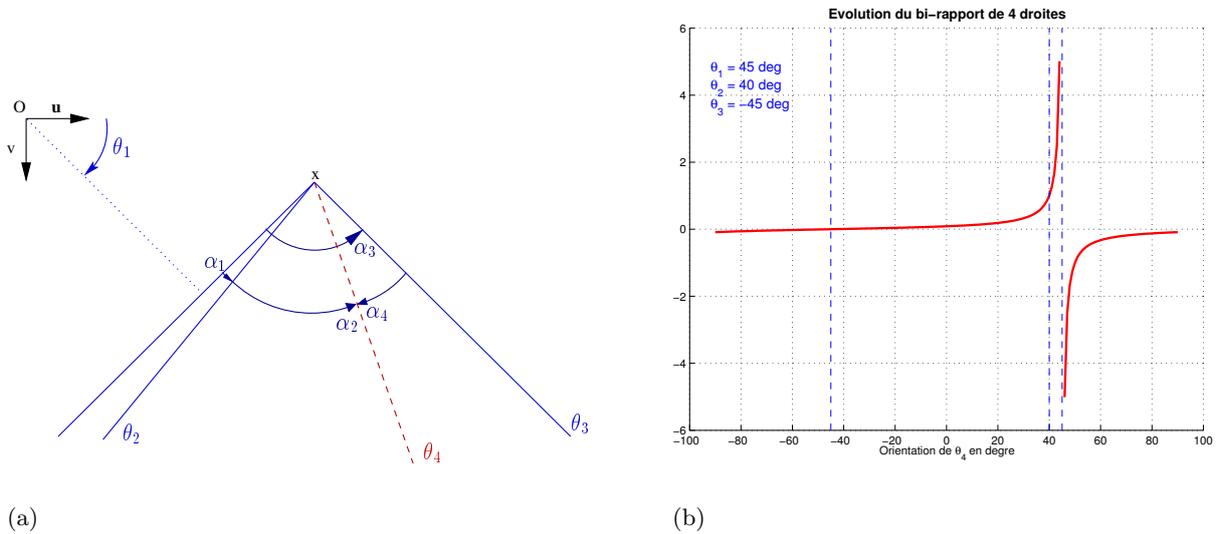


FIG. 3.3 – Le bi-rapport étant invariant par homographie, nous calculons le bi-rapport d'un faisceau de quatre droites dont l'orientation de trois d'entre elles sont fixées à -45° , -40° et 45° . La pente du bi-rapport évolue grandement en fonction de la position relative de la 4^{ème} droite en fonction des trois autres, ce qui justifie qu'un critère d'invariance fixe ne permettra pas forcément d'identifier deux représentations différentes d'un même faisceau de quatre droites.

Le bi-rapport évolue entre $] -\infty, +\infty[$ lorsqu'il est calculé à partir d'un faisceau de quatre VLs dont trois sont fixes en faisant varier l'orientation de la quatrième VL entre $[-90^\circ, 90^\circ]$. Le tracé de la valeur du bi-rapport (Fig. 3.3(b)) illustre que suivant la position relative de la quatrième droite en fonction des trois autres, la variation de la pente bi-rapport est comprise entre une valeur quasiment nulle ($[-90^\circ; 0^\circ]$) et une pente quasi-infinie (dans le voisinage de 45°). L'invariance du bi-rapport ne se justifie donc qu'en cas de disposition particulière des quatre droites : dans l'exemple proposé, elle sera vérifiée que si la 4^{ème} droite se situe avec une orientation dans le voisinage proche de -45° .

Le bi-rapport admet six résultats différents $\{\nu, \frac{1}{\nu}, 1 - \nu, \frac{1}{1-\nu}, \frac{\nu}{\nu-1}\}$ pour un même jeu de quatre primitives. La valeur obtenue dépend de l'ordre dans lequel sont appelées les quatre primitives.

3.2 Les modèles géométriques de caméras

3.2.1 Le modèle sténopé

La projection d'une scène sur le plan image nécessite un modèle conforme avec le formalisme de la géométrie projective. Le modèle sténopé ou projection centrale correspond au modèle de caméra le plus couramment utilisé. Il est utilisé depuis le *XVII^{ème}* siècle : il se compose d'un

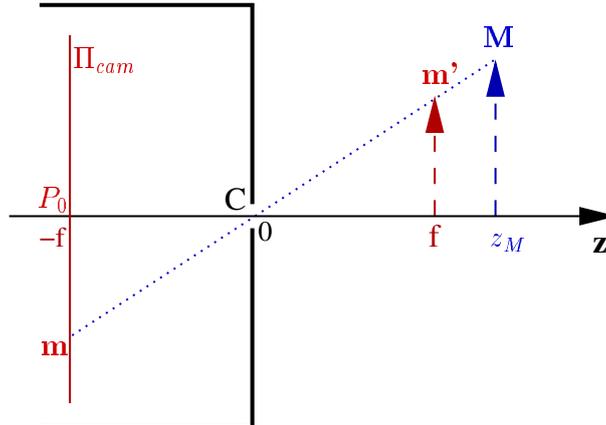


FIG. 3.4 – Projection m du point M sur le plan image selon un modèle de caméra sténopé.

plan image, appelé rétine, Π_{cam} et d'un centre de projection ou centre optique C . Tout point M de l'espace projectif admet une projection m sur le plan rétinien par l'interception de ce dernier avec le rayon optique associé CM , d'où la terminologie anglaise de modèle pinhole. Seul le centre optique C n'a pas de projection sur le plan image.

Le plan image est situé à une distance f , appelée focale, du centre de projection C selon l'axe optique, perpendiculaire au plan image passant par C . Le repère lié à la caméra \mathcal{R}_c est donc formé des axes horizontal, vertical du plan image et de l'axe optique.

Nous définissons alors C comme l'origine du repère caméra \mathcal{R}_c . La normale au plan image passant par C s'appelle l'axe principal ou axe optique. Cet axe intercepte le plan image au point principal de coordonnées $P_0 = [0, 0, f]^t$ où f est la focale. Le repère lié à la caméra \mathcal{R}_c est donc formé des axes horizontal X , vertical Y du plan image et de l'axe optique. Ce modèle donne lieu à une image inversée (voir Fig3.4) mais il est courant d'employer par commodité l'image virtuelle qui est, elle, non-inversée.

3.2.2 Paramètres intrinsèques et extrinsèques

La projection d'un point de l'espace sur le plan image d'une caméra nécessite trois transformations :

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

3.2.2.1 Le changement de repère

Quelle que soit la base orthonormée \mathcal{R}_0 de l'espace projectif, il est d'usage de considérer le repère \mathcal{R}_c dont l'origine se situe au centre de projection de la caméra et l'axe Z_0 corresponde à la normale au plan de l'image. Ce changement de repère se justifie par des simplifications notables des expressions des deux transformations suivantes.

Soit un point M de l'espace projectif associé au repère \mathcal{R}_0 . Les coordonnées de ce point, exprimées dans le repère \mathcal{R}_c sont alors obtenues en introduisant les paramètres extrinsèques de la caméra : Le changement de repère dans \mathcal{P}^3 est représenté par :

- la position représentée par la translation $\overline{\mathbf{T}}$, correspondant au vecteur $-\mathbf{OC}$, exprimé dans la repère \mathcal{R}_c ,
- l'orientation qui exprime la rotation du repère \mathcal{R}_0 vers \mathcal{R}_c . Elle est représentée par une matrice orthogonale $\mathbf{R}[3,3]$ de déterminant unitaire dont les colonnes sont les coordonnées des vecteurs $(\mathbf{X}_0, \mathbf{Y}_0, \mathbf{Z}_0)$ de \mathcal{R}_0 exprimées dans la nouvelle base \mathcal{R}_c .

Soient $\mathbf{M}_0 = [x_0, y_0, z_0]^t$ et $\mathbf{M}_c = [x_c, y_c, z_c]^t$, les coordonnées du point \mathbf{M} exprimées respectivement dans les bases \mathcal{R}_0 et \mathcal{R}_c . La transformation \mathbf{P}_b relative au changement de base est :

$$\begin{aligned} \mathbf{CM}_c &= \mathbf{R} \cdot (\mathbf{OM}_0 - \mathbf{OC}) \\ \overline{\mathbf{M}}_c &= \mathbf{R} \cdot (\overline{\mathbf{M}}_0 - \overline{\mathbf{C}}) \\ &= \mathbf{R} \cdot \overline{\mathbf{M}}_0 + \overline{\mathbf{T}} \end{aligned} \quad (3.16)$$

en introduisant $\overline{\mathbf{T}} = -\mathbf{R} \cdot \overline{\mathbf{C}}$ de façon à retrouver l'expression courante $\overline{\mathbf{M}}_c = \mathbf{R} \cdot \overline{\mathbf{M}}_0 + \overline{\mathbf{T}}$. La transformation équivalente à ce changement de repère est un déplacement (cf 3.1.2.4) tel que $\mathbf{M}_c = \mathbf{P}_D \cdot \mathbf{M}_0$:

$$\mathbf{P}_D = \begin{bmatrix} \mathbf{R} & \overline{\mathbf{T}} \\ \mathbf{0} & 1 \end{bmatrix} \quad (3.17)$$

3.2.2.2 La projection perspective

Nous employons désormais les coordonnées homogènes des points, vecteurs, droites et plans cités. Ainsi le point \mathbf{M} de l'espace projectif s'exprime dans le repère caméra \mathcal{R}_c :

$$\mathbf{M} = \begin{bmatrix} X_M \\ Y_M \\ Z_M \\ T_M \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{M}} \\ T_M \end{bmatrix} \text{ avec } T_M = \begin{cases} 0 & \text{si } \mathbf{M} \in \Pi_\infty \\ 1 & \text{autrement} \end{cases} \quad (3.18)$$

Les coordonnées de la projection $\tilde{\mathbf{m}}$ de ce point \mathbf{M} sur le plan image sont alors obtenues en interceptant le rayon $\mathbf{CM} = [\overline{\mathbf{M}}, 0]^t$ avec le plan rétinien $\Pi_{cam} = [0, 0, f, 0]^t$:

$$\tilde{\mathbf{m}} = \mathbf{CM} \cap \Pi_{cam} = \begin{bmatrix} X_M \\ Y_M \\ Z_M \\ 0 \end{bmatrix} \cap \begin{bmatrix} 0 \\ 0 \\ f \\ 0 \end{bmatrix} = \begin{bmatrix} X_M \cdot f / Z_M \\ Y_M \cdot f / Z_M \\ f \\ 1 \end{bmatrix} \quad (3.19)$$

Nous introduisons alors la matrice dite de projection centrale $\mathbf{P}_c[4,3]$ qui à tout point \mathbf{M} de l'espace projectif fait correspondre le point image $\tilde{\mathbf{m}}$ dans le plan rétinien, tous deux exprimés en coordonnées homogènes. Il s'agit d'une application de l'espace euclidien \mathbb{R}^3 dans l'espace euclidien \mathbb{R}^2 , la 3^{eme} coordonnée de $\tilde{\mathbf{m}}$ étant imposée à 1 :

$$\begin{bmatrix} X_M \\ Y_M \\ Z_M \\ T_M \end{bmatrix} \mapsto \begin{bmatrix} X_M \cdot f / Z_M \\ Y_M \cdot f / Z_M \\ 1 \end{bmatrix} = \frac{1}{Z_M} \cdot \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_M \\ Y_M \\ Z_M \\ T_M \end{bmatrix} \quad (3.20)$$

soit,

$$\tilde{\mathbf{m}} = \frac{1}{Z_M} \cdot \mathbf{P}_c \cdot \mathbf{M} \quad (3.21)$$

avec

$$\mathbf{P}_c = \mathbf{P}_f \cdot [I_3, \mathbf{0}] \quad \text{et} \quad \mathbf{P}_f = \text{diag}(f, f, 1) \quad (3.22)$$

Tous les points du rayon \mathbf{CM} se projetant sur le plan image en un même point, l'égalité précédente est vérifiée au facteur $1/Z_M$ près.

3.2.2.3 La transformation caméra-image

Les coordonnées des points projetés sur le plan rétinien selon le repère caméra \mathcal{R}_c sont désormais exprimées en pixel selon le repère lié à l'image \mathcal{R}_i . Soit $\mathbf{P}_{\text{cal}}[3,3]$, la matrice dite de calibration qui permet le passage du repère caméra au repère image :

$$\mathbf{m} = \mathbf{P}_{\text{cal}} \cdot \widetilde{\mathbf{m}} \quad (3.23)$$

$$\begin{vmatrix} u \\ v \\ 1 \end{vmatrix} = \begin{bmatrix} k_u & -k_u \cdot \cot \xi & u_0 \\ 0 & k_v \cdot \sin \xi & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{vmatrix} \widetilde{x}_m \\ \widetilde{y}_m \\ 1 \end{vmatrix}$$

où nous avons introduit les cinq paramètres suivant :

- k_u et k_v sont respectivement les facteurs d'échelle horizontale et verticale exprimés en pixel/mm,
- ξ représente l'angle exprimé en radian qui sépare les deux vecteurs \mathbf{u} et \mathbf{v} pour former la base normée $(\mathbf{i}, \mathbf{u}, \mathbf{v})$ du repère image \mathcal{R}_i avec \mathbf{i} choisi arbitrairement.
- $[u_0, v_0]^t$ sont les coordonnées du point principal \mathbf{P}_0 exprimées en pixel dans le repère image,

Ces paramètres sont qualifiés d'intrinsèques car ils sont liés aux caractéristiques de la caméra. La matrice \mathbf{P}_{cal} telle qu'elle est présentée est qualifiée de modèle de caméra projectif. Lorsque $\xi = \pi/2$, le modèle est qualifié d'anne. Les coordonnées pixeliques de la projection \mathbf{m} du point \mathbf{M} de l'espace projectif sur le plan image, lorsque l'origine du repère se trouve en \mathbf{C} , sont alors fournies par l'équation suivante :

$$\mathbf{m} = \frac{1}{Z_M} \cdot \mathbf{K} \cdot [I_3, \mathbf{0}] \cdot \mathbf{M} \quad (3.24)$$

avec

$$\mathbf{K} = \mathbf{P}_{\text{cal}} \cdot \mathbf{P}_f = \begin{bmatrix} f \cdot k_u & -f \cdot k_u \cdot \cot \xi & u_0 \\ 0 & f \cdot k_v \cdot \sin \xi & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Nous poserons par la suite $\alpha_u = f \cdot k_u$ et $\alpha_v = f \cdot k_v$ car la détermination de la valeur de la focale f n'a pas d'intérêt en soi.

3.2.2.4 La transformation perspective

La transformation perspective réunit sous la même appellation les trois transformations détaillées précédemment. Les coordonnées \mathbf{m} de la projection d'un point de l'espace projectif \mathbf{M} sur le plan de la caméra Π_{cam} sont liées par $\mathbf{m} \simeq \mathbf{P}_{\text{pers}} \cdot \mathbf{M}$ avec :

$$\begin{aligned} \mathbf{P}_{\text{pers}} &= \mathbf{P}_{\text{cal}} \cdot \mathbf{P}_c \cdot \mathbf{P}_D \\ &= \mathbf{K} \cdot [\mathbf{R}, \overline{\mathbf{T}}] \end{aligned} \quad (3.25)$$

[Faugeras, 1993] a donné une interprétation particulièrement intéressante de la matrice de projection P_{pers} . Elle est vue comme une matrice de changement de repère entre le repère euclidien \mathcal{R}_0 et le repère caméra \mathcal{R}_c . Les trois premières colonnes de P_{pers} représentent alors les coordonnées des trois VPs $\mathbf{x}_{\mathbf{X}_0}, \mathbf{x}_{\mathbf{Y}_0}, \mathbf{x}_{\mathbf{Z}_0}$, associés aux trois directions du repère euclidien. La quatrième colonne correspond aux coordonnées dans l'image \mathbf{o} de la projection du centre \mathbf{O} du repère \mathcal{R}_0 .

$$P_{\text{pers}} = [\mathbf{x}_{\mathbf{X}_0}, \mathbf{x}_{\mathbf{Y}_0}, \mathbf{x}_{\mathbf{Z}_0}, \mathbf{o}] \quad (3.26)$$

3.2.3 Géométrie d'une paire stéréo

Nous considérons le cas où deux prises de vues \mathcal{I}_1 et \mathcal{I}_2 différentes de la même scène sont disponibles. Deux possibilités sont à envisager pour obtenir un tel résultat : soit une paire de caméras stéréo réalise une prise de vues instantanée, soit une caméra mobile photographie à deux instants différents une même scène considérée comme statique durant la durée du déplacement. Ces deux situations sont géométriquement équivalentes.

En posant $\gamma \in [1, 2]$, indice indifférencié lié aux caractéristiques des images \mathcal{I}_1 et \mathcal{I}_2 , nous pouvons vérifier que l'emploi d'une paire de caméras stéréo permet d'établir le cadre géométrique indispensable pour successivement résoudre les problèmes de :

- correspondance géométrique : quelles sont les contraintes sur les coordonnées des projections \mathbf{m}_2 de points \mathbf{M} dans l'image \mathcal{I}_2 connaissant les coordonnées des projections \mathbf{m}_1 dans l'image \mathcal{I}_1 ?
- déplacement : quelles sont les matrices de projection P_γ lorsque sont connues les projections \mathbf{m}_γ de plusieurs points \mathbf{M} de l'espace projectif dans les deux images \mathcal{I}_γ ?
- structure : quelles sont les coordonnées des points \mathbf{M} de l'espace projectif lorsque nous disposons de leurs projections \mathbf{m}_γ dans les images \mathcal{I}_γ et des matrices de projection P_γ ?

Chacune des deux caméras est caractérisée par un centre de projection \mathbf{C}_γ , un repère caméra $\mathcal{R}_{c\gamma}$ ($\mathbf{C}_\gamma, \mathbf{X}_\gamma, \mathbf{Y}_\gamma, \mathbf{Z}_\gamma$) et une matrice de projection P_γ . Nous considérons que les vecteurs des deux bases forment des repères orthonormés. Tout point \mathbf{M} de l'espace projectif, privé de \mathbf{C}_γ , se projette dans l'image \mathcal{I}_γ selon l'équation projective $\mathbf{m}_\gamma \simeq P_\gamma \cdot \mathbf{M}$ avec :

$$P_\gamma = [\tilde{P}_\gamma, \mathbf{p}_\gamma] \quad (3.27)$$

où P_γ représente la projection du centre du repère \mathcal{R}_γ dans l'image \mathcal{I}_γ .

La projection perspective n'étant pas définie pour le centre optique \mathbf{C}_γ de chaque caméra, nous en déduisons $P_\gamma \cdot \mathbf{C}_\gamma = \mathbf{0}$. Nous obtenons alors les coordonnées des centres de projection :

$$\overline{\mathbf{C}_\gamma} = -\tilde{P}_\gamma^{-1} \cdot \mathbf{p}_\gamma \quad (3.28)$$

L'épipôle \mathbf{e}_2 (resp. \mathbf{e}_1) correspond à la projection dans l'image \mathcal{I}_2 (resp. \mathcal{I}_1) du centre de projection \mathbf{C}_1 (resp. \mathbf{C}_2). Leurs coordonnées dans les plans projectifs sont donc :

$$\mathbf{e}_2 = P_2 \cdot \mathbf{C}_1 = [P_2, \mathbf{p}_2] \cdot \begin{bmatrix} \overline{\mathbf{C}_1} \\ 1 \end{bmatrix} = -\tilde{P}_2 \tilde{P}_1^{-1} \cdot \mathbf{p}_1 + \mathbf{p}_2$$

$$\mathbf{e}_1 = -\tilde{P}_1 \tilde{P}_2^{-1} \cdot \mathbf{p}_2 + \mathbf{p}_1$$

Afin de simplifier les expressions qui vont suivre, nous considérons \mathcal{R}_{c1} comme origine de l'espace projectif. La généralisation éventuelle des formulations à partir d'un repère \mathcal{R}_o

$(\mathbf{O}, \mathbf{X}_0, \mathbf{Y}_0, \mathbf{Z}_0)$ ne nécessitera que la prise en compte du déplacement de \mathcal{R}_o à \mathcal{R}_{c1} . Ainsi, la matrice de projection est égale à : $\mathbf{P}_1 = [I_3, \mathbf{0}]$. Nous en déduisons alors les coordonnées simplifiées des centres de projection et des épipôles des deux caméras :

$$\mathbf{C}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{C}_2 = \begin{pmatrix} \mathbf{e}_1 \\ 1 \end{pmatrix}, \quad \mathbf{e}_1 \simeq -\tilde{\mathbf{P}}_2^{-1} \cdot \mathbf{e}_2, \quad \mathbf{e}_2 \simeq \mathbf{p}_2 \quad (3.29)$$

3.2.4 La géométrie épipolaire

La géométrie épipolaire dépend de la position relative des deux caméras et de leurs paramètres intrinsèques. Nous allons donc identifier quelles sont les contraintes imposées entre les points de l'espace projectif et leurs projections dans chacune des deux images.

Tout point \mathbf{M} de l'espace projectif induit un plan, appelé épipolaire $\Pi_{\mathbf{M}}$, contenant des centres de projection \mathbf{C}_γ de chaque caméra. Les projections \mathbf{m}_γ correspondent à l'intersection des rayons $\mathbf{C}_\gamma \mathbf{M}$ avec les deux plans projectifs \mathcal{I}_γ . L'intersection du plan épipolaire avec le plan projectif de l'image \mathcal{I}_2 définit la ligne épipolaire \mathbf{l}_2 . Cette ligne comprend par définition l'épipôle \mathbf{e}_2 et la projection \mathbf{m}_2 du point \mathbf{M} . Nous remarquons que la droite $\mathbf{C}_1 \mathbf{C}_2$, appelée ligne de base, est commune à tous les plans épipolaires de l'espace projectif : elle est donc à l'origine du faisceau de plans épipolaires $\Pi_{\mathbf{M}}$.

Nous vérifions que les épipôles \mathbf{e}_γ appartenant à la ligne de base et aux plans projectifs \mathcal{I}_γ , sont indépendantes du point \mathbf{M} considéré. De même, nous remarquons que les lignes épipolaires, issues de l'intersection d'un faisceau de plan et du plan projectif, forment un faisceau de droites qui convergent vers les épipôles.

Il est à noter qu'un mouvement de translation fronto-parallèle à l'axe des caméras consécutif à un mouvement de rotation dont l'axe est perpendiculaire aux plans des caméras imposent le renvoi à l'infini des épipôles et par conséquent induit des lignes épipolaires parallèles entre elles.

3.2.4.1 La matrice fondamentale F

La matrice fondamentale \mathbf{F} est une représentation algébrique de la géométrie épipolaire : tout point \mathbf{m}_1 du plan projectif \mathcal{I}_1 admet un correspondant \mathbf{m}_2 dans l'image \mathcal{I}_2 qui appartient à la ligne épipolaire \mathbf{l}_2 .

$$\mathbf{l}_2 \simeq \mathbf{F} \cdot \mathbf{m}_1 \quad (3.30)$$

La ligne épipolaire \mathbf{l}_2 est donc définie par l'épipôle \mathbf{e}_2 et le point image \mathbf{m}_2 :

$$\mathbf{l}_2 \simeq \mathbf{e}_2 \wedge \mathbf{m}_2 \quad (3.31)$$

La matrice fondamentale dépend seulement de la configuration des caméras (paramètres intrinsèques et extrinsèques) et en aucun cas de l'objet visualisé. Géométriquement, \mathbf{F} représente une application de l'image \mathcal{I}_1 dans le faisceau de lignes épipolaires passant par \mathbf{e}_2 , soit une application du plan projectif \mathcal{P}^2 vers une droite projective \mathcal{P}^1 . La matrice \mathbf{F} [3,3] est donc de rang 2.

Le principal intérêt de la matrice fondamentale est qu'elle permet de vérifier la correspondance entre les points images \mathbf{m}_γ . Nous avons en effet introduit la ligne épipolaire \mathbf{l}_2 comme la droite qui passait par la projection \mathbf{m}_2 du point \mathbf{M} , soit $\mathbf{m}_2^t \cdot \mathbf{l}_2 = 0$. Nous en déduisons alors

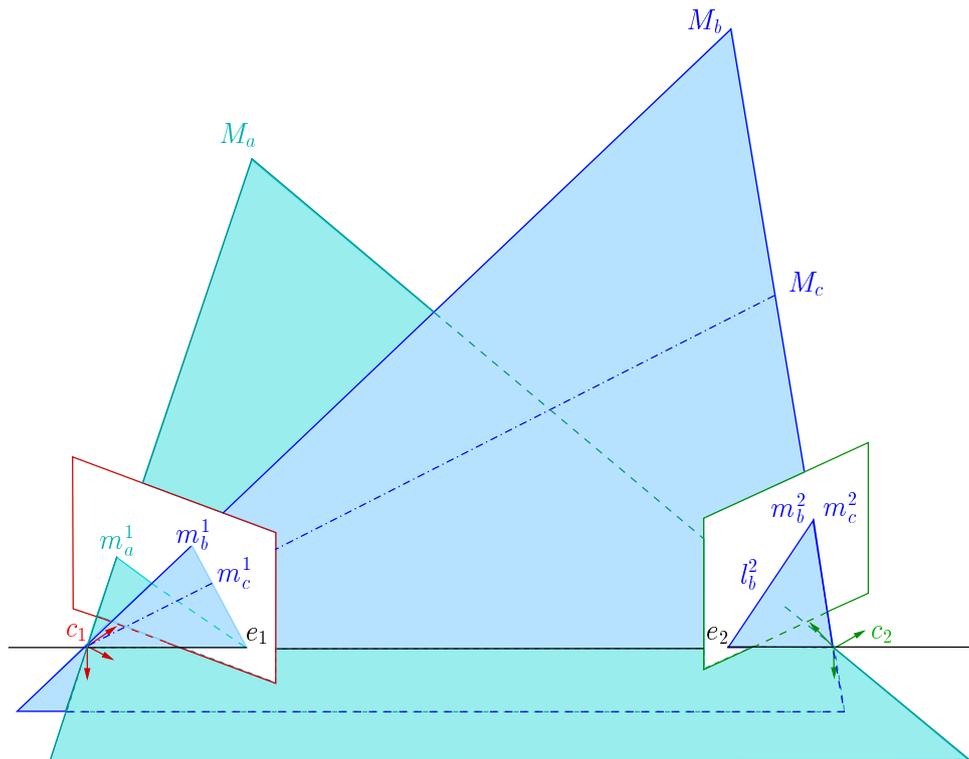


FIG. 3.5 – Les points M_a et M_b de l'espace projectif définissent avec la ligne de base (C_1C_2) les plans épipolaires Π_{M_a} et Π_{M_b} . L'intersection des plans épipolaires avec les plans images forment les lignes épipolaires (l_b^2 correspond à l'intersection du plan épipolaire Π_{M_b} avec l'image \mathcal{I}_2). L'intersection de la ligne épipolaire avec la ligne de base et un rayon forme respectivement l'épipôle et le point image. Le segment $[e_2m_b^2]$ représente la projection du rayon (C_1M_b) dans l'image \mathcal{I}_2 . Le point M_c appartenant au rayon (C_2M_b) , sa projection dans l'image \mathcal{I}_2 est confondue avec celle du point M_b .

que les points \mathbf{m}_γ sont en correspondance ou représentent les projections du point \mathbf{M} dans les deux plans projectifs \mathcal{I}_γ à condition que leurs coordonnées vérifient la contrainte épipolaire :

$$\mathbf{m}_2^t \cdot \mathbf{F} \cdot \mathbf{m}_1 = 0 \quad (3.32)$$

Les épipôles sont définies grâce à la matrice fondamentale. Pour tout point \mathbf{m}_1 de l'image \mathcal{I}_1 (différent de \mathbf{e}_1), la ligne épipolaire correspondante est $\mathbf{l}_2 = \mathbf{F} \cdot \mathbf{m}_1$ contient l'épipôle \mathbf{e}_2 . L'épipôle \mathbf{e}_2 satisfait $\mathbf{e}_2^t \cdot (\mathbf{F} \cdot \mathbf{m}_1) = (\mathbf{e}_2^t \cdot \mathbf{F}) \cdot \mathbf{m}_1 = 0$ quel que soit \mathbf{m}_1 . Il en résulte que $\mathbf{e}_2^t \cdot \mathbf{F} = 0$. L'épipôle \mathbf{e}_2 représente donc le noyau de l'application \mathbf{F} . De manière identique, l'épipôle \mathbf{e}_1 représente donc le noyau de l'application \mathbf{F}^t qui régit la projection des points de l'image \mathcal{I}_2 vers l'image \mathcal{I}_1 ($\mathbf{m}_1^t \cdot \mathbf{F}^t \cdot \mathbf{m}_2 = 0$).

3.2.4.2 La matrice essentielle E

La matrice essentielle \mathbf{E} représente la duale de la matrice fondamentale dans le cas particulier où les paramètres intrinsèques des deux caméras (\mathbf{K}_γ) sont connus. Dès lors, les transformations projectives \mathbf{P}_γ et les coordonnées des épipôles sont de la forme :

$$\mathbf{P}_1 = \mathbf{K}_1 \cdot [\mathbf{I}_3, \mathbf{0}] \quad \text{et} \quad \mathbf{P}_2 = \mathbf{K}_2 \cdot [\mathbf{R}, \bar{\mathbf{T}}] \quad (3.33)$$

les expressions de l'équation 3.29 deviennent :

$$\mathbf{e}_1 \simeq -\mathbf{K}_1 \cdot \mathbf{R} \cdot \bar{\mathbf{T}} \quad \text{et} \quad \mathbf{e}_2 \simeq \mathbf{K}_2 \cdot \bar{\mathbf{T}} \quad (3.34)$$

Soient $\widetilde{\mathbf{m}}_\gamma = \mathbf{K}_\gamma^{-1} \cdot \mathbf{m}_\gamma$, les coordonnées euclidiennes, exprimées en mètres, des projections \mathbf{m}_γ . La matrice essentielle peut alors être définie à partir de l'équation 3.32 :

$$[\mathbf{K}_2 \cdot \widetilde{\mathbf{m}}_2]^t \cdot \mathbf{F} \cdot [\mathbf{K}_1 \cdot \widetilde{\mathbf{m}}_1] = 0 \quad (3.35)$$

$$\widetilde{\mathbf{m}}_2^t \cdot \mathbf{E} \cdot \widetilde{\mathbf{m}}_1 = 0 \quad \text{avec} \quad \mathbf{E} = \mathbf{K}_2^t \cdot \mathbf{F} \cdot \mathbf{K}_1 \quad (3.36)$$

Dès lors, en égalisant les équations 3.30 et 3.31, relatives à la droite épipolaire \mathbf{l}_2 , nous obtenons :

$$\mathbf{F} \cdot \mathbf{m}_1 \simeq \mathbf{e}_2 \wedge \mathbf{m}_2$$

Soit, en reportant les expressions des équations 3.34 et 3.36 :

$$\mathbf{K}_2^{-t} \cdot \mathbf{E} \cdot \mathbf{K}_1^{-1} \simeq \mathbf{K}_2 \cdot \bar{\mathbf{T}} \wedge \mathbf{K}_2 \cdot \mathbf{R} \cdot \mathbf{K}_1^{-1}$$

Or, pour toute matrice \mathbf{Q} , $\mathbf{Q} \cdot \mathbf{a} \wedge \mathbf{Q} \cdot \mathbf{b} = \det(\mathbf{Q}) \cdot \mathbf{Q}^{-t} \cdot (\mathbf{a} \wedge \mathbf{b})$, nous retrouvons ainsi la contrainte de rigidité :

$$\mathbf{E} \simeq [\bar{\mathbf{T}}]_{\times} \mathbf{R} \quad (3.37)$$

Nous obtenons une expression de la matrice essentielle \mathbf{E} qui s'avère uniquement dépendante du déplacement \mathbf{D}_{12} réalisé entre les centres de projection \mathbf{C}_γ des deux caméras. La matrice essentielle $\mathbf{E}[3,3]$ a des propriétés identiques à celles de \mathbf{F} : elle est de rang deux et admet les mêmes valeurs propres que la matrice fondamentale. Le noyau de l'application correspondante correspond au vecteur de translation $\bar{\mathbf{T}}$ qui n'est autre qu'un facteur d'échelle près que l'expression de l'épipôle \mathbf{e}_2 . Enfin, la matrice \mathbf{R} étant orthogonale, $\mathbf{E} \cdot \mathbf{E}^t \simeq -[\bar{\mathbf{T}}]_{\times}^2$.

3.3 L'homographie planaire

Un plan Π de l'espace projectif se projette dans une image \mathcal{I}_1 selon une relation d'équivalence point à point appelée homographie planaire \mathbf{H}_1 car induite par un plan. L'observation de ce même plan à partir d'une seconde caméra est à l'origine d'une nouvelle homographie \mathbf{H}_2 dans l'image \mathcal{I}_2 . Les projections du plan dans chacune des images sont alors liées par une troisième homographie \mathbf{H}_{12} qui dépend de la transformation géométrique qui sépare les deux repères caméras ainsi que des paramètres intrinsèques des caméras. Calculer l'homographie induite par un plan entre deux images permet donc d'estimer le déplacement entre les deux repères caméras.

3.3.1 Description

3.3.1.1 Projection d'un plan sur une image

La transformation projective de \mathcal{P}^2 qui associe les points d'un plan Π à ceux de son image sur le plan rétinien est une homographie planaire. Les coordonnées homogènes des points \mathbf{M} du plan Π et de l'image \mathbf{m} sont définis selon trois paramètres. L'homographie planaire dépend donc de 8 paramètres (9-1 de facteur d'échelle) et peut donc être déterminée de façon unique grâce à la résolution d'un système linéaire, obtenu à partir de quatre points en correspondance, non-collinéaires trois à trois.

Considérons en effet que le plan Π contienne les deux premiers axes du repère euclidien. Chaque point $\mathbf{M} = [X_M, Y_M, 0, T_M]^t$ du plan Π fournit deux équations indépendantes selon les coordonnées relatives de sa projection dans l'image $\mathbf{M} = [u_M, v_M, w_M]^t$:

$$\frac{u_M}{w_M} = \frac{\mathbf{H}_{11}.X_M + \mathbf{H}_{12}.Y_M + \mathbf{H}_{13}.T_M}{\mathbf{H}_{31}.X_M + \mathbf{H}_{32}.Y_M + \mathbf{H}_{33}.T_M} \quad \text{et} \quad \frac{v_M}{w_M} = \frac{\mathbf{H}_{21}.X_M + \mathbf{H}_{22}.Y_M + \mathbf{H}_{23}.T_M}{\mathbf{H}_{31}.X_M + \mathbf{H}_{32}.Y_M + \mathbf{H}_{33}.T_M}$$

3.3.1.2 Transformations entre images de scène planaire

Nous considérons que le repère euclidien est \mathcal{R}_{c1} . Dès lors les homographies qui, à tout point \mathbf{M} de Π font correspondre les points \mathbf{m}_1 et \mathbf{m}_2 (voir Fig. 3.6) sont les transformations projectives \mathbf{P}_1 et \mathbf{P}_2 de la forme :

$$\mathbf{m}_1 = [\mathbf{l}_3, \mathbf{0}].\mathbf{M} \quad \text{et} \quad \mathbf{m}_2 = [\tilde{\mathbf{P}}_2, \mathbf{p}_2].\mathbf{M}$$

Soit $\mathbf{N} = [X_N, Y_N, Z_N, T_N]^t$, la normale au plan Π d'équation :

$$\mathbf{N}^t.\mathbf{M} = 0 \quad \forall \mathbf{M} \in \mathcal{P}^3 \quad (3.38)$$

Le rayon $\mathbf{C}_1\mathbf{M} = [\overline{\mathbf{M}}^t, 0]^t$ intercepte le plan image \mathcal{I}_1 au point de coordonnées \mathbf{m}_1 . Ce point est obtenu par projection selon \mathbf{P}_1 et nous considérons sans perte de généralité que $\mathbf{m}_1 \simeq \mathbf{P}_1.\mathbf{M}$ soit :

$$\mathbf{m}_1 \simeq \overline{\mathbf{M}}$$

Les coordonnées du point \mathbf{M} appartenant au plan Π peuvent s'exprimer en fonction du point \mathbf{m}_1 et de la normale au plan \mathbf{N} , toutes deux exprimées dans le repère de la caméra \mathcal{R}_{c1} . Le rayon intercepte le plan Π au point \mathbf{M} . Les coordonnées du point \mathbf{M} vérifient donc l'équation 3.38 $\overline{\mathbf{N}}^t.\mathbf{m}_1 + T_N.T = 0$. Nous en déduisons la valeur du paramètre T (coordonnée homogène du point \mathbf{M}) qui détermine la position du point \mathbf{M} le long du rayon :

$$T = -\frac{\overline{\mathbf{N}}^t.\mathbf{m}_1}{T_N}$$

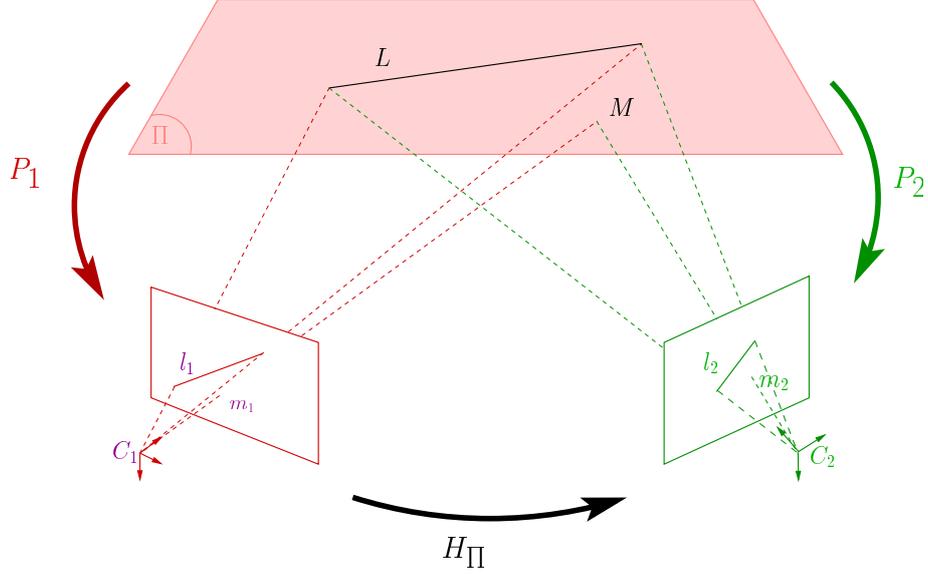


FIG. 3.6 – L'homographie planaire H qui lie l'image \mathcal{I}_1 à l'image \mathcal{I}_2 est obtenue par composition des homographies planaires P_1 et P_2 entre le plan de référence et les plans images. Les droites l_1, l_2 et les points p_1, p_2 sont les projections de la droite l_0 et du point p_0 sur les plans images des caméras de centre C_1 et C_2 .

Les coordonnées de la projection \mathbf{m}_2 du point \mathbf{M} par la projection P_2 sont alors :

$$\mathbf{m}_2 \simeq [\tilde{P}_2, \mathbf{p}_2] \cdot \begin{vmatrix} \mathbf{m}_1 \\ -\bar{\mathbf{N}}^t \cdot \mathbf{m}_1 \\ T_N \end{vmatrix}$$

L'égalité projective qui lie les images \mathbf{m}_1 et \mathbf{m}_2 selon les projections P_1 et P_2 du point \mathbf{M} est donc une homographie de \mathcal{P}^2 , induite par le plan Π , de la forme :

$$\mathbf{m}_2 = H_\Pi \cdot \mathbf{m}_1 \quad (3.39)$$

avec :

$$H_\Pi = [\tilde{P}_2 - \frac{\mathbf{p}_2 \cdot \bar{\mathbf{N}}^t}{T_N}] \quad (3.40)$$

Il existe donc une famille de plans $\bar{\mathbf{N}}$ de \mathcal{P}^3 et inversement il existe une famille d'homographies H_Π entre deux vues induite par un plan Π .

3.3.1.3 Cas particulier des caméras calibrées

L'homographie planaire exprimée dans le repère euclidien se distingue de celle exprimée dans l'espace projectif car elle est représentée par une égalité au sens strict. La connaissance des matrices de calibration des deux caméras implique des expressions des transformations projectives conformes à celles des relations 3.33.

$$H_\Pi = K_2 \cdot [R - \frac{\bar{\mathbf{T}} \cdot \bar{\mathbf{N}}^t}{T_N}] \cdot K_1^{-1} \quad (3.41)$$

L'homographie planaire qui lie les coordonnées euclidiennes des projections est alors $\widetilde{\mathbf{m}}_2 = H_e \cdot \widetilde{\mathbf{m}}_1$ avec :

$$H_e = R - \frac{\overline{\mathbf{T}} \cdot \overline{\mathbf{N}}^t}{T_N} \quad (3.42)$$

En posant $\mathbf{t} = \overline{\mathbf{T}}$ les coordonnées du repère \mathcal{R}_{c2} , $\mathbf{n} = \overline{\mathbf{N}}$, les coordonnées de la normale au plan, tous deux exprimées dans le repère \mathcal{R}_{c1} , et $d = T_N$ la distance qui sépare le plan du repère \mathcal{R}_{c2} , on retrouve l'écriture usuelle de l'homographie :

$$H_e = R - \frac{\mathbf{t} \cdot \mathbf{n}^t}{d}. \quad (3.43)$$

Cette matrice possède une valeur propre $\lambda = 1$: il est donc possible de normaliser la matrice H_e à partir de la recherche de ses valeurs propres. D'autre part, les matrices $[\overline{\mathbf{N}}_1]_{\times}$ et $[\overline{\mathbf{N}}_2]_{\times}$, où $\overline{\mathbf{N}}_1$ et $\overline{\mathbf{N}}_2$ représentent le vecteur normal au plan Π , exprimés dans les repères \mathcal{R}_{c1} et \mathcal{R}_{c2} , vérifient la relation :

$$[\overline{\mathbf{N}}_2]_{\times}^k \cdot H_e^{-t} = H_e \cdot [\overline{\mathbf{N}}_1]_{\times}^k \quad (3.44)$$

Pour $k=1$, cette relation permet d'exprimer les coordonnées du vecteur normal au plan Π en fonction de l'homographie euclidienne avec $\det(M) \cdot M \cdot [\mathbf{p}]_{\times} M^t = [M^{-t} \mathbf{p}]_{\times}$:

$$\overline{\mathbf{N}}_2 = \frac{1}{\det(H_e)} H_e^{-t} \cdot \overline{\mathbf{N}}_1 \quad (3.45)$$

3.3.1.4 Relations entre géométrie épipolaire et homographie planaire

Les projections \mathbf{m}_1 et \mathbf{m}_2 d'un point \mathbf{M} de \mathcal{P}^3 selon les projections P_1 et P_2 respectent la contrainte épipolaire 3.32. Si, de surcroît, les points en correspondance sont liés par une homographie planaire H_{Π} alors l'homographie est dite consistante ou compatible avec F . L'homographie contraint la matrice fondamentale car H_{Π} permet de placer \mathbf{m}_2 , correspondant du point \mathbf{m}_1 , sur la ligne épipolaire décrite par F .

L'expression de la contrainte épipolaire à partir de l'égalité projective $\mathbf{m}_2 \simeq H_{\Pi} \cdot \mathbf{m}_1$ fournit : $(H_{\Pi} \cdot \mathbf{m}_1)^t \cdot F \cdot \mathbf{m}_1 = \mathbf{m}_1 \cdot H_{\Pi}^t \cdot F \cdot \mathbf{m}_1 = 0$. Par conséquent, la condition nécessaire et suffisante pour qu'une homographie soit compatible avec la matrice fondamentale est que la matrice $H_{\Pi}^t \cdot F$ doit être anti-symétrique :

$$H_{\Pi}^t \cdot F + F \cdot H_{\Pi} = 0 \quad (3.46)$$

La formulation de la ligne épipolaire \mathbf{l}_2 (relation 3.31) en fonction de l'homographie planaire permet l'établissement d'une nouvelle expression de la matrice fondamentale :

$$F \simeq [\mathbf{e}_2]_{\times} \cdot H_{\Pi} \quad (3.47)$$

La matrice fondamentale ayant 5 degrés de liberté, il suffit donc de mettre en correspondance les projections de trois points de l'espace projectif pour paramétrer les 8 degrés de liberté d'une homographie planaire. Ces trois points forment un plan Π de \mathcal{P}^3 défini par trois paramètres : deux pour sa normale $\overline{\mathbf{N}}$ normalisée et un pour la distance T_N qui le sépare du centre du repère. Les deux derniers points nécessaires à l'établissement de H_{Π} sont les deux épipôles.

3.3.1.5 Homographie du plan à l'infini Π_∞

L'homographie du plan à l'infini H_∞ correspond à cas particulier de l'expression 3.40. En effet, le plan considéré se situe à une distance T_N qui tend vers l'infini, on obtient alors l'expression :

$$H_\infty = \lim_{T_N \rightarrow +\infty} H_\Pi = \lim_{T_N \rightarrow +\infty} \left[\tilde{P}_2 - \frac{\mathbf{p}_2 \cdot \overline{\mathbf{N}}^t}{T_N} \right] = \tilde{P}_2 \quad (3.48)$$

soit, dans le cas où les caméras sont calibrées (3.41) :

$$H_\infty = K_2^{-1} \cdot R \cdot K_1 \quad (3.49)$$

Les VPs et VLs sont des projections sur les plans images d'éléments appartenant au plan à l'infini. Il est donc envisageable de déterminer l'homographie du plan à l'infini H_∞ à partir du mouvement dans les images des coordonnées de trois VPs ou d'une VL et d'un VP, distinct de cette dernière.

Enfin, la connaissance de l'homographie à l'infini permet de réduire la région de recherche de la correspondance d'un point image entre deux caméras stéréo. En effet, le correspondant du point \mathbf{m}_1 dans l'image \mathcal{I}_2 appartient à la ligne épipolaire \mathbf{l}_2 (cf 3.30). Or, le point \mathbf{M} observé se trouve entre la paire de caméras stéréo et le plan à l'infini Π_∞ voir Fig. 3.7. Par conséquent, le point \mathbf{m}_2 se trouve entre l'épipôle \mathbf{e}_2 et le projeté par H_∞ de \mathbf{m}_1 .

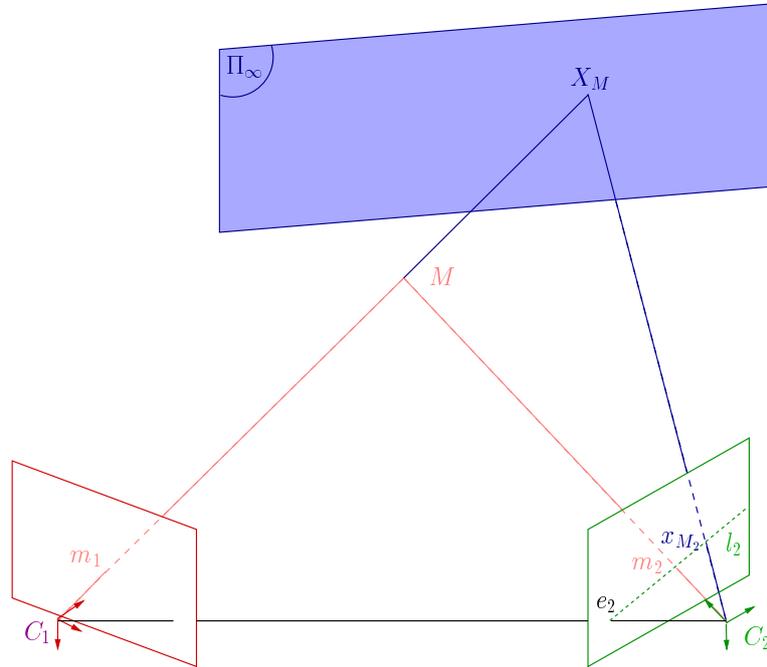


FIG. 3.7 – Projection d'un point dans deux images. La projection x_{M_2} du point m_1 par H_∞ permet d'obtenir l'extrémité de la projection dans l'image \mathcal{I}_2 du rayon passant par le point M . Le correspondant m_2 du point m_1 se trouve donc entre l'épipôle e_2 et $H_\infty \cdot m_1$.

3.3.1.6 Relations entre deux homographies planaires

Nous considérons que la scène comporte un second plan π qui induit une homographie planaire H_π entre les points des deux images. La décomposition de F selon la relation 3.47 n'est pas unique : tout plan π de l'espace projectif la vérifie. Soit H_0 une transformation projective telle que $[\mathbf{e}_2]_\times H_0 = 0$, il existe alors un vecteur \mathbf{v} qui vérifie $H_0 = \mathbf{e}_2 \cdot \mathbf{v}^t$. Dès lors, toute homographie planaire H_π peut se décomposer en fonction d'une homographie de référence H_Π et de H_0 :

$$H_\pi \simeq H_\Pi + \mathbf{e}_2 \cdot \mathbf{v}^t \quad (3.50)$$

Par conséquent, la relation induite entre les points en correspondance entre deux images est une famille d'homographies planaires, paramétrée selon le plan considéré. Les projections selon H_Π et H_π dans l'image \mathcal{I}_2 d'un point \mathbf{M}_c appartenant à l'intersection des deux plans sont confondues $\mathbf{m}_c^2 \simeq H_\Pi \cdot \mathbf{m}_c^1 \simeq H_\pi \cdot \mathbf{m}_c^1$ ce qui implique $\mathbf{v}^t \cdot \mathbf{m}_c^1 \simeq 0$. Le vecteur \mathbf{v} est donc le support d'une droite qui passe par le point \mathbf{m}_c^1 . Cette droite n'est autre que la projection dans \mathcal{I}_1 de l'intersection des deux plans Π et π .

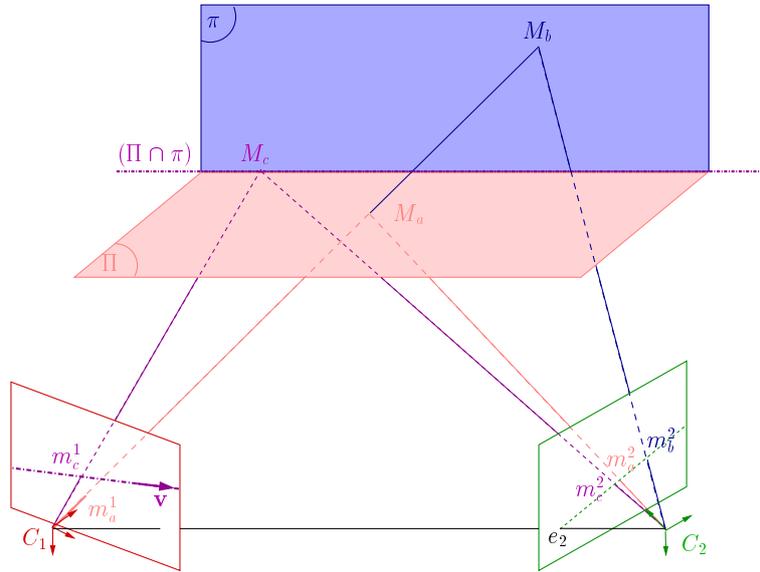


FIG. 3.8 – Projections de points appartenant à des plans différents. Le point M_a appartient à Π , M_b appartient à π et M_c appartient à l'intersection des deux plans. Les points M_a et M_b appartenant au même rayon, les projections $m_a^2 \simeq H_\Pi \cdot m_a^1$ et $m_b^2 \simeq H_\pi \cdot m_a^1$ de ces deux points selon P_2 sont alignées avec l'épipôle e_2 . Les projections des trois points dans les deux images vérifient l'homographie planaire H_Π , en prenant garde de tenir compte du parallaxe projectif du point M_b qui n'appartient à Π . Le vecteur v supporte la projection de la droite, intersection des plans Π et π .

Nous pouvons utiliser la formule 3.50 pour déterminer la position relative de points par rapport au plan de référence Π . Considérons le point \mathbf{M}_b de la figure 3.8 qui appartient au rayon $(\mathbf{C}_1 \mathbf{M}_a)$: nous avons $\mathbf{m}_a^1 = \mathbf{m}_b^1$. En considérant que l'épipôle \mathbf{e}_2 et l'homographie H_Π soient connus, nous en déduisons :

$$\mathbf{m}_b^2 \simeq \tilde{P}_2 \cdot \mathbf{m}_a^1 + \lambda_b \cdot \mathbf{p}_2$$

$$\begin{aligned}
&= \left[\tilde{\mathbf{P}}_2 - \frac{\mathbf{p}_2 \cdot \bar{\mathbf{N}}^t}{T_N} \right] \cdot \mathbf{m}_a^1 + \frac{\mathbf{p}_2 \cdot \bar{\mathbf{N}}^t}{T_N} \cdot \mathbf{m}_a^1 + \lambda_b \cdot \mathbf{p}_2 \\
&\simeq \mathbf{H}_\Pi \cdot \mathbf{m}_b^1 + \mathbf{p}_2 \cdot \left[\frac{\bar{\mathbf{N}}^t \cdot \mathbf{m}_a^1}{T_N} + \lambda_b \right]
\end{aligned} \tag{3.51}$$

Le point \mathbf{M}_a appartenant au plan Π est tel que : $\bar{\mathbf{N}}^t \cdot \mathbf{m}_a^1 = 0$. Ainsi la projection d'un point \mathbf{M}_b en dehors du plan de référence Π s'exprime en fonction de la projection \mathbf{m}_a^1 , intersection du rayon $\mathbf{C}_2\mathbf{M}_b$ et du plan Π et de l'épipôle $\mathbf{p}_2 = \alpha \cdot \mathbf{e}_2$ (relation 3.29) :

$$\mathbf{m}_b^2 \simeq \mathbf{H}_\Pi \cdot \mathbf{m}_b^1 + \lambda_b \cdot \alpha \cdot \mathbf{e}_2 \tag{3.52}$$

Les points $\mathbf{e}_2, \mathbf{m}_a^2, \mathbf{m}_b^2$ sont donc alignés et le produit $\lambda_b \cdot \alpha$ est appelé parallaxe projectif. Cette valeur peut être calculée à partir de la correspondance entre les points \mathbf{m}_b^1 et \mathbf{m}_b^2 , connaissant \mathbf{H}_Π et \mathbf{e}_2 . Le parallaxe projectif planaire est proportionnel à la distance qui sépare le point \mathbf{M}_b du plan Π et inversement proportionnel à la profondeur du point \mathbf{M}_b au centre optique \mathbf{C}_2 . Le signe du parallaxe indique la position relative du point par rapport au plan : il est nul si \mathbf{M}_b appartient à Π .

3.3.2 L'estimation de l'homographie

La mise en oeuvre du calcul de l'homographie induite par un plan suivis dans deux images nécessite la résolution d'un système linéaire entre les coordonnées des primitives en correspondance. En faisant abstraction du problème de la mise en correspondance des primitives, leur nombre, leur répartition spatiale ainsi que la résolution associée à leurs coordonnées (la proximité des éléments détectés) influencent grandement le conditionnement du système linéaire à résoudre. Nous présentons dans ce qui suit les précautions indispensables à la mise en oeuvre de l'estimation de l'homographie. Cette partie est largement inspirée du chapitre 4 du livre [Hartley and Zisserman, 2003] que nous recommandons vivement de lire à toute personne confrontée à ce délicat problème.

3.3.2.1 Algorithme de transformation linéaire direct (DLT)

Nous considérons dans ce qui suit des primitives ponctuelles mais le même raisonnement peut être tenu à partir de droites ou combinaison de points et de droites. Soient \mathcal{I}_1 et \mathcal{I}_2 deux images d'un plan π sur lesquelles deux jeux de FPs en correspondance sont détectés. L'homographie \mathbf{H} qui lie les projections de la k^{eme} primitive \mathbf{p}_{ak} et \mathbf{p}_{bk} est telle que :

$$\mathbf{p}_{bk} \propto \mathbf{H} \cdot \mathbf{p}_{ak} \quad \text{soit} \quad \mathbf{p}_{bk} \wedge \mathbf{H} \cdot \mathbf{p}_{ak} = \mathbf{0} \tag{3.53}$$

ce qui peut se réécrire en introduisant le vecteur colonne $\mathbf{H}_{col} = [\mathbf{H}_1^t, \mathbf{H}_2^t, \mathbf{H}_3^t]^t$ et les coordonnées du FP $\mathbf{p}_{bk} = [u_{bk}, v_{bk}, w_{bk}]^t$:

$$\left[\begin{array}{ccc} \mathbf{0}^t & -w_{bk} \mathbf{p}_{ak}^t & v_{bk} \mathbf{p}_{ak}^t \\ w_{bk} \mathbf{p}_{ak}^t & \mathbf{0}^t & -u_{bk} \mathbf{p}_{ak}^t \\ -v_{bk} \mathbf{p}_{ak}^t & u_{bk} \mathbf{p}_{ak}^t & \mathbf{0}^t \end{array} \right] \cdot \left| \begin{array}{l} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \mathbf{H}_3 \end{array} \right. = \mathbf{0} \tag{3.54}$$

Chaque couple de FPs en correspondance est donc à l'origine de trois équations linéaires en fonction des colonnes de la matrice \mathbf{H} . Seulement deux d'entre elles sont linéairement indépendantes :

la suppression de la troisième ligne du système ainsi constitué ne modifie pas le système. L'homographie étant définie à un facteur projectif près, quatre couples de primitives en correspondance suffisent donc à son estimation. Le terme primitive fait ici référence à un point ou une droite. Le système Equ. 3.54 devient :

$$\widetilde{\mathbf{A}}_k \cdot \mathbf{H}_{\text{col}} = \mathbf{0} \quad (3.55)$$

soit :

$$\begin{bmatrix} \mathbf{0}^t & -w_{bk}\mathbf{p}_{ak}^t & v_{bk}\mathbf{p}_{ak}^t \\ w_{bk}\mathbf{p}_{ak}^t & \mathbf{0}^t & -u_{bk}\mathbf{p}_{ak}^t \end{bmatrix} \cdot \begin{vmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \mathbf{H}_3 \end{vmatrix} = \mathbf{0}. \quad (3.56)$$

L'estimation de la matrice d'homographie se résume donc à la résolution du système homogène

$$\mathbf{A} \cdot \mathbf{H}_{\text{col}} = \mathbf{0} \quad (3.57)$$

où la matrice \mathbf{A} représente la concaténation des contraintes formulées pour au moins quatre couples de primitives en correspondance. Le nombre de couples en correspondance N est généralement bien supérieur à 4, le système linéaire \mathbf{A} résultant est alors sur-déterminé ($[2N; 9]$). La chaîne de mesure introduisant un bruit supposé additif, l'estimation de l'homographie recherchée nécessite d'introduire une fonction de coût \mathcal{C} qui permet de se prémunir de la solution évidente $\mathbf{H}_{\text{col}} = \mathbf{0}$.

L'homographie appartenant au groupe SL^3 , certains proposent de résoudre le système Equ. 3.57 en vérifiant que le vecteur associé à la matrice \mathbf{H} vérifie $|\mathbf{H}_{\text{col}}| = 1$. Une autre possibilité est de fixer à 1 un des éléments de \mathbf{H}_{col} puisque l'homographie est définie à un facteur d'échelle près. Il est en effet tentant de poser $\mathbf{H}_{33} = \mathbf{H}_{\text{col}}(9) = 1$ ce qui rend le système inhomogène, à condition que les coordonnées w soient toutes unitaires ($w_{ak} = w_{bk} = 1$). Le système obtenu est alors :

$$\begin{bmatrix} 0 & 0 & 0 & \vdots & -u_{ak} & -v_{ak} & -1 & \vdots & v_{bk} \cdot u_{ak} & v_{bk} \cdot v_{ak} \\ u_{ak} & v_{ak} & 1 & \vdots & 0 & 0 & 0 & \vdots & -u_{bk} \cdot u_{ak} & -u_{bk} \cdot v_{ak} \end{bmatrix} \cdot \begin{vmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \mathbf{H}_{13} \\ \mathbf{H}_{23} \end{vmatrix} = \begin{vmatrix} -v_{bk} \\ u_{bk} \end{vmatrix} \quad (3.58)$$

L'intérêt de cette nouvelle écriture demeure néanmoins restreint car elle suppose qu'aucune primitive du plan à l'infini Π_∞ ($w = 0$) ne soit mise en correspondance, ce qui n'est pas notre cas lorsque nous prenons en compte les coordonnées du point de fuite dominant. La solution qui nous paraît la plus appropriée est celle qui minimise la distance entre les projections des FPs et leur correspondant. Il nous faut pour cela considérer le bruit qui corrompt les coordonnées des primitives.

3.3.2.2 Modèle de bruit

Si le bruit suit une loi de distribution gaussienne (hypothèse généralement admise par la majorité des auteurs), il est naturel d'utiliser une approche stochastique et d'en déduire une estimation de l'homographie au sens du maximum de vraisemblance. Nous considérons donc que les coordonnées idéales $\tilde{\mathbf{p}}$ des FPs qui vérifient l'estimation de l'homographie sont perturbées par un bruit $\Delta \mathbf{p}$ dont la distribution de probabilité est une gaussienne de moyenne nulle, de variance σ^2 . Les coordonnées détectées d'un FP \mathbf{p} sont telles que $\mathbf{p} = \tilde{\mathbf{p}} + \Delta \mathbf{p}$ avec :

$$Pr(\mathbf{p}) = \frac{1}{2\pi\sigma^2} \cdot \exp(-d(\tilde{\mathbf{p}}, \mathbf{p})^2/2\sigma^2) \quad (3.59)$$

L'estimation de l'homographie sera alors optimisée de façon à ce que la fonction de coût \mathcal{C} qui représente les erreurs de projection soit minimale. Nous considérons alors que l'estimation de H est vérifiée par les coordonnées idéales $\widetilde{\mathbf{p}}$ des FPs en correspondance. Nous avons seulement accès aux coordonnées détectées \mathbf{p} , la fonction de coût à minimiser revient à considérer les distances des erreurs de transfert dans la première image via H^{-1} et les erreurs de transfert dans la seconde image via H conformément à la Fig 3.9 :

$$\mathcal{C} = \sum_k d(\mathbf{p}_{ak}, H^{-1}\mathbf{p}_{bk})^2 + d(H\mathbf{p}_{ak}, \mathbf{p}_{bk})^2. \quad (3.60)$$

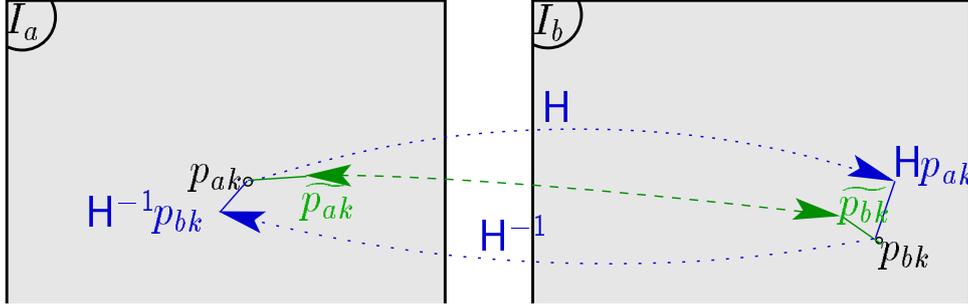


FIG. 3.9 – Les différents types d'erreurs de projection, d'après Hartley : l'erreur de transfert symétrique bleue représente la distance qui sépare chaque estimée de sa détection indépendamment de la transformée inverse tandis que l'erreur de projection verte correspond à la distance qui sépare les coordonnées idéales des extraites.

La densité de probabilité de bruit précédente est alors modifiée par la prise en compte du premier terme de l'expression Equ. 3.60 qui généralise la densité de probabilité précédente :

$$Pr(\mathbf{p}_{ak}, \mathbf{p}_{bk} \setminus H, \widetilde{\mathbf{p}}_{ak}) = \prod_k \frac{1}{2\pi\sigma^2} \cdot \exp^{-(d(\mathbf{p}_{ak}, \widetilde{\mathbf{p}}_{ak})^2 + d(\mathbf{p}_{bk}, \widetilde{\mathbf{p}}_{bk})^2)/2\sigma^2} \quad (3.61)$$

L'estimée au sens du maximum de vraisemblance de l'homographie recherchée maximise le logarithme de l'expression précédente :

$$\log(Pr(\mathbf{p}_{ak}, \mathbf{p}_{bk} \setminus H, \widetilde{\mathbf{p}}_{ak})) = -\frac{1}{2\sigma^2} \sum_k (d(\mathbf{p}_{ak}, \widetilde{\mathbf{p}}_{ak})^2 + d(\mathbf{p}_{bk}, \widetilde{\mathbf{p}}_{bk})^2) + cte \quad (3.62)$$

et minimise par conséquent au sens du maximum de vraisemblance les distances de reprojection entre FPs détectés et idéaux de l'expression Equ. 3.60 :

$$\mathcal{C} = \sum_k d(\mathbf{p}_{ak}, \widetilde{\mathbf{p}}_{ak})^2 + d(\mathbf{p}_{bk}, \widetilde{\mathbf{p}}_{bk})^2. \quad (3.63)$$

3.3.2.3 Configurations dégénérées

La disposition des primitives dans les images influence grandement la résolution de l'estimation de l'homographie. Ainsi, le système linéaire A $[2N, 9]$ admet 8 degrés de liberté. Par conséquent, si le nombre de primitives est minimal, nous devons nous assurer que les FPs ne sont pas alignés ou que des triplets de VLs ne sont pas linéairement dépendantes, soit en d'autres termes qu'elles n'aient pas un point d'intersection commun. De tels cas sont qualifiés de

dégénérés car le nombre de contraintes imposées par les coordonnées des primitives est inférieur au nombre de degrés de liberté du système. La résolution du système n'est pas unique : il existe une famille de solutions parmi lesquelles se trouve la solution recherchée si les primitives étaient linéairement indépendantes.

Les configurations de primitives présentées en Fig. 3.10 sont équivalentes et dégénérées. Bien qu'elles soient formées de cinq primitives, trois d'entre elles sont linéairement liées. Ainsi, l'estimation du système A ne peut être contrainte par un seul faisceau de droites car les configurations minimales requises pour estimer de manière unique une homographie sont : quatre FPs non-alignés trois à trois ou quatre droites n'ayant aucune intersection commune pour plus de deux d'entre elles ou toute autre configuration à partir de ces deux extrêmes.

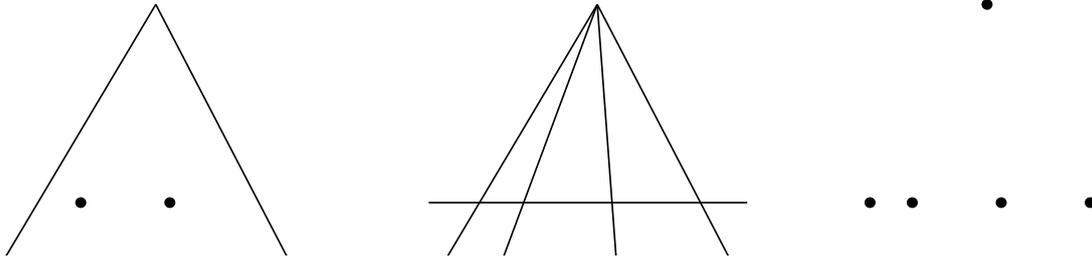


FIG. 3.10 – *Équivalences géométriques entre FPs et faisceau de VLs. Les estimations d'homographies obtenues à partir de ces configurations sont dégénérées car les primitives sont linéairement dépendantes : intersection commune de trois VLs ou alignement de trois FPs. L'homographie solution appartient alors à une famille d'homographies solutions à un paramètre.*

3.3.2.4 Normalisation du système

La normalisation ou pré-conditionnement du système est une opération indispensable à la résolution du système formé par l'homographie si l'on souhaite que la solution soit indépendante de l'origine du repère et/ou de facteurs d'échelles. En effet, la disparité des coefficients qui forment les colonnes de la matrice A confirme la liaison entre le conditionnement du système et l'amplitude des coordonnées (abscisses, ordonnées, profondeurs) qui constituent chacune des colonnes.

Reprenons les deux jeux de FPs en correspondance $\mathbf{p}_{\mathbf{ak}}, \mathbf{p}_{\mathbf{bk}}$ via l'homographie H et considérons désormais les transformations $\mathcal{T}_a, \mathcal{T}_b$ telles que :

$$\begin{cases} \mathbf{p}'_{\mathbf{ak}} &= \mathcal{T}_a \mathbf{p}_{\mathbf{ak}} \\ \mathbf{p}'_{\mathbf{bk}} &= \mathcal{T}_b \mathbf{p}_{\mathbf{bk}} \end{cases} \quad (3.64)$$

Les jeux de FPs $\mathbf{p}'_{\mathbf{ak}}, \mathbf{p}'_{\mathbf{bk}}$ vérifient donc l'homographie H' définie par : $H' = \mathcal{T}_b \cdot H \cdot \mathcal{T}_a^{-1}$. Soit $\epsilon_{\mathbf{k}} = \mathbf{p}_{\mathbf{bk}} \wedge H \mathbf{p}_{\mathbf{ak}}$ le résidu de l'estimation de H calculée par DLT. Calculons le résidu associé au k^{eme} couple de FPs correspondants exprimés dans les nouvelles bases suite aux transformations $\mathcal{T}_a, \mathcal{T}_b$:

$$\begin{aligned} \epsilon'_{\mathbf{k}} &= \mathbf{p}'_{\mathbf{bk}} \wedge H' \mathbf{p}'_{\mathbf{ak}} = \mathcal{T}_b \mathbf{p}_{\mathbf{bk}} \wedge (\mathcal{T}_b \cdot H \cdot \mathcal{T}_a^{-1}) \cdot \mathcal{T}_a \mathbf{p}_{\mathbf{ak}} \\ &= \mathcal{T}_b \mathbf{p}_{\mathbf{bk}} \wedge \mathcal{T}_b H \mathbf{p}_{\mathbf{ak}} = \mathcal{T}_b^* (\mathbf{p}_{\mathbf{bk}} \wedge H \mathbf{p}_{\mathbf{ak}}) \\ &= \mathcal{T}_b^* \epsilon_{\mathbf{k}} \end{aligned}$$

où \mathcal{T}_b^* représente la matrice de cofacteurs d'après le lemme : $M\mathbf{x}_1 \wedge M\mathbf{x}_2 = M^*(\mathbf{x}_1 \wedge \mathbf{x}_2)$. Ainsi, le résidu ϵ'_k n'est pas simplement lié au résidu ϵ_k . Chacun des produits $\tilde{A}_k \mathbf{H}_{\text{col}}$ et $A_k^* \mathbf{H}'_{\text{col}}$ est représenté par les deux premières composantes des vecteurs ϵ_k et ϵ'_k . Dès lors, si l'estimation de \mathbf{H} a été calculée de manière à minimiser l'erreur de projection $\tilde{A}_k \mathbf{H}_{\text{col}}$, $A_k^* \mathbf{H}'_{\text{col}}$ ne peut être minimal que si $\mathcal{T}_b^* = I_3$.

Nous rappelons que l'estimée \mathbf{H} est calculée sous la contrainte $\|\mathbf{H}_{\text{col}}\| = 1$. La distance algébrique entre les couples de FPs en correspondance $\mathbf{p}_{bk}, \mathbf{H}\mathbf{p}_{ak}$ et $\mathbf{p}'_{bk}, \mathbf{H}'\mathbf{p}'_{ak}$ est donc minimisée. Cependant cette dernière relation s'oppose à l'égalité $\|\mathbf{H}'_{\text{col}}\| = 1$. Par contre, dans l'éventualité où la transformation \mathcal{T}_b s'avère être une transformation euclidienne (sans changement d'échelle), si \mathbf{H} minimise l'erreur géométrique de projection pour le jeu de FPs en correspondance, alors \mathbf{H}' minimise l'erreur géométrique de projection entre les couples de FPs transformés. L'erreur de projection d'une homographie est donc invariante suite à une transformation euclidienne qui n'est pas une similarité. Il existe par conséquent une transformation optimale \mathcal{T}_b qui permet de minimiser l'erreur de projection $\|\tilde{A}'\mathbf{H}'_{\text{col}}\|$.

Notre objectif est de déterminer quel est le changement de repère permettant de réduire l'influence des coordonnées des jeux de FPs en correspondance dans l'estimation de l'homographie. Cette normalisation augmente de surcroît la répétabilité de l'estimation en s'affranchissant de phénomènes d'échelles (effet de zoom) ou d'images de dimensions différentes. Les auteurs proposent de centrer chacun des jeux de FPs autour de l'origine en imposant une distance moyenne d'un FP à l'origine de $\sqrt{2}$ pixels dans le cas où la caméra est non-calibrée. L'estimation de l'homographie \mathbf{H}' est alors entreprise à partir de ces deux jeux de FPs transformés et l'homographie initiale est obtenue par $\mathbf{H} = \mathcal{T}_b \cdot \mathbf{H}' \cdot \mathcal{T}_a^{-1}$.

La résolution du système linéaire Equ. 3.57 est calculé par SVD : $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^t$ avec $\mathbf{U}\mathbf{U}^t = \mathbf{V}\mathbf{V}^t = \mathbf{I}_8$ et $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_9)$. En raison du bruit de mesure qui perturbe les coordonnées des FPs, le système admet non plus 8 mais 9 degrés de libertés. La solution \mathbf{H}_{col} est donc le vecteur colonne de \mathbf{V}^t associé à la valeur propre la plus réduite (λ_9), théoriquement nulle.

Cette méthode de résolution agit de manière isotrope sur les deux dimensions de l'image : le facteur d'échelle affecté aux abscisses et ordonnées est identique. Certaines configurations de jeu de FPs nécessitent des facteurs d'échelles dédiés à chacun des axes. En effet, le conditionnement du système $\tilde{\mathbf{A}}$ est le rapport entre la première et l'avant dernière valeur propre (la dernière étant nulle) lorsque celles-ci sont ordonnées dans le sens décroissant. Le système sera d'autant mieux conditionné que ce rapport tendra vers 1, soit que la disparité entre la norme du plus grand et du plus petit des éléments qui forment le vecteur solution \mathbf{H}_{col} sera réduite. Or, les divers produits croisés entre coordonnées qui constituent les colonnes du système originel \mathbf{H} ne sont pas du même ordre lorsque l'on considère les coordonnées pixeliques des FPs : $w_{ak}w_{bk} = 1$ tandis que $1 < u_{ak}u_{bk} < l^2$ et $1 < v_{ak}v_{bk} < h^2$, pour une image de dimensions $[l; h]$. Une normalisation non-isotrope peut être employée mais nécessite de déterminer précisément quelles sont les paramètres à optimiser.

L'emploi d'une approche au sens des moindres carrés globaux (Total Least Squares) est recommandée par [Mester and Muehlich, 2001] lorsque l'on considère que les erreurs de mesures sont portées par un seul jeu de primitives. Le système à résoudre est inchangé mais la méthode minimise alors la distance de Mahalanobis entre les coordonnées des primitives et leurs projections. Les hypothèses sur l'origine des erreurs étant moins réalistes qu'une prise en compte symétrique, nous ne développerons pas plus cette méthode.

3.3.2.5 Estimation itérative de la matrice d'homographie

L'estimation robuste d'une homographie H^n nécessite d'être menée de manière itérative. En effet, l'étape de mise en correspondance des primitives étant effectuée à partir de connaissances a priori sur la scène et d'une prédiction de l'homographie, son estimation ne peut fournir des résultats corrects en une unique passe. Trop d'indéterminations demeurent sur les opérations de prédiction de H , mise en correspondance des couples de primitives et résolution du système Equ. 3.57. C'est pourquoi, une estimation itérative s'impose : en supposant que ${}^1H^n$ soit meilleure que ${}^0H^n$, l'introduction de cette nouvelle estimation dans l'étape de mise en correspondance de primitives doit permettre d'identifier de nouveaux couples de primitives et d'en rejeter d'autres retenus par erreur lors de la première mise en correspondance.

La condition d'arrêt d'un tel processus est difficile à déterminer. La minimisation de l'erreur de transfert est, certes, un objectif louable mais dans notre cas d'observation de la route par une paire stéréo embarquée, nous pouvons mettre à profit l'existence de contraintes plus spécifiques. En effet, la région de l'image où apparaît le plan est relativement limitée : sa surface n'excède pas le tiers de l'image, contrairement aux exemples fournis dans la littérature. De plus, parmi la trentaine de primitives supposées coplanaires détectées en moyenne sur chaque image, une minorité d'entre elles sont détectées dans la partie basse de l'image. Or, ces primitives, détectées au premier plan, sont celles qui contraignent le plus le déplacement entre deux images mais subissent aussi le plus les déformations dues à la distorsion des images.

La qualité de l'estimation de l'homographie entre deux images est dans notre cas réduite en raison de la faible quantité de primitives en correspondance, de leurs dispositions dans l'image et de l'incertitude qui demeurent sur leurs coordonnées. Il nous paraît donc à ce stade peu probable d'obtenir des résultats d'estimation corrects à partir d'une telle configuration. Nous proposons donc de rechercher tous les couples de primitives susceptibles d'être coplanaires grâce à l'estimation de l'homographie entre deux images, plutôt que d'implémenter un algorithme de type Levenberg-Marquardt qui permettrait de minimiser au sens du maximum de vraisemblance l'erreur de transfert.

Minimiser la distance entre les coordonnées théoriques $\hat{\mathbf{p}}$ et $\hat{\mathbf{l}}$ des primitives qui vérifient une homographie estimée et leurs coordonnées détectées \mathbf{p} et \mathbf{l} ne nous apparaît donc pas opportun à ce stade de l'estimation. En effet, nous avons pu vérifier expérimentalement que se focaliser sur la minimisation du critère de l'erreur de transfert est à l'origine d'estimations incorrectes qui écartent des couples de primitives coplanaires qui dépassent le seuil de rejet.

Une recherche itérative est par définition coûteuse en temps de calcul et peut converger vers un minimum local. Nous affinons néanmoins la détection des primitives coplanaires de manière itérative en considérant que la recherche doit s'arrêter dès que le groupe de couples de primitives en correspondance cesse d'évoluer. Le critère d'acceptation utilisé est l'erreur de transfert symétrique qui ne fait certes pas référence à l'introduction d'un processus d'optimisation statistique mais fournit des renseignements cohérents sur la probabilité qu'un couple en correspondance soit effectivement coplanaire. L'estimation de l'homographie à partir de laquelle nous extrayons les informations relatives au déplacement de la caméra est alors correcte mais peut être améliorée en prenant en compte les liaisons croisées qui existent entre plusieurs vues.

3.4 Les déplacements rigides

3.4.1 Base projective d'une paire stéréo à partir d'une homographie plane

Une base projective d'une paire stéréo est définie par cinq points qui représentent les caractéristiques géométriques extrinsèques des deux caméras. Les matrices de projection sont :

$$P_1 = [I_3, \mathbf{0}] \quad (3.65)$$

$$P_2 = [H_\Pi, a \cdot \mathbf{e}_2] \quad (3.66)$$

avec $H_\Pi \simeq H_\infty + \mathbf{e}_2 \cdot \bar{\mathbf{a}}^t$ et $[\bar{\mathbf{a}}^t, a]^t$ un vecteur arbitraire de l'espace projectif. Les deux centres de projection étant définis (3.29), nous recherchons alors trois points représentatifs de la géométrie épipolaire. Une famille d'homographies plane de type $H_\Pi + \mathbf{e}_2 \cdot \mathbf{v}^t$ vérifie la même géométrie épipolaire quel que soit le vecteur \mathbf{v} arbitraire. En imposant, $\mathbf{v} = \mathbf{0}$, nous complétons alors la base projective à l'aide de trois points de Π qui vérifient la géométrie épipolaire. Les coordonnées de ces points sont de la forme $\mathbf{p}_k = [\bar{\mathbf{p}}_k^t, \theta]^t$ avec $k \in [1, 2, 3]$.

3.4.2 Dualité entre les déplacements de caméras et d'objets

Soit un point \mathbf{M} de l'espace projectif, observé à partir d'une paire stéréo. Nous considérons que les deux caméras \mathcal{I}_1 et \mathcal{I}_2 sont liées par un mouvement rigide D_{12} et que leurs paramètres internes K_γ (quel que soit $\gamma \in [1, 2]$) sont invariants. Les coordonnées projectives \mathbf{m}_γ et euclidiennes $\widetilde{\mathbf{m}}_\gamma$ d'un point \mathbf{M} de l'espace projectif sont liées par l'homographie H_{pe} quelle que soit l'image \mathcal{I}_γ considérée :

$$\widetilde{\mathbf{m}}_\gamma \simeq H_{pe} \cdot \mathbf{m}_\gamma \quad (3.67)$$

Sachant que $\mathbf{m}_2 \simeq D_{12} \cdot \mathbf{m}_1$, il nous est alors possible de déterminer la transformation projective H_{12} qui lie les deux caméras lorsqu'elles ne sont pas calibrées :

$$H_{12} = \alpha \cdot H_{pe}^{-1} \cdot D_{12} \cdot H_{pe} \quad (3.68)$$

La détermination du déplacement projectif normalisé est envisageable car les déterminants et traces des transformations projectives sont invariants par similitudes.

$$trace(H_{12}) = trace(\alpha \cdot D_{12}) = 2\alpha(1 + \cos \theta)$$

$$det(H_{12}) = det(\alpha \cdot D_{12}) = \alpha^4$$

où θ représente l'angle de rotation associé au déplacement D_{12} . Etant donné que $1 + \cos \theta \geq 0$, le signe du scalaire α est égal à celui de $trace(H_{12})$. Nous en déduisons ainsi :

$$\alpha = sign(trace(H_{12})) \cdot \sqrt[4]{det(H_{12})} \quad (3.69)$$

Contrairement à la plupart des applications de la robotique où nous considérons qu'une caméra mobile se déplace dans un environnement statique, il est parfois intéressant de considérer une caméra fixe observant un objet qui réalise des déplacements rigides. Le mouvement apparent équivalent au déplacement de la caméra D_{12} s'avère être l'inverse du conjugué de celui de l'objet D_{12}^o (voir Fig 3.11) :

$$D_{12} = P_{pers}^{-1} \cdot D_{12}^o \cdot P_{pers} \quad (3.70)$$

où P_{pers} représente la transformation projective qui lie l'objet et la caméra, tous deux en positions initiales.

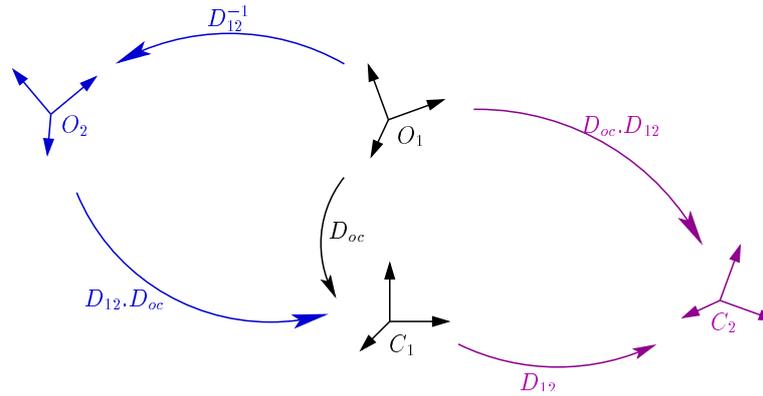


FIG. 3.11 – Le déplacement rigide d'un objet devant une caméra fixe est équivalent au déplacement d'une caméra mobile dans un environnement statique.

3.4.3 Analyse algébrique du mouvement rigide

Nous cherchons désormais à caractériser la transformation projective H_{12} à partir du déplacement rigide D_{12} [4*4]. La similitude H_{pe} peut être considérée comme un changement de base entre les repères projectif et euclidien. Dès lors, les valeurs propres $\{e^{j\theta}, e^{-j\theta}, 1, 1\}$ de la transformation euclidienne sont communes à la transformation projective.

L'ordre de multiplicité algébrique de la valeur propre $\lambda = 1$ est égal à deux sauf lorsque le déplacement est une translation pure où il est égal à 4. L'ordre de multiplicité géométrique d'une valeur propre est lié à la dimension du vecteur propre associé. Ainsi, l'ordre de multiplicité géométrique de la valeur propre $\lambda = 1$ dépend du type de mouvement D_{12} réalisé :

- il est égal à 1 si le mouvement est général,
- il est égal à 2 si le mouvement est plan : la translation est perpendiculaire à l'axe de rotation,
- il est égal à 4 si le déplacement est une translation pure ($\theta = 0$).

Il nous est donc possible de déterminer le type de déplacement réalisé à partir de la transformation projective. Il suffit pour cela de déterminer son ordre de multiplicité de la valeur propre $\lambda = 1$, c-à-d calculer le rang de la matrice $(H_{12} - I_3)$.

3.5 Conclusion

Ce chapitre nous a permis de rappeler les différents outils mathématiques de la géométrie projective que nous allons utiliser dans les prochains chapitres. Le système de vision que nous employons étant embarqué, les déplacements estimés seront supposés ceux du véhicule. La perte de l'information de profondeur due la projection de la scène 3D dans l'image 2D est le principal inconvénient du capteur de vision. L'introduction du plan à l'infini permet de justifier l'observation de spécificités propres à la géométrie projective telles que "les projections de deux droites parallèles convergent en un point" ou "les projections des points du plan à l'infini sont insensibles aux mouvements de translation de la caméra".

La modélisation de la caméra que nous employons est celle adoptée par la majorité des auteurs. Nous utilisons un banc stéréo non calibré ce qui nous permet de nous affranchir de contraintes matérielles et d'éventuelles dérives essentiellement dues à des vibrations. La contrepartie de cette robustification est que l'emploi d'un banc stéréo ne permet pas de retrouver

l'information de profondeur de points observés dans les deux images.

La géométrie épipolaire rend compte des contraintes qui lient les projections de points de la scène à partir de deux points de vues différents. Les matrices fondamentales et essentielles (dans le cas de caméra(s) calibrée(s)) sont une représentation des mouvements de translation et rotation qui lient les deux repères caméras des images considérées. La connaissance de l'une de ces matrices permet d'introduire des contraintes fortes sur les projections d'un point de la scène dans les deux images : à un point d'une image correspond une droite de l'autre image sur laquelle se situe la projection recherchée.

L'estimation de matrice fondamentale étant très sensible aux bruits de mesure, il est courant de s'affranchir de ce calcul très instable en calculant le mouvement des caméras à partir de l'observation d'un ou plusieurs plan(s). La restriction de l'environnement à une scène plane simplifie le problème car la transformation entre les deux repères caméras est une homographie, soit la transformation qui permet de passer de la projection du plan de la première image à la seconde. La contre-partie du calcul de l'homographie pour estimer les déplacements entre deux caméras est l'introduction de trois nouveaux paramètres relatifs à la projection du plan observé dans l'une des images. Le calcul d'homographie à partir de données réelles nécessite une mise en oeuvre précautionneuse afin d'optimiser le conditionnement du système linéaire formé par les coordonnées des primitives en correspondance.

Nous n'avons volontairement pas abordé dans ce chapitre la partie qui traite de la calibration et de la reconstruction de l'environnement. En effet, nous n'avons pas abordé cette question au cours de cette étude bien qu'elle apparaisse désormais une suite logique à la poursuite de ces travaux.

Chapitre 4

Caractérisation des plans principaux par stéréovision

La vision appliquée à la robotique mobile permet actuellement de faire rouler un véhicule sur plusieurs centaines de kilomètres sur autoroute ou de déplacer un robot en milieu inconnu de manière complètement autonome. En revanche, la navigation de robots mobiles en milieu urbain n'était que peu étudiée jusqu'alors en raison du caractère dynamique de l'environnement. Or, les progrès accomplis dans le domaine de la vision et la nécessité d'apporter de nouvelles solutions aux problèmes de déplacements en milieu urbain ont permis de valoriser ce type de recherches en le considérant comme un axe prioritaire.

[Gibson, 1950] (1950) a écrit "there is literally no such thing as a perception of space without the perception of a continuous background surface". Cette citation justifie à elle seule l'intérêt que nous portons à détecter les frontières qui délimitent le "sol" d'un véhicule et d'estimer par la suite son déplacement relatif au cours du temps. En effet, lors d'une tâche de navigation, la caractérisation de l'espace est avant tout une tâche de localisation du sol et de la ligne d'horizon plus que des éléments qui le constituent.

L'objectif du présent chapitre est de segmenter dans l'image les éléments structurants qui délimitent l'espace de navigation d'un véhicule en ville. Suite à une modélisation polygonale, nous identifierons dans la première partie, les plans principaux qui structurent l'environnement urbain et plus particulièrement celui de la route. Nous rechercherons dans une deuxième partie des primitives dans les régions segmentées de façon à caractériser les plans observés. La mise en correspondance des primitives détectées entre les images d'un banc stéréo nous permettra dans une troisième partie d'identifier quelles sont les primitives réellement coplanaires. Nous terminerons ce chapitre par la présentation de résultats expérimentaux à partir de différentes séquences vidéo enregistrées en environnement urbain.

Dans ce chapitre, nous allons présenter les diverses informations sur la scène qui peuvent être obtenues à partir du traitement d'un couple d'images stéréo.

4.1 Segmentation des plans principaux

L'environnement urbain est structuré par différents éléments qui délimitent l'espace de navigation des véhicules : le marquage horizontal des voies de circulation, les trottoirs, les terre-pleins représentent les frontières de la route mais le mobilier urbain (panneaux de circulation, feux-rouges, lampadaires, abri-bus) peut nous être utile en tant qu'amer naturel pour effectuer une

tâche de navigation. Cependant, les dimensions de ces derniers étant relativement réduites, seule leur structure verticale permet de les détecter puis de les identifier si leur position dans la scène est connue. Une telle approche est étudiée actuellement par [Kais et al., 2004].

Dans notre cas, nous nous sommes focalisés sur le corps de la structuration des rues : les façades des bâtiments. Les différents types de constructions (hangars, pavillons, maisons de pied, immeubles) qui longent les routes en milieu urbain sont dans leur quasi-intégralité formés de larges surfaces géométriques. Une modélisation polygonale de l'environnement permet d'extraire des surfaces planaires, facilement détectables au-dessus de l'horizon, suite à une étape de détection de contours. Les plans ainsi détectés jouent alors le rôle d'amers naturels utilisés dans un processus de localisation par estimation d'homographies planaires. Le suivi de plusieurs plans permet, en effet, de calculer la position du véhicule par rapport à chacun d'eux et d'estimer le déplacement du véhicule.

L'observation des constructions urbaines a de nombreux avantages : elles sont généralement alignées avec la direction de la route, elles se distinguent les unes des autres par des contours rectilignes orientés selon les trois directions principales que sont l'horizontale, la verticale et leur perpendiculaire. Dans le cas de route rectiligne par morceaux, il est d'ailleurs courant que la troisième direction n'est autre que celle de la route. De plus, les surfaces planaires qui constituent les façades sont rarement uniformes : de nombreuses primitives peuvent être détectées à partir de coins qui délimitent des éléments de construction (éléments de structures porteuses, portes, fenêtres, balcons) ou d'identification (enseignes, texture).

Cependant, en raison des nombreux obstacles qui réduisent le champ visuel en ville, l'observation continue de plusieurs façades statiques qui permettraient une tâche de localisation d'un véhicule est difficile. La solution que nous proposons d'employer est de se focaliser dans un premier temps sur la route, considérée plane au premier plan, comme unique support aux amers naturels, puis de rechercher dans la région de l'image extérieure à la route d'éventuels nouveaux plans qui permettront de compléter, le cas échéant, les informations extraites à partir du seul plan de la route.

4.1.1 Segmentation de la région apparentée à la route

En milieu urbain, la route représente la région de l'image la plus homogène située dans la partie basse de l'image lorsque l'axe d'une caméra embarquée est orienté selon l'axe de déplacement du véhicule. La matérialisation du milieu urbain favorise l'emploi d'un détecteur de contours pour segmenter les différents éléments qui structurent la scène et par conséquent délimiter la région de l'image apparentée à la route.

Les constructions qui bordent la route sont généralement parallèles aux limites de la route qui les traverse. Une scène urbaine contient donc de nombreux contours alignés avec les frontières de la route. En introduisant les notions de point de fuite dominant (*Dominant Vanishing Point*) et de lignes de fuite (*Vanishing Lines*), la segmentation dans l'image du ruban à bords parallèles que forme la route se résumera à déterminer parmi un faisceau de droites quelles sont les droites qui représentent ses frontières.

4.1.1.1 Extraction de contours

L'extraction de contours d'une image a pour vocation d'identifier les zones de fortes variations d'intensité des niveaux de gris dans l'image de façon à caractériser, suite à une étape de segmentation de contours, les divers éléments structurants de la scène. La détermination de frontières ou de variations d'un signal est étroitement liée à la recherche des extréma locaux du

gradient ou des zéros du Laplacien. Comme tout problème de traitement du signal, la recherche de contours dans une image est complexifiée par la présence de bruits d'origines diverses : la discrétisation spatiale et celle des niveaux de gris, les non-linéarités de la chaîne d'acquisition (disparité des optiques, bruit d'échantillonnage, etc), les irrégularités des objets observés.

Afin de déterminer le filtre adapté aux particularités des images de scènes urbaines, il est nécessaire de définir quels sont les types de contours qui seront détectés et un modèle du bruit, généralement considéré additif. La présence de bruit dans le signal impose dans un premier temps un pré-filtrage qui consiste à augmenter le rapport signal/bruit avant l'étape de différentiation. Le pré-traitement que nous opérons consiste à égaliser les niveaux de gris des images afin de rehausser le contraste. L'éclairage en milieu urbain est particulièrement complexe car il ne peut être considéré uniforme en raison de la présence de nombreuses sources lumineuses d'origine naturelle et artificielle (éclairage public, panneaux publicitaires, projecteurs de vitrines, feux de croisement des véhicules) observées de manière directe ou indirecte (reflets et ombrages) dus aux éléments statiques et dynamiques de la scène.

L'opération de dérivation pouvant être réalisée à partir de filtres linéaires, il est courant d'employer des filtres séparables récursifs pour extraire des contours car ce sont des filtres simples d'implémentation qui fournissent des résultats satisfaisants, à condition d'avoir affaire à un bruit de moyenne nulle. La séparabilité permet de traiter des signaux de manière indépendante à leurs dimensions tandis que la récursivité est à l'origine de filtres à réponse impulsionnelle infinie qui ont un coût algorithmique réduit. De plus, le filtrage est une opération de convolution entre un filtre $f(x)$ et un signal $s(x)$, choisi à une dimension pour l'exemple, dans le domaine temporel. L'opérateur de différentiation D appliqué au signal filtré est équivalent à appliquer l'opérateur différentiation au filtre avec lequel est convolué le signal :

$$D(s \otimes f) = s \otimes D(f) \quad (4.1)$$

ce qui a l'avantage d'appliquer un unique traitement à l'intégralité du signal.

Les différents type d'extracteurs de contours Les deux approches gradient et Laplacien fournissent des résultats sensiblement comparables mais s'avèrent optimales pour des contextes différents. L'approche Laplacien étant calculée à partir de dérivées secondes a tendance à arrondir les angles tout en maintenant une localisation exacte des coins. Elle est cependant plus sensible au bruit qu'une approche gradient.

Afin de valider notre choix, nous avons extrait 8 images que nous considérons représentatives de la diversité des cas de figures qui peuvent être rencontrés lors de déplacements en milieu urbain pour comparer les résultats des différents filtres. Les images retenues représentent différents types de route : de la simple ruelle en sens unique à des doubles voies de pénétrantes, la présence d'obstacles (piétons, véhicules) ou des conditions de luminosité extrêmes sont également prises en compte.

Nous invitons le lecteur curieux de connaître l'origine des extracteurs de contours de lire le chapitre deux du livre de [Horaud, 1998] qui relate fort bien la problématique et les innovations qui ont été développées en ce domaine ces vingt dernières années. Pour notre part, nous avons appliqué trois extracteurs de contours Sobel, Laplacien et Canny à chacune de ces images et reporté le résultat dans le Tab.4.1. Dans un premier temps, notre objectif est de déterminer lequel de ces filtres est le mieux adapté à la future étape de segmentation de contours de manière à discriminer sans ambiguïté les différents éléments structurants des diverses scènes.

Le choix des seuils a été calculé de manière automatique pour chaque image. Nous reportons dans le Tab. 4.1 les valeurs retenues : la disparité entre les seuils varient suivant les filtres entre

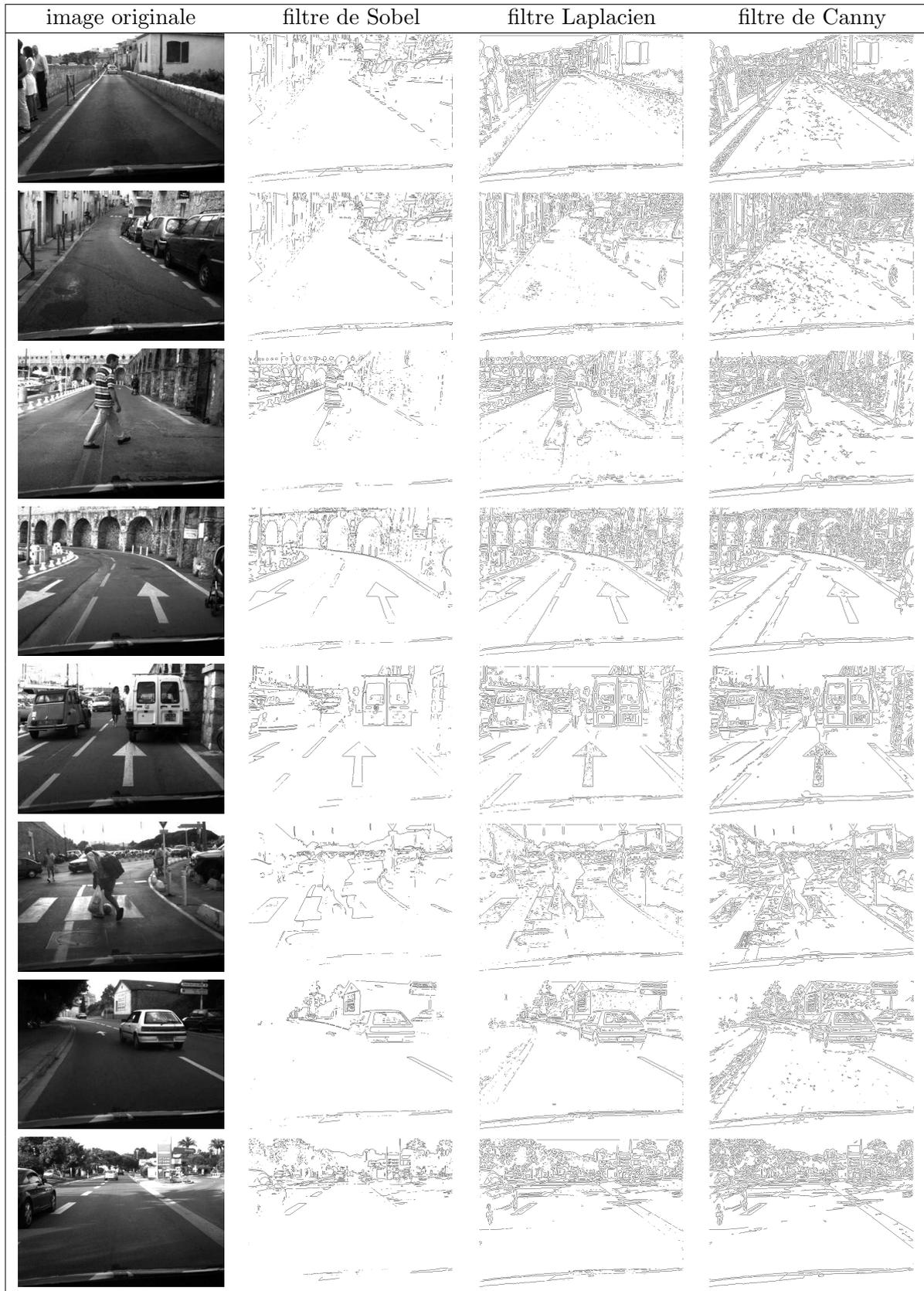


FIG. 4.1 – Comparaison des résultats de différents extracteurs de contours sur différentes images de scènes urbaines.

25% et 50%. La qualité des extractions de contours est d'ailleurs difficilement comparable. Il apparaît néanmoins que la détermination d'un seuil fixe adaptés aux diverses images est délicat : trop bas des éléments de texture et de structures sont détectés, trop élevé certaines arêtes sont peu ou mal distinguées.

Images	Sobel	LOG	Canny	
<i>seuils de ref.</i>	27,745	1,3844	0,0375	0,0937
L180015.0775	0,74	0,67	0,5	0,5
L180441.0370	0,91	0,85	0,83	0,83
L180841.0482	0,95	0,93	0,83	0,83
L180841.0645	1	1	1	1
L180841.0905	0,97	0,89	0,67	0,67
L180841.1005	0,83	0,76	0,67	0,67
L184340.0510	0,78	0,6	0,5	0,5
L181328.1835	0,93	0,82	0,67	0,67

TAB. 4.1 – *Disparité relative des seuils utilisés pour chacune des images-tests en fonction du filtre employé. Les seuils ont été déterminés de manière automatique pour chacun des filtres. Les seuils obtenus pour la 4^{ème} image étant les plus élevés, ils nous servent de référence. Les variations des seuils fluctuent respectivement de 25%, 40% et 50 % selon les trois filtres proposés. Les filtres de Sobel et LOG (Laplacien) ne conservent que les contours dont les valeurs dépassent respectivement les seuils de la première et 2^{ème} colonne. Les contours de Canny sont extraits à partir du franchissement par hystérésis des deux seuils (bas et haut) de la 3^{ème} et 4^{ème} colonne.*

Afin de faciliter la comparaison entre les résultats des trois filtres, nous nous focalisons sur l'extraction des contours du marquage de la signalisation horizontale. Le filtre de Sobel exploitant le résultat de deux filtrages selon les directions horizontale et verticale parvient difficilement à extraire les contours obliques qui constituent la majorité des marquages. Le filtre Laplacien fournit de meilleurs résultats, ils sont d'ailleurs comparables à ceux obtenus par filtrage de Canny pour les éléments qui structurent la route.

Le détecteur de Canny-Deriche Notre objectif premier étant d'extraire les contours du marquage de la signalisation horizontale, essentiellement représentés par des bandes rectilignes, notre choix s'est finalement porté sur un détecteur de type [Canny, 1986] car ce détecteur s'avère optimal en cas de bruit gaussien et optimise les trois critères de performances que sont :

- la détection : l'opérateur doit donner une réponse au voisinage d'un contour,
- la localisation : le contour doit être localisé avec précision,
- l'unicité de la réponse : un contour doit provoquer une seule réponse de l'opérateur.

Outre la comparaison précédente sur les images test, l'intérêt d'employer un tel filtre comparé à son concurrent direct qu'est le filtre de Sobel, plus simple d'implémentation et donc plus rapide, est la possibilité de recueillir une information sur la direction du gradient de niveau de gris (clair/foncé ou foncé/clair) et de pouvoir chaîner les segments de contours extraits grâce à l'emploi d'un seuillage à hystérésis et reconstituer des contours fermés (cf 2^{ème} colonne du Tab. 4.2). Nous avons cherché à optimiser les différents paramètres de configuration du filtre de Canny-Deriche pour obtenir une segmentation la plus claire possible des contours du marquage au sol. Nous avons retenu les seuils bas de trois *pixels* et haut de 40 *pixels* avec une variance

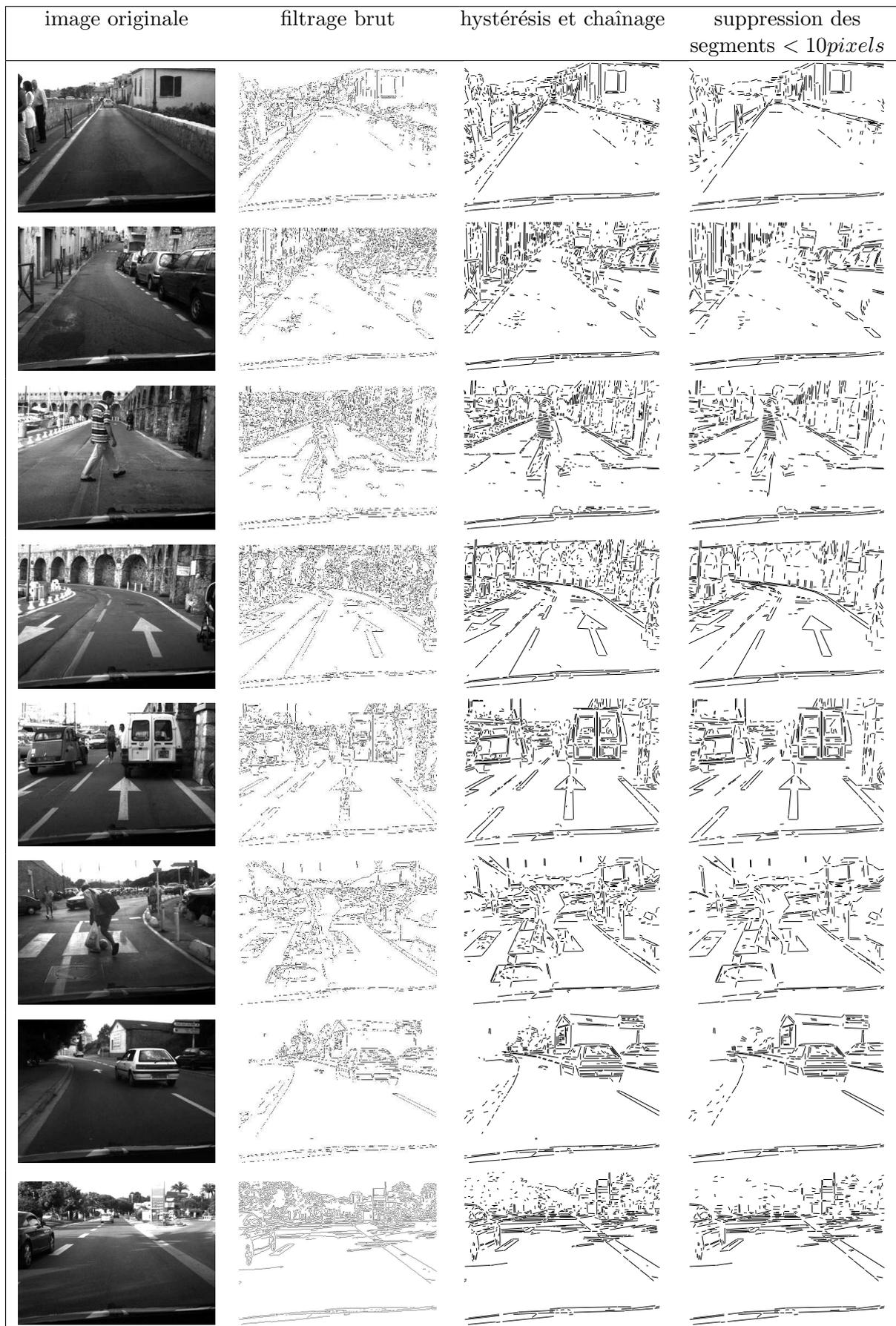


FIG. 4.2 – Détections de contours par filtrage de Canny-Deriche.

du filtre gaussien de lissage égale à 1 pixel. L'ensemble fournit des résultats très satisfaisants. Nous supprimons dans la dernière colonne les segments de longueur inférieure à 10 *pixels*.

Le filtre de Canny-Deriche implémenté dans notre algorithme a pour équation :

$$\begin{cases} f(x) &= k (1 + \alpha|x|) \exp^{-\alpha|x|} \\ k &= \frac{(1-e^{-\alpha})^2}{1+2\alpha e^{-\alpha}-e^{-2\alpha}} \end{cases} \quad (4.2)$$

où α représente la largeur du filtre, exprimée en pixel, soit le compromis détection-localisation désiré. Il représente une solution exacte de l'équation de Canny et est donc couramment employé en tant que filtre de référence. Cependant, comme nous le verrons plus bas, la qualité des contours extraits est perfectible. La recherche de point de fuite communs à plusieurs contours orientés selon la même direction s'impose afin d'améliorer le résultat de la détection de contours.

4.1.1.2 Le point de fuite dominant

De nombreux auteurs exploitent le fait de travailler en environnement structuré pour rechercher les points de fuite selon les trois axes principaux pour affiner l'orientation des segments détectés et éventuellement réaliser une auto-calibration de la caméra. Ce n'est hélas pas envisageable dans notre cas : nous disposons généralement de nombreuses verticales mais rares sont les cas où le système de vision pourra observer avec certitude des contours orientés selon l'axe perpendiculaire au déplacement du véhicule. De tels contours sont observables en présence d'un passage piéton (cf Fig4.3) ou de croisements de routes perpendiculaires, plus répandues dans les métropoles nord-américaines qu'européennes.

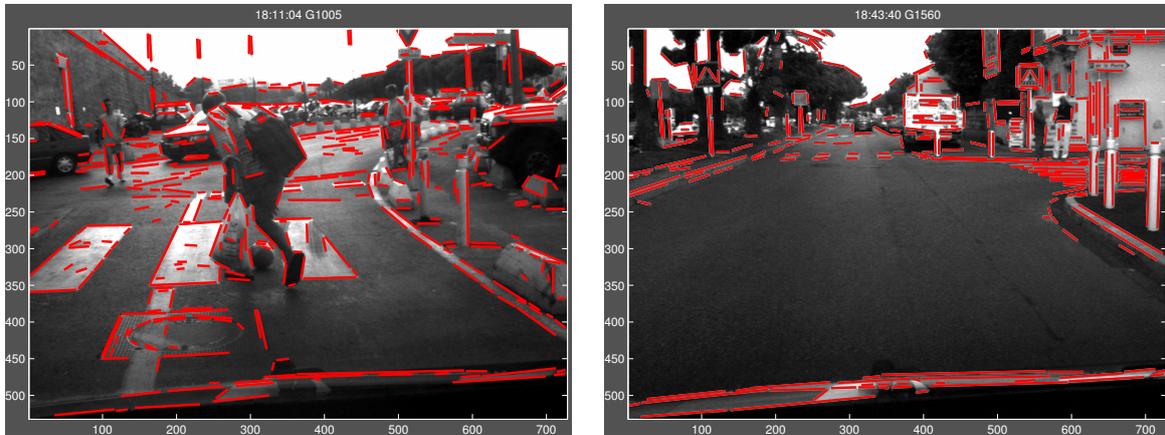


FIG. 4.3 – Exemples de rares scènes où des contours orientés selon la direction perpendiculaire à la direction de la route peuvent être détectés. A gauche, présence d'un passage piéton, à droite intersection de deux routes perpendiculaires.

La segmentation dans l'image de la projection P_{roj} des arêtes de l'environnement parallèles à la direction $[\mathbf{D}, 0]^t$ est facilitée par une particularité de la vision projective : le point de fuite \mathbf{x}_D . Ce point est la projection sur le plan image du point \mathbf{X}_D du plan à l'infini Π_∞ vers lequel converge toutes les droites-supports de direction commune \mathbf{D} (cf 3.1.3.3). Ce point remarquable est invariant aux mouvements de translations de la caméra et renseigne donc sur ses seuls mouvements de rotations R : la 4^{ème} coordonnée homogène des points du plan Π_∞ étant nulle,

la contribution de la translation \mathbf{T} lors de la projection sur le plan image est annulée :

$$\mathbf{x}_D \simeq P_{\text{roj.}} \cdot \mathbf{X}_D \simeq K \left[\begin{array}{c|c} R & \bar{\mathbf{T}} \\ \mathbf{0}^t & 1 \end{array} \right] \left| \begin{array}{c} \bar{\mathbf{D}} \\ 0 \end{array} \right. \simeq KR \cdot \bar{\mathbf{D}} \quad (4.3)$$

Nous supposons qu'une majorité de contours de la scène sont orientés selon la direction de la route. Le point de fuite associé à cette direction est donc qualifié de dominant (*Dominant Vanishing Point*). L'estimation de ses coordonnées s'impose alors pour faciliter et rendre plus robuste la tâche de sélection des contours parallèles à la route, sachant que par hypothèses, ses coordonnées évoluent lentement du fait qu'elles sont liées au déplacement relatif de la route dans l'image.

L'extraction du point de fuite dominant est donc soumis à deux hypothèses : une majorité des segments extraits convergent vers sa position, qui n'évolue que si le véhicule subit des mouvements de rotation. L'exploitation des coordonnées du point de fuite dominant est donc double : il permet une segmentation rapide de la scène et l'estimation simultanée des mouvements de rotation de la caméra.

L'approche que nous proposons se déroule en deux étapes :

1. la recherche des droites-supports parallèles à la direction de la route à partir des segments extraits qui semblent converger vers la prédiction du DVP,
2. l'estimation des coordonnées du DVP en sélectionnant par un processus itératif d'optimisation les contours dont les droites-supports des projections sont les plus proches du DVP.

4.1.1.3 Extraction de droites-supports

Le résultat de la détection de contours sur une image représentative de l'horizon observable dans le cadre de scène urbaine est proposée en Fig. 4.4. La détection de contours autour des bandes linéaires et continues du marquage de la signalisation horizontale est formée de plusieurs segments de tailles diverses. Les caractéristiques ρ et θ mesurées sont proches mais ont des variances parfois importantes. La longueur et l'orientation des segments extraits sont bruitées. Dans la majorité des cas, seul le milieu du segment se situe sur le contour détecté. La précision des coordonnées des extrémités des segments est au mieux pixellique en raison de la discrétisation du plan image. Néanmoins, nous remarquons que les segments les plus longs sont les mieux localisés.

Nous vérifions sur la même figure, que le résultat de la détection de contours linéaires diffère d'une image à l'autre. En raison de l'emploi de seuils à hystérésis et de l'approximation polygonale de l'étape de chaînage, un contour linéaire est rarement détecté par un unique segment dont les extrémités correspondent avec celles du contour. Il est également courant qu'un contour de l'environnement soit représenté par un segment dans une image et plusieurs dans l'autre. Il n'est donc pas envisageable de chercher à mettre en correspondance des contours des éléments structurants de la scène du fait du manque de fiabilité de leurs caractéristiques.

Approche proposée L'ensemble de ces constatations nous a conduit à la réflexion suivante : nous avons tout intérêt à regrouper les segments issus d'un même contour linéaire de manière à former un segment-composé dont les caractéristiques seront calculées en fonction de celles des segments regroupés. Nous améliorons ainsi la résolution des caractéristiques des droites-supports formées tout en augmentant la représentativité de chacune d'elles. La représentativité

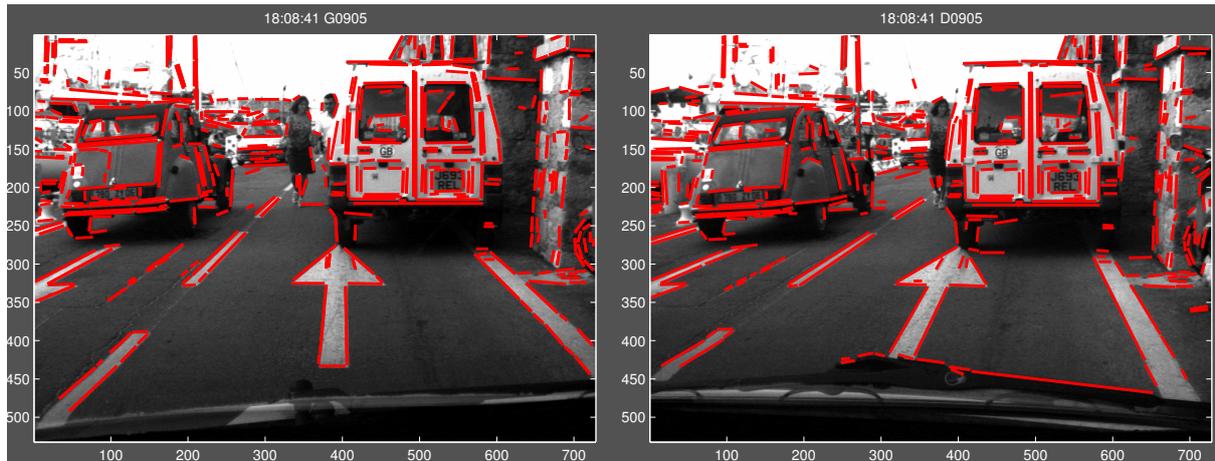


FIG. 4.4 – Contours détectés par le filtrage de Canny sur une paire d’images stéréo. Si la flèche de la voie de droite est correctement détectée à l’aide de contours quasiment fermés, il n’en est pas de même pour les bandes blanches centrale et droite qui délimitent les voies de circulation. Bien que rectilignes, les contours sont détectés par de multiples segments non-superposables. La plupart des écarts constatés apparaissant au 1^{er} plan, il est probable qu’il s’agisse d’un problème de distorsion des optiques.

est définie comme la somme des longueurs des segments qui composent une droite-support : c’est une image de la fiabilité que l’on peut accorder à ses caractéristiques.

La plupart des auteurs qui utilisent une segmentation par contours emploient la transformée de Hough, méthode robuste mais réputée coûteuse en temps de calcul, pour distinguer les arêtes parallèles de l’environnement 3D. Nous proposons d’employer une méthode alternative qui exploite le fait que les projections dans l’image de ces arêtes parallèles convergent vers le DVP. Les coordonnées du DVP \mathbf{x} sont calculées de manière itérative en minimisant au sens des moindres carrés pondérés la distance qui sépare les droites-supports des coordonnées estimées $r\mathbf{x}$ (r étant le numéro de l’itération), toutes exprimées dans le repère image. Nous exploitons pour cela deux connaissances a priori sur l’environnement :

- d’une image à l’autre, l’évolution des paramètres liés à la route évoluent lentement : nous pourrions donc utiliser une prédiction du DVP $\tilde{\mathbf{x}}$ et des coordonnées des droites-supports qui convergent dans sa direction,
- le premier plan étant libre d’obstacles, les contours qui sont détectés dans cette région appartiennent généralement à des éléments structurants statiques de la scène.

La qualité des segments extraits est mauvaise comme le montrent les différentes figures précédentes : les coordonnées de leurs extrémités sont souvent éloignées de plusieurs *pixels* du contour qu’elles représentent. Dans l’éventualité où nous parviendrions à identifier des contours représentatifs d’arêtes de la scène parallèles à la direction de la route, l’intersection de leurs droites-supports ne sera pas unique et fournira un nuage de points plus ou moins dispersé autour du DVP réel. Une étape de fusion des segments correspondants aux mêmes contours physiques est donc indispensable afin de faire émerger les droites-supports représentatives de la direction de la route.

Méthodologie Notre objectif est d’identifier les contours de l’environnement orientés dans la direction de la route. Nous supposons que parmi les segments détectés, plusieurs d’entre eux

correspondent à des contours de la scène identiques. Nous allons expliciter une méthode qui permet de reconstituer les droites-supports, faisceau de droites formées de segments extraits qui convergent vers le DVP. Chaque segment S_k issu de la détection de contours est caractérisé par son orientation θ_k et sa distance à l'origine ρ_k :

$$\rho_k = \cos\theta_k \cdot u + \sin\theta_k \cdot v \quad (4.4)$$

où (u, v) sont les coordonnées en *pixels* d'un point de l'image, l'origine du repère se situant dans le coin supérieur gauche de l'image. Les caractéristiques des droites-supports recherchées sont alors calculées en tant que droites qui minimisent au sens des moindres carrés la distance des extrémités des groupes de contours, pondérés par la longueur des segments, exprimée en *pixels*.

Nous distinguons trois groupes parmi les segments extraits en fonction de leur orientation : les (quasi-)horizontaux, les (quasi-)verticaux, les autres. Nous ne tenons pas compte des segments (quasi-)horizontaux car la probabilité de pouvoir observer la ligne d'horizon en ville est très faible. Les segments (quasi-)verticaux ont principalement deux origines. Il s'agit soit de projections de contours verticaux d'obstacles (statiques ou mobiles), soit dans l'éventualité où les segments sont localisés entre les deux limites du bord de route, de la projection d'une bande du marquage au sol. Fort de ce constat, seuls les segments verticaux dont une des extrémités est suffisamment proche de l'abscisse de la prédiction du DVP \tilde{x} seront conservés. Ces deux groupes de contours à orientations particulières interceptent en effet naturellement les supports des autres contours et peuvent engendrer des erreurs dans le processus de détection du DVP.

La reconstitution de toutes les droites-supports de l'image est inutile et fastidieuse, nous nous intéressons qu'à celles qui sont susceptibles de converger vers le DVP. Une phase de présélection est opérée parmi les contours restants : elle consiste à ne conserver que les segments qui sont orientés en direction de la prédiction des coordonnées du DVP. Le plus juste serait de calculer l'écart angulaire qui sépare l'orientation de chaque segment de sa droite-support théorique qui passe par le milieu du segment et les coordonnées prédites du DVP. Cette méthode nécessite un calcul d'inverse d'angle ou de tangente qui ne se justifie aucunement à ce stade du fait de l'incertitude qui demeure sur les coordonnées des extrémités des segments et de la prédiction du DVP.

La présélection s'effectue donc en mesurant la distance qui sépare chaque droite-support de la prédiction du DVP. Le seuil retenu pour présélectionner les contours est très lâche ($5 \cdot dist_max \text{ pixels}$) de manière à n'écarter que les segments qui ne convergent pas dans la région de la prédiction du DVP. La spécificité des images enregistrées dans des conditions de trafic urbain nous contraint de surcroît à éliminer des segments qui se situent dans la région du DVP. Nous éliminons donc les segments qui sont entièrement contenus dans une zone de $seuil_secu = 100 \text{ pixels}$ autour de la position prédite du DVP et dont l'orientation ne correspond pas à celle d'une droite-support détectée à l'itération précédente. Ces segments ont en effet une forte probabilité de représenter les arêtes d'obstacles et peuvent perturber l'estimation du DVP.

Le regroupement des différents segments présélectionnés en droites-supports s'effectue selon une procédure établie qui cherche à reconstituer les contours réels de l'image à partir de segments qui ont des caractéristiques proches sans pour autant s'appuyer sur des contraintes de voisinage. Le cas le plus difficile auquel nous sommes confrontés correspond à la reconstitution des droites-supports qui délimitent les bandes blanches discontinues. Ces dernières sont formées de contours dont les extrémités peuvent être éloignées de plusieurs dizaines de *pixels*.

De plus, il n'est pas possible d'assurer un niveau de gris uniforme de part et d'autre des segments extraits pour deux raisons. La première, les segments traversent souvent le contour

qu'ils représentent. La seconde, sur des contours de plusieurs centaines de *pixels*, la luminosité peut évoluer de manière graduelle dans chacune des deux régions. C'est pourquoi, nous regroupons de manière itérative les segments retenus lors de la présélection initiale en appliquant les critères de sélection suivants :

- les segments les plus longs sont les plus fiables,
- la direction du gradient des niveaux de gris,
- l'orientation θ_k ne doit pas varier de plus d'un seuil fixé $2.seuil_theta = 20^\circ$,
- la position par rapport à l'horizon,
- le respect du critère de compatibilité : les segments qui présentent des coordonnées communes sont qualifiés d'incompatibles. Les segments restants sont regroupés en recherchant parmi les incompatibles un segment susceptible de représenter la même droite-support (cf Fig.4.5),
- les distances entre une droite-support et les extrémités d'un segment compatible doivent être faibles.

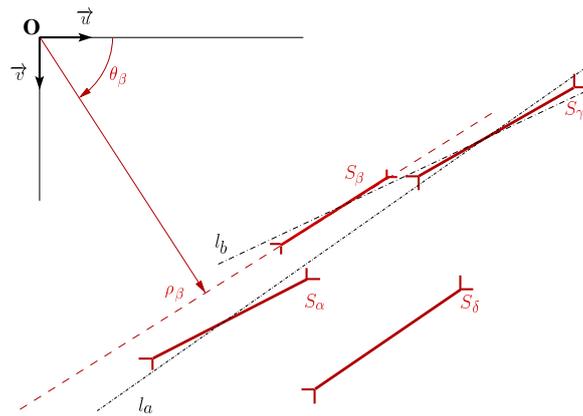


FIG. 4.5 – Regroupement des segments représentant un même contour. Les 4 segments présentés ont un gradient identique. Les trois segments ($S_\alpha, S_\beta, S_\delta$) ayant des coordonnées (abscisses et/ou ordonnées) communes, ils ne peuvent représenter le même contour de la scène : ils sont incompatibles. Par contre, le segment S_γ présente des caractéristiques (ρ_k, θ_k) similaires à ceux des segments S_α et S_β . Il est probable que S_γ appartienne à l'un des contours représentés par S_α et S_β . Seul le calcul des coordonnées des droites-supports composées l_a et l_b permet alors d'identifier le segment référent en supposant que la prise en compte de S_γ modifie les caractéristiques de la droite-support dans des proportions limitées.

Les segments issus de la présélection initiale sont répartis en groupes de segments distincts différenciés en fonction du gradient et de la position en fonction de la ligne d'horizon. L'agrégation de chacun des quatre groupes de segments est régie par l'algorithme suivant :

tant que le groupe de segments n'est pas vide :

1. recherche du segment référent (le plus long du groupe),
2. sélection des segments compris dans le tube de $2 \cdot dist_max$ de rayon, centré sur le segment référent, et ayant une orientation comparable à ($2 \cdot seuil_theta$) près,
3. recherche des couples de segments incompatibles puis formation de groupes de segments compatibles,
4. calcul de la distance qui sépare les extrémités des segments restants avec les coordonnées des droites-supports,
5. agrégation itérative des segments restants en fonction de la distance qui sépare leurs extrémités des droites-supports,
6. les segments non-agrégés dont les distances mesurées sont supérieures au seuil $2 \cdot dist_max$ sont considérés comme des droites-supports à un unique segment,
7. suppression des segments affectés du groupe.

fin tant que

Estimation des coordonnées du point de fuite dominant Les coordonnées courantes du DVP sont recherchées de manière itérative à partir du jeu de droites-supports formé précédemment. Deux groupes se distinguent parmi ces dernières : les droites-supports dites de confiance avec lesquelles est initié le calcul des coordonnées du DVP et les autres. Les droites-supports de confiance sont supposées appartenir à l'environnement statique de la scène (un des segments qui les constitue a une extrémité dans la région de confiance) et/ou ont une orientation proche d'une droite-support qui convergeait à l'itération précédente vers le DVP et qui était mise en correspondance dans les deux images du couple stéréo.

Le calcul des coordonnées du DVP \mathbf{x} consiste à minimiser la distance entre chaque droite-support \mathbf{l}_m retenue. Nous résolvons pour cela le système linéaire suivant au sens des moindres carrés pondérés :

$$\min_{\mathbf{x}} \sum_{m=1}^M (w_m \cdot \mathbf{l}_m^t \cdot \mathbf{x})^2 \quad (4.5)$$

avec $\mathbf{l}_m^t = [\cos(\theta_m), \sin(\theta_m), -\rho_m]$ et $\mathbf{x}^t = [u_x, v_x, 1]$ et w_m correspond à la somme des longueurs des segments, exprimée en *pixels*, qui forment les segments composés. Cette étape achevée, une vérification est menée sur chaque droite-support candidate en calculant leur résidu, c-à-d la distance qui les sépare de l'estimée du DVP. Seules les candidates ayant un résidu inférieur à $dist_max(10 \text{ pixels})$ sont conservées pour une nouvelle résolution du système. Ce seuil permet de tenir compte de l'imprécision des contours détectés. L'algorithme rompt la boucle itérative dès que plus aucune droite-support n'est ajoutée ou supprimée de la liste des M candidates.

Le nombre d'itérations est toutefois limité à 4 de manière à s'affranchir d'éventuel problème lié à l'oscillation entre deux valeurs distinctes de l'estimée du DVP. Nous n'avons jamais été confronté jusqu'alors à un tel cas car l'extraction de droites-supports au-dessus de l'horizon permet généralement d'identifier des contours supplémentaires qui lèvent toute indétermination.

En cas de forts mouvements de rotation du véhicule entre deux images de la séquence, la prédiction $\tilde{\mathbf{x}}$ peut être éloignée de la position du DVP \mathbf{x} . Or, la présélection des segments retenus pour former les droites-supports dépend de la qualité de la prédiction $\tilde{\mathbf{x}}$. C'est pourquoi,

l'intégralité des procédures précédentes est alors reprise tant que prédiction et estimation seront distantes de plus de $dist_max$.

Les droites-supports qui ont permis l'estimation des coordonnées du DVP ne convergent pas exactement vers ce dernier. Afin de minimiser les erreurs de mesure et de détection, les coordonnées théoriques de chaque droite-support sont calculées de façon à former un réel faisceau de droites. La projection théorique de chaque droite-support est représentée par la droite qui passe par le DVP et le milieu du segment extrait le plus long.

La séquence des opérations nécessaires à l'extraction des droites-supports qui convergent vers le DVP à partir des segments issus de la détection de contours peut être résumée grâce à l'algorithme fonctionnel suivant :

1. *extraction des contours,*
2. **tant que** $dist(\mathbf{x}, \tilde{\mathbf{x}}) \geq dist_max$,
 - (a) *présélection des contours en fonction de leur position dans l'image, leur orientation et leur distance de $\tilde{\mathbf{x}}$,*
 - (b) *recherche des droites-supports, puis sélection des droites-supports de confiance,*
 - (c) *estimation des coordonnées \mathbf{x} du DVP à partir des droites-supports de confiance,*
- fin tant que,**
3. *sélection des droites-supports qui convergent vers l'estimée \mathbf{x} du DVP,*
4. *calcul des coordonnées théoriques des droites-supports.*

Récapitulatif de l'extraction des contours à l'estimation des coordonnées du DVP

Le détail des différentes opérations effectuées entre la segmentation de contours et l'extraction des droites-supports ainsi que les coordonnées du DVP est illustré sur la Fig.4.6 :

1. une présélection des contours détectés par le détecteur de Canny est réalisée (Fig. 4.6(a)). Seuls les contours rouges sont susceptibles d'appartenir à des arêtes parallèles à la direction de la route,
2. les droites-supports extraites à partir de ces segments (Fig. 4.6(b)) se classent en deux catégories : celles dites de confiance (en vert) car au moins un contour est détecté dans la zone de confiance sous l'horizon et les autres (en bleu). Les coordonnées du DVP calculées à partir des droites-supports de confiance ('*' rouge) étant distantes de plus de $dist_max$ des coordonnées prédites ('+' rouge), l'extraction des segments et des droites-supports est reprise en employant comme prédiction l'estimation courante du DVP,
3. de nouveaux contours apparaissent, tandis que d'autres sont rejetés (Fig. 4.6(c)),
4. une estimation fine des coordonnées du DVP est alors calculée (Fig. 4.6(d)) à partir des droites-supports vertes dont la distance à l'estimation courante du DVP est inférieure à $seuil_dx$, matérialisé par le cercle rouge. Les droites-supports dont l'orientation théorique s'écarte de plus de $seuil_theta/2$ (5°) sont rejetées,
5. les coordonnées théoriques (orange) des droites supports (Fig. 4.6(e)) sont substituées aux mesurées (bleues ou vertes),
6. l'ordonnancement du faisceau de droites-supports en fonction des droites-supports de gradient positif (tirets rouges) et négatif (tirets bleus) permet de distinguer des bandes (suc-

cession de lignes continues rouge puis bleue). Ces bandes (Fig. 4.6(f)) seront détectées dans les deux images du couple stéréo et vont nous aider à mieux caractériser les faisceaux extraits.

Il est à noter qu'à l'image, des contours peuvent être dirigés en direction du DVP sans que leurs supports dans l'environnement 3D soient effectivement orientés dans la direction de la route. C'est le cas de la première droite support sous l'horizon en partant de la gauche, le contour détecté appartient à l'essuie-glace du véhicule en stationnement. Ce contour est donc perpendiculaire à la direction de la route mais n'a pas été rejeté en tant que segment horizontal car il ne vérifiait pas les critères correspondants.

Cette erreur d'interprétation de l'image sera néanmoins sans répercussions notoires tant que les fausses détections de ce type de contours demeureront minoritaire en comparaison de vrais contours parallèles à la direction représentée par le DVP. Nous pourrions en effet vérifier la cohérence de chaque contour retenu en vérifiant leur convergence vers le DVP au cours du temps. Seuls les contours parallèles à la direction de la route convergent de manière continue vers le DVP.

L'estimation des coordonnées du DVP par la présente méthode repose sur l'hypothèse forte qu'il est possible d'extraire dans les séquences urbaines des contours relatifs à la direction de la route suivie : que ce soit les délimitations naturelles du plan de la route (trottoir, terre-plains) ou artificielles (marquage au sol). Il va de soi que si aucun des deux types de contours n'est détectable, l'application de la méthode est alors contestable.

Cependant dans le cas d'une configuration de type canyon urbain, les contours détectés sur l'ensemble de l'environnement peuvent permettre de compenser une perte temporaire de contours liés au plan de la route, pour peu qu'ils soient alignés avec celle-ci. La contre-partie de la prise en compte des contours au dessus du plan de la route est d'estimer un point de fuite qui ne correspond plus à la direction de la route si le nombre de contours orientés selon la direction de la route s'avèrent minoritaires. Afin de réduire la probabilité de cette éventualité, nous présentons dans la prochaine partie comment privilégier les contours extraits sur le plan de la route.

4.1.1.4 Les lignes de fuite

La terminologie de lignes de fuite (Vanishing Lines) est normalement réservée à la projection sur le plan image de l'intersection d'un plan avec le plan à l'infini. La ligne de fuite la plus commune étant l'horizon. Dans notre cas, la probabilité d'observer des projections du plan à l'infini est faible. C'est pourquoi, nous appellerons abusivement lignes de fuite les droites-supports du faisceau de droites qui convergent vers le DVP et qui appartiennent au plan de la route. Ces droites-supports particulières, projections de droites parallèles et coplanaires, jouent un rôle important dans la suite du processus. Elles vont premièrement nous permettre de segmenter la région de la route dans l'image puis, dans un second temps, seront utilisées comme primitives coplanaires pour faciliter les estimations d'homographies entre couples d'images.

Toutefois, afin de respecter la cohérence de notre propos, nous utiliserons la terminologie de VL lorsque nous serons assuré que les représentations d'une droite-support dans les images courantes \mathcal{I}_g^n et \mathcal{I}_d^n vérifie l'homographie stéréo \mathbf{H}_{st}^n .

La notion de bandes La caractérisation des VLs est difficile car seules les informations de gradient et d'orientation sont exploitables. La recherche de couples de VLs opposées par le gradient qui forment une "bande" permet une sectorisation avantageuse du faisceau de VLs,

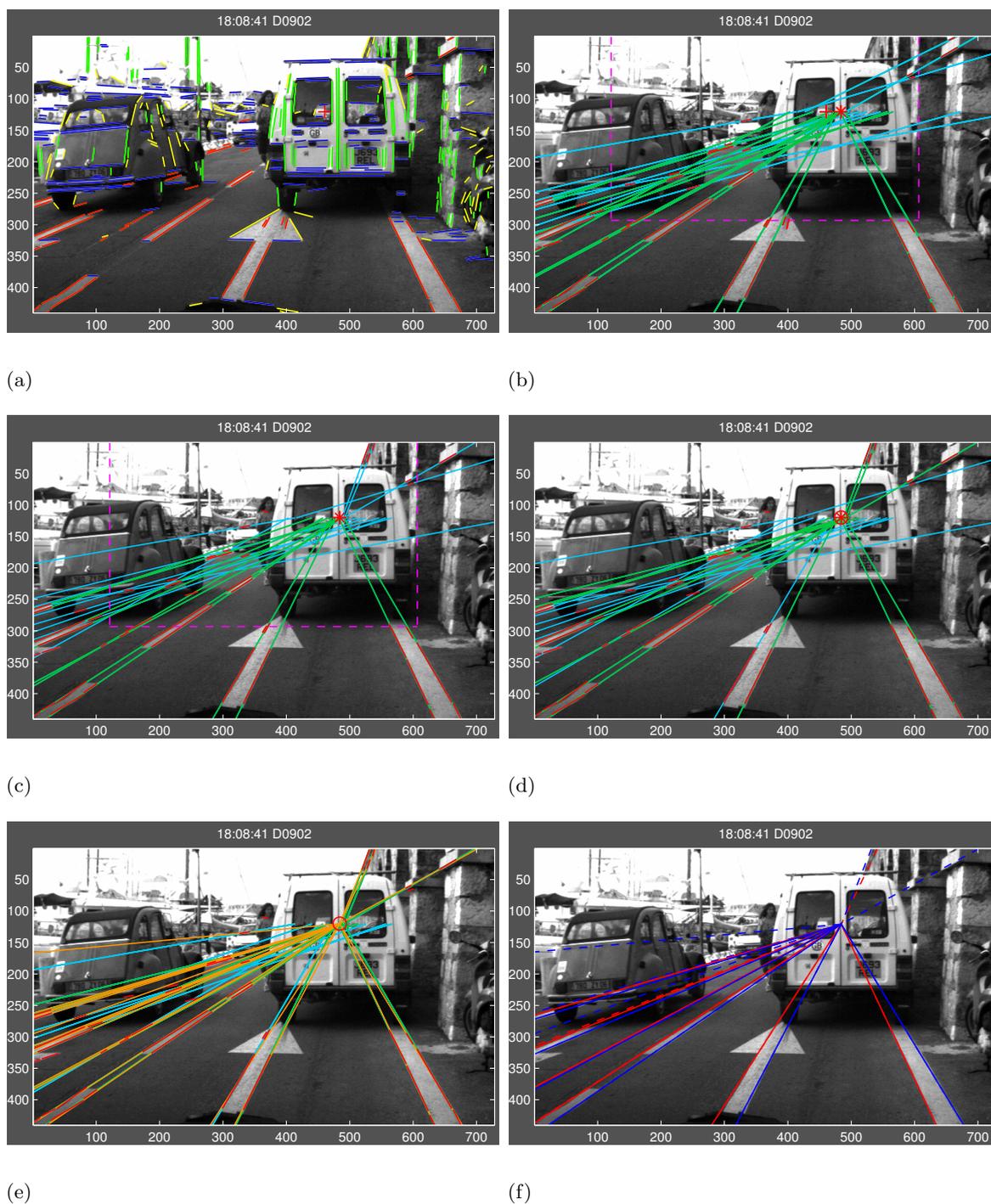


FIG. 4.6 – *Détail des différentes opérations nécessaires pour extraire le faisceau de droites-supports à partir des contours détectés par le filtre de Canny.*

d'autant plus que la majorité des VLs sont issues de la détection de contours autour de bandes du marquage au sol. Une bande est donc formée de deux droites-supports consécutives dont les gradients sont opposés et l'écart angulaire inférieur à seuil $d\theta_{max}$ que nous allons préciser. Cette description relativement sommaire permet d'identifier les projections dans le plan image de la plupart des bandes de la scène et notamment celles détectées sur la route.

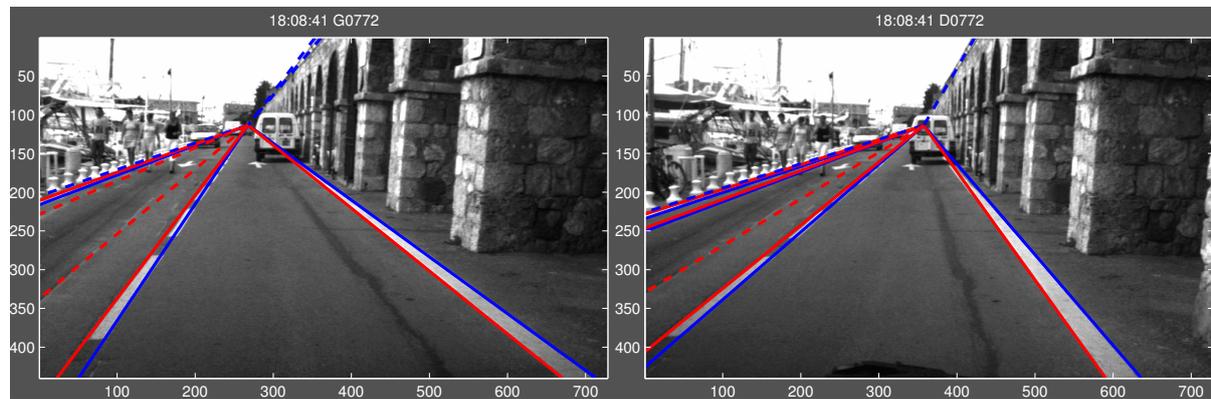


FIG. 4.7 – Représentation à l'aide des coordonnées théoriques des droites-supports qui convergent vers le DVP. Les droites-supports sont représentées en fonction du gradient des segments qui les composent : en rouge transitions de type foncée-claire, en bleu transitions de type claire-foncée. Les bandes sont représentées en traits continus, les droites-supports isolées en pointillés.

n°	1	2	3	4	5	6	7	8	9	10
$grad$	0	1	0	1	0	1	1	0	1	0
θ	72.61	72.43	71.42	69.66	69.11	58.961	50.77	48.90	-35.55	-40.53
$\Sigma[pix]$	67,00	174,26	235,96	29,73	69,02	104,01	221,11	227,2	337,9	358,97
$n^\circ bande$		1		2			3		4	
$d\theta[^\circ]$		1,01		0,55			1,87		4,98	
$Rapport[\%]$		26,1		57,2			2,6		5,3	

TAB. 4.2 – Caractéristiques des droites supports puis des bandes extraites de l'image D0772. Seuls l'orientation et le gradient des droites-supports permettent d'identifier de manière fiable les bandes d'un faisceau de VLs. Le rapport des représentativités des droites-supports qui forment les bandes évolue entre 2,6% et 57,2% : cette mesure ne peut être utilisée comme un critère de regroupement des droites-supports de gradient opposé pour former une bande.

Nous présentons sur la Fig. 4.7 le résultat de la décomposition de bandes à partir du faisceau de droites-supports. Les droites-supports isolées correspondent à des bandes dont un contour n'est pas détecté (cas de la deuxième bande en partant de la gauche de l'image droite) ou à des contours dont l'ouverture est trop importante pour représenter une bande de la signalisation horizontale. Aucune recherche de bandes n'est menée à partir des droites-supports situées au-dessus de l'horizon car nous ne disposons pas à ce stade d'homographies qui permettraient de les mettre en correspondance.

Les caractéristiques (gradient, orientation et représentativité) des droites-supports extraites de l'image de la caméra droite sont reportées dans le Tab. 4.2. Bien que les bandes de la signalisation horizontale soient de largeur comparable, les écarts angulaires de leurs représentations dans l'image (bandes 1,3 et 4) varient en fonction de leur proximité de la caméra. Le calcul des

caractéristiques des bandes détectées permet de vérifier que les droites-supports qui forment des bandes peuvent être considérées comme deux faisceaux de droites-supports quasiment superposables, de gradient opposés, décalés de quelques degrés.

Bandes principales et supplémentaires Parmi l'ensemble des droites-supports détectées, il est possible d'identifier rapidement celles associées au marquage de la signalisation horizontale. Les bandes peintes, généralement blanches ou jaunes selon les pays, sont en effet de largeur réglementée. Parmi les bandes détectées, nous nous intéressons particulièrement à celles que nous qualifierons désormais de principales qui délimitent les voies de circulation.

Les bandes de la signalisation horizontale sont à l'origine de bandes coplanaires détectées dans chacune des images et mises en correspondance grâce à l'homographie du plan de la route qui lie les images du couple stéréo. Cette dernière étant inconnue à l'initialisation de l'algorithme, l'identification ne peut être entreprise qu'à partir de la deuxième itération ($n = 2$). Nous exploitons pour cela la connaissance a priori de la route que le véhicule va suivre. Les bandes peintes sont de largeur réglementée, de même pour les largeurs des voies de circulation, considérées elles aussi constantes.

Ne pouvant faire appel à des considérations métriques quant à la représentation des projections des bandes peintes dans les images car la calibration des caméras et de la paire stéréo sont inconnues, l'identification des bandes principales s'effectue sur des portions de séquences où la région de la route est libre de tout obstacle. Par facilité, nous imposons que le nombre de bandes principales soit connu et supposé constant durant une même séquence. Suivant le type de voie empruntée par le véhicule, ce nombre est généralement égal à deux ou 3, voire plus dans des cas particuliers. L'identification des deux premières bandes principales qui délimitent la voie empruntée par le véhicule est facilitée grâce à leur orientation de signe opposé. L'identification des bandes principales restantes s'effectue en précisant le nombre de bandes à rechercher à l'extérieur des deux premières.

Les bandes principales s'avèrent alors être les bandes coplanaires dont la représentativité (somme des représentativités des VLs qui forment les couples) sont les plus élevées et ont au moins un segment dont une extrémité se situe dans la région de confiance. Une configuration telle qu'elle est présentée dans la Fig. 4.7 permet d'identifier les bandes principales dès le début de la deuxième itération ($n = 2$).

Les bandes coplanaires détectées ont deux origines. Dans la grande majorité des cas, les contours détectés qui sont orientés dans l'axe de la route ont une cause accidentelle (rainurages, défauts du revêtement, ombres portées). Il est d'ailleurs rare de pouvoir associer deux VLs de gradients opposés pour former une bande à partir de tels contours. De manière plus anecdotique, les autres bandes coplanaires sont issues des projections dans les images de bandes peintes autres que celles des voies de circulation (flèches, passages piétons, délimitations de voies supplémentaires, de places de stationnement). Dans les deux cas, les représentativités de telles bandes sont réduites en comparaison des bandes principales. Elles forment alors le jeu de bandes supplémentaires.

Rejet de détections de contours erronées La détection de segments isolés entre deux droites-supports d'une même bande perturbe l'identification des contours réels de la bande. La Fig. 4.8 illustre cette nouvelle difficulté où l'on peut observer qu'entre les droites-supports qui représentent les contours d'une bande principale, la détection des segments composés $\{173,191\}$ en vert et $\{213\}$ risque de perturber la détection de la bande. En effet, si les droites qui supportent ces segments composés vérifient le critère de distance *seuil_dx* par rapport au DVP,

elles seront considérées comme de véritables droites-supports. Après calcul des coordonnées théoriques, ces droites-supports seront intercalées entre les projections des contours de la bande qui ne pourra plus être détectée comme telle.

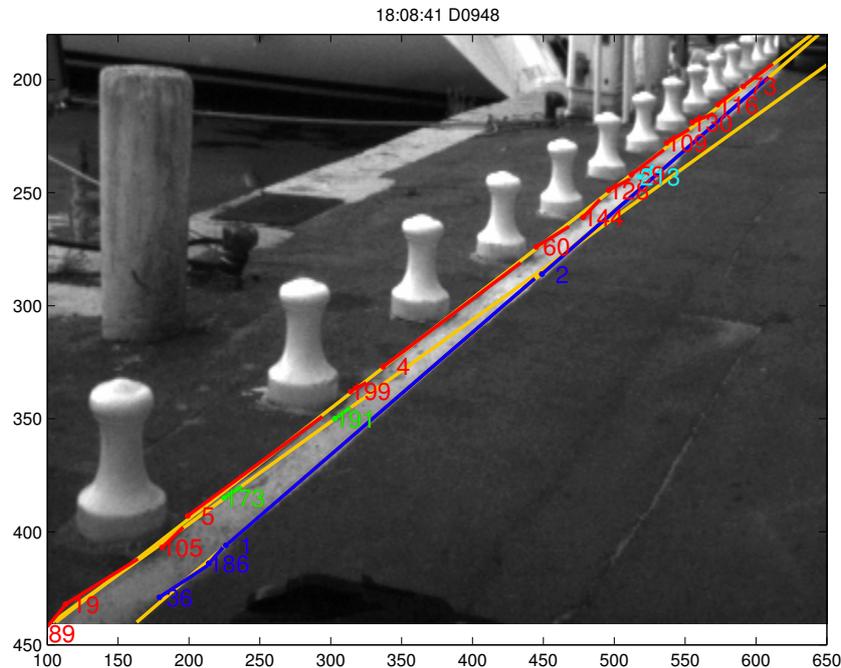


FIG. 4.8 – Résultat de la détection de contours autour d'une longue bande rectiligne. La recherche des segments alignés permet d'identifier quatre droites-supports dont seulement trois sont représentées en jaune par soucis de clarté. La quatrième droite-support est formée de l'unique segment 213, colorié en cyan, situé dans le coin supérieur droit de l'image.

VL	gauche										
n°	89	19	105	5	199	4	6	144	128	59	
ρ [pix]	413,6	443,4	442,1	440,2	424,9	439,8	417,3	482,3	393,0	439,2	
θ [°]	53,13	68,58	59,03	65,14	70,01	64,62	68,62	56,30	71,56	64,44	
long. [pix]	20,00	54,78	17,49	104,7	11,70	107,3	24,70	14,42	15,81	25,49	

gauche				{173,191}	{213}	droite				
109	130	116	73	173	191	213	2	36	186	1
402,0	354,6	433,5	437,3	441,9	444,0	449,0	465,7	464,8	461,8	464,6
69,44	75,06	64,98	64,53	67,38	65,55	63,43	61,46	67,61	55,00	61,57
17,08	15,52	16,55	23,26	13,00	12,08	11,18	182,1	36,77	12,20	247,88

TAB. 4.3 – Caractéristiques des segments détectés autour de la bande principale. Leur ordonnancement correspond à l'ordre d'apparition dans l'image de gauche à droite pour chacune des droites-supports. Un unique contour peut être représenté par plusieurs segments dont les écarts entre orientations peut dépasser 20° comme c'est le cas entre les segments 89 et 130 de la VL de la bande supérieure. Si les deux autres groupes de segments vérifient le critère de distance en fonction du DVP, elles seront considérées comme des droites-supports qu'il sera difficile de rejeter pour identifier les contours réels de la bande principale.

Nous avons reporté dans le Tab. 4.3 les caractéristiques des segments détectés autour de cette bande principale afin de rendre compte de la disparité des caractéristiques des segments qui représentent un même contour. La droite-support gauche est formée de 14 segments rouges dont l'écart entre les orientations de deux segments peut dépasser 20° . Le regroupement des segments en droites-supports est par conséquent une opération délicate qui nécessite une méthodologie rigoureuse que nous ne développerons pas dans ce manuscrit car il ne s'agit que de difficultés algorithmiques. A titre de comparaison, l'autre contour de la bande est détectée à partir de seulement quatre segments bleus dont la somme des longueurs est néanmoins comparable à celle obtenue à l'aide des 14 segments.

L'extraction des segments composés réussie, il nous faut maintenant identifier parmi l'ensemble des droites-supports celles qui représentent les projections des bandes de la scène. L'élimination des détections de contours parasites ou erronés s'impose afin de segmenter de manière certaine ces droites-supports particulières. La variabilité de la qualité de la détection de contours autour d'une même arête suivant l'image considérée ne permet pas d'exploiter la représentativité d'une droite-support comme critère d'identification. Cependant, les segments parasites extraits entre les droites-supports qui représentent des VLs sont généralement de taille réduite. La représentativité peut donc être employée comme un critère de rejet en considérant que la représentativité des droites-supports d'origine parasite est toujours inférieure à celle des droites-supports qui délimitent de "vraies" bandes.

L'élimination des droites-supports parasites s'effectue donc suite à la formation des bandes. Si une bande est formée d'une droite-support d'origine parasite, cela signifie que la droite-support parasite et celle à laquelle elle se substitue ont des caractéristiques proches (orientation), voire semblables (gradient). Cependant, la représentativité de la droite-support parasite est inférieure à celle qui forme le vrai contour de la bande. La recherche de droites-supports parasites est systématique sur tout jeu de trois droites consécutives dont les extrêmes sont de gradients opposés. La droite-support centrale sera considérée comme parasite et éliminée du jeu de droites si les droites-supports extrêmes vérifient le critère sur l'écart angulaire et si la représentativité de la bande augmente. De manière plus explicite, si les bandes recherchées sont des droites-supports dont les gradients sont de type '10', la recherche de droite-support parasite sera entreprises pour tout jeu de trois droites-supports consécutives dont les gradients sont de la forme '110' ou '100'.

Dans le cas de la Fig. 4.8, l'identification de la bande à partir des droites-supports formées des segments rouges et bleus ne sera possible qu'après élimination de la droite-support formée par le segment cyan 213. La droite-support formée des segments verts {173,191} n'est finalement pas retenue en tant que telle car l'écart angulaire qui la sépare de sa représentation théorique (calculé à partir des coordonnées estimées du DVP) dépasse le seuil de tolérance $seuil_theta/2$ fixé qui permet de tenir compte des incertitudes des coordonnées des droites-supports et du DVP.

4.1.1.5 Mise en correspondance de lignes de fuite

La mise en correspondance de contours Entreprendre une mise en correspondance entre les droites-supports détectées dans les images de la paire stéréo à ce stade de l'algorithme est une gageure. Les droites-supports étant formées de segments discontinus, les méthodes fondées sur des mesures de corrélation sont peu envisageables à partir d'une unique fenêtre de corrélation car les caractéristiques d'une telle fenêtre (longueur et position le long de la droite-support) devraient être adaptées aux différents segments qui forment la droite-support.

Or, à ce niveau de l'algorithme, nous sommes à peine parvenus à détecter les projections des

contours de la scène orientés dans la direction de la route. Les orientations des droites-supports retenues ont pu être modifiées dans une tolérance de $seuil_theta/2$ pour ne conserver que les coordonnées théoriques des droites-supports. Nous soulignons que ce seuil est dans la majorité des cas supérieur à l'écart angulaire qui sépare les orientations des droites-supports qui forment des bandes principales. Il sera par conséquent difficile d'identifier sans ambiguïté des bandes consécutives dont l'écart angulaire est inférieur à ce seuil.

Les droites-supports extraites étant formées de segments disjoints, l'emploi de méthodes basées sur la corrélation des niveaux de gris sur des régions qui entourent les droites-supports comme dans [Aubert et al., 1990; Aufrère et al., 2001] ne peuvent aboutir. En effet, il est peu probable que le calcul d'auto-corrélation d'un patch de forme à définir (rectangulaire, trapézoïdal, autre ?) le long d'une même droite-support fournisse des résultats cohérents.

De l'utilisation du bi-rapport Afin de réaliser l'appariement des droites-supports entre les images \mathcal{I}_g^n et \mathcal{I}_d^n de la paire stéréo, nous avons étudié différentes approches. Nous avons essayé d'employer le bi-rapport (paragraphe 3.1.3.4) qui est le second invariant avec le respect de la disposition géométrique des primitives aux transformations projectives et de ce fait aux transformations homographiques. Bien que robuste, cette méthode est peu utilisée car la précision des coordonnées des VLs extraites ne permet pas de vérifier de manière certaine la mise en correspondance de faisceaux de quatre VLs coplanaires.

Afin de minimiser les erreurs de précision des VLs et faciliter la mise en correspondance par utilisation du bi-rapport, nous avons d'abord introduit les coordonnées théoriques des VLs. L'idée était de réaliser de véritables faisceaux de VLs convergeant en un unique point de manière à affiner le résultat du calcul du bi-rapport. Les coordonnées théoriques d'une VL sont calculées à partir des coordonnées du DVP et du milieu du segment le plus long qui forme une VL. Les écarts entre les orientations mesurées et extraites sont inférieures à $seuil_theta/2$ sans quoi la VL détectée n'est pas retenue.

L'amélioration obtenue en prenant soin de calculer le bi-rapport à partir de faisceaux de droites convergeant en un même point nous a satisfait pendant une courte période. L'écart angulaire entre les VLs les plus extrêmes du faisceau est souvent inférieur au seuil toléré. Nous avons constaté de nombreuses erreurs de mise en correspondance, essentiellement sur ces VLs les moins bien caractérisées (voir Fig 4.9).

Nous nous étions fixé une tolérance de $5e-3$ sur l'invariance du bi-rapport, cette valeur nous permettait de mettre en correspondance les VLs qui délimitent les bandes principales centrales et droites sur plus d'une centaine d'images consécutives. Nous avons reporté dans le Tab. 4.4, les valeurs de quelques uns des bi-rapports pouvant être calculés à partir du jeu des 6 VLs à notre disposition. Les bi-rapports sont calculés en vérifiant l'ordre d'appartenance des VLs (de gauche à droite, numérotées de 1 à 6). Si l'invariance à partir du seuil que nous nous sommes fixé permet de vérifier que les bandes principales sont en correspondance deux à deux (première, quatrième et dernière colonne), nous pouvons remarquer qu'un respect strict de ce critère ne permettrait pas de vérifier de manière certaine la mise en correspondance suivant les droites choisies.

Mise en correspondance de bandes grâce à la prédiction de l'homographie stéréo
Le banc stéréo étant solidaire du véhicule, la transformation qui lie les deux repères caméras peut être considérée comme rigide et invariante au cours du temps. Dès lors, sans dévoiler le développement ultérieur de la section 4.2.3, nous pouvons supposer que nous disposons d'une prédiction de l'homographie stéréo \tilde{H}_{st}^n relativement proche de la future estimée.

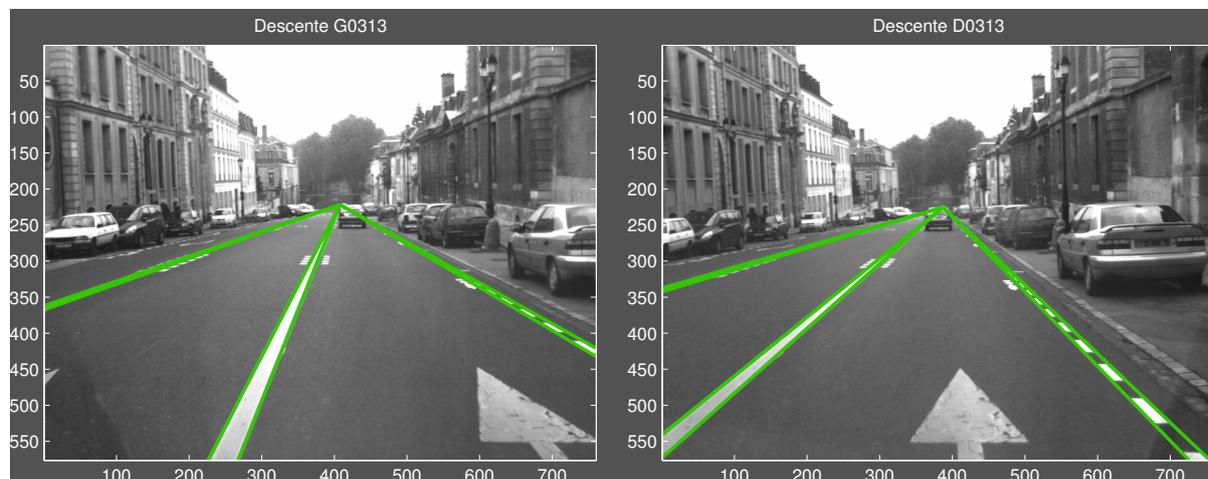


FIG. 4.9 – Bandes principales extraites pour calculer le bi-rapport. En opérant un zoom, on peut s'apercevoir que dans les deux images, la première des six VLs est légèrement éloignée du contour qu'elle représente.

VLs	{1, 2, 3, 4}	{1, 2, 3, 5}	{1, 2, 3, 6}	{1, 2, 5, 6}
BRg	0,0052	9,0039	9,2955	0,0505
BRd	0,0052	8,9279	9,1051	0,0499
$ BRg - BRd $	0	0,0760	0,1904	0,0006

{1, 3, 4, 6}	{1, 3, 5, 6}	{1, 4, 5, 6}	{2, 4, 5, 6}	{3, 4, 5, 6}
0,0016	0,0056	0,0340	0,0324	0,0031
0,0010	0,0553	0,0208	0,0198	0,0020
0,0006	0,0006	0,0132	0,0126	0,0011

TAB. 4.4 – Bi-rapport des bandes principales. Les bi-rapports BRg et BRd ont été calculé à partir des orientations estimées des 6 VLs qui délimitent les 3 bandes principales dans les images \mathcal{I}_g^n et \mathcal{I}_d^n . Les bi-rapports dont l'écart entre les deux images est inférieur à $5e-3$ ont été mis en évidence en gras.

Exploiter cette prédiction pour mettre en correspondance les deux faisceaux de droites-supports extraits dans les images du couple stéréo n'est hélas pas possible comme le montre la Fig. 4.10(a). Nous avons superposé aux faisceaux de droites-supports extraits (traits continus) les faisceaux projetés grâce à la prédiction \tilde{H}_{st}^n (pointillés). Les coordonnées des DVPs extraits et prédits ne se superposant pas, la prédiction de l'homographie s'avère peu précise. L'identification des droites-supports en correspondance n'est pas possible : les droites-supports projetées ne correspondent pas aux extraites.

La Fig. 4.10(b) a été réalisée en utilisant cette fois-ci une estimée de l'homographie stéréo H_{st}^n . La mise en correspondance des droites-supports qui représentent les mêmes contours est alors immédiate : la vérification de manière croisée que les écarts angulaires entre droites-supports extraites et projetées sont inférieurs à $seuil_theta/5(2^\circ)$ permet d'identifier toutes les droites-supports en correspondance dans les deux images. Ce seuil relativement faible permet néanmoins de compenser les erreurs d'estimations des coordonnées théoriques des droites-supports et des incertitudes liées à l'estimation de H_{st}^n .

Mise en correspondance des bandes principales grâce au suivi de leurs caractéristiques

Les méthodes fondées sur l'utilisation du bi-rapport ou d'une prédiction de l'homographie stéréo entre les droites-supports sont insatisfaisantes essentiellement en raison du mouvement apparent que subit le faisceau de droites-supports d'une image du couple stéréo à l'autre. Or, ce mouvement est beaucoup plus réduit entre deux images successives de la même caméra : les caractéristiques du plan de la route évoluent peu par hypothèses entre deux acquisitions à la cadence vidéo.

Nous supposons dans ce qui suit que nous disposons d'une prédiction de la projection des bandes principales dans les images courantes du couple stéréo. La mise en oeuvre de cette prédiction est un problème de suivi de primitives au cours du temps que nous développerons en détail dans la section 5.1.3 du prochain chapitre. L'identification des nouvelles projections des bandes principales s'effectue donc de manière séparée pour chacune des images \mathcal{I}_g^n et \mathcal{I}_d . La recherche consiste à déterminer parmi les bandes extraites de chacun des faisceaux, celles dont les droites-supports présentent des orientations égales aux prédictions en tolérant un écart angulaire inférieur à $seuil_theta/5$ (cf Fig.4.11).

Comme nous le verrons dans le chapitre 4, les caractéristiques des bandes principales prédites généralement suffisamment fiables pour identifier sans ambiguïté leur nouvelle projection dans les images courantes. Néanmoins, la recherche des projections courantes des bandes principales peut échouer dans deux cas. Soit la projection de la bande principale est cachée dans sa quasi-intégralité par un obstacle, soit de nombreux contours proches des caractéristiques prédites de la bande sont détectés ce qui peut perturber son identification. Nous disposons toutefois d'une prédiction qu'il sera possible d'affiner suite à l'estimation de l'homographie H_{st}^n si sa duale est détectée dans l'autre image du couple stéréo.

Mise en correspondance des lignes de fuite grâce à une estimation de l'homographie stéréo

Méthodologie La mise en correspondance effective des VLs détectées sur la route entre les deux images du couple stéréo n'est entreprise que lorsque l'estimation de l'homographie H_{st}^n est disponible. Ce dernier point signifie que l'étape que nous décrivons dans ce paragraphe intervient dans la séquence des tâches réalisées par l'algorithme après la détection des points d'intérêt coplanaires dans les deux images du couple stéréo, décrite dans la section 4.2.

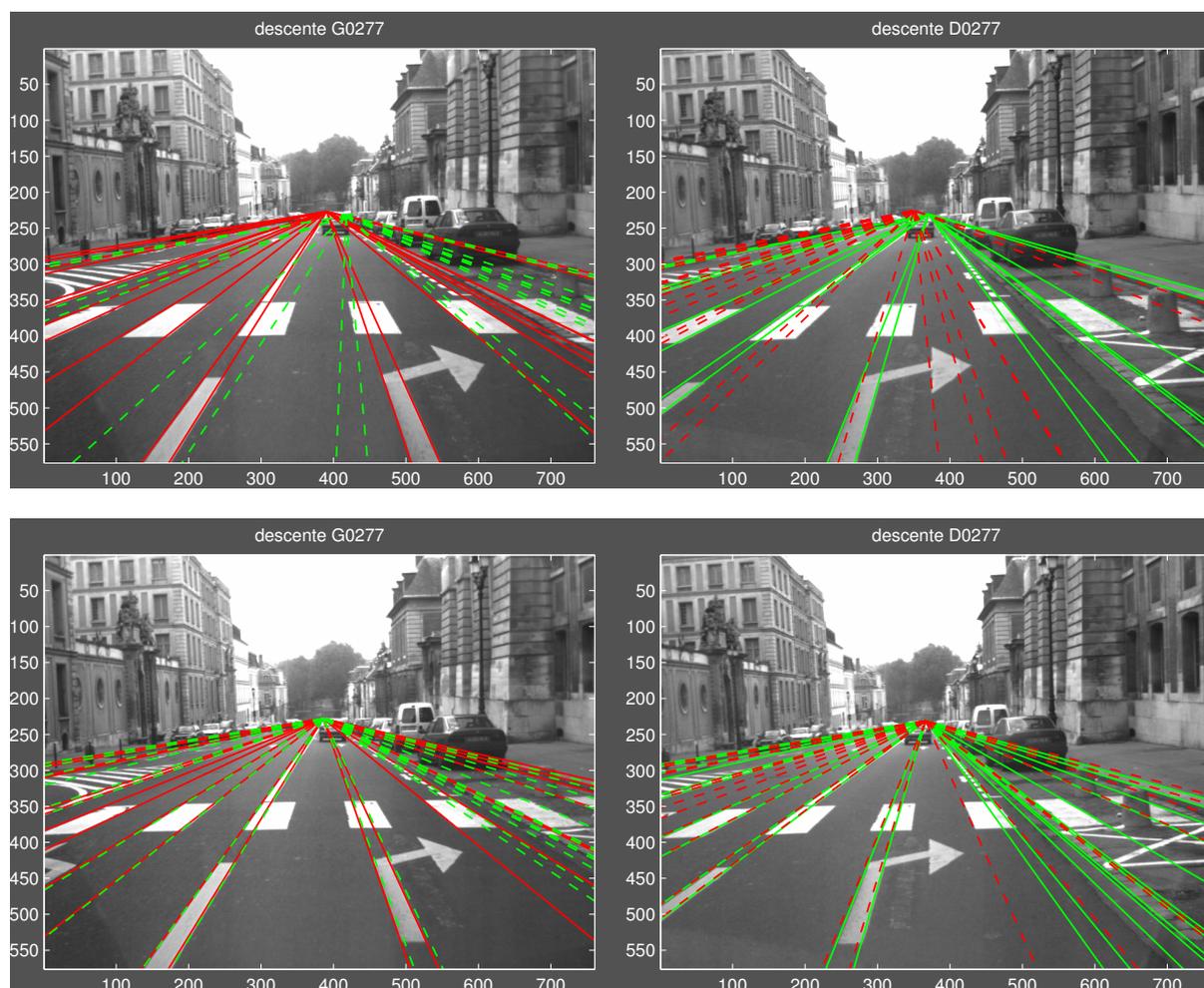


FIG. 4.10 – Projection des bandes extraites des faisceaux de droites grâce à une prédiction et une estimation de l'homographie stéréo. Les primitives des images gauche et droite sont respectivement représentées en rouge et vert, les bandes originales en traits continus et les prédictions en pointillés. En haut, les faisceaux de bandes prédits sont relativement éloignés des originaux, tandis qu'en bas, la mise en correspondance des faisceaux projetés grâce à une estimation courante de l'homographie stéréo permet de mettre en correspondance facilement les bandes en supposant un écart angulaire inférieur à 2° entre bandes prédites et projetées.

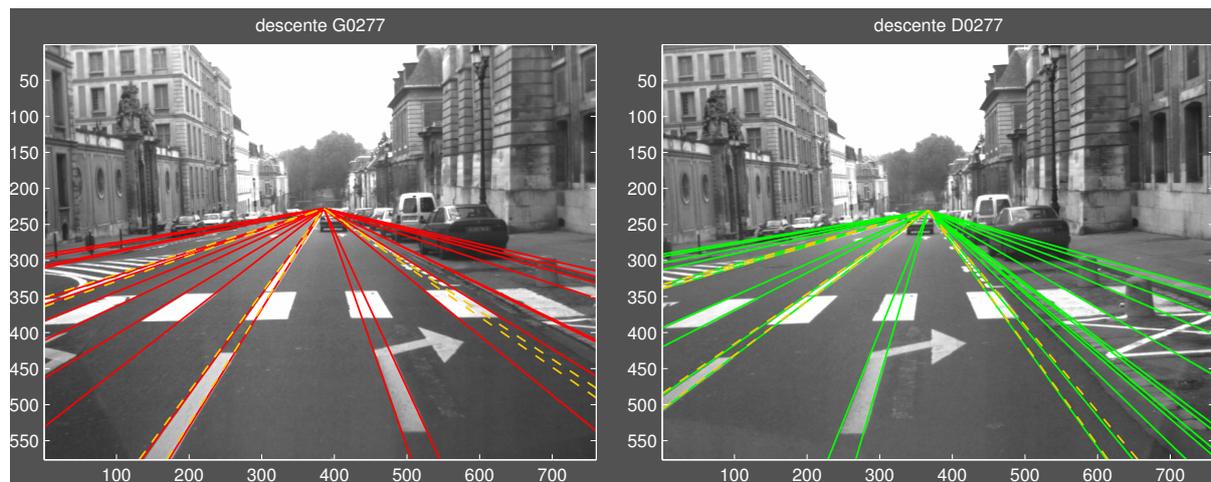


FIG. 4.11 – *Superposition des bandes extraites des deux faisceaux de droites et des prédictions des bandes principales. Les prédictions sont représentées en pointillés oranges. La recherche des projections courantes des bandes principales parmi les bandes détectées est immédiate en considérant un écart d'orientation inférieur à $\text{seuil_theta}/5$.*

En supposant que l'estimée H_{st}^n soit suffisamment précise, les projections des droites-supports de \mathcal{I}_g^n via H_{st}^n se superposent avec leur correspondante dans \mathcal{I}_d^n et inversement. L'identification des VLs s'effectue de manière hiérarchisée de façon à limiter les ambiguïtés. Nous rappelons en effet que l'orientation apparaît comme l'unique critère d'identification fiables à ce niveau de l'algorithme. Cependant, la vérification croisée de l'écart angulaire et le respect de la contrainte d'ordre d'apparition (les droites-supports en correspondance sont ordonnées dans les deux images) permettent de contraindre efficacement cette tâche de mise en correspondance.

La hiérarchisation de la mise en correspondance des droites-supports supposées coplanaires signifie que la recherche croisée des droites-supports des deux images dont les orientations sont comparables s'effectue en deux passes : une première à partir des droites-supports qui forment des bandes, une seconde à partir des droites-supports restantes en prenant soin d'éliminer celles qui ont été mises en correspondance en tant que bande. Cette précaution est nécessaire car malgré une tolérance relativement faible, plusieurs droites-supports projetées peuvent correspondre avec une même droite-support, comme c'est le cas sur la Fig. 4.10(b). La distinction entre droites-supports qui forment des bandes et droites-supports isolées permet de considérablement réduire les ambiguïtés surtout autour des bandes extrêmes car les écarts entre les orientations des droites-supports sont généralement de l'ordre du degré.

Une mise en correspondance plus rigoureuse nécessiterait de connaître la calibration des caméras. Ainsi, à partir du mouvement du DVP, nous pourrions en déduire les mouvements de rotation de la caméra et par conséquent estimer le mouvement relatif de chaque VL qui forme le faisceau de droites-supports.

4.1.1.6 Détection des bords de la route : la région d'intérêt

La route étant supposée localement plane et à bords parallèles, sa projection au 1^{er} plan de l'image s'apparente à un triangle dont le sommet principal est le DVP et la base est formée de la limite basse de l'image interceptée par les projections des bords de la route, soit les caractéristiques de la première et la dernière bande principale que nous venons d'identifier.

Le champ de vision étant rarement dégagé en milieu urbain, la région d'intérêt (*Region Of Interest*) représentant la projection du plan de la route dans l'image sera volontairement tronquée sur sa partie supérieure et formera donc un parallélogramme. Cette troncature pourrait matérialiser la région de libre espace si un détecteur d'obstacle était implémenté. Ce n'est pas le cas actuellement, cette limitation correspond donc au champ de vision supposé libre et se justifie pour trois raisons :

- il est peu probable que l'horizon soit observable en milieu urbain,
- la route est supposée plane sur une distance finie,
- la limitation au 1^{er} plan de l'image valide l'approximation linéaire de la route employée même lorsque son profil n'est pas rectiligne.

L'utilisation de la ROI va nous permettre de rechercher des points d'intérêt dans une région limitée de l'image. Nous considérons, en effet, qu'au-delà de cette zone les régions observées n'appartiennent plus au plan de la route. Néanmoins, la région de la route délimitée par la ROI peut contenir des obstacles qui vont nous compliquer la tâche.

Deux nouveaux points sont à considérer pour déterminer les limites de la ROI. Le premier, les régions basses des images peuvent présenter des reflets du tableau de bord sur le pare-brise ou l'essuie-glace. Ces éléments n'apportant aucune information, les régions de l'image correspondantes seront désormais supprimées. Le second, la mise en correspondance des points d'intérêt utilisant une mesure de corrélation à partir d'une fenêtre Q carrée de $2.q$ pixels de côté suppose que les points d'intérêt soient situés à une distance q du bord de l'image.

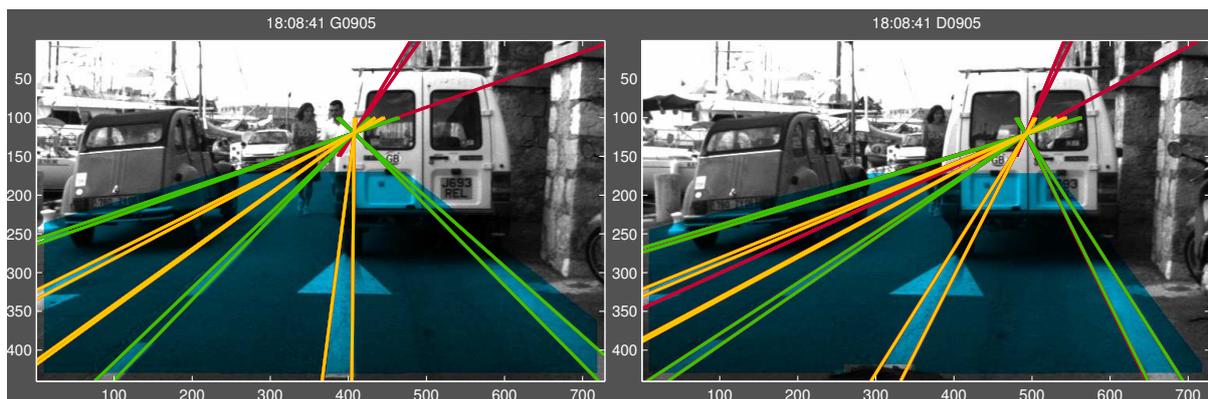


FIG. 4.12 – Création des ROIs suite à l'identification des bandes principales (vertes). Les bandes supplémentaires (jaunes) en correspondance permettront de contraindre l'estimation finale de l'homographie stéréo. Les droites-supports restantes (rouges) se trouvent soit en dehors du plan de la route, soit n'ont pas de correspondance dans l'autre image. La ROI est délimitée par une horizontale située à $10 \cdot dist_max$ sous l'ordonnée de l'estimée du DVP, les deux VLs extrêmes des bandes principales gauche et droite et le bas de l'image.

4.2 Extraction de points d'intérêt appartenant à la route

4.2.1 Extraction de points d'intérêt

Depuis [Moravec, 1977], il apparaît évident que tous les points d'une image n'ont pas le même niveau informatif : un point situé dans une région homogène est peu caractéristique comparé à un point situé dans une région de forte variation d'intensité. Ces points caractéristiques d'une

image et par conséquent de la scène ont reçu la dénomination de points d'intérêt. Il existe deux grandes familles de détecteurs :

- celle qui considère les FPs comme les lieux de forte variation de la fonction intensité de l'image dans des directions différentes,
- celle qui exploite une représentation directe des coins de formes connues 'T', 'Y' et 'L'.

L'emploi d'un détecteur de la seconde famille est tentant car l'environnement urbain regorge d'éléments structurants délimités par des coins saillants, généralement à l'intersection d'arêtes linéaires. Cependant, comme nous l'avons exprimé plus haut, notre objectif est de détecter un maximum de primitives sur la route en sus de celles que nous détecterons grâce au marquage de la signalisation horizontale. Les détecteurs de la première famille seront donc mieux adaptés à la recherche de défauts du revêtement, de coins d'ombres portés, d'intersection d'obstacles avec la route.

[Schmid, 1996] a comparé dans sa thèse différents extracteurs de points d'intérêt. Des critères d'évaluation tels que la répétabilité et l'invariance des points d'intérêt détectés aux changements de luminosité, d'échelles et de position de caméras ont été introduits. La répétabilité est la propriété de détecter les projections des mêmes points 3D, en considérant une erreur de détection bornée, entre deux images. Le taux de répétabilité correspond au rapport entre le nombre de points d'intérêt correctement mis en correspondance entre les deux images et le minimum des points d'intérêts extraits dans la partie commune des deux images. Parmi les méthodes étudiées, le filtre de Harris présente globalement des caractéristiques équivalentes sinon meilleures sur le critère de répétabilité à la plupart de ses concurrents.

[Espiau, 2002] a repris à son compte les travaux de comparaison précédents en se plaçant dans le cadre de l'environnement sous marin, milieu naturel où la détection de points d'intérêt robustes et correctement localisés est réputée difficile. Il confirme la qualité des points d'intérêt extraits grâce à un filtre de Harris en comparaison des extracteurs CSS (représentation multi-échelle du rayon de courbure des contours extraits à partir d'un filtre de Canny) et Susan (détection de points d'intérêt à partir de masques qui ne nécessite aucun calcul de dérivées) qui fournissent des points d'intérêt moins stables ou réclament le réglage de nombreux seuils.

4.2.2 Le détecteur de Harris-Stephens

4.2.2.1 Rappel

Le détecteur de Harris-Stephens est une amélioration de la formulation de base de Moravec qui emploie une mesure d'auto-corrélation. L'idée fort simple est néanmoins ingénieuse : les points d'intérêt se distinguent de leurs voisins car ils représentent un maximum local d'une mesure d'auto-corrélation. Le calcul d'auto-corrélation d'une image \mathcal{I} consiste à sommer la différence de niveaux de gris pixel à pixel entre une fenêtre Q de référence centrée en $\mathbf{p}(u,v)$ et une autre décalée de $\mathbf{d} = [d_u, d_v]^t$:

$$Corr(\mathbf{p}) = \sum_{w \in Q} [\mathcal{I}(q) - \mathcal{I}(q + d)]^2 \quad (4.6)$$

Considérons le développement limité à l'ordre 1 de la fonction image autour du point \mathbf{p} et en introduisant le gradient ∇I , où \mathcal{I}_u et \mathcal{I}_v représentent les dérivées partielles de l'image I selon les deux axes horizontal et vertical :

$$\mathcal{I}(\mathbf{p} + \mathbf{d}) = \mathcal{I}(\mathbf{p}) + \mathbf{d}^t \nabla \mathcal{I}(\mathbf{p}) + \varepsilon(\mathbf{p}). \quad (4.7)$$

En négligeant, le résidu $\varepsilon(\mathbf{p})$, la valeur du coefficient d'auto-corrélation peut donc s'écrire :

$$Corr(\mathbf{p}) = \sum_{q \in Q} [\mathbf{d}^t \cdot \nabla I(\mathbf{q})]^2 \quad (4.8)$$

$$= \sum_{q \in Q} [\mathbf{d}^t \cdot \nabla I(\mathbf{q}) \cdot \nabla I(\mathbf{q})^t \cdot \mathbf{d}] \quad (4.9)$$

$$= \mathbf{d}^t \cdot \sum_{q \in Q} \begin{bmatrix} \mathcal{I}_u^2(q) & \mathcal{I}_u \mathcal{I}_v(q) \\ \mathcal{I}_u \mathcal{I}_v(q) & \mathcal{I}_v^2(q) \end{bmatrix} \cdot \mathbf{d} \quad (4.10)$$

$$= \mathbf{d}^t \cdot \mathbf{C} \cdot \mathbf{d} \quad (4.11)$$

Les valeurs propres de la matrice \mathbf{C} permettent de caractériser le type de point d'intérêt étudié. Trois cas sont envisageables :

- les deux valeurs propres sont faibles : le point appartient à une région homogène,
- une valeur propre est dominante : le vecteur associé est une direction privilégiée de l'auto-corrélation et le point considéré appartient à une arête,
- les deux valeurs propres sont fortes, le point considéré se situe à l'intersection de deux arêtes : il représente un maximum local : le point est caractéristique soit d'un coin, soit d'un motif texturé.

Le calcul de l'auto-corrélation doit être entrepris pour des valeurs de \mathbf{d} particulières : l'emploi d'une fenêtre Q circulaire permet d'obtenir un comportement isotrope du détecteur. L'isotropie est une propriété indispensable au détecteur car elle permet de s'affranchir des mouvements de rotation de l'image. La difficulté étant alors de définir une telle fenêtre dans l'espace image discrétisé. L'idée est de remplacer l'opération d'intégration sur une fenêtre Q par le produit de convolution d'une fonction symétrique qui décroît de manière isotrope autour du point considéré. La gaussienne est généralement retenue, elle a pour équation pour un point \mathbf{r} de R^2 :

$$g(\mathbf{r}, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{\mathbf{r}^2}{2\sigma^2}} \quad (4.12)$$

L'introduction de la convolution par la gaussienne permet d'entreprendre le calcul d'auto-corrélation sur une fenêtre Q de rayon infini. Sachant que $f \otimes g(r) = \int \int (q) \cdot g(r-q) \cdot dq$, le calcul d'auto-corrélation s'exprime donc désormais dans R^2 :

$$Corr(\mathbf{p}) = g(\mathbf{p}, \sigma) \otimes [\mathcal{I}(\mathbf{r}) - \mathcal{I}(\mathbf{r}+\mathbf{d})]^2 \quad (4.13)$$

$$= \int_{R^2} g(\mathbf{p}-\mathbf{r}, \sigma) \cdot [\mathcal{I}(\mathbf{r}) - \mathcal{I}(\mathbf{r}+\mathbf{d})]^2 \cdot d\mathbf{r} \quad (4.14)$$

$$= \int_{R^2} g(\mathbf{p}-\mathbf{r}, \sigma) \cdot [\mathbf{d}^t \nabla \mathcal{I}(\mathbf{r})]^2 \cdot d\mathbf{r} \quad (4.15)$$

$$= \int_{R^2} \mathbf{d}^t \cdot [g(\mathbf{p}-\mathbf{r}, \sigma) \cdot \nabla \mathcal{I}(\mathbf{r}) \cdot \nabla \mathcal{I}(\mathbf{r})^t] \cdot \mathbf{d} \cdot d\mathbf{r} \quad (4.16)$$

$$= \mathbf{d}^t \cdot g(\mathbf{p}, \sigma) \otimes \cdot [\nabla \mathcal{I}(\mathbf{r}) \cdot \nabla \mathcal{I}(\mathbf{r})^t] \cdot \mathbf{d} \quad (4.17)$$

L'expression intégrale représentée par la matrice $\mathbf{C}(\mathbf{p})$ précédente s'écrit alors :

$$g(\mathbf{p}, \sigma) \otimes \cdot \begin{bmatrix} \mathcal{I}_u^2(q) & \mathcal{I}_u \mathcal{I}_v(q) \\ \mathcal{I}_u \mathcal{I}_v(q) & \mathcal{I}_v^2(q) \end{bmatrix} = \begin{bmatrix} g(\mathbf{p}, \sigma) \otimes \cdot \mathcal{I}_u^2(q) & g(\mathbf{p}, \sigma) \otimes \cdot \mathcal{I}_u \mathcal{I}_v(q) \\ g(\mathbf{p}, \sigma) \otimes \cdot \mathcal{I}_u \mathcal{I}_v(q) & g(\mathbf{p}, \sigma) \otimes \cdot \mathcal{I}_v^2(q) \end{bmatrix} \quad (4.18)$$

Cette nouvelle formulation permet de localiser en un unique point les extréma locaux et permet donc de s'affranchir d'effets de plateaux, de fixer la taille de la fenêtre Q par la variance σ et de permettre une implémentation rapide de la convolution par filtre récursif.

4.2.2.2 Implémentation

Le calcul des valeurs propres en tous les points de l'image est très coûteux en temps de calcul. Il est d'usage de calculer l'expression suivante dans laquelle interviennent les valeurs propres de manière non-explicite :

$$K = \det(A) - \lambda \cdot \text{tr}^2(A) \quad (4.19)$$

avec λ compris entre $[0,04; 0,06]$. Les points d'intérêt sont obtenus pour les valeurs de K positives.

Le score originel du détecteur de Harris est peu satisfaisant : la détermination de la constante λ est arbitraire. [Benhimane and Malis, 2004] propose de fusionner les scores de Forstner et Noble afin de tenir compte de leurs avantages respectifs. Le premier, $s_F = 4\det/\text{tr}^2$, a la particularité d'être normalisé : le score tend vers 1 si les valeurs propres sont proches et vers 0 si $\lambda_1 \gg \lambda_2$. Mais ce score nécessite l'emploi de deux nouveaux seuils à définir : le premier pour déterminer si la valeur de la trace est suffisante pour entreprendre le calcul du score, le second pour déterminer la présence d'un point d'intérêt en fonction du score normalisé. Le second score $s_N = \det/\text{tr}$ ne nécessite plus qu'un unique seuil. Le score est grand si $\lambda_1 \approx \lambda_2$, faible si $\lambda_1 \gg \lambda_2$ mais une indétermination demeure lorsque $\lambda_1 \gg \lambda_2 \geq \text{seuil}$, ce qui est le cas quand le calcul est effectué sur un contour où le gradient est prononcé.

$$s_B \propto s_F * s_N \quad (4.20)$$

$$= \frac{\det^2}{\text{tr}^3} \quad (4.21)$$

$$= \frac{(\lambda_1 \lambda_2)^2}{(\lambda_1 + \lambda_2)^3} \quad (4.22)$$

La sélection du seuil de détection des points d'intérêt demeure empirique. Nous avons appliqué le score de Benhimane sur les images test utilisées dans l'étude sur l'extraction de contours. Deux possibilités nous sont offertes : l'emploi d'un seuil invariant ou celui d'un seuil relatif où sont alors considérés comme points d'intérêt (*Feature Points*) tous les points extraits dont le score est supérieur à une fraction du score du plus élevé. Nous avons choisi de privilégier la répétabilité en imposant un seuil fixe : tous les points de l'image dont le score de Benhimane est supérieur à $\text{seuil_score} = 100$ sont considérés comme des points d'intérêt (cf Fig.4.13). Le choix empirique du seuil a été motivé par la détection du plus grand nombre de coins du marquage de la signalisation horizontale. Une réduction de ce seuil favoriserait la détection de coins situés dans la partie basse de l'image mais engendrerait par la même occasion des détections de blobs ou tâches, souvent peu robustes.

L'extraction de points d'intérêt sur les images test confirme que la route apparaît comme un région homogène où à quelques rares exceptions près, seul le marquage de la signalisation horizontale permet d'extraire des points d'intérêt fiables. Aux vus des résultats, l'extracteur de points d'intérêt employé détecte peu de coins sur la route. Les points d'intérêt de plus haut score sont généralement des saillances aiguës qui séparent des régions de fort contraste ou à l'intersection de trois régions de textures différentes.

Le contenu informatif des points d'intérêt situés au premier plan est nettement supérieur à ceux détectés au-delà. En effet, la surface de la route associée à un *pixel* dépend de la distance qui sépare la région observée du foyer de la caméra. La Fig. 4.14 représente une coupe transversale des angles solides formés par des régions de même dimension qui sont interceptés par le plan



FIG. 4.13 – Résultats de l'extraction de points d'intérêt avec un filtre de Harris et le score de Benhimane. Les points d'intérêt sont classés en fonction de leur score : entre $[100,300]$ en rouge, entre $[300,700]$ en orange, au-delà en vert.

de l'image. En supposant la surface lambertienne, il apparaît alors clairement que le flux de lumière intercepté par l'écran sera proportionnel à la surface de l'intersection avec les angles solides. Ce dernier point justifie que le score de Harris des projections des coins de la scène les plus éloignés aient les scores les plus élevés (cf. Fig. 4.15).

Ainsi, la contrepartie à la précision des FPs détectés au premier plan est la difficulté inhérente à les détecter en raison d'une part de la double discrétisation spatiale en intensité et d'autre part le lissage introduit par la gaussienne. Un coin en "L", fortement contrasté sera atténué suite à la convolution avec la fenêtre gaussienne lors du calcul des termes du second ordre. Du fait de leur mouvements apparents importants et de la résolution de leur localisation, les FPs du premier plan doivent être extraits en priorité.

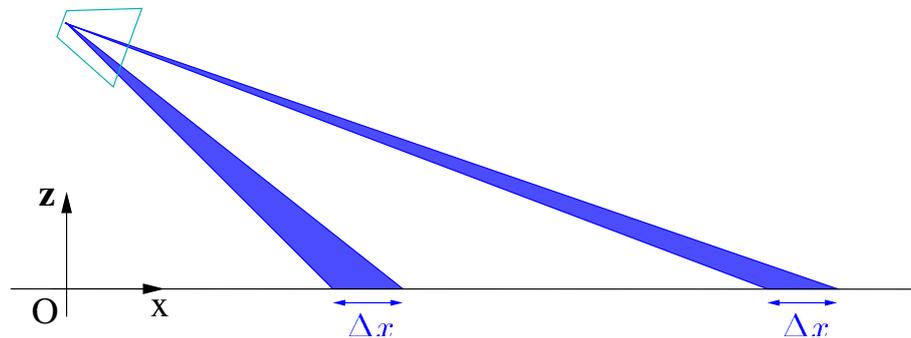


FIG. 4.14 – La résolution des pixels dépend de la distance qui sépare la scène observée du foyer de la caméra. Il est alors évident que les points caractéristiques du premier plan seront localisés de manière plus précise que d'autres plus éloignés de la caméra.

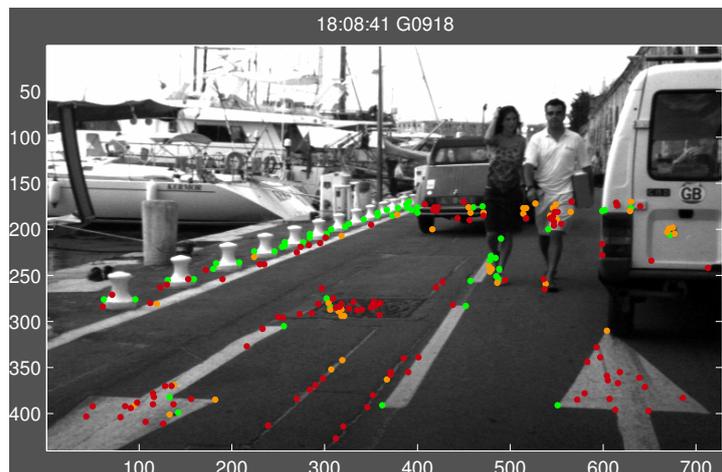


FIG. 4.15 – Caractérisation des points d'intérêt détectés au premier plan. La prise en compte des points d'intérêt les mieux caractérisés (en vert), ayant un score de Harris supérieur à 100, s'avère trop réducteur car de nombreux coins du premier plan, pourtant clairement identifiables, ont des scores inférieurs à ce seuil critique. Les points d'intérêt coloriés en orange ont un score de Harris compris dans l'intervalle $[50; 100]$ alors que ceux en rouge ont un score $[20, 50]$.

4.2.3 Exploitations des contraintes de rigidité et de coplanarité

L'emploi d'un couple stéréo embarqué se justifie pleinement pour identifier parmi l'ensemble des primitives détectées celles qui appartiennent à l'environnement statique de la scène et sont à la fois coplanaires. En supposant que le mouvement du banc stéréo est réduit entre deux acquisitions, les variations des caractéristiques de projection de la route dans les images évoluent lentement. La route étant considérée comme localement plane, sa projection sur les plans images du couple stéréo induit une homographie planaire H_{st} (st pour stéréo).

L'homographie stéréo dépend par définition des paramètres intrinsèques des caméras (K_l et K_d), des mouvements de rotation et de translation entre les deux repères, de la représentation de la normale au plan de la route et de la distance qui sépare le repère de référence (dans notre cas celui de la caméra gauche) du plan de la route. En supposant la transformation géométrique (rotation et translation) quasi-rigide entre les repères des caméras du banc stéréo, les homographies stéréo aux instants $(n - 1)$ et n se développent de la sorte :

$$H_{st}^{n-1} = K_d \cdot \left[R_{st} - \frac{\mathbf{T}_{st}^t \cdot \bar{\mathbf{N}}^{n-1}}{d^{n-1}} \right] \cdot K_g^{-1} \quad (4.23)$$

$$H_{st}^n = K_d \cdot \left[R_{st} - \frac{\mathbf{T}_{st}^t \cdot \bar{\mathbf{N}}^n}{d^n} \right] \cdot K_g^{-1} \quad (4.24)$$

avec $\mathbf{N} = [\bar{\mathbf{N}}^t, d]^t$, coordonnées de la normale du plan de la route dans le repère de la caméra gauche et \mathbf{T}_{st} et R_{st} la transformation rigide qui permet de passer du repère de la caméra gauche à celui de la caméra droite.

L'évolution de l'homographie stéréo entre deux positions du banc stéréo ne dépend donc que de la représentation de la normale \mathbf{N} , exprimée dans le repère de la caméra gauche, et de la distance d du centre de la caméra gauche C_g au plan de la route, conformément à la Fig. 4.16. Or, les variations de la représentation de la normale entre deux positions d'une même caméra ne sont dus qu'aux mouvements de rotation R_g^n qui permettent au repère C_g de passer de la position $(n - 1)$ à la position n . Dans notre cas, le mouvement du véhicule étant supposé plan, seule la variation de l'angle d'inclinaison α du véhicule est source de variation de la représentation de \mathbf{N} . D'autre part, la distance d est liée à l'angle d'inclinaison via la hauteur h , altitude constante du repère C_g , exprimée dans le repère absolu, lié à la route :

$$h = d \cdot \cos(\alpha) \quad (4.25)$$

Ainsi, pour de faibles variations de l'angle d'inclinaison du véhicule entre deux instants d'acquisitions, nous pouvons considérer en première approximation que la représentation de la normale \mathbf{N} et la distance d du centre du repère caméra C_g au plan de la route évoluent peu conformément à la représentation de la Fig. 4.17 :

$$\bar{\mathbf{N}}^n \simeq \bar{\mathbf{N}}^{n-1} \quad \text{et} \quad d^n = d^{n-1} = h. \quad (4.26)$$

Autrement dit, l'homographie stéréo H_{st}^{n-1} estimée à l'itération $(n - 1)$ entre les images \mathcal{I}_g^{n-1} et \mathcal{I}_g^{n-1} peut être considérée comme une prédiction acceptable de l'homographie H_{st}^n qui lie les images \mathcal{I}_g^n et \mathcal{I}_g^n :

$$\widetilde{H}_{st}^n = H_{st}^{n-1}. \quad (4.27)$$

Il est alors envisageable d'identifier facilement les primitives coplanaires lors d'une nouvelle itération en recherchant parmi les primitives détectées dans le couple d'images stéréo celles qui

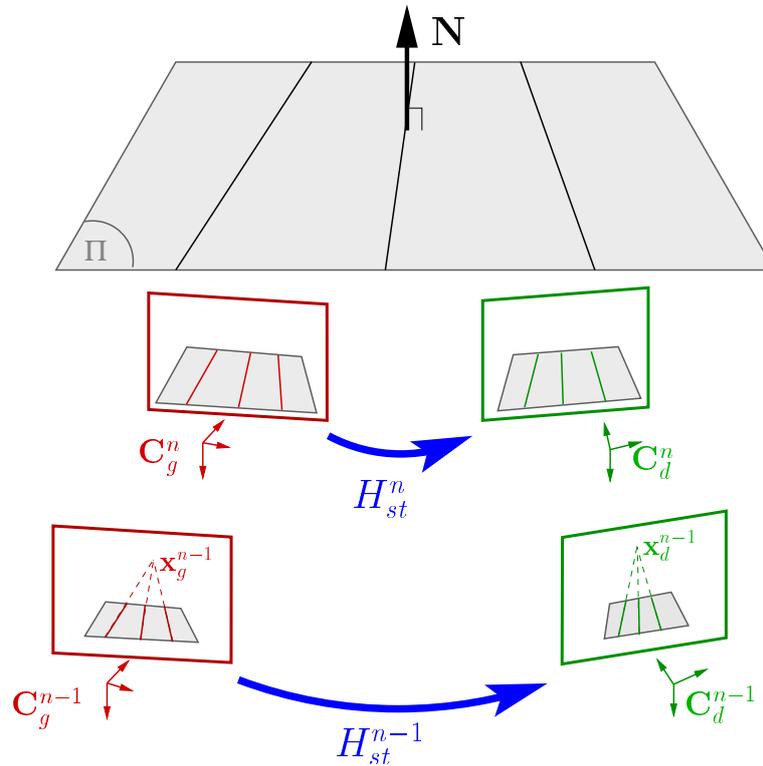


FIG. 4.16 – La paire stéréo étant supposée quasi-rigide, les mouvements de translation et rotation entre les deux repères caméras peuvent être considérés comme invariants. Les homographies H_{st}^{n-1} et H_{st}^n , induites par le plan π entre les plans images du couple stéréo aux instants $(n - 1)$ and n , dépendent donc essentiellement de la représentation de \mathbf{N} et de la hauteur d de la caméra gauche par rapport au plan π . Le déplacement du véhicule étant supposé plan, seul le paramètre d'inclinaison entre deux positions est à l'origine de variations de la projection de la normale \mathbf{N} du plan π .

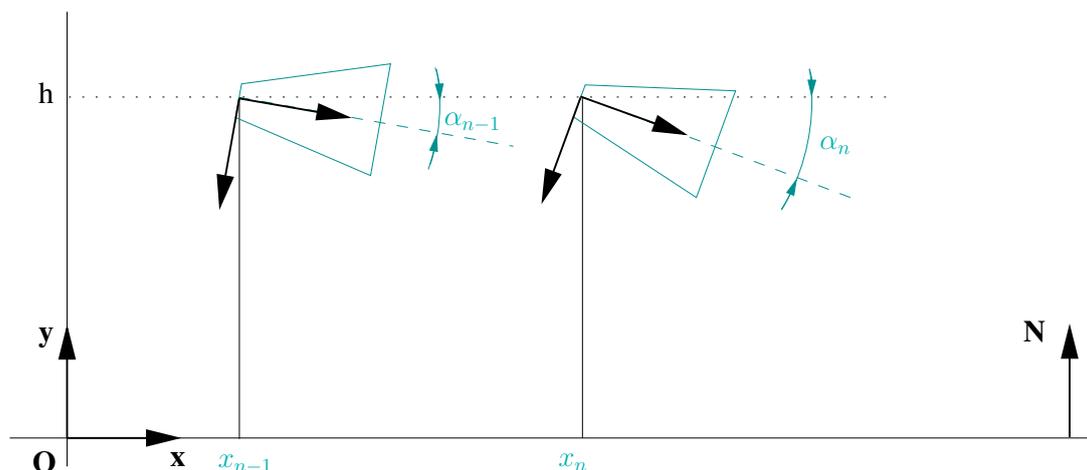


FIG. 4.17 – Evolution de l'angle d'inclinaison α d'une caméra entre deux positions successives. La distance d qui sépare le repère de la caméra du plan peut être considérée comme invariante et égale à h en première approximation.

vérifient l'homographie stéréo de l'itération précédente. Le bruit sur les coordonnées des primitives extraites, l'imprécision de l'estimation des homographies et les approximations contenues dans \widehat{H}_{st}^n impliquent néanmoins une recherche itérative des primitives coplanaires de l'itération courante. Nous procédons pour cela par étapes :

1. pré-segmentation de la route grâce aux bandes principales de la signalisation horizontale, appariées entre les deux images,
2. estimation de l'homographie stéréo en recherchant les couples de points d'intérêt coplanaires sur la route,
3. mise en correspondance des faisceaux de VLs à partir de l'estimée courante de l'homographie stéréo.

4.2.4 Mise en correspondance de points d'intérêt

Contrairement aux droites-supports, il est possible de déterminer avec précision si les FPs détectés dans deux images sont les représentations d'un même point de la scène. Sous l'hypothèse que l'éclairage de la scène est constant, la surface lambertienne et les deux caméras du couple stéréo à distance à peu près égale du point observé, alors les distributions de niveaux de gris d'un même point de la scène observé par les deux caméras seront similaires. D'autre part, du fait du phénomène de vision rasante, mentionné précédemment, la valeur du coefficient de corrélation normalisé (compris entre $[-1; 1]$) entre deux projections d'un même point dépendra l'éloignement de ce point sur la route.

4.2.4.1 Les points d'intérêt détectés sur la route

Afin d'illustrer la relation entre la géométrie de la tête stéréo et la disparité observée dans l'image, nous présentons en Fig. 4.18 un cas d'école où le champ de vision des deux caméras du banc stéréo est dégagé et de nombreux FPs coplanaires ('+' rouges à gauche, verts à droite) sont détectés dans les ROIs. La répartition des FPs peut être considérée comme idéale du fait de la présence de FPs aux premier et second plans. Le tracé des mouvements apparents entre les coordonnées des FPs mis en correspondance permet de vérifier que le déplacement entre les repères C_g et C_d est essentiellement un mouvement de rotation selon l'angle de lacet et un autre moins important selon l'angle de tangage. Les FPs en correspondance au premier plan présentent des déplacements réduits tandis que ceux proches de l'horizon, éloignés des deux caméras ont des projections qui subissent des forts déplacements. Les coordonnées des projections du DVP sont éloignées de plus de 80 *pixels* entre les deux images.

Parmi les configurations de caméras envisageables pour entreprendre l'étude que nous menons, il nous faut aussi tenir compte du cas où les axes optiques des caméras sont quasiment parallèles comme dans la Fig. 4.19. Dans une telle configuration, le déplacement des FPs en correspondance est différent de ceux présentés jusqu'alors. Le mouvement apparent du DVP qui est un FP particulier est quasiment nul (ce point appartenant au plan à l'infini, son déplacement ne dépend que des rotations), ce qui conduit à des variations d'abscisses des FPs en correspondance qui évoluent de manière conjointe avec leur ordonnée, contrairement à la configuration présentée auparavant. A titre d'exemple, l'écart entre les coordonnées du FP en correspondance le plus bas dans les images est de 263 *pixels* en abscisses et seulement trois en ordonnée. Nous avons tracé sur les figures Fig. 4.18 et Fig. 4.19 les mouvements apparents des FPs mis en correspondance.

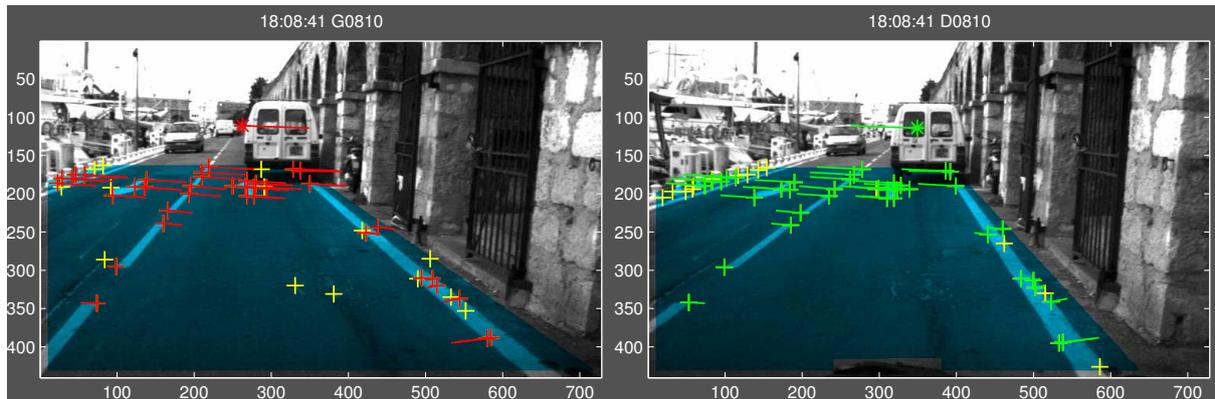


FIG. 4.18 – Représentation des déplacements des FPs coplanaires mis en correspondance entre deux images de la paire stéréo, en rouge image gauche, en vert image droite. Le déplacement de la projection des coins vérifie nos hypothèses : les ordonnées sont sensiblement égales, les translations et rotations entre les deux repères caméra (déplacement relativement important du DVP $'^*$) sont à l'origine des disparités essentiellement portées par l'axe des abscisses. Le deuxième FP en correspondance en partant du coin inférieur gauche des images semble proche du centre instantané de rotation de la transformation qui lie les deux repères caméra car son déplacement est très faible.

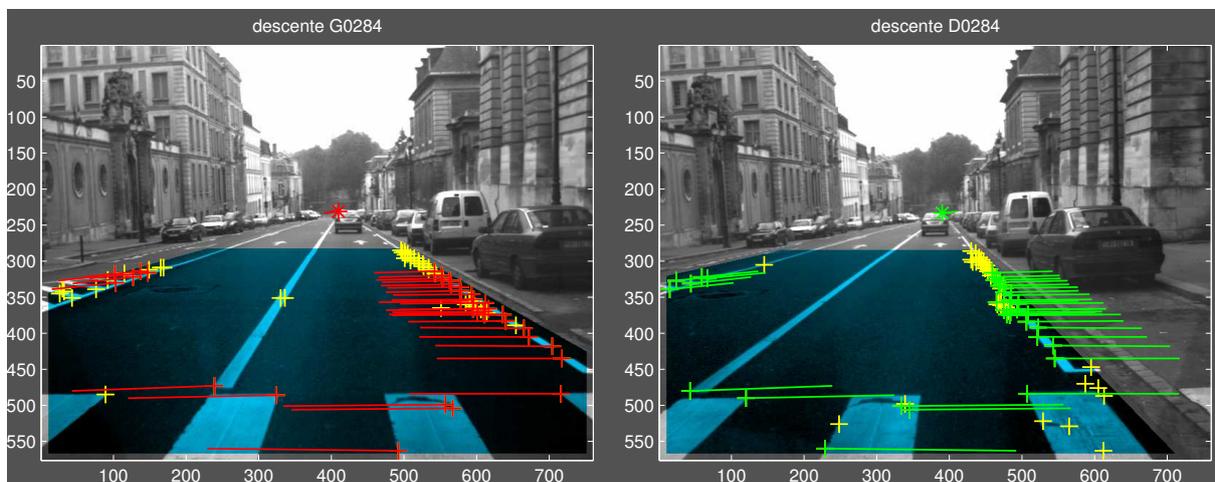


FIG. 4.19 – Représentation des déplacements entre FPs en correspondance lorsque les axes optiques des caméras du banc stéréo sont alignés. Les FPs détectés dans les ROIs sont représentés en jaunes. Les FPs coloriés en rouge et vert ont été mis en correspondance suite au calcul de l'homographie stéréo H_{st}^n . La transformation rigide qui lie les deux repères caméras est essentiellement une rotation autour de l'axe vertical. Les axes optiques des deux caméras sont proches de l'horizontale car les ordonnées des DVPs $'^*$, respectivement $(409,8;231,1)$ et $(390,1;233)$, sont légèrement supérieures à la moitié de la hauteur de l'image (568 pixels) alors que le véhicule suit une route en pente.

4.2.4.2 Mise en correspondance de points d'intérêt par la méthode de Pilu

Origine de la méthode L'étape de mise en correspondance est réalisée grâce à la méthode développée par [Pilu, 1997] pour des images stéréo. Il s'agit d'une amélioration de l'algorithme proposé par [Scott and Longuet-Higgins, 1991] qui permet de mettre en correspondance deux ensembles de FPs en exploitant une propriété de la décomposition en valeurs singulières d'une matrice de proximité $\text{Dist}[n_a; n_b]$. Cette matrice est une image de la distance qui sépare chacun des FPs, pondérée par une gaussienne :

$$\text{Dist}(i, j) = \exp\left(-\frac{\text{dist}^2(\mathbf{p}_i, \mathbf{p}_j)}{2 \cdot \sigma^2}\right). \quad (4.28)$$

Tous les éléments de Dist sont compris dans l'intervalle $[0; 1]$. La matrice de proximité permet de rechercher des couples de FPs susceptibles d'être en correspondance en fonction de leur proximité. La décomposition par SVD de cette matrice $\text{Dist} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^t$ permet de sélectionner dans un rapport 1 pour 1 les couples de FPs en correspondance en remplaçant les n_a premiers éléments de la matrice diagonale \mathbf{D} par une matrice identité \mathcal{I}_{n_a} . La matrice Dist_{svd} ainsi formée demeure une matrice de proximité particulière où s'applique le principe d'exclusion : si un élément s'avère le maximum simultané d'une ligne et d'une colonne, il signale la présence d'un couple de FPs en correspondance.

L'implémentation d'une telle méthode pour effectuer la mise en correspondance de FPs entre deux images présente quelques erreurs d'appariements liées à la proximité de FPs ne représentant pas le même point de la scène. L'algorithme de Scott et Longuet-Higgins correspond en effet à la minimisation de l'expression $\mathbf{A} \cdot \mathbf{Q} \mathbf{B}$ avec \mathbf{Q} matrice orthogonale et \mathbf{A} et \mathbf{B} sont les coordonnées des ensembles de FPs. Ainsi, c'est la somme des distances entre couples de FPs, soit $\text{tr}(\mathbf{Q}^t \mathbf{B}^t \mathbf{A})$, qui est minimisée, la matrice \mathbf{Q} étant directement liée à la décomposition SVD de $\mathbf{B}^t \mathbf{A}$.

L'hybridation des critères de distance et corrélation L'apport de Pilu a été d'introduire le résultat de l'étape de corrélation Corr entre les deux ensembles de FPs en complément de la matrice de proximité. La matrice hybride \mathbf{L} , formée par le produit membre à membre de deux matrices, nécessite une mise à l'échelle de Corr , à l'origine bornée entre $[-1; 1]$:

$$\mathbf{L}(i, j) = \frac{1 + \text{Corr}(i, j)}{2} \cdot \text{Dist}(i, j) \quad (4.29)$$

La sélection des couples de FPs en correspondance nécessite le calcul de la matrice \mathbf{L}_{svd} conformément à la méthode employée pour obtenir Dist_{svd} à partir de Dist . La détermination du seuil d'acceptation en dessous duquel le résultat de la décomposition de la matrice de proximité hybride \mathbf{L}_{svd} n'est pas pris en compte est une affaire de compromis. Trop élevé, le nombre de couples susceptibles d'être en correspondance sera faible. Un seuil trop faible tolère un nombre plus important de possibilités de correspondance pour un même FP, ce qui réduit la probabilité de trouver une correspondance un pour un pour chacun des FPs potentiellement en correspondance.

4.2.4.3 Adaptation de l'algorithme de Pilu à notre cas

Les algorithmes de mise en correspondance sont généralement conçus pour rechercher les couples de FPs se situant dans un proche voisinage. Notre cas est exactement à l'opposé de cette hypothèse. Nous proposons donc de se ramener au cas général de manière à appliquer la méthode décrite précédemment.

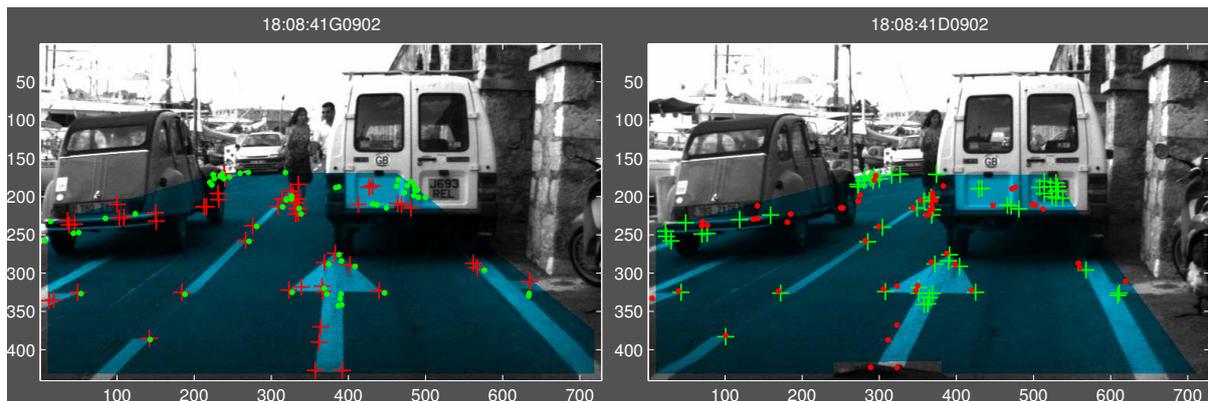


FIG. 4.20 – Représentation des FPs détectés dans les ROIs. Les primitives relatives aux images gauche et droite sont respectivement représentées en rouge et en vert. Les prédictions '+' des projections des FPs ('+') calculées grâce à la prédiction H_{st}^n de l'homographie stéréo, montrent que seules celles des FPs coplanaires se situent dans le voisinage de leur correspondant. Le calcul de la matrice de proximité $Dist$ est donc réalisé à partir de la distance qui sépare les prédictions des mesures, plutôt que de la distance qui sépare les coordonnées des FPs dans les deux images.

Prise en compte des connaissances a priori La disparité des écarts qui séparent les coordonnées des FPs en correspondance entre les deux images est telle qu'il n'est pas envisageable d'obtenir des résultats cohérents en utilisant directement ces données dans le processus décrit précédemment. En effet, les algorithmes de mise en correspondance ne traitent pas de tels écarts. En revanche, la connaissance d'une prédiction de l'homographie stéréo, aussi grossière soit-elle, permet de ramener les deux images dans un même repère. Substituer les coordonnées extraites par les prédites au moyen de l'homographie revient à considérablement réduire les distances qui séparent les couples de FPs potentiellement en correspondance. Les FPs ont alors des prédictions distantes de l'ordre de la dizaine de *pixels* de leurs correspondants comme le montre la Fig. 4.20.

Application directe de l'algorithme

Choix des seuils Les coefficients de corrélation des FPs détectés dans des régions peu texturées présentent parfois des variances importantes. Nous avons en effet remarqué que le coefficient de corrélation pouvait varier du simple au double en considérant un pixel voisin au pixel extrait. C'est pourquoi, le coefficient de corrélation retenu entre deux FPs potentiellement en correspondance est celui le maximum parmi un FP détecté et ses 8 plus proches.

La variance σ est choisie de telle sorte qu'un écart de projection maximal ($dist_max$) induise une contribution du facteur de proximité de l'ordre de 0,6. Ainsi, une prédiction se situant à moins de $\sigma/2$ aura un facteur de proximité égal 0,78. L'utilisation d'un critère de distance relativement lâche permet de compenser les erreurs de prédiction des FPs projetés d'une image vers l'autre en raison de l'imprécision de la prédiction de l'homographie stéréo disponible. Nous fixons donc $\sigma = dist_max$ dans un premier temps, puis $\sigma = dist_max/2$ dès que le nombre de couples et leur répartition dans les plans images sont suffisamment contraignants pour affiner l'identification.

Le calcul de corrélation est entrepris à partir de deux imagettes carrées de ($2 \cdot q \text{ pixels}$) de

coté, centrées en chacun des FPs potentiellement en correspondance. En fixant q à $dist_max$, écart maximal toléré entre la prédiction d'un FP et son correspondant, le seuil critique $seuil_score$ en dessous duquel deux FPs ne peuvent être considérés en correspondance a été choisi volontairement bas $seuil_score = 0,5$ de manière à ce que les FPs qui subissent le plus de déplacements (et par conséquent de déformation) entre les deux images du banc stéréo soient pris en compte. Les caractéristiques des FPs observés au premier plan ne vérifient plus forcément les conditions d'invariance que nous avons émises en hypothèse. Les coins en "L" à la base de la flèche peinte sur la route dans la figure précédente ont en effet des saillances qui évoluent en fonction de la position relative de la caméra.

Cas d'un exemple Fort des remarques du paragraphe précédent, l'implémentation de l'algorithme de Pilu est réalisée en substituant à la matrice de proximité $Dist$, la matrice $Dist_H$ qui représente les écarts exprimés dans un repère image unique. Cette dernière est calculée en faisant la différence entre les coordonnées prédites d'une image avec les coordonnées mesurées dans l'autre. Le résultat de la première mise en correspondance entre les deux images du couple stéréo est présenté en Fig. 4.21. Tous les couples de FPs identifiés sont correctement mis en correspondance. Cependant, sept d'entre eux sont détectés sur les véhicules, en dehors du plan de la route.

L'étude des caractéristiques des couples de FPs identifiés, regroupées dans le Tab. 4.5, permet de vérifier nos hypothèses. Les FPs détectés ont des scores élevés (supérieur à 100) pour la majorité d'entre eux : ils sont fortement caractérisés, des erreurs de mise en correspondance sont peu probables. Les distances qui séparent les prédictions des FPs de leur correspondant sont faibles ($\leq 8, 48\ pixels$) pour les FPs coplanaires et relativement élevées pour les FPs détectés sur les véhicules ($\geq 18, 63\ pixels$) à l'exception du $n^{\circ}5$ dont la mise en correspondance est erronée. Les coefficients de corrélation sont tous largement supérieurs à $seuil_score$, quel que soit le niveau de caractérisation (score de Harris) des FPs.

Réduction de l'espace de recherche des appariements Le calcul systématique des coefficients de corrélation entre tous les doublets de FPs s'avère lourd en temps de calcul et surtout inutile. En effet, la prédiction de l'homographie permet de réduire l'ensemble des FPs susceptibles d'être en correspondance. En fonction du seuil de distance choisi entre la prédiction et les FPs potentiellement candidats, l'algorithme de mise en correspondance sera plus ou moins tolérant aux faux appariements. Un seuil faible favorisera la qualité des appariements au détriment de leur nombre. A contrario, un seuil trop important risque de mettre en correspondance des FPs ne vérifiant pas l'homographie induite par le plan de la route comme c'est le cas des quatre couples de FPs détectés sur les véhicules (Fig. 4.21).

L'algorithme de Pilu a été développé de façon à mettre des FPs non-planaires en correspondance en n'ayant aucun a priori sur le type de mouvement de la caméra. La prise en compte d'une prédiction de l'homographie recherchée permet de restreindre considérablement l'espace de recherche entre les deux images : le nombre de FPs potentiellement en correspondance est alors limité.

Nous proposons donc d'exploiter au mieux la connaissance d'une prédiction de l'homographie recherchée. La présélection des FPs susceptibles d'être en correspondance s'effectue de manière symétrique pour les images du couple stéréo. Les prédictions des coordonnées des primitives sont calculées à partir de la prédiction d'homographie stéréo. Tous les FPs de la seconde image qui vérifient le critère de distance avec une prédiction d'un FP de la première image sont retenus. Seuls les couples de FPs dont les écarts entre les prédictions et les coordonnées extraites ont

été vérifiés de manière croisée entre les deux images sont présélectionnés. Leur coefficient de corrélation est alors calculé. Les couples susceptibles d'être en correspondance sont ceux dont le coefficient de corrélation est supérieur à $seuil_{corr}$.

Le fait de découpler les deux critères de sélection est à l'origine d'une matrice de proximité hybride creuse. La mise en correspondance par l'algorithme de Pilu aura alors pour rôle non plus de trouver les FPs en correspondance mais plutôt de discerner le meilleur correspondant parmi les possibilités restreintes préalablement sélectionnées.

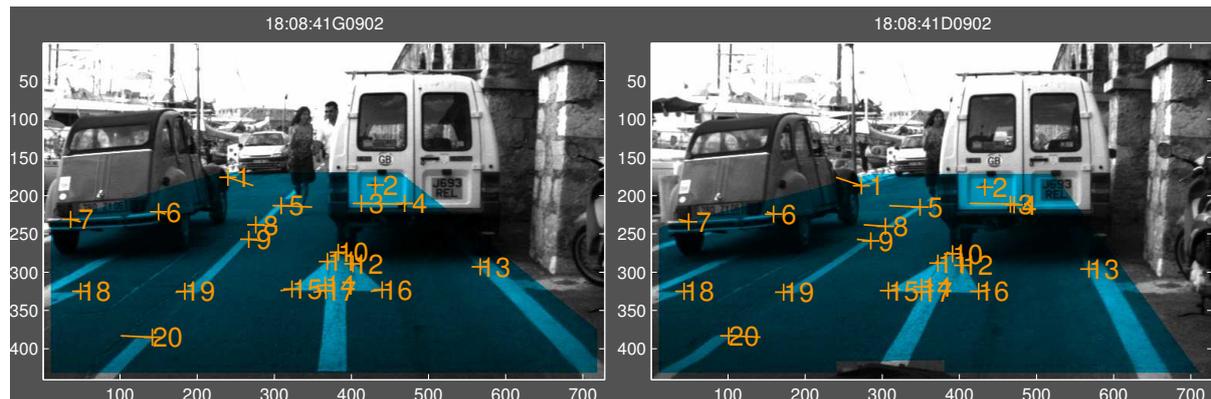


FIG. 4.21 – Détail de la mise en correspondance des FPs en utilisant par l'algorithme de Pilu. Les couples de FPs extraits sont correctement mis en correspondance à l'exception du n°3. Les segments représentent le déplacement apparent de chaque FP entre leurs coordonnées dans les deux images. Mais la prise en compte des couples de FPs 1 à 7 détectés sur les véhicules pose problème : les couples recherchés sont supposés appartenir au plan de la route.

n°	1	2	3	4	5	6	7	8	9	10
$score(p_g)$	378.55	151.58	82.05	111.57	341.70	251.47	97.17	145.34	182.30	174.86
$score(p_d)$	177.72	115.35	137.48	221.90	210.32	193.36	93.31	120.35	115.92	111.53
$dist$	22.34	43.78	18.63	31.86	5.43	26.09	21.76	4.66	3.64	4.32
$corr$	0.93	0.89	0.77	1.00	0.81	0.99	0.98	0.86	0.85	0.96

	11	12	13	14	15	16	17	18	19	20
	75.82	55.43	34.87	107.55	484.68	165.61	219.38	90.57	167.56	55.82
	58.99	51.31	39.63	86.34	348.79	204.89	532.08	84.01	88.80	30.22
	4.87	5.81	8.48	3.09	4.00	6.00	4.62	3.81	3.77	1.74
	0.91	0.81	0.88	0.90	0.92	0.87	0.89	0.92	0.88	0.88

TAB. 4.5 – Caractéristiques des FPs mis en correspondance grâce à l'algorithme de Pilu. Les lignes 2 et 3 représentent les scores de Harris des FPs, la ligne 4 la distance qui sépare les prédictions, calculées grâce à la prédiction de l'homographie stéréo, la ligne 5 le coefficient de corrélation normalisé entre les coordonnées des FPs à partir de fenêtres de 20 pixels de coté.

Pondération du critère de distance lors de la présélection Les FPs observés au premier plan ont un rôle prépondérant dans l'estimation de l'homographie : leurs mouvements apparents sont les plus importants, ils contraignent donc fortement le mouvement apparent de l'intégralité

du plan sur lequel ils sont détectés. Or, l'emploi de la prédiction de l'homographie \widetilde{H}_{st} permet de réduire la distance entre les FPs en correspondance lorsqu'ils sont exprimés dans le même repère image. Hélas, nous avons vérifié de manière expérimentale que cet écart a tendance à augmenté avec l'ordonnée des FPs.

Etant donné que la mise en oeuvre de la méthode de présélection à partir d'un seuil de distance fixe élimine la plupart des FPs détectés au premier plan, nous proposons de pondérer le seuil de comparaison en fonction de l'ordonnée de chaque FP. La pondération obéit à une loi linéaire :

$$pds(v) = 1 - 0,3 \cdot \frac{v - v_{dvp}}{v_{max} - v_{dvp}} \quad (4.30)$$

où v représente l'ordonnée du FP, v_{dvp} l'ordonnée du DVP et v_{max} l'ordonnée du bas de l'image, toutes exprimées en *pixels*. Ainsi, les distances entre les coordonnées prédites et mesurées des FPs seront comparées à un seuil inférieur à $2 \cdot dist_max$ lorsque le FP a une ordonnée proche de celle du bas de l'image, tandis qu'elle sera légèrement supérieur à $dist_max$ pour les FPs dont l'ordonnée est proche de celle du DVP.

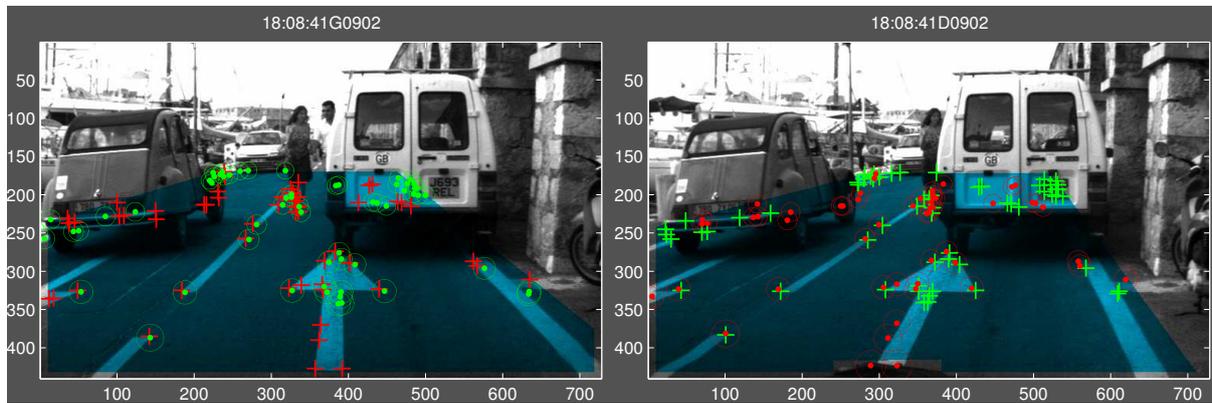


FIG. 4.22 – Mise en oeuvre de la pondération sur le critère de distance ($dist_max$) appliqué aux matrices de proximité lors de la présélection. Les FPs extraits dans une image sont susceptibles d'être en correspondance avec une prédiction de l'autre image s'ils se situent dans son voisinage, représenté par un cercle. Le rayon de chaque cercle est inversement proportionnel à l'ordonnée du FP de la prédiction considérée. L'introduction du facteur de pondération permet de s'affranchir de l'inexactitude de la prédiction de l'homographie stéréo qui a tendance à éloigner les FPs en correspondance au premier plan.

La présélection s'effectuant de manière symétrique sur les deux images, considérons les FP \mathbf{p}_1 et \mathbf{p}_2 des images \mathcal{I}_1 et \mathcal{I}_2 et H_{21} l'homographie planaire induite par la route dans ces deux images. Les FPs seront considérés comme potentiellement en correspondance s'ils vérifient simultanément :

$$\begin{cases} pds1 \cdot pds2 \cdot dist(H_{21} \cdot p_1, p_2) \leq dist_max \\ pds1 \cdot pds2 \cdot dist(p_1, H_{21}^{-1} \cdot p_2) \leq dist_max \end{cases} \quad (4.31)$$

où $pds1$ et $pds2$ sont respectivement les poids des FPs \mathbf{p}_1 et \mathbf{p}_2 . Dans notre cas, les FPs en correspondance ayant des ordonnées semblables, les rayons des régions autour de chaque prédiction sont tels que $r = dist_max / pds1 * pds1$. Chacun des poids étant compris entre $[0,7;1]$, le produit $pds1 * pds2$ permet quasiment de doubler le critère de distance des FPs du premier plan en comparaison de ceux situés au niveau de la ligne d'horizon, comme le montre la Fig. 4.22.

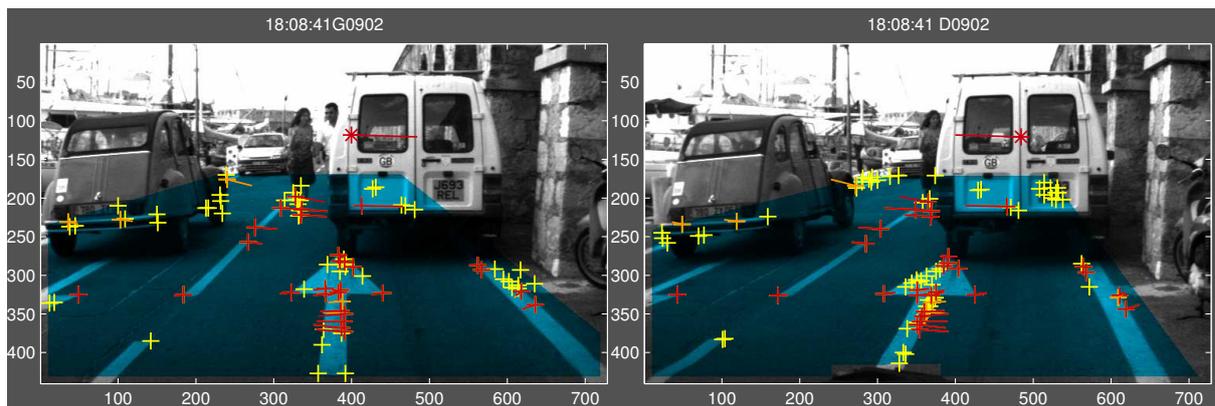


FIG. 4.23 – Comparaison entre les résultats fournis par l'application directe de la méthode de Pilu (FPs orange) et la méthode proposée (FPs rouges) qui restreint l'espace de recherche lors de la mise en correspondance des FPs à partir de la prédiction de H_{st} . Les coordonnées des prédictions des projections des FPs des deux images ont été calculées à partir de H_{st}^n et H_{st}^{n-1} . La prise en compte massive des distances pondérées et les corrélations entre tous les FPs de chaque image induisent en erreur l'algorithme de mise en correspondance un pour un : parmi les couples de FPs oranges, quatre couples ne sont pas détectés sur la route. Par contre, la sélection de couples de FPs qui vérifient les critères de distance et de corrélation limitent l'espace des possibilités : plus qu'un seul couple mis en correspondance est détecté en dehors du plan de la route.

La mise en oeuvre de la présélection des couples de FPs qui vérifient le critère de distance est directe. Il suffit de calculer la matrice de proximité $Dist_g$ (resp. $Dist_d$) qui représente les distances qui séparent les prédictions des FPs de l'image \mathcal{I}_g^n (resp. \mathcal{I}_d^n) des FPs extraits dans \mathcal{I}_d^n (resp. \mathcal{I}_g^n). La pondération s'effectue en multipliant membre à membre $Dist_g$ (resp. $Dist_d$) avec une matrice poids obtenue en multipliant le vecteur colonne correspondant au poids des FPs de \mathcal{I}_g^n avec son dual de \mathcal{I}_d^n sous forme de vecteur ligne.

La décomposition par SVD de la matrice creuse de proximité hybride L n'introduit plus de manière erronée de nouveaux couples dont le coefficient de proximité hybride est supérieur au seuil toléré (Fig. 4.23). Ainsi, trois des quatre couples de FPs détectés sur les véhicules ne sont plus retenus. La contre-partie de l'emploi d'une matrice L creuse est que les couples de FPs fournis par l'algorithme de mise en correspondance un pour un sont choisis parmi les couples présélectionnés. La détection de nouveaux couples à partir de la nouvelle estimée de l'homographie nécessite la mise en oeuvre d'une méthode itérative.

Cas où la prédiction de l'homographie stéréo est mal définie Nous devons cependant préciser que le couple d'images présenté a été choisi pour sa représentativité des conditions de trafic en milieu urbain : le champ libre de vision est réduit à quelques mètres. Par contre, la forte densité de FPs détectés en raison d'un marquage au sol important et de qualité (à fort contraste) est, elle, très rare en comparaison à l'ensemble des séquences vidéos en notre possession enregistrées dans différentes villes. La principale conséquence de cette forte densité de FPs coplanaires est de fournir une prédiction de l'homographie stéréo de grande qualité qui autorise une discrimination partielle des FPs en correspondance à partir du simple critère de distance.

L'utilisation de ce seul critère de distance ne s'avère pas assez discriminant en général lorsque

la prédiction de l'homographie stéréo est peu fiable comme c'est le cas dans les figures Fig. 4.24(a) et Fig. 4.24(b). Ceci est essentiellement dû à une répartition non-uniforme des FPs dans les ROIs. La distance qui sépare les prédictions des coordonnées des FPs par l'homographie $\widehat{\mathbf{H}}_{st}^n$ de leur correspondant peut excéder les seuils pondérés initialement calculés. Les FPs extraits '+' sont alors en dehors des cercles de tolérance des prédictions '.' correspondantes.

L'algorithme de mise en correspondance parvient néanmoins à fournir de nombreux couples de FPs choisis parmi les couples de FPs présélectionnés en fonction des critères de distance et de corrélation. Il en résulte une mise en correspondance erronée ayant deux origines distinctes : la prédiction de l'homographie n'est pas fiable et la duplication du motif de bandes blanches pointillées fausse le résultat car l'hypothèse que les FPs en correspondance présente un maximum du coefficient de corrélation n'est pas respectée.

L'intégralité de l'étape de présélection et de sélection de FPs potentiellement en correspondance est alors reprise en doublant le critère de distance $dist_max$. Si l'ambiguïté demeure sur la mise en correspondance des FPs extraits sur les coins des bandes blanches pointillées, elle est fortement réduite sur les autres FPs (par exemple autour des lettres de "payant") dont la caractérisation par le coefficient de corrélation est plus évidente. Nous présentons en Fig. 4.24(c) et Fig. 4.24(d) l'évolution du nombre de couples de FPs mis en correspondance après la première et la dernière itération du processus. Les images sont formées par le warping de l'image \mathcal{I}_g^n dans \mathcal{I}_d^n suite à l'estimation de l'homographie calculée avec les couples de FPs en correspondance. Le processus itératif parvient à mettre tous les FPs extraits en correspondance bien que de nombreux couples de FPs étaient erronés lors de la première itération.

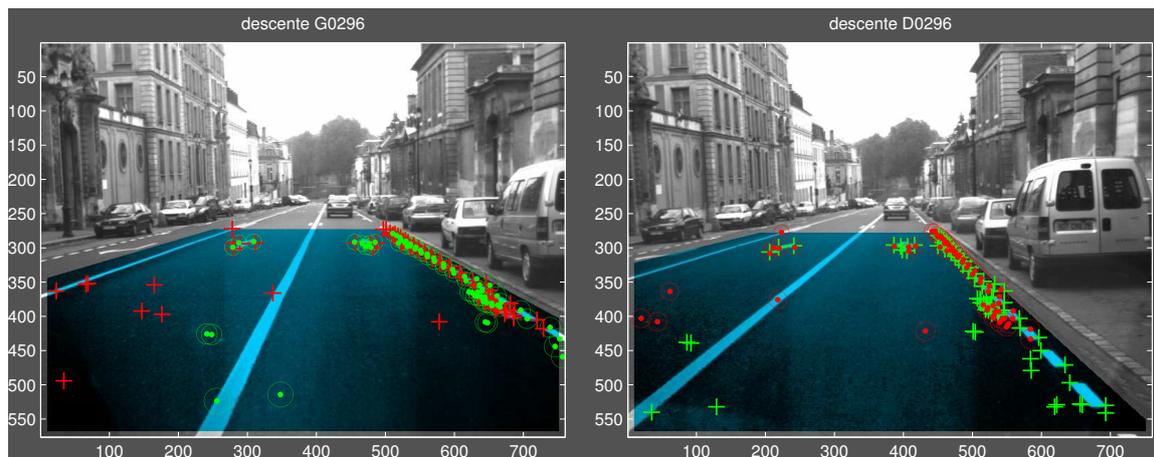
4.2.4.4 Récapitulatif de la méthode employée

La mise en correspondance des FPs à partir de la prédiction de $\widetilde{\mathbf{H}}_{st}^n$ s'effectue de manière itérative. Chacune des étapes peut se résumer de la sorte :

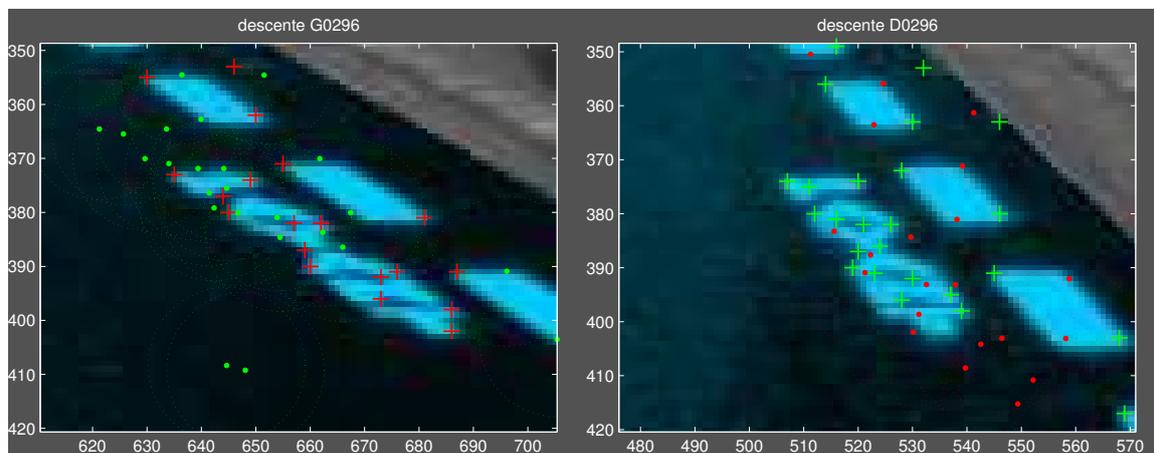
tant que les couples de FPs en correspondance évoluent,

1. calcul des coordonnées prédites des FPs des deux images grâce aux estimations courantes de \mathbf{H}_{st}^n et \mathbf{H}_{st}^{n-1} , puis calcul des matrices de proximité Dist_g et Dist_d entre coordonnées prédites et extraites ;
2. mesure de l'écart en ordonnée entre les coordonnées extraites des deux images ;
3. présélection des couples qui vérifient le critère de distance de manière croisée ;
4. calcul des coefficients de corrélation normalisé ;
5. sélection des couples dont le coefficient de corrélation est supérieur à $seuil_corr$;
6. mise en correspondance des points d'intérêt dans un rapport un pour un en décomposant par SVD la matrice hybride de $Pilu$ creuse ;
7. estimation de \mathbf{H}_{st}^n à partir des couples de FPs en correspondance ;

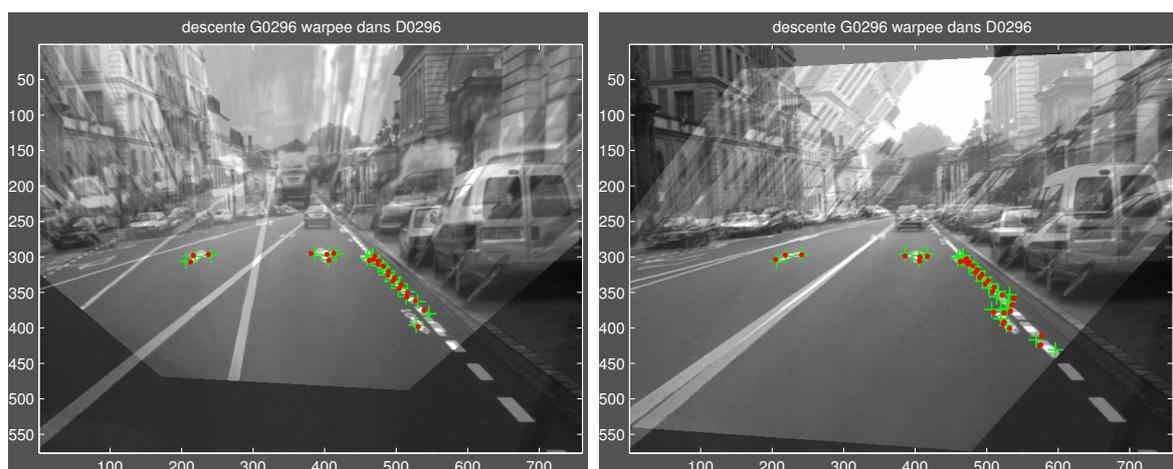
fin tant que



(a)



(b)



(c)

(d)

FIG. 4.24 – Recherche de couples de FPs en correspondance lorsque la prédiction de l'homographie stéréo est parfaite. Les FPs '+' extraits et leurs prédictions '.' sont respectivement représentés en rouge et vert pour les primitives des images gauche et droite du banc stéréo. Au centre, le détail montre que la plupart des couples de FPs sélectionnés à partir du critère de distance ne contiennent pas leur correspondant. En bas, les images hybrides sont obtenues par le warping de \mathcal{I}_g dans \mathcal{I}_d . Les homographies sont calculées à partir des couples de FPs extraits lors de la première et de la dernière itération (r) de l'opérateur de mise en correspondance : le nombre de couples de FPs correctement identifiés a augmenté ce qui permet une meilleure

4.3 Estimation de l'homographie

4.3.1 Difficulté de mise en oeuvre

L'estimation de l'homographie à partir de primitives extraites d'images de scènes réelles est délicate pour plusieurs raisons ;

- la distorsion introduite par les optiques des caméras ne satisfait pas le modèle à projection centrale,
- les données sont bruitées par la quantification et l'échantillonnage,
- les conditions de luminosité peuvent être très variables dans une même image et ne pas être compensée par le contrôle automatique de gain de la caméra. L'égalisation des niveaux de gris opérée dans la ROI avant l'extraction des FPs apportant une amélioration tangible peut dans certains cas être insuffisante,
- les extracteurs de primitives fournissent des résultats ne correspondant pas exactement à la localisation réelle des FPs et des contours,
- la région apparentée à la route ne représente qu'une fraction de l'image,
- les primitives extraites dans la région apparentée à la route ne sont pas distribuées de manière uniforme.

La résolution du système linéaire formé par les caractéristiques des primitives extraites réserve encore une difficulté notoire : les caractéristiques des FPs et des VLs vérifient des systèmes linéaires transposés inverses (cf 3.3) :

$$\mathbf{p}_a \simeq \mathbf{H}\mathbf{p}_b \quad \text{et} \quad \mathbf{l}_b \simeq \mathbf{H}^t\mathbf{l}_a \quad (4.32)$$

Il est toutefois possible de rassembler les relations entre les coordonnées des deux types de primitives en correspondance dans un unique système linéaire. Le conditionnement de ce système est alors notablement détérioré du fait des différences de dimensions entre les coefficients des points et des droites. Plusieurs remarques s'imposent de manière à déterminer la meilleure option qui permettra d'obtenir un système linéaire dont le conditionnement est élevé :

- le faisceau de VLs coplanaires qui converge vers le DVP n'est pas suffisant au calcul de l'homographie entre deux images, contrairement à quatre FPs coplanaires, non-alignés trois à trois. Il s'agit en effet d'un cas dégénéré que nous traitons en détail en 3.3.2.3,
- la qualité des VLs extraites à partir de notre algorithme dépend grandement de la présence de structures dans la scène. En cas de courbure de la route, seuls les contours du 1^{er} plan seront exploitables ce qui rendra l'extraction de droites-supports inefficace et les caractéristiques déduites seront donc peu fiables,
- l'intersection des VLs à l'image forment le DVP qui n'est autre qu'un FP particulier. Il est systématiquement intégré à tous les couples de FPs lors des estimations d'homographie,
- le nombre de FPs est dans la majorité des cas plus important que celui des VLs et nombre d'entre eux représentent des coins à l'extrémités de contours, à l'origine de VLs.

C'est pourquoi, nous avons décidé d'améliorer le conditionnement du système en vue de sa résolution en faisant abstraction des caractéristiques des VLs. Cependant, les VLs continueront à jouer leur rôle primordial lors de la segmentation de la région de la route et dans les homographies par le biais du DVP.

4.3.2 Résolution du système linéaire

L'homographie stéréo recherchée est celle qui minimise l'erreur de re-projection entre les différents couples de FPs identifiés comme coplanaires. Le calcul de l'homographie à partir

des coordonnées homogènes des projections dans les deux images des éléments caractéristiques détectés sur la route consiste à résoudre un système linéaire de 8 inconnues et $2n_{match}$ équations (n_{match} étant le nombre de couples de FPs en correspondance). Malgré son apparente simplicité, le système linéaire ainsi formé nécessite, entre autres, une normalisation de façon à améliorer son conditionnement (détails dans la partie 3.3.2).

Quantifier la qualité de l'estimation d'une homographie s'avère difficile : nous avons tout d'abord utilisé le conditionnement du système formé par les coordonnées des couples de FPs en correspondance. Nous employons à présent la variance de la distance de projection qui représente un critère moins abstrait. Dans les deux cas, le nombre de primitives en correspondance ainsi que leur représentativité en terme de contraintes sur le déplacement estimé n'est pas pris en compte.

4.3.3 Initialisation de l'algorithme en début de séquence

Lors de la première itération ($n = 1$), nous utilisons une prédiction de l'homographie stéréo établie à partir des hypothèses suivantes. Nous supposons une matrice de paramètres intrinsèques générique $K_0 = K_g = K_d$ et considérons que les caméras ont des plans images parallèles et que la ligne de base, formée par les deux centres optiques, est orientée selon l'axe horizontal \mathbf{u} . Les caméras sont supposées dans un plan perpendiculaire à la route, situées à une altitude de d^0 . La matrice d'homographie stéréo initiale H_{st}^0 est donc égale à :

$$H_{st}^0 = K_d \cdot \left\{ R + \frac{\mathcal{I}_3 + T \cdot N^t}{d} \right\} \cdot K_g^{-1} \quad (4.33)$$

$$= \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{1}{d^0} \cdot \begin{bmatrix} T_u^0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \end{bmatrix} \right\} \cdot \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \quad (4.34)$$

avec $[u_0, v_0]$ coordonnées du point central et f focale des caméras supposées identiques.

Une alternative est alors envisageable pour estimer H_{st}^1 . Si le champ de vision est obstrué par des obstacles, l'algorithme nécessite l'intervention d'un opérateur de manière à clairement identifier la région des images qui correspond à la route. Dans le cas inverse (celui que nous utilisons en priorité), les bandes principales détectées ne pouvant être vérifiées par projection selon H_{st}^0 , la ROI est exceptionnellement prise comme la région du bas de l'image, bornée par les droites-supports extrêmes détectées sous l'horizon. Cette segmentation grossière englobe des régions de la scène autre que la route, dans lesquelles des FPs non-coplanaires sont détectés.

La mise en correspondance des FPs entre les deux images du banc stéréo s'effectue en supposant que H_{st}^0 joue le rôle de prédiction de l'homographie stéréo. Les couples de FPs mis en correspondance sont en majorité correctement appariés, même si nombre d'entre eux n'appartiennent pas au plan de la route. L'estimation de l'homographie H_{st}^1 ne peut être calculée que si le nombre de couples de FPs associés au plan de la route est majoritaire dans la ROI.

L'emploi d'une méthode robuste RANSAC (cf. Fischler and Bolles [1981]) permet de déterminer quels sont les couples de FPs coplanaires. L'algorithme recherche alors le quadruplet de FPs en correspondance qui permet de projeter correctement le plus de couples de FPs entre les deux images. La recherche s'effectue à l'aide d'un nombre fixé de tirages aléatoires de quadruplets. La procédure qui permet de déterminer le nombre de tirages aléatoires est développée en B. La représentativité de chaque homographie estimée est vérifiée en comptabilisant le nombre de couples de FPs dont l'erreur de projection est inférieure au seuil de distance $seuil_{proj_ransac} = \sqrt{5}, 99\sigma \text{ pixels}$, d'après [Hartley and Zisserman, 2003].

4.3.4 Décomposition de l'homographie

La décomposition de l'homographie stéréo n'est pas à première vue indispensable à notre objectif final qui est d'estimer le déplacement de la tête stéréo entre deux acquisitions. Il peut être cependant intéressant de vérifier que les homographies stéréo estimées soient cohérentes les unes par rapport aux autres. Nous avons en effet vu que l'évolution de l'homographie stéréo au cours du temps ne dépendait que de l'inclinaison de la paire stéréo par rapport au plan de la route. La transformation qui lie les deux repères caméras étant rigide, l'évolution brutale des paramètres liés à la géométrie du banc stéréo est alors un témoin que des couples de primitives coplanaires mis en correspondance sont erronés.

En effet, considérons le cas où des couples de FPs mis en correspondance n'appartiennent pas au plan de la route. Tout se passe comme si le plan qui induit l'homographie correspondait à un plan moyen qui minimise la distance entre les points de la scène dont les projections sont en correspondance. Outre la normale, la distance d^n qui sépare le repère de la caméra gauche C_g^n du plan de la route subira aussi une variation. Dans ce cas, les trois paramètres liés à la translation entre les deux repères subiront une variation commune.

La détermination de cas où l'homographie stéréo présente de telles variations n'est toutefois pas aussi évidente : le plan moyen s'avère très proche de celui de la route car les points de la scène qui ne sont pas détectés sur cette dernière se situent généralement à des altitudes de quelques centimètres. Dans la prochaine section, nous présentons l'évolution de la décomposition de l'homographie stéréo sur une séquence. Nous allons voir que l'estimation de l'homographie stéréo nécessite de détecter des primitives réparties sur l'ensemble de la ROI sans quoi les résultats de la décomposition s'avèrent incohérents.

4.4 Résultats : évolution de la géométrie du banc stéréo au cours du temps

4.4.1 La séquence du port d'Antibes

Nous avons enregistré plusieurs séquences vidéo de résolution [728; 440] *pixels*, dans les rues du Vieil-Antibes et les abords du port. Nous avons utilisé une paire de caméras stéréo non-calibrées, fixées derrière le pare-brise. Les acquisitions ont été enregistrées à 25 images/s avec une vitesse du véhicule qui n'a pas dépassé 10 m/s. Le détail de la séquence est fourni en A.1.

Les résultats présentés correspondent à une séquence dans des conditions de trafic favorables : la route est dégagée, très structurée (signalisation horizontale abondante) avec un revêtement homogène. Nous avons pu cependant vérifier le bon fonctionnement de la détection de VLs et DVPs sur des séquences plus délicates comme des trajectoires en courbes et/ou la proximité d'obstacles qui réduisent le champ de vision.

Bien que la scène soit une route en ligne droite, la trajectoire du véhicule n'est pas rectiligne. Le véhicule se déplace en ligne droite, ralentit en se décalant sur la droite devant un premier véhicule en stationnement pour se laisser dépasser par un véhicule, redémarre en contournant l'obstacle et se rabat rapidement pour éviter un second véhicule en stationnement, garé sur la voie opposée, puis poursuit son déplacement en ligne droite ¹.

¹Une animation est disponible sur <http://www-sop.inria.fr/icare/personnel/nsimond/index.html>

4.4.2 Décomposition de l'homographie stéréo

Nous présentons en Fig. 4.25 la décomposition de l'évolution de la matrice stéréo H_{st}^n au cours de la séquence du port. Ne connaissant pas les paramètres intrinsèques des caméras, nous avons été contraint d'introduire une matrice générique K_0 que nous supposons proche des paramètres réels internes des caméras qui composent le banc stéréo. En considérant le plan image quasiment perpendiculaire à la route et sa normale orientée dans la direction de cette dernière, la représentation de la normale du plan de la route observée dans l'image est essentiellement portée par l'axe des ordonnées, ce qui revient dans notre cas à ne considérer que la solution dont la normale est la plus proche du vecteur $N_0^t = [0; -1; 0]$, exprimé en coordonnées non-homogènes.

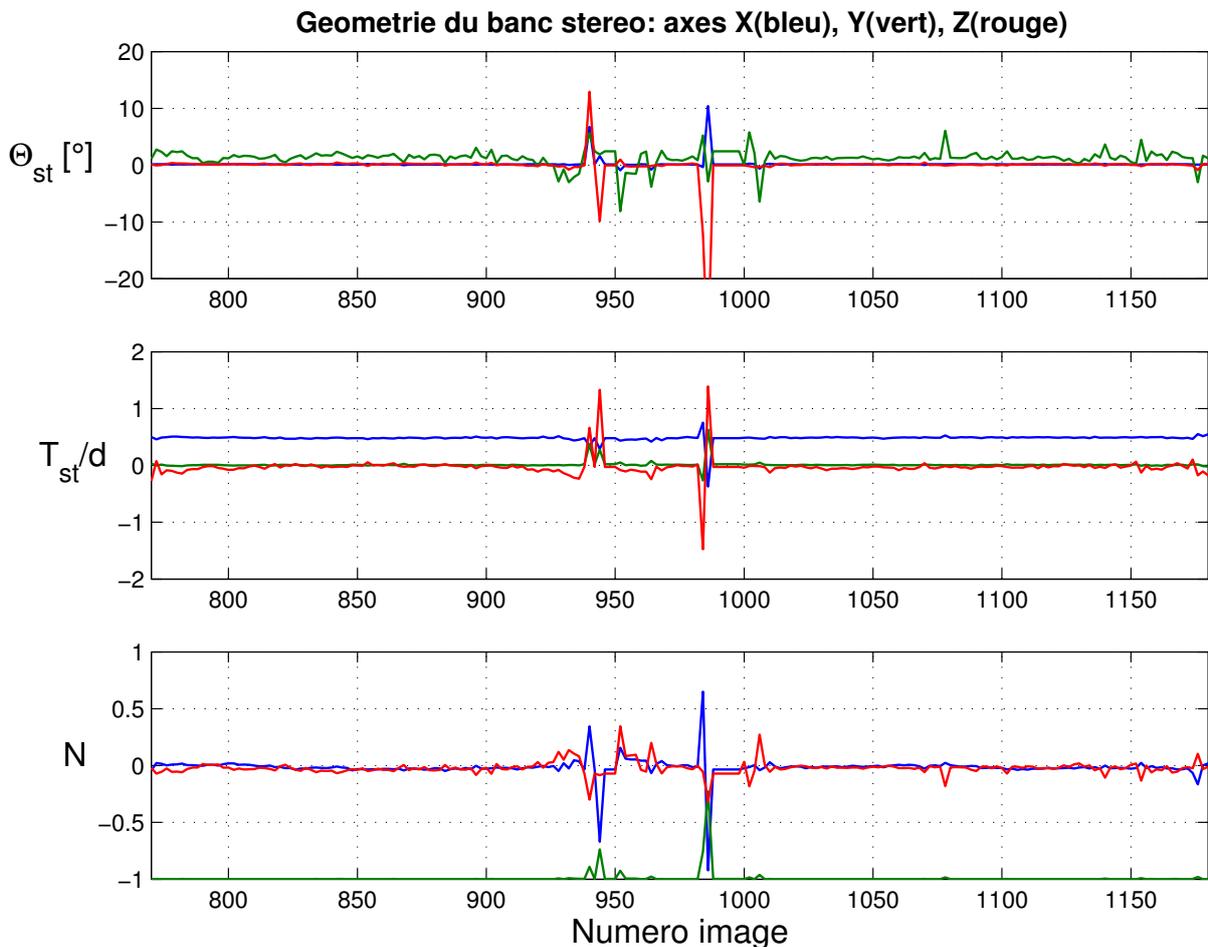


FIG. 4.25 – Décomposition des mouvements de rotation θ_{st} et translation normalisée T_{st}/d entre les repères C_g et C_d et évolution de la normale au plan de la route N . Les paramètres intrinsèques des caméras étant inconnus, les valeurs numériques, obtenues à partir d'une matrice générique, sont données à titre indicatif. L'hypothèse de rigidité de la paire stéréo est vérifiée sur la majeure partie de la séquence. Les écarts observés entre les images [936,948] et [986,998] sont dues à une mauvaise répartition des primitives coplanaires détectées sur le plan de la route.

Les mouvements de translation et rotation représentés sont ceux du repère de la caméra droite C_d exprimés dans le repère de la caméra gauche C_g . Les composantes du mouvement de

translation sont normalisées en fonction de la distance d qui sépare le repère \mathbf{C}_g du plan de la route. Les chronogrammes montrent que l'hypothèse de rigidité de la transformation qui lie les repères des caméras gauche et droite est vérifiée. Le comportement stationnaire de l'évolution de chacun des paramètres de mouvement est cependant fortement perturbé par deux fois entre les images [936,948] et [986,998]. La portion de route visible dans les images (Fig. 4.26) est alors réduite à l'équivalent d'une voie, observée de surcroît de biais. Aucun FP n'est détecté dans le voisinage des bandes blanches peintes continues, la répartition des primitives coplanaires durant ces instants est dégénérée. L'homographie stéréo étant recherchée comme la solution du système linéaire formé par les coordonnées des primitives en correspondance, son estimation ne peut être qu'erronée car le système linéaire est mal conditionné voire sous-déterminé.



FIG. 4.26 – Cas dégénérés où la répartition des FPs en correspondance n'est pas suffisamment uniforme dans les ROIs pour contraindre correctement l'homographie stéréo induite par le plan de la route. De plus, les quelques FPs détectés sur le véhicule en stationnement perturbent l'estimation de l'homographie stéréo. Lorsque leur nombre devient équivalent à ceux extraits sur la route, le plan qui induit l'homographie estimée correspond à un plan moyen se situant entre la route et le capot de la voiture.

En faisant abstraction de ces situations particulières, ces résultats montrent que la transformation qui lie le repère \mathbf{C}_d au repère \mathbf{C}_g est quasi-rigide. La caméra droite se situe dans le prolongement de l'axe des abscisses x de la caméra gauche à une distance de $0,49d$ avec d estimé à $1,2m$. L'axe optique de la caméra droite est légèrement pincé ($1,5^\circ$) en direction de celui de la caméra gauche. De manière plus précise, il apparaît un décalage de $-0,05.d$ selon

l'axe \mathbf{Z} . L'orientation de la normale au plan de la route observée dans le repère \mathbf{C}_g connaît les mêmes perturbations que celles observées sur les déplacements. Nous rappelons que les valeurs numériques obtenues dépendent des matrices de paramètres intrinsèques que nous avons imposées au préalable.

La mauvaise estimation de l'homographie stéréo sur la partie délicate de la séquence n'est due qu'à la répartition non-uniforme des FPs détectés, supposés être sur la route. La prise en compte de \mathbf{H}_{st}^{n-1} en tant que prédiction de la future estimée \mathbf{H}_{st}^n mène à une divergence de l'algorithme aux alentours de l'itération 940. A ce stade aucun processus temporel de filtrage n'est mis en oeuvre. La mise en correspondance correcte des FPs supposés détectés sur la route entre les images [936;998] n'a pu être réalisée qu'en imposant en tant que prédiction de l'homographie stéréo la matrice $\mathbf{H}_{st}^n = \mathbf{H}_{st}^{10}$, estimée grâce au dixième couple d'images de la séquence.

4.5 Conclusion

Nous avons présenté dans ce chapitre l'intérêt d'utiliser un banc stéréo dans l'objectif de segmenter les plans principaux lorsque le système de vision est embarqué dans un véhicule qui se déplace en milieu urbain. La région de l'image apparentée à la route est généralement facilement détectable grâce à la signalisation horizontale qui délimitent les différentes voies de circulation. En condition de trafic urbain, ces délimitations sont souvent masquées par des véhicules en mouvement.

Cependant, en tenant compte du fait que la largeur des voies est constante, la projection des contours de la scène orientés dans la direction de la route convergent dans l'image vers le point de fuite dominant. En cas de virage, la supposition précédente ne demeure valable que pour les contours extraits au premier plan de l'image. Le point de fuite dominant correspond à la projection dans l'image de l'intersection de la direction de la route avec le plan à l'infini. Ses coordonnées ne dépendent que des mouvements de rotation de la caméra. A la cadence vidéo, les coordonnées de ce point particulier évoluent lentement dans les images dans des conditions de conduite normales.

L'identification des bandes principales, projections des frontières de la route dans l'image à partir du faisceau de droites qui convergent vers le point de fuite dominant, est facilitée par la prise en compte d'une prédiction. La région d'intérêt relative à la route est segmentée en considérant le parallélogramme délimité par la première et dernière bande principale, le bas de l'image et une horizontale se situant à une distance fixe du point de fuite.

L'identification des points d'intérêt et par la suite des lignes de fuite détectées sur la route, considérée localement plane, s'effectue en calculant l'homographie stéréo qui lie les projections du plan de la route entre les images gauche et droite du banc stéréo. En considérant que la transformation qui lie les deux repères caméras est quasi-rigide et que les paramètres intrinsèques des caméras sont invariants, cette homographie ne dépend que de la représentation de la normale au plan considérée et de la distance qui sépare le plan de la caméra qui sert de référence.

Dans le cas des séquences en notre possession, nous avons vérifié que les fluctuations mesurées sur ces deux éléments étaient réduites entre deux images successives. Ainsi, l'estimation de l'homographie stéréo induite par le plan de la route entre les deux images du banc stéréo peut être considérée comme une prédiction fiable de l'homographie stéréo estimée à l'itération suivante et faciliter la mise en correspondance des primitives coplanaires détectées dans les régions d'intérêt.

Les résultats de la décomposition de l'homographie stéréo obtenus dans des conditions difficiles demeurent toutefois incorrects. Nous proposons dans le chapitre suivant de réduire l'in-

certitude liée à ces cas en exploitant les aspects temporels des séquences d'images.

Chapitre 5

Suivi des plans principaux par vision dynamique

La segmentation puis la recherche des primitives qui vérifient l'homographie stéréo H_{st}^n décrite dans la section précédente a pour seule vocation de discriminer parmi l'ensemble des primitives extraites celles qui appartiennent aux plans principaux de la scène et en particulier celui de la route. Comme nous l'avons vu, des erreurs subsistent : des primitives coplanaires sont oubliées et d'autres n'appartenant pas à la route sont détectées comme telles. En dépit de ces aléas, inhérents à la détection de primitives fiables et stables en environnement extérieur, la majorité des primitives qui vérifient l'homographie stéréo sont effectivement coplanaires.

Le problème que nous abordons dans ce chapitre est l'estimation du déplacement du système stéréo, et par conséquent du véhicule, entre deux poses successives à partir du mouvement apparent des plans segmentés dans les images. Nous nous appuyons pour cela sur les primitives coplanaires détectées grâce à la mise en correspondance stéréo décrite au chapitre précédent. La difficulté de la réalisation de cette tâche de suivi dans le cas d'un véhicule se déplaçant en milieu urbain est liée au nombre et à la répartition spatiale des primitives coplanaires susceptibles d'être observées au cours du temps. En effet, la projection des plans segmentés grâce à la mise en correspondance entre les images stéréo évoluent grandement au cours du temps en fonction de la vitesse du véhicule et du fait des nombreux obstacles statiques et dynamiques qui obstruent le champ de vision.

Le plan de la route observé au premier plan représente alors le seul élément structurant commun à toutes les scènes urbaines libre de toute occultation lorsque les distances de sécurité sont respectées entre deux véhicules. Dans l'éventualité où le nombre de primitives coplanaires détectées grâce au marquage au sol s'avère insuffisant pour estimer le déplacement du véhicule, nous supposons pouvoir compenser la perte d'observabilité temporaire en fusionnant les estimations de déplacement de la caméra, estimés à partir des autres plans principaux segmentés.

La méthode employée pour entreprendre le suivi de tels plans est identique à celle utilisée sur le plan de la route. Toutefois, la méthode proposée étant développée dans un esprit générique, aucun a priori sur la configuration de l'environnement urbain n'est imposé. Le suivi de façades verticales n'est pas nécessaire à l'estimation du déplacement du véhicule.

Nous rappelons que les séquences vidéo à notre disposition ont été enregistrées à partir de caméras non-calibrées. La transformation qui lie les repères caméras de la paire stéréo est supposée rigide. Le véhicule se déplace sur une route plane par morceaux, tout du moins au premier plan : le mouvement est supposé approximativement plan puisque un mouvement d'incli-

raison lors de phases d'accélération et de freinage est observé. La vitesse du véhicule dépend de la séquence considérée, sur certaines d'entre elles, elle dépasse largement la vitesse maximale du cadre de cette étude, à l'origine limitée à 10 m/s soit un déplacement de 40 cm entre deux images enregistrées à la cadence vidéo de 25 images/s.

Dans une première partie, nous présentons comment segmenter le plan de la route dans l'image de manière robuste à partir d'un simple modèle linéaire. Dans la seconde partie, nous expliquons comment calculer le déplacement d'une caméra à partir du mouvement apparent des primitives extraites du plan de la route.

5.1 Segmentation du plan de la route par filtrage temporel

Le modèle linéaire de projection du plan de la route dans l'image que nous avons retenu est caractérisé par les bandes principales qui convergent vers le point de fuite dominant. Le point de fuite dominant a la particularité d'être à l'intersection des bandes principales et de toutes les droites supports parallèles à la direction de la route. Le suivi temporel de ces deux types de caractéristiques introduit une redondance des informations relatives à la projection de la route dans l'image.

Nous proposons désormais de substituer un modèle à vitesse constante au modèle d'évolution stationnaire employé jusqu'alors pour estimer les caractéristiques des bandes principales car ce dernier présentait quelques faiblesses lors des mouvements de rotation du véhicule. Comme nous le montrerons par la suite, reformuler le problème dans le cadre du filtrage permettra de s'affranchir de pertes momentanées de certaines observations et de fournir des prédictions plus précises.

Il est important d'extraire correctement le plus en amont possible les limites de la route. Les contraintes épipolaires ne sont, en effet, pas suffisantes pour discriminer parmi le faisceau de droites-supports celles qui appartiennent au plan de la route des autres, tout du moins à l'aide des seuils que nous utilisons. Une ROI mal choisie sera à l'origine d'erreurs sur la nature des FPs détectés : trop large, les FPs auront une probabilité faible d'appartenir au plan de la route, trop réduite, des FPs coplanaires seront écartés alors que le nombre que nous parvenons à extraire est généralement limité.

5.1.1 Filtrage de Kalman

5.1.1.1 Rappel

Notre objectif dans cette partie est de rappeler les principaux résultats du filtrage de Kalman afin de justifier notre choix d'employer ce type de filtre sur diverses variables qui composent notre système.

Le filtrage de Kalman est un estimateur récursif optimal dans le cas des processus gaussiens centrés. Il a été développé au début des années 60 afin de déterminer la trajectoire d'un satellite (coordonnées x, y, z) au cours du temps à partir de l'observation des seuls angles de site et d'azimut (θ, ϕ) . Kalman a conceptualisé le filtrage dans le domaine temporel grâce à la notion de modèle d'état en exploitant une description interne des processus aléatoires dite "représentation gaussienne-markovienne". L'intérêt d'un tel filtre repose sur sa récursivité : l'estimée ne dépend que de sa valeur précédente et de la valeur courante de l'observation. Contrairement aux méthodes concurrentes où la plupart des estimations sont effectuées a posteriori en

considérant l'ensemble des mesures, l'estimation du filtre de Kalman est mise à jour à chaque nouvelle mesure.

Le filtrage de Kalman s'applique à des processus continus et discrets. Nous ne traiterons dans ce qui suit que du cas discret qui correspond à notre application. Considérons un processus à temps discret gouverné par une équation stochastique linéaire ayant pour entrée la variable \mathbf{U} , l'état $\mathbf{\Gamma}$ et la mesure \mathbf{Y} :

$$\begin{cases} \mathbf{\Gamma}_n = \mathbf{A}\mathbf{\Gamma}_{n-1} + \mathbf{B}\mathbf{U}_{n-1} + \mathbf{W}_{n-1} \\ \mathbf{Y}_n = \mathbf{C}\mathbf{\Gamma}_n + \mathbf{V}_n \end{cases} \quad (5.1)$$

Les variables aléatoires \mathbf{W} et \mathbf{V} représentent respectivement les bruits de modèles et les bruits de mesure du processus. Ces bruits sont considérés décorrélés, blancs avec des lois de distributions gaussiennes :

$$p(\mathbf{W}) \simeq \mathcal{N}(0, \Phi) \quad (5.2)$$

$$p(\mathbf{V}) \simeq \mathcal{N}(0, \Psi) \quad (5.3)$$

avec Φ et Ψ matrices de covariances qui caractérisent la distribution associée à chacun des signaux.

Soient $\mathbf{\Gamma}_{n \setminus n-1}$ et $\mathbf{\Gamma}_{n \setminus n}$ les estimations a priori et a posteriori de l'état à l'instant n . Le filtre de Kalman a été conçu de manière à minimiser la moyenne de l'erreur quadratique des signaux entre l'état $\mathbf{\Gamma}$ et chacune de ces deux estimées :

$$\mathbf{E}_{n \setminus n-1} \equiv \mathbf{\Gamma}_n - \mathbf{\Gamma}_{n \setminus n-1} \quad (5.4)$$

$$\mathbf{E}_{n \setminus n} \equiv \mathbf{\Gamma}_n - \mathbf{\Gamma}_{n \setminus n} \quad (5.5)$$

Les erreurs précédentes étant des variables aléatoires, les précisions associées sont définies comme les moments d'ordre 2 :

$$\mathbf{P}_{n \setminus n-1} = E[\mathbf{E}_{n \setminus n-1} \mathbf{E}_{n \setminus n-1}^t] \quad (5.6)$$

$$\mathbf{P}_{n \setminus n} = E[\mathbf{E}_{n \setminus n} \mathbf{E}_{n \setminus n}^t] \quad (5.7)$$

L'estimation a posteriori de l'état est calculée à partir d'une correction de la prédiction par l'innovation, différence entre la prédiction et la mesure :

$$\mathbf{\Gamma}_{n \setminus n} = \mathbf{\Gamma}_{n \setminus n-1} + \mathbf{K}(\mathbf{Y}_n - \mathbf{C}\mathbf{\Gamma}_{n \setminus n-1}) \quad (5.8)$$

où \mathbf{K} représente le gain de Kalman et \mathbf{Y}_n la mesure. Bien que le processus soit fondé sur une loi aléatoire, le modèle linéarisé imposé en Equ.5.1 parvient à suivre fidèlement la dynamique du processus. Les matrices \mathbf{A} , \mathbf{B} , \mathbf{C} du modèle sont considérées invariantes et c'est le gain \mathbf{K} qui compense les dérives entre le processus et le modèle : \mathbf{K}_n est donc mis à jour à chaque itération.

La mesure de l'innovation $(\mathbf{Y}_n - \mathbf{C}\mathbf{\Gamma}_{n \setminus n-1})$ différence entre la mesure courante et celle prédite par le filtre permet de vérifier la conformité du processus au modèle. Une innovation qui tend vers 0 signifie en effet que les prédictions et observations sont en accord.

La matrice \mathbf{K} est calculée de sorte à minimiser la covariance de l'erreur a posteriori $\mathbf{P}_{n \setminus n}$. Cela revient à rechercher la valeur de \mathbf{K} qui annule la dérivée de la trace de l'expression Equ. 5.6 dans laquelle $\mathbf{E}_{n \setminus n}$ est remplacée par $\mathbf{K}(\mathbf{Y}_n - \mathbf{C}\mathbf{\Gamma}_{n \setminus n-1})$. Une expression du gain courant est :

$$\mathbf{K}_n = \mathbf{P}_{n \setminus n-1} \mathbf{C}^t (\mathbf{C} \mathbf{P}_{n \setminus n-1} \mathbf{C}^t + \Psi)^{-1} \quad (5.9)$$

Les deux valeurs extrêmes vers lesquelles tend K sont obtenues en calculant son comportement à la limite quand les covariances de l'erreur de mesure Ψ et de l'estimée a priori $P_{n \setminus n-1}$ tendent vers 0 :

$$\lim_{\Psi_{n \rightarrow 0} K_n} = C^{-1} \quad (5.10)$$

$$\lim_{P_{n \setminus n-1} \rightarrow 0} K_n = 0 \quad (5.11)$$

D'après Welch et Bishop¹, l'influence du gain de Kalman est donc réduite quand l'estimée a priori correspond à l'estimée a posteriori et annule l'effet de l'innovation lorsque la mesure \mathbf{Y}_n est en accord avec le modèle. Vu différemment, tout se passe comme si la pondération du gain K quantifiait le niveau de pertinence du modèle en fonction des mesures réalisées. Lorsque la covariance Ψ tend vers 0, la confiance en la mesure est de plus en plus importante par rapport à la mesure prédite $C\Gamma_n$ et inversement.

Le filtre de Kalman a été développé de manière à maintenir les deux premiers moments de la distribution du vecteur d'état aléatoire :

$$E[\Gamma_n] = \Gamma_{n \setminus n} \quad (5.12)$$

$$E[(\Gamma_n - \Gamma_{n \setminus n})(\Gamma_n - \Gamma_{n \setminus n})^t] = P_{n \setminus n} \quad (5.13)$$

$$(5.14)$$

La covariance de l'estimée a posteriori de l'état est comparable à celle de l'état, son second moment est donc non-centré bien que son premier moment soit normalement distribué si les conditions 5.2 et 5.3 sont vérifiées, soit :

$$p(\Gamma_n \setminus \mathbf{Y}_n) \simeq \mathcal{N}(E[\Gamma_n], E[(\Gamma_n - \Gamma_{n \setminus n})(\Gamma_n - \Gamma_{n \setminus n})^t]) \quad (5.15)$$

$$= \mathcal{N}(\Gamma_{n \setminus n}, P_{n \setminus n}) \quad (5.16)$$

5.1.1.2 Implémentation

Le cas général L'implémentation d'un filtre de Kalman peut être vue comme un système contre-réactionné où la chaîne directe fournit une estimée a priori $\Gamma_{n \setminus n-1}$ tandis que la chaîne de retour exploite la mesure bruitée \mathbf{Y}_n pour vérifier a posteriori l'estimation précédente $\Gamma_{n-1 \setminus n-1}$. L'action du filtre se partage donc en deux opérations distinctes :

- la mise à jour des équations temporelles permet le calcul d'une prédiction $\Gamma_{n \setminus n-1}$ grâce à la projection de l'état $\Gamma_{n-1 \setminus n-1}$ et de sa précision $P_{n-1 \setminus n-1}$ associée.

$$\begin{cases} \Gamma_{n \setminus n-1} &= A\Gamma_{n-1 \setminus n-1} + B\mathbf{U}_{n-1} \\ P_{n \setminus n-1} &= AP_{n-1 \setminus n-1}A^t + \Phi \end{cases} \quad (5.17)$$

- la mise à jour de l'estimation a posteriori de l'état $\Gamma_{n \setminus n}$ et de sa précision $P_{n \setminus n}$ nécessitent une correction qui dépend du calcul de la valeur courante du gain induit K_n par l'écart observé entre l'estimée prédite de l'état et sa mesure.

$$\begin{cases} K_n &= P_{n \setminus n-1}C^t(CP_{n \setminus n-1}C^t + \Psi)^{-1} \\ \Gamma_{n \setminus n} &= \Gamma_{n \setminus n-1} + K_n(\mathbf{Y}_n - C\Gamma_{n \setminus n-1}) \\ P_{n \setminus n} &= (I - K_nC)P_{n \setminus n-1} \end{cases} \quad (5.18)$$

¹An introduction to the Kalman filter, <http://cs.unc.edu/~welch>

L'initialisation du filtre nécessite les matrices de covariances Φ et Ψ des signaux \mathbf{W} et \mathbf{V} qui représentent les bruits associés aux modélisations du processus et de la chaîne de mesure. Tous deux influencent la rapidité de convergence du filtre et sa disposition à fournir des résultats lorsque le processus s'éloigne du modèle. Le bruit de mesure peut être estimé de manière empirique à l'aide d'une étape de calibration de la chaîne de mesure. Par contre, la détermination de Φ demeure conceptuellement difficile car il est peu probable d'avoir accès à l'état du processus et de pouvoir mesurer l'erreur de modèle. Expérimentalement, il est alors courant d'imposer des valeurs relativement élevées de manière à s'affranchir de possibles erreurs de modélisation.

Ces deux matrices sont généralement considérées constantes ce qui accélère la convergence du filtre. Dans l'éventualité où le processus peut s'écarter grandement du modèle imposé, il est envisageable d'entreprendre un second filtrage de Kalman sur chacune de ces deux matrices soit hors ligne (quand c'est possible), soit en ligne de façon à suivre leurs évolutions tout en les considérant constantes pour le filtre initial.

Les modèles d'évolution La modélisation d'un système est par définition réductrice de la complexité qui le gouverne. Les écarts observés entre les sorties du système et son modèle sont généralement considérés comme un bruit de modèle qui se superpose au bruit de mesure, tous deux supposés additifs. La performance d'un filtre se mesure alors à sa capacité de suivre une évolution qui n'est pas en conformité avec son modèle. La difficulté de la modélisation consiste à déterminer un modèle dont l'évolution est la plus proche possible de la réalité en prenant soin de quantifier les différentes composantes de bruit.

Le filtrage de Kalman est un filtre linéaire, deux types d'évolutions sont généralement associés à son implémentation : un modèle à évolution stationnaire et celui à vitesse constante. L'évolution de ces deux modèles exprimée pour un système à temps discret admet une écriture commune :

$$\begin{cases} \mathbf{\Gamma}_{n+1} = \mathbf{A}\mathbf{\Gamma}_n + \mathbf{B}\mathbf{U}_n \\ \mathbf{Y}_{n+1} = \mathbf{C}\mathbf{\Gamma}_n + \mathbf{D}\mathbf{U}_n \end{cases} \tag{5.19}$$

où seule la matrice de transition \mathbf{A} se distingue. Elle est égale à l'identité de dimension celle du vecteur d'état dans le cas d'une évolution stationnaire. En revanche, dans le cas où le système modélisé a une vitesse constante, le vecteur d'état nécessite une adaptation en introduisant les mesures différences entre deux itérations de chaque état. L'Equ. 5.19 appliquée à un système à deux états x_1, x_2 s'écrit dans le cas d'un processus à vitesse constante :

$$\begin{vmatrix} x1_{n+1} \\ x2_{n+1} \\ x1_{n+1} - x1_n \\ x2_{n+1} - x2_n \end{vmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{vmatrix} x1_n \\ x2_n \\ x1_n - x1_{n-1} \\ x2_n - x2_{n-1} \end{vmatrix} + b \begin{vmatrix} u1 \\ u2 \\ u3 \\ u4 \end{vmatrix} \tag{5.20}$$

où Δt représente la période d'échantillonnage entre deux mesures.

5.1.2 Estimation du point de fuite dominant

L'intégralité des résultats présentés dans cette partie et la suivante ont été obtenus à partir de la séquence du port d'Antibes. Cette séquence a en effet le double intérêt de présenter de forts mouvements de rotation lors du contournement d'un véhicule en stationnement sur la voie empruntée et présente de nombreuses occlusions du plan de la route dues à la présence d'obstacles statiques et dynamiques.

5.1.2.1 Particularités de la primitive point de fuite dominant

La qualité de l'estimation du DVP dépend essentiellement de la qualité de sa prédiction. En effet, suite à l'extraction de contours, la présélection des droites-support conservées pour l'estimation du DVP s'effectue en calculant la distance en *pixels* qui les sépare de la prédiction du DVP. Seules les droites-supports dont la distance est inférieure au seuil $4.dist_max$ sont retenues. Nous supposons que parmi les droites-supports présélectionnées figurent deux types de contours : ceux qui représentent des arêtes alignées selon la direction de la route et d'autres, sans rapport mais dont la projection à l'image est orientée dans la direction du DVP. La qualité de la prédiction permet de discriminer parmi les droites-support présélectionnées celles qui convergent effectivement vers le DVP.

Ce second jeu de droites-support a pour origine des contours parallèles entre eux et orientés dans une direction proche de celle de la route : le cas le plus simple est l'observation d'un véhicule au premier plan légèrement décalé par rapport l'axe de la route. La détection de contours fournit alors de nombreux segments orientés vers un nouveau point de fuite sur la ligne d'horizon, proche du DVP. D'autres possibilités sont envisageables : il suffit que les hypothèses sur le modèle de l'environnement ne soient pas respectées. C'est le cas lorsque des façades non alignées avec la route sont observées ou en présence de zébras de la signalisation horizontale, etc. Ces artefacts induisent les erreurs de sélection lors de la recherche initiale de contours, qui est peu discriminante dans un premier temps du fait que la prédiction des coordonnées du DVP est peu précise.

Les coordonnées du DVP sont calculées de manière itérative en résolvant le système linéaire pondéré par la représentativité de chaque droite-support retenue. La représentativité des VLs peut être réduite pour cause d'occultation par un obstacle ou en cas de virage car le modèle linéaire de route n'est alors plus vérifié que par les contours détectés observés au premier plan. Dans de tels cas, l'estimation des coordonnées du DVP peut être grandement influencée par le second groupe de contours extraits surtout si la représentativité des contours orientés selon la direction de la route est moins importante que celle des contours parasites.

5.1.2.2 Filtrage temporel des coordonnées du point de fuite dominant

Analyse de l'évolution du mouvement relatif du point fuite dominant Les contraintes imposées par la géométrie épipolaire permettent de vérifier la cohérence des estimations des coordonnées du DVP pour peu que la prédiction de l'homographie stéréo \widehat{H}_{st}^n , disponible en début de chaque nouvelle itération n soit fiable. Cependant, lorsque les conditions d'observabilité de l'ensemble du système se détériorent comme c'est le cas sur la séquence du port d'Antibes entre les images [934;946], la projection des coordonnées des DVPs d'une image vers l'autre en utilisant l'estimation courante H_{st}^n peut conduire à des erreurs de plusieurs dizaines de *pixels*. Ces écarts dus à une estimation erronée de l'homographie stéréo s'avèrent bien plus importants que les erreurs mesurées entre les prédictions et les estimations erronées dues à une prise en compte intempestive de contours non-orientés dans la direction de la route.

Afin d'affiner la précision des prédictions, nous avons implémenté un filtre de Kalman, basé sur une hypothèse d'évolution stationnaire pour l'ordonnée et à vitesse constante pour l'abscisse du DVP. Le véhicule se déplaçant sur une route considérée plane, l'ordonnée de l'horizon ne peut évoluer qu'avec les mouvements d'inclinaison du véhicule. La stationnarité de l'ordonnée du DVP n'est mise à mal qu'en cas de mouvements d'inclinaison de la caméra. Les variations que nous avons mesurées dépassent alors dans de très rares cas la dizaine de *pixels*. A l'opposé, en cas de mouvement de lacet, les variations en abscisses dépassent couramment la dizaine de

pixels et même dépasser la vingtaine de *pixels*, ce qui justifie l'emploi d'un modèle d'évolution à vitesse constante. L'algorithme ayant pour vocation de fonctionner sans a priori sur la scène ou les conditions matérielles, les variances des bruits de modèles et mesures ont été fixées de manière empirique à 10 *pixels*.

Lors des déplacements en ligne droite du véhicule, nous pouvons extraire quelques informations sur la géométrie de la paire stéréo en analysant le chronogramme de la Fig. 5.1 représentant le suivi des coordonnées du DVP dans les deux images du couple stéréo. Dans cette séquence, le véhicule se déporte sur le coté gauche de la route afin de dépasser un véhicule en stationnement. L'hypothèse de rigidité du banc stéréo est vérifiée : l'écart entre les coordonnées du DVP dans les deux images peut être considéré comme stationnaire. Les horizons sont quasiment alignés ce qui signifie que l'angle d'élévation des caméras est comparable et les axes optiques sont légèrement pincés sans quoi les abscisses des DVPs seraient identiques lorsque le véhicule est aligné selon l'axe de la route.

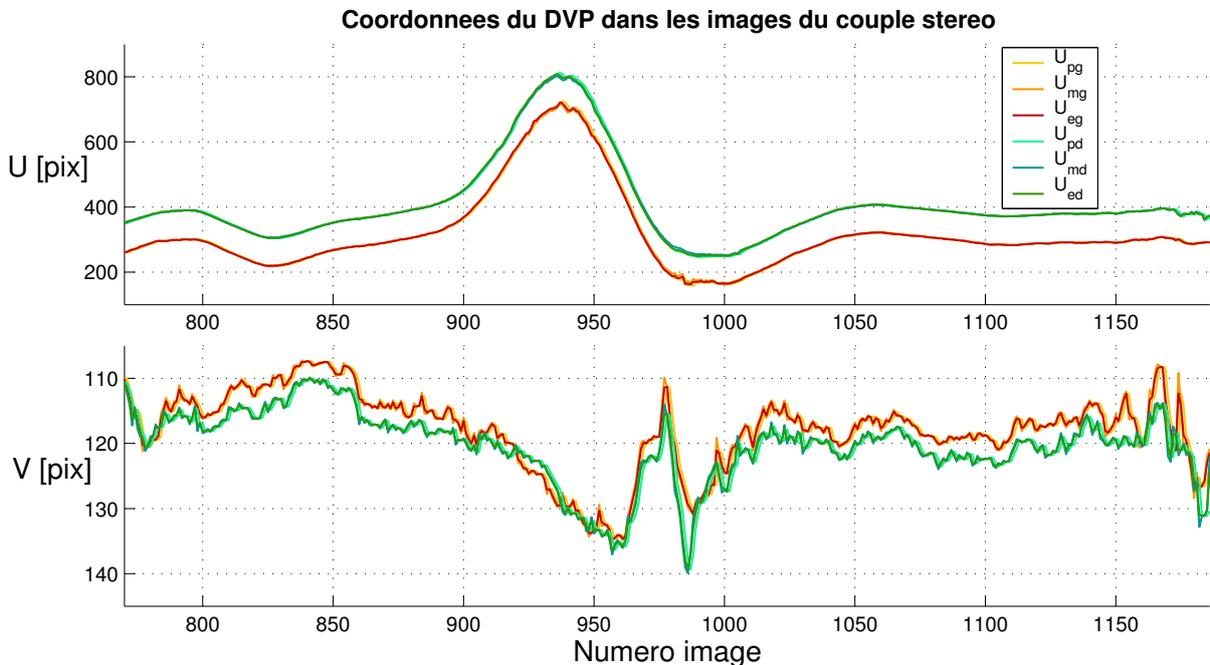


FIG. 5.1 – Chronogrammes des coordonnées prédites (p), mesurées (m) et estimées (e) du DVP extrait des images gauche(g) et droite (d), fournies par le filtre prédictif de Kalman. Les coordonnées avec une dominante de rouge caractérisent celles du DVP de l'image gauche, celles avec une dominante de vert, l'image de droite. Les variances de bruits de modèle et de mesure sont fixées à 10 *pixels*. La superposition des trois sources permet de vérifier que la mesure et l'estimation sont conformes à la prédiction fournie par le filtre malgré des déplacements des caméras peu en accord avec le modèle de vitesse constante imposé au système.

La régularité de l'évolution de l'abscisses du DVP de manière conjointe dans les deux images du couple stéréo est un gage de la robustesse de la méthode mise en oeuvre. La route suivie étant droite, les variations des abscisses sont induites par un mouvement de rotation de lacet du véhicule, sa trajectoire peut être retrouvée à partir de la seule lecture des abscisses :

- [770; 798] le véhicule roule de manière légèrement décentré par rapport à l'axe de la route,
- [800; 824] décalage rapide sur la droite,

- [826; 900] centrage lent sur la voie de droite,
- [902; 940] déboîtement en direction de la voie gauche,
- [842; 998] rabattement sur la voie de droite,
- [1000; 1050] ré-alignement dans l'axe de la route
- [1052; 1190] déplacement en ligne droite.

L'évolution conjointe des ordonnées des DVPs est tout aussi satisfaisante que celle observée en abscisse. Cependant, l'interprétation des variations mesurées est plus délicate. Nous savons que l'ordonnée du DVP nous renseigne sur la position de l'horizon dans l'image. La route étant plane et horizontale, les variations mesurées sont uniquement dues aux mouvements de tangage du véhicule, eux mêmes ayant pour cause les différentes accélérations et décélérations filtrées par les amortisseurs.

Cependant, s'il ne fait aucun doute que durant la deuxième partie de la séquence le véhicule roule à vitesse constante vue la stationnarité des ordonnées, il en est tout autrement pour la première. Nous devons tout d'abord préciser que l'évolution aussi brutale qu'originale des ordonnées entre les images [958 ; 1000] n'est en aucun cas due à une erreur de mesure. Le nombre de VLs détectées et leur distribution dans les images sont des plus contraignants comme le montre en détail la Fig. 5.2.

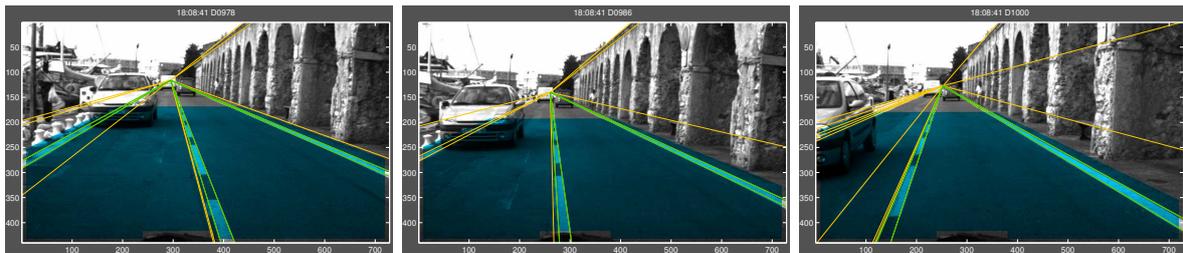


FIG. 5.2 – Droites-supports extraites lors des brusques mouvements d'inclinaison du véhicule entre les itérations [958 ; 1000]. Les droites-supports sont représentées en jaune, les bandes principale en vert. Les images présentées sont extraites de la caméra droite aux itérations 978, 986 et 1000, échantillons pour lesquels l'ordonnée du DVP présente de fortes variations. Les coordonnées extraites des DVPs sont contraintes par de nombreuses droites-supports, les fortes variations des DVPs en ordonnée qui ne sont donc pas dues à des erreurs de mesure.

Nous pouvons néanmoins tenté de fournir une explication aux brusques variations des ordonnées des DVPs. La croissance des ordonnées des DVPs jusqu'à l'itération 850 signifie que le véhicule ralentit au point d'être quasiment à l'arrêt pour laisser passer le véhicule rattrapant. A partir de cet instant, le véhicule-test redémarre lentement puis accélère jusqu'à ce qu'il parvienne à hauteur stationné qu'il est en train de contourner. Au-delà de l'itération 950, le brusque arrêt de l'accélération se fait ressentir jusqu'à l'itération 980 où une nouvelle accélération est observée durant une dizaine d'images pour rabattre rapidement le véhicule-test sur sa voie de circulation. Après quoi, il se déplace à vitesse constante.

Justification du choix des modèles d'évolution des coordonnées du DVP La modélisation du mouvement relatif des coordonnées du DVP qui fait appel à des modèles d'évolution différents pour les abscisses et les ordonnées est pleinement satisfaisante. Ce choix est justifié par les plages de variations de chacune des coordonnées qui sont respectivement de 500 pixels pour les abs-

cisses et seulement 40 *pixels* pour les ordonnées.

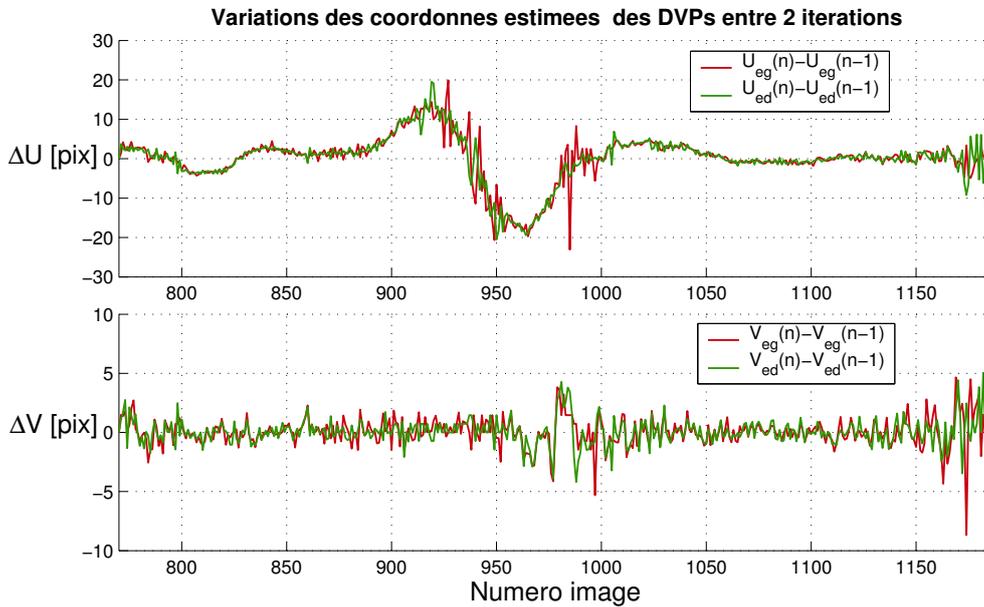


FIG. 5.3 – *Écarts entre les coordonnées mesurées et prédites des DVPs, fournies par le filtre de Kalman. Le comportement du filtre prédictif est conforme à nos attentes : la prédiction fournie est proche de la mesure précédente. Les variations observées sont donc proches du mouvement apparent du DVP entre deux images et peuvent atteindre quelques dizaines de pixels en abscisse et moins de 5 pixels en ordonnée, en faisant exception des erreurs de mesures.*

Les estimées fournies par le filtre en fonction de la mesure et de sa prédiction permettent un suivi effectif des coordonnées du DVP comme l'illustre la Fig. 5.3. L'écart important observé pour les coordonnées du DVP dans l'image \mathcal{I}_g^{980} est dû à une limitation de la variation en ordonnée qui est à l'origine d'une erreur importante en abscisse. Les variations mesurées en abscisse sont supérieures à $dist_max$ seulement durant le contournement du véhicule stationné (images [900;980]). En ordonnée, les écarts sont inférieurs à 5 *pixels* ($dist_max/2$) à de rares exceptions près qui sont dues, eux, à une perte d'observabilité du système entre la fin de la phase du déboîtement et en début du rabattement (images [936;952]).

La mise en oeuvre d'une recherche itérative des VLs qui convergent vers la prédiction du DVP dès que l'écart entre prédiction et mesure est supérieur à $dist_max$ (voir paragraphe 4.1.1.3) permet de compenser les erreurs de prédictions quand le système s'éloigne du modèle simplifié que nous avons imposé. Moins de 10% des images nécessitent une deuxième itération pour parvenir à présélectionner les nouvelles représentations des VLs qui avaient permis l'estimation des coordonnées du DVP à l'image précédente.

Constat de l'inobservabilité du système Nous nous intéressons désormais à l'intervalle [914;952] de la séquence que nous avons volontairement passé sous silence jusqu'alors. Il s'agit de la période durant laquelle le véhicule déboîte en direction de la seconde voie pour contourner le véhicule utilitaire. Le champ de vision de la route vue par les caméras est alors des plus réduits (cf. Fig. 5.4) : le véhicule étant orienté de manière oblique par rapport à l'axe de la route, la scène observée se décompose par une portion de la seconde voie de circulation, le quai

dans son prolongement et la poupe des bateaux amarrés. A partir de l'image 914, la bande principale droite n'est plus visible dans le couple d'images, à l'image 936, c'est au tour de la bande principale centrale de disparaître jusqu'à l'image 952.

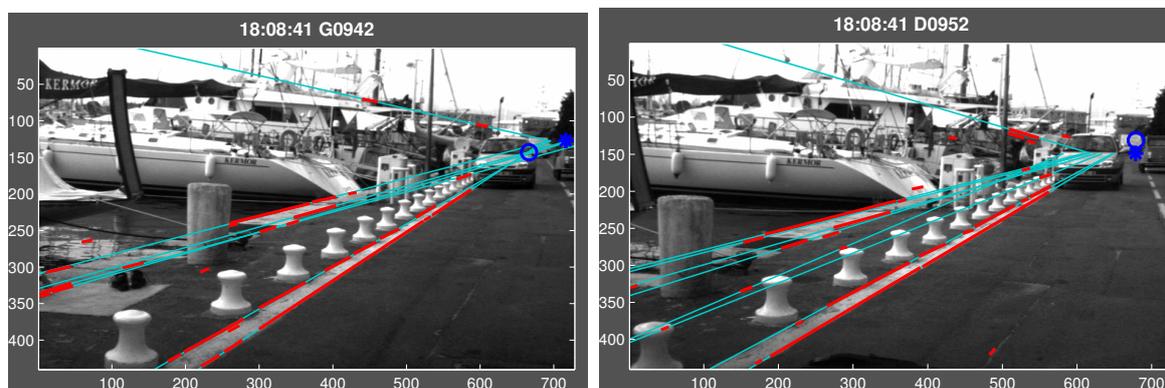


FIG. 5.4 – Cas où l'ordonnée de la prédiction du DVP est éloignée de la valeur réelle. Les droites-supports cyans sont formées à partir des segments détectés rouges qui sont orientés vers la prédiction '*' bleue des coordonnées du DVP. L'estimation directe 'o' bleu des coordonnées du DVP est obtenue à partir des droites-supports situées sous l'horizon. Le point milieu de ces deux points fournit la 3^{ème} ordonnée avec lesquelles est entrepris la mesure d'abscisse minimale entre la première et la dernière droite-support qui forment le faisceau.

Les quelques contours détectés sur le quai permettent d'extraire quelques VLs supplémentaires dont la représentativité est faible : elle ne dépasse pas 200 *pixels* ce qui induit des droites-supports dont les orientations sont mal définies et par conséquent peu fiables. De plus les VLs extraites durant ce passage sont détectées sur une région de la scène qui ne dépasse pas 2m de large, les orientations des VLs ont de surcroît toutes le même signe. La mauvaise qualité, voire l'absence des mesures impose de propager la prédiction du filtre de Kalman à la place de l'estimée. Hélas, c'est durant cette même période que l'amplitude des mouvements relatifs du DVP sont parmi les plus importants de la séquence, or le modèle imposé ne correspond pas à des mouvements aussi importants.

Le filtre de Kalman parvient à jouer correctement son rôle prédictif tant que la bande principale centrale est visible, la distribution des VLs étant encore suffisamment contraignante. Au-delà, l'évolution des coordonnées du DVP dans l'image droite n'est alors plus observable. L'extraction des coordonnées du DVP à partir de VLs, pourtant correctement présélectionnées, fournit des estimées dont l'ordonnée est généralement distante de plus 10 *pixels* avec la prédiction. Nous avons en effet pu mesurer des écarts en ordonnée entre la prédiction et une estimation "à la main" de plus de 17 *pixels* dans certains cas, ce qui induit des écarts encore plus importants en abscisse. Or, nous avons vu au paragraphe précédent, que même en cas de brusques accélérations, l'évolution de l'ordonnée ne dépasse pas 10 *pixels* entre deux images. Le modèle stationnaire d'évolution mis en oeuvre sur le filtrage de l'ordonnée du DVP permet alors la propagation d'une prédiction plus cohérente que la mesure.

5.1.3 Identification et suivi des bandes principales

L'opération de suivi du marquage au sol tient un rôle important dans l'algorithme. Elle permet, d'une part, de segmenter rapidement la région de l'image apparentée à la route et, fournit, d'autre part, des limites précises de la projection du plan de la route combien même celles-ci seraient occultées par des obstacles.

5.1.3.1 Identification des bandes principales

Modélisation des bandes principales L'estimation des coordonnées du DVP réalisée, notre objectif est de segmenter dans l'image la projection du plan de la route qui représente la région d'intérêt. Nous supposons pour cela qu'elle est délimitée sur les cotés par les bandes extrêmes du marquage de la signalisation horizontale. L'intérêt de l'emploi de ces bandes est double : elles sont communes à toutes les scènes urbaines et sont aisément détectables pour peu que les premiers mètres devant le véhicule soient dégagés. Notre objectif est donc d'identifier parmi les droites-supports extraites celles qui délimitent des bandes et nous focaliser sur celles qui constituent les bords de la route.

Les bandes extrêmes de la signalisation horizontale qui délimitent les frontières de la route sont généralement voisines d'accotements dont les variations de matériau (bitume/béton), voire de surfaces par des délimitations marquées (route/caniveau ou trottoir ou terre-plein) sont à l'origine de bandes de droites-supports dont les caractéristiques sont proches de celles des bandes principales. Afin de discriminer parmi ces bandes celles qui représentent le marquage au sol, nous proposons d'affiner la prédiction de la position des bandes principales grâce au suivi de leurs caractéristiques. Les résultats du filtrage de Kalman appliqués au suivi des coordonnées des DVPs étant satisfaisant, nous proposons d'employer le même type de filtre pour réaliser le suivi des paramètres des bandes principales.

Les bandes principales, projections dans l'image du marquage de la signalisation horizontale, sont modélisées grâce à une approximation du 1^{er} ordre. Nous ne considérons que la partie linéaire de leur projection, c'est à dire celle observée au premier plan. Ainsi, les caractéristiques des bandes principales modélisées sont celles des droites-supports théoriques calculées dès l'estimation des coordonnées du DVP obtenue :

$$\cos(\theta_k).u + \sin(\theta_k).v = \rho_k \quad (5.21)$$

où $(u; v)$ représentent les coordonnées d'un point du plan image et θ_k et ρ_k , respectivement l'orientation et le rayon de la k^{eme} droite-support exprimés dans le repère cylindrique.

Le faisceau de droites-supports convergeant en direction du DVP peut être paramétré de plusieurs manières (cf 2.4). Le DVP faisant déjà l'objet d'un filtrage sur ses coordonnées, l'orientation θ caractérise pleinement la projection de chaque VL du faisceau de droites qui convergent en sa direction. La comparaison des caractéristiques de gradient, d'écart angulaire et, dans une moindre mesure, de représentativité permet de limiter le nombre de VLs d'un image susceptibles d'être en correspondance avec une autre.

Considérer l'ensemble des paramètres relatifs à chaque VL qui délimite une bande s'avère inutile car l'information serait redondante. Les VLs sont ordonnées selon l'orientation décroissante : une bande est, entre autres, caractérisée par deux VLs successives dont les gradients seront '1' puis '0' (délimitations de régions foncée/claire/foncée). Les couples de VLs qui représentent des projections de bandes délimitent des secteurs angulaires que nous pouvons caractériser grâce aux paramètres suivants :

- les orientations moyennes qui représentent les orientations des médianes des secteurs angulaires,
- les écarts angulaires,
- les représentativités en tant que somme des représentativités des couples de VLs.

L'hypothèse du paramètre orientation Lorsque les caméras ne sont pas calibrées, la caractérisation des bandes principales à partir de leur orientation ne peut être effectuée. En effet, l'amplitude de la variation de l'orientation des VLs suite à un mouvement de rotation de la caméra dépend essentiellement de leur orientation initiale. Ainsi, les simulations de la Fig. 5.5 montrent l'influence d'un mouvement de rotation selon l'angle d'inclinaison et de lacet pour un faisceau de trois bandes principales. Il apparaît clairement que le mouvement d'inclinaison vers le bas de la caméra (le DVP a un mouvement ascendant) a tendance à dilater le faisceau vers les extérieurs de l'image tandis que le mouvement latéral du DVP indique le signe de la variation des orientations des bandes principales. Les amplitudes des variations des orientations dépendent essentiellement de la bande considérée : elles sont négligeables pour la bande gauche et plus importante pour les deux autres dont les orientations en valeurs absolues sont plus faibles.

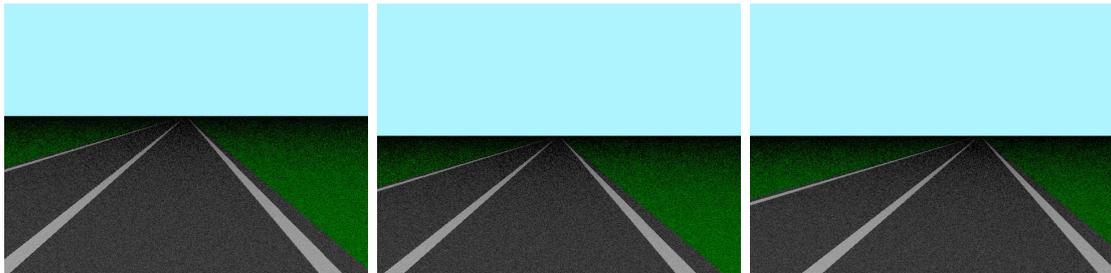


FIG. 5.5 – Evolution du faisceau de bandes principales en fonction des mouvements de rotation de la caméra. L'axe optique de l'image au centre est parallèle à la direction de la route et aligné sur l'horizon. A gauche, un mouvement d'inclinaison de la caméra de 5° vers le bas écarte chacune des bandes vers les extérieurs de l'image. A droite, un mouvement de lacet de 10° vers la gauche réduit l'orientation signée des bandes principales : elles ont un mouvement relatif vers la droite.

La détermination de l'amplitude des variations en orientation de chaque VL qui forme le faisceau nécessite donc de connaître la calibration de la caméra. Les mouvements de rotation de la caméra entre deux poses, estimés par la mesure des variations des coordonnées du DVP, permettraient alors de calculer une prédiction de l'orientation de chaque VL. Etant donné que nous ne disposons pas de cette information indispensable, nous avons cherché une représentation du faisceau de VLs plus conforme à notre modèle de route.

Le paramètre discriminant des bandes principales : l'abscisse Nous considérons depuis le début que les bandes principales peuvent être modélisées par une approximation linéaire au premier plan de l'image, même en cas de virage. Cela signifie que les intersections des bandes principales avec la limite basse de l'image représentent non seulement l'ordonnée où l'approximation linéaire est la plus justifiée mais aussi que les écarts entre les intersections des bandes principales doivent quasiment être constants en supposant l'effet des mouvements d'inclinaison de la caméra négligeable.

Nous avons donc mis en oeuvre un filtre de Kalman prédictif dont l'état est formé des abscisses des intersections des médianes des bandes principales avec le bas de l'image : $\mathbf{\Gamma abs}^t = [u_{1g}, u_{2g}, u_{3g}, u_{1d}, u_{2d}, u_{3d}]$ dans le cas d'une route à deux voies de circulation, séparées par trois bandes blanches. Le modèle de variation des abscisses retenu est à vitesse constante. Nous distinguerons dans ce qui suit chacune d'elles grâce aux couleurs rouge (gauche), bleu (centrale), vert (droite). Les caractéristiques de l'image gauche du couple stéréo sont représentées avec les couleurs foncées tandis que celles de l'image droite avec les claires, conformément à la Fig. 5.6.

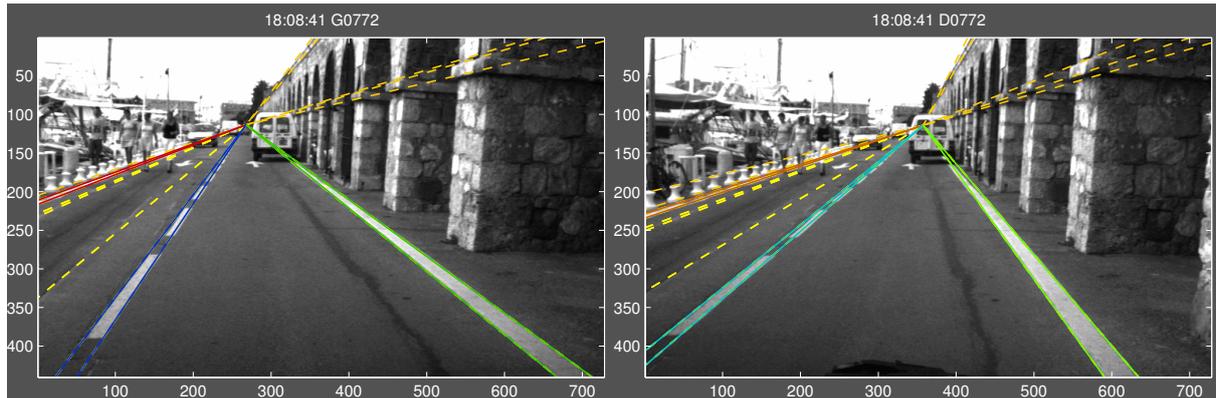


FIG. 5.6 – Code de couleurs employé pour identifier les caractéristiques de trois bandes principales (gauche, centrale et droite) dans chacune des images du couple stéréo. Les caractéristiques orientations et abscisses correspondent aux médianes des VLs mises en évidence. Les droites-supports oranges et jaunes pointillées, fournies à titre indicatif, ont respectivement permis l'estimation du DVP et sont en correspondance via l'homographie stéréo.

n°	Bp_{1g}	Bs_{1g}	Bp_{2g}	Bp_{3g}	Bp_{1d}	Bs_{1d}	Bp_{2d}	Bp_{3d}
θ [°]	69,5	66,32	35,3	-52,2	71,9	69,4	49,9	-37,9
$d\theta$ [°]	0,98	0,64	3,23	2,96	0,82	0,55	1,77	4,82
u [pix]	-605,5	-480	36,7	689,9	-637,4	-509,8	-28,9	612,5
du [pix]	45,8	23,1	27,6	45	48,1	25,5	24,3	44,2

TAB. 5.1 – Orientations et abscisses des bandes coplanaires en correspondance dans le couple d'images stéréo. Nous distinguons les trois bandes principales (Bp) qui représentent les projections des bandes du marquage au sol de la supplémentaire (Bs_1) qui est détectée à partir d'une fissure du revêtement. L'abscisse de chaque bande est le paramètre discriminant de la modélisation des bandes. Les bandes Bp_1 et Bs_1 pourtant relativement proches selon leur orientation présentent des écarts en abscisses de plus de 125 pixels. La même comparaison effectuée cette fois-ci entre l'angle d'ouverture de chaque bande et leur largeur mesurée au bas de l'image confirme que les mesures relatives à l'orientation dépendent de la bande considérée contrairement aux mesures relatives aux abscisses.

Les caractéristiques des bandes principales de la Fig. 5.6 sont regroupées dans le Tab. 5.1. La caractérisation des bandes principales à partir de l'abscisse de leur médiane se justifie en comparant les ordres de grandeur des orientations θ et des abscisses u correspondantes. Il est en effet légitime de se demander si nous sommes capable de caractériser précisément des VLs dont l'écart angulaire est inférieur au degré. Nous avons en effet insisté dans la partie 4.1.1.3

sur la difficulté d'extraire des contours fiables et avons précisé que l'écart entre les coordonnées extraites et théoriques des VLs retenues pouvait varier dans une tolérance de 5° .

En fait, la détection des contours des bandes les plus étroites ne pose aucune difficulté, contrairement à leur identification. En effet, lorsque plusieurs bandes sont détectées dans un même voisinage, la caractérisation des VLs par leur orientation ou leur écart angulaire s'avère difficile : l'écart angulaire qui sépare deux bandes est souvent inférieur à la précision des caractéristiques extraites. Ainsi, l'écart angulaire qui sépare les orientations des médianes des bandes principale gauche (B_{p1}) et supplémentaire (B_{s1}) est inférieur à 3° tandis que leurs abscisses sont séparées de plus de 125 *pixels*.

Parallèlement, la caractérisation de l'ouverture des secteurs angulaires des bandes principales a tout intérêt à être représentée par la largeur des bandes, mesurée en *pixels* toujours au bas de l'image. Ces largeurs correspondent à la différence d'abscisses entre les VLs qui délimitent les bandes. Les largeurs *du* des bandes principales peuvent être assimilées à une constante pour les deux images, contrairement à leur écart angulaire $d\theta$. La bande principale centrale B_{p2} se distingue des extérieures car sa largeur est plus réduite (15 cm au lieu de 25 cm pour les deux autres).

5.1.3.2 Mise en correspondance des mesures et des prédictions des bandes principales

Au début de chaque nouvelle itération n , l'estimation des coordonnées du DVP permet d'extraire un faisceau de droites-supports convergeant vers ce dernier pour chaque image du couple stéréo. L'objectif de cette étape est de rechercher parmi les couples de VLs qui forment des bandes, les nouvelles projections des bandes principales. Nous disposons pour cela d'une prédiction de l'abscisse u de la médiane de chaque bande principale $\Gamma_{abs_{n-1}\setminus n}$, fournie par le filtre de Kalman, ainsi que les caractéristiques de largeur et de représentativité de chacune d'elles, mesurées à l'image précédente. En supposant que les caractéristiques de la projection du plan de la route évoluent peu entre deux images successives, l'identification des nouvelles représentations des bandes principales s'effectue en recherchant parmi le faisceau de droites-supports extrait les bandes dont les caractéristiques sont les plus proches de celles enregistrées à l'image précédente ($n - 1$).

L'identification de la projection d'une bande principale dans une nouvelle image peut échouer dans deux cas. Le premier cas, la prédiction fournie par le filtre est relativement éloignée de la position réelle de la bande dû à un mouvement de la caméra plus complexe que celui imposé par le modèle de notre filtre. Il est alors possible d'affiner la prédiction de l'abscisse de la bande non-identifiée dès qu'une des bandes principales a été mise en correspondance. Nous supposons pour cela que les écarts qui séparent les abscisses des bandes principales sont stationnaires. En posant n_{id} et $\overline{n_{id}}$ les numéros des médianes identifiées et celles qui ne le sont pas encore, les prédictions affinées des abscisses u des médianes non-identifiées sont calculées de la sorte :

$$\Delta u = \begin{bmatrix} \Gamma_{abs_{n-k}\setminus n-k}(n_{id}) - \Gamma_{abs_{n-k}\setminus n-k}(\overline{n_{id}}) \\ \vdots \\ \Gamma_{abs_{n-1}\setminus n-1}(n_{id}) - \Gamma_{abs_{n-1}\setminus n-1}(\overline{n_{id}}) \end{bmatrix} \quad (5.22)$$

$$\Gamma_{abs_{n\setminus n-1}}(\overline{n_{id}}) = med(\Gamma_{abs_{n\setminus n}}(n_{id}) - \Delta u) \quad (5.23)$$

où le calcul de la médiane des écarts s'effectue par exemple sur les $k = 5$ dernières images.

Dans le second cas, l'estimation de contours fournissant des résultats de qualité variable, lorsqu'une bande a une faible représentativité (due à une occlusion ou à sa sortie du champ de

vision), il est courant de ne détecter qu'une fraction voire aucun des contours qui délimitent une bande. L'algorithme recherche alors parmi les VLs "isolées", une VL dont les caractéristiques correspondent à la prédiction fournie par le filtre. Dans de tels cas, il est courant de mettre en correspondance une seule des deux VLs recherchée. Nous supposons alors que la seconde VL qui compléterait la bande n'a pas été détectée. Les caractéristiques de la bande mesurée correspondent alors aux prédites, à l'exception de l'abscisse qui est, elle, calculée à partir de la VL mise en correspondance.

La possibilité d'observer plusieurs bandes principales permet de calculer des prédictions affinées lorsque certaines ne sont pas identifiées ou détectées. Néanmoins, il est possible de prédire les abscisses de toutes les bandes à partir d'une seule correctement identifiée : les prédictions sont alors moins fiables et sujettes à la qualité de la détection de la bande suivie. Les prédictions fournies par le filtre demeureront cohérentes tant que l'estimation des coordonnées du DVP sera correcte. En revanche, si aucune bande principale n'est visible dans les images, le filtre est incapable de fournir des prédictions cohérentes.

5.1.3.3 Récapitulatif de l'étape de suivi des bandes principales

L'intégralité des opérations successives nécessaires au suivi des bandes principales en début de chaque nouvelle image (n) peut se résumer de la manière suivante :

1. calcul de la prédiction des abscisses des médianes $\Gamma_{\text{abs}_{n \setminus n+1}}$ des bandes principales des images courantes,
2. extraction des faisceaux de droites-supports qui convergent vers les DVPs,
3. sélection des doublets de droites-supports qui forment des bandes,
4. **pour** l'image gauche puis l'image droite courante,
 - (a) calcul de l'orientation prédite des médianes des bandes principales à partir de $\Gamma_{\text{abs}_{n \setminus n+1}}$ et de l'estimée des coordonnées du DVP Γ_{dvp}^n ,
 - (b) recherche parmi les bandes extraites les plus proches des caractéristiques prédites des bandes principales,
 - (c) affinage de la prédiction de l'abscisse, si l'identification d'au moins une bande fait défaut,
5. mise en correspondance des deux faisceaux de droites-supports entre les images du couple stéréo en tenant compte des bandes principales identifiées et de la prédiction de l'homographie stéréo H_{st}^n ,
6. recherche parmi les droites-supports en correspondance de VLs dont l'identification en tant que bande principale n'a pas été possible car le second contour ne vérifiait pas les critères de sélection,
7. enregistrement des caractéristiques abscisse, largeur et représentativité de chaque bande principale pour calculer l'estimée courante de l'abscisse $\Gamma_{\text{abs}_{n \setminus n}}$.

5.1.3.4 Résultats du suivi des bandes principales sur la séquence du port d'Antibes

Le trajet du véhicule test sur la séquence du port d'Antibes consiste à un déplacement en ligne droite avec le contournement d'un véhicule en stationnement au premier quart de la séquence. De nombreux obstacles statiques et dynamiques sont sources d'occlusions partielles et totales de certaines bandes principales au cours du temps. Le suivi des bandes principales sur cette séquence peut être scindée en 6 étapes :

- [770, 890] déplacement en ligne droite puis décélération jusqu'à l'arrêt pour se laisser doubler par le véhicule en mouvement,
- [892, 916] dépassement du véhicule en mouvement qui cache la bande principale gauche,
- [918, 946] déboîtement du véhicule de test qui contourne le premier véhicule en stationnement, le champ de vision va se réduire au point que la voie de navigation utilisée jusqu'alors devienne hors-champ (disparition des bandes principales gauches et centrales),
- [948, 986] rabattement du véhicule test sur la voie de circulation gauche (réapparition progressive des bandes principales hors-champ),
- [988, 1018] passage au niveau du second véhicule en stationnement sur la voie de droite (nouvelle occultation de la bande principale gauche),
- [1020, 1190] reprise de la trajectoire en ligne droite, à vitesse quasi-constante.

Les chronogrammes de la Fig. 5.7 représentent les évolutions des paramètres des bandes principales lorsqu'elles ont pu être mesurées. Lorsque une bande n'est pas identifiée, ses caractéristiques précédentes sont propagées à l'exception de l'abscisse de la médiane (et par déduction l'orientation), qui est, elle, filtrée. Les caractéristiques présentées sont de haut en bas pour chaque bande principale :

1. l'abscisse correspondant à l'intersection de la médiane avec le bas de l'image,
2. la largeur de la bande, mesurée au bas de l'image,
3. l'orientation,
4. l'écart angulaire,
5. la représentativité.

Il apparaît clairement que l'écart entre les abscisses des médianes des bandes principales est régi par une loi stationnaire quelle que soit la position et l'orientation de la caméra par rapport à la direction de la route. De manière moins évidente, la largeur des bandes mesurée au bas de l'image dépend seulement de l'orientation de la caméra par rapport à la direction de la route (Fig. 5.8). Les variations mesurées dans notre cas sont minimes mais elles permettent de justifier les variations observées pour l'ouverture de l'écart angulaire formé par la projection des bandes dans les images.

L'étude de la représentativité de chaque bande permet de retrouver les différentes étapes de la séquence grâce aux occlusions partielles ou totales que subisse chacune d'elles. Les images pour lesquelles des bandes ne sont pas identifiées correspondent à celles où les largeurs des bandes sont propagées. A noter, la périodicité de la représentativité de la bande principale centrale permet d'évaluer la vitesse du véhicule en supposant la longueur des bandes peintes constante. Si le franchissement d'une bande requiert une vingtaine d'images avant l'image 822 ou après la 1010, le masquage de la bande par le véhicule utilitaire ne permet plus de détecter le franchissement de la bande peinte entre les images [824;912].

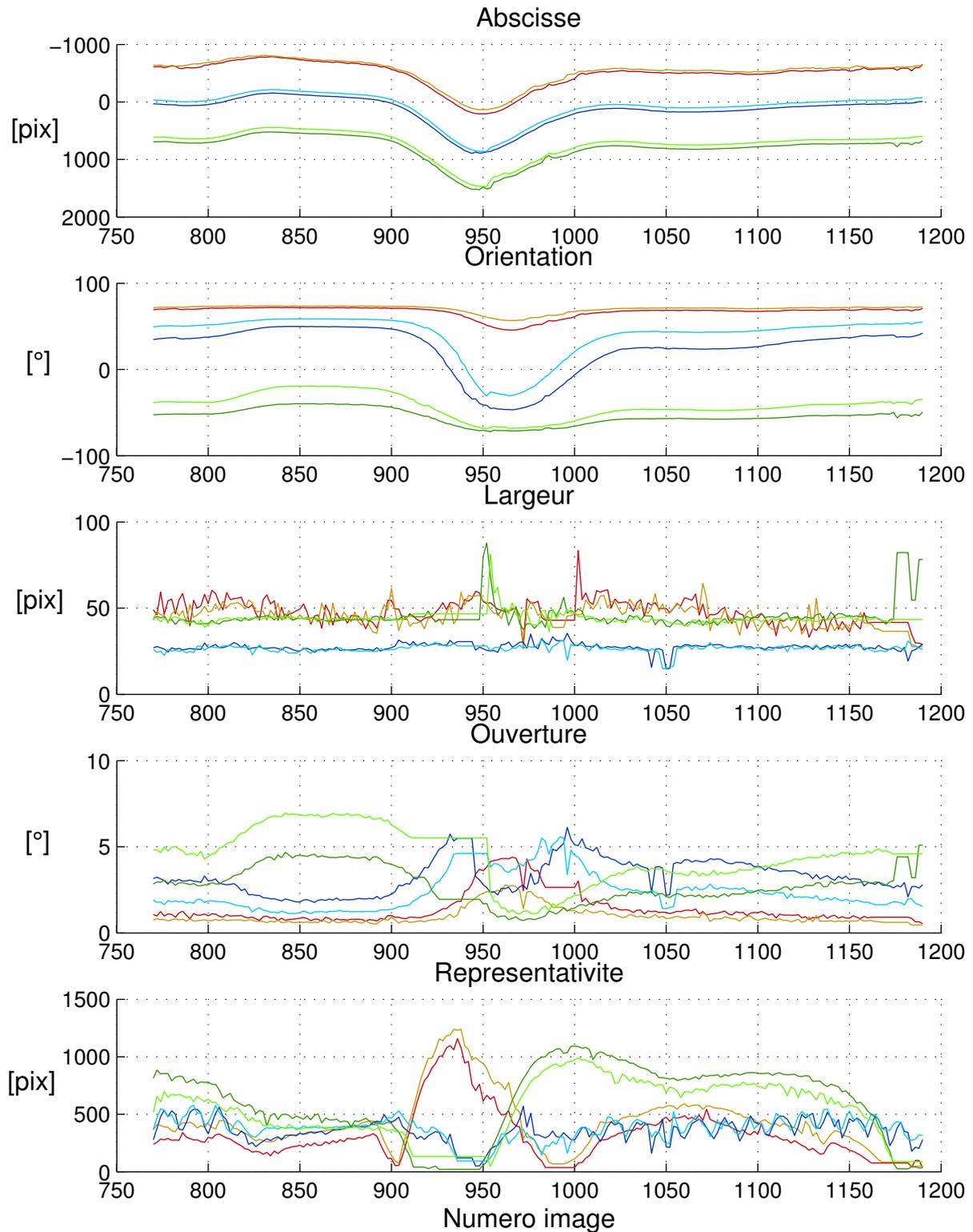


FIG. 5.7 – Evolution des caractéristiques des bandes principales au cours de la séquence du port d'Antibes. Parmi les caractéristiques représentées (abscisse, orientation, largeur, ouverture et représentativité), seule l'abscisse des médiane des bandes principales est filtrée par un filtre de Kalman, fondé sur une hypothèse d'évolution à vitesse constante. Lorsqu'une bande n'est pas identifiée, la prédiction du filtre est propagée pour l'abscisse et par calcul pour l'orientation tandis que les caractéristiques de largeur, d'ouverture et de représentativité sont supposées invariantes.

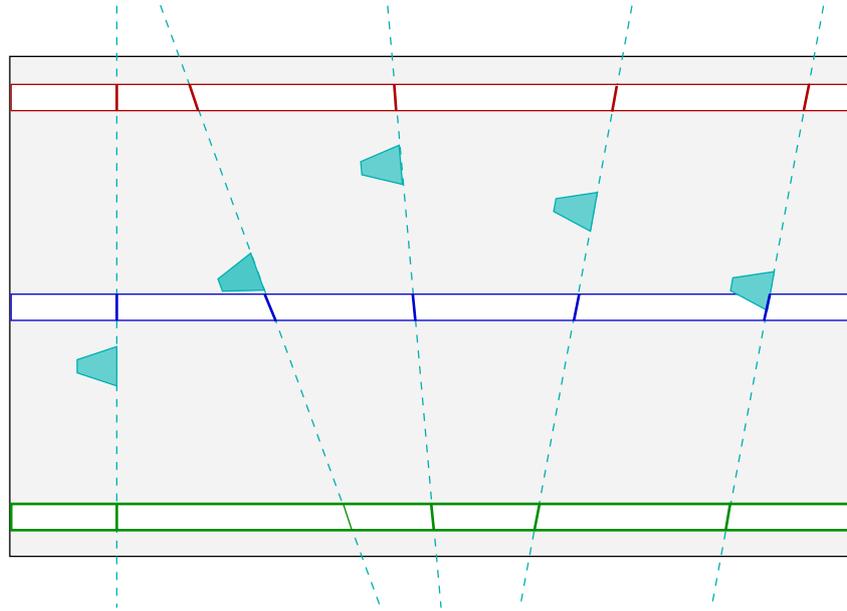


FIG. 5.8 – Evolution de la largeur des bandes mesurée en fonction de l'azimut de la caméra par rapport à celui de la route. Le trait pointillé cyan représente le plan image de la caméra qui intercepte les bandes principales.

Détermination des critères d'identification Le choix des critères d'identification des bandes selon les 5 paramètres précédents doit tenir compte de deux points :

- la dépendance de la variable en fonction de l'orientation et de la position latérale de la caméra par rapport à la route,
- l'incertitude des mesures.

Afin d'éclairer notre choix, nous proposons en Fig. 5.9, l'évolution de ces cinq paramètres sur la séquence descente, enregistrée à Versailles dont le détail est fourni en A.2. Le paramètre abscisse se détache très nettement des quatre autres par la continuité de son évolution au cours de la séquence et l'invariance des écarts qui séparent les abscisses de deux bandes. Ce résultat encourageant nécessite néanmoins d'identifier avec précision quelles sont parmi les VLs détectées celles qui représentent les limites de chaque bande principale.

De manière générale, la largeur des bandes évolue peu même en cas de forts mouvements de rotation du véhicule (images ([900;950]) de la séquence antiboise). Cependant, les bandes les plus éloignées de la pose de la caméra, comme la bande principale gauche de la séquence versaillaise, peuvent présenter des disparités telles qu'il ne paraît pas possible de pouvoir utiliser ce paramètre. Néanmoins, les variations observées sont essentiellement dues à une extraction difficile des contours qui forment ces bandes : la représentativité et l'écart angulaire sont réduits et les abscisses sont localisées à plusieurs centaines de *pixels* en dehors du cadre de l'image.

Nous proposons donc de réaliser l'identification des VLs qui forment les bandes principales en distinguant celles dont les abscisses se situent dans les limites de la largeur de l'image avec une tolérance de 50%. Ainsi les VLs qui forment une bande et vérifient les caractéristiques suivantes seront considérées comme les nouvelles projections des bandes principales.

- 100 *pixels* de variations entre l'abscisse prédite par le filtre de Kalman et l'abscisse mesurée,
- 50% de variation entre les rapports des largeurs des bandes,

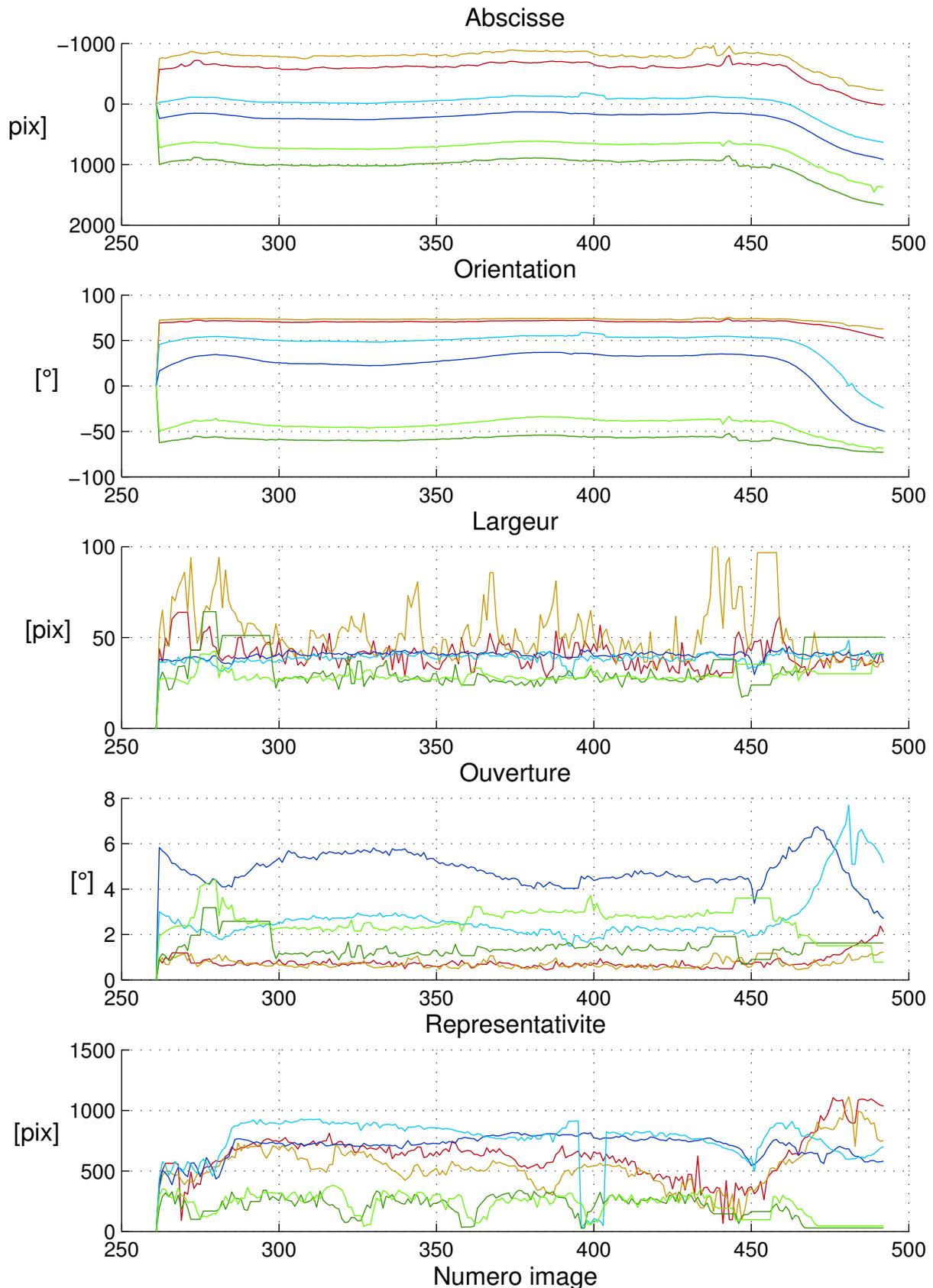


FIG. 5.9 – Relevé des caractéristiques des bandes principales pour la séquence *Descente*. Le véhicule effectue un déplacement en ligne droite durant les 4/5 de la séquence, puis se déporte sur la voie de gauche. Nous retrouvons les propriétés décrites précédemment : constance entre les abscisses des médianes, évolution des représentations des bandes en fonction de l'angle de lacet que fait la caméra avec la route. Cependant, la disparité de la largeur de la bande gauche observée à partir de la caméra droite est surprenante.

Pour les bandes principales dont l'abscisse est éloignée des cadre de l'image, le critère selon la largeur est remplacé par un identique selon l'orientation.

Suivi des bandes principales sur la séquence Le chronogramme de la Fig. 5.10 indique le niveau d'identification de chacune des trois bandes principales suivies au cours du temps dans les deux images du couple stéréo sur la séquence du port d'Antibes. Un point vert indique que le couple de VLs qui forment la bande principale a été correctement identifié, un point orange qu'une seule des deux VLs est détectée et un point rouge qu'aucune VL supposée représenter le contour de la bande n'a été identifiée.

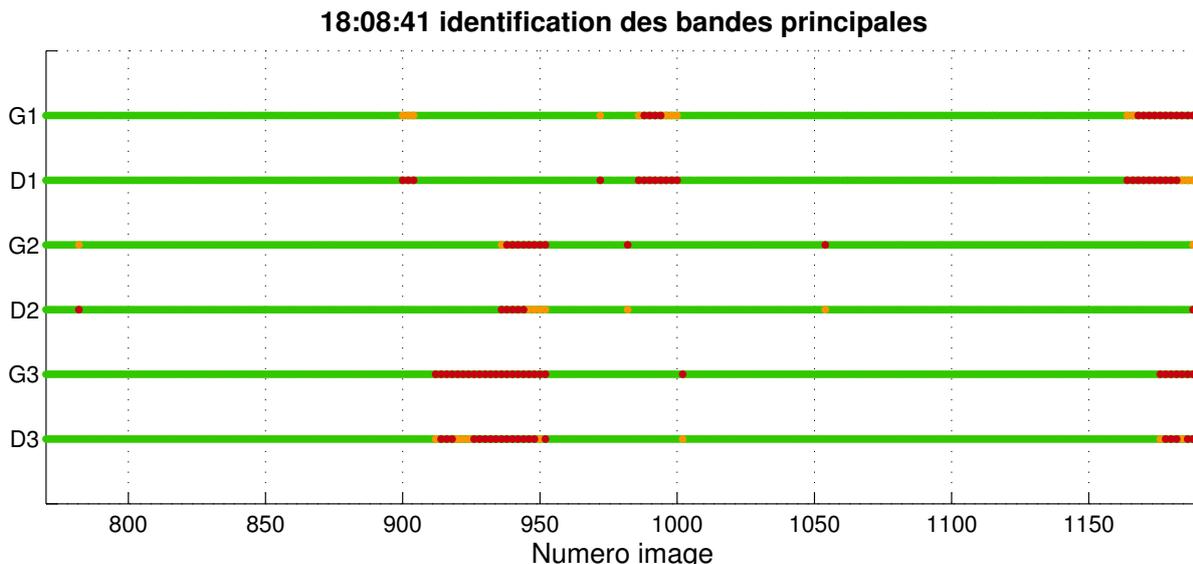


FIG. 5.10 – *Dénombrement des cas où l'identification des projections des trois bandes principales a posé problèmes. En orange, une seule VL qui forme le contour d'une bande a été détectée, en rouge aucune.*

La robustesse de la modélisation mise en oeuvre est illustrée en Fig. 5.12. Nous montrons que la simplicité de la modélisation permet néanmoins de propager correctement les caractéristiques des bandes principales malgré un mouvement de rotation important selon l'angle de lacet. Les projections des bandes principales centrales et droites sont immédiatement identifiées bien que leurs caractéristiques aient été propagées durant plus d'une dizaine d'itérations et que leurs caractéristiques ne correspondent plus aux dernières mesurées.

Quelques aléas demeurent néanmoins sur chacune des caractéristiques des bandes principales. Les erreurs d'identification sont généralement dues à un non-respect du modèle que nous avons imposé. La détection de plusieurs VLs avec des caractéristiques très proches peut en effet perturber le processus d'identification, tels les exemples de la Fig. 5.11.

La majorité des discontinuités des caractéristiques mesurées des bandes principales pourraient être éliminées en employant des seuils d'identification moins lâches en substituant aux mesures les prédictions fournies par le filtre. L'inconvénient de cette apparente continuité sur les caractéristiques suivies est de s'opposer à une identification rapide des bandes principales centrales et droites de l'exemple précédent. En restreignant à 75 pixels et 30% les critères utilisés, l'identification de la bande principale droite ne s'effectue alors plus qu'à l'itération 976 au lieu de 956 actuellement. Les discontinuités observées n'ayant aucune incidence fâcheuse sur

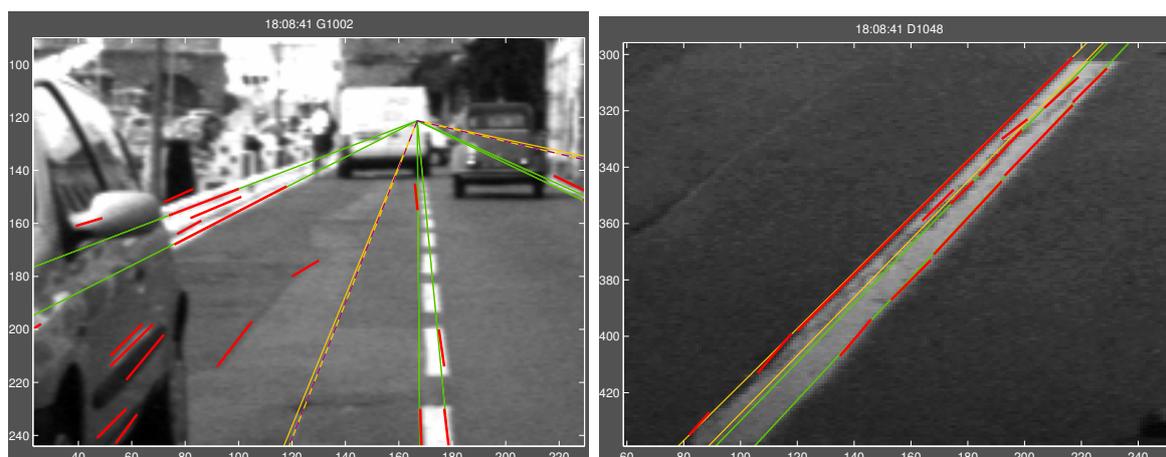


FIG. 5.11 – *Détail de détection de contours qui posent problèmes. A gauche, la bande principale gauche est proche des plots blancs dont la juxtaposition des contours est généralement à l'origine d'une bande. A droite, une fissure sur le bitume dégrade l'uniformité de la bande peinte. Des contours sont détectés tout au long de la fissure, deux bandes sont détectées.*

l'algorithme, nous avons opté pour des seuils qui permettent de s'affranchir des erreurs de dérive lorsque la propagation des caractéristiques dure pendant plusieurs itérations.

5.2 Suivi des points d'intérêt coplanaires

Dans cette section, nous délaissions les primitives linéaires pour nous intéresser aux primitives ponctuelles. Le suivi des projections des bandes blanches fournit un jeu de primitives coplanaires insuffisant pour calculer l'homographie induite par le plan de la route entre deux poses de chaque caméra. Nous proposons désormais de réaliser une étape de suivi des points d'intérêt coplanaires de façon à faciliter leur mise en correspondance entre deux images successives dans l'objectif d'estimer le déplacement inconnu de la caméra entre deux poses.

La mise en correspondance des primitives est entreprise de manière indépendante pour chacune des deux caméras qui forment le banc stéréo. Nous aborderons donc dans cette partie de manière indifférenciée le traitement des images gauches et droites en utilisant le suffixe ' $g(d)$ ' pour caractériser chacun des éléments propres aux deux caméras de la paire.

Le problème de la mise en correspondance des primitives coplanaires entre deux images successives est comparable à celui de la mise en correspondance de primitives coplanaires entre deux images d'un couple stéréo, décrite dans la partie 4.2.4. Les résultats obtenus lors de l'estimation de l'homographie stéréo étant satisfaisants, nous comptons transposer cette même méthode au cas de deux images consécutives. Deux nuances distinguent la tâche de mise en correspondance de primitives entre les images d'un couple stéréo du cas qui nous intéresse désormais. Premièrement, le déplacement entre deux poses de la caméra est inconnu. Nous supposons cependant que le véhicule est soit à l'arrêt, soit avance. Le cas du déplacement en marche arrière n'est pas considéré car le système serait alors inobservable avec la disposition de caméras orientées dans la direction de la marche avant du véhicule. Deuxièmement, nous connaissons la majorité des primitives que nous devons mettre en correspondance car il s'agit de celles qui ont permis les estimations des homographies stéréo H_{st}^{n-1} et H_{st}^n .

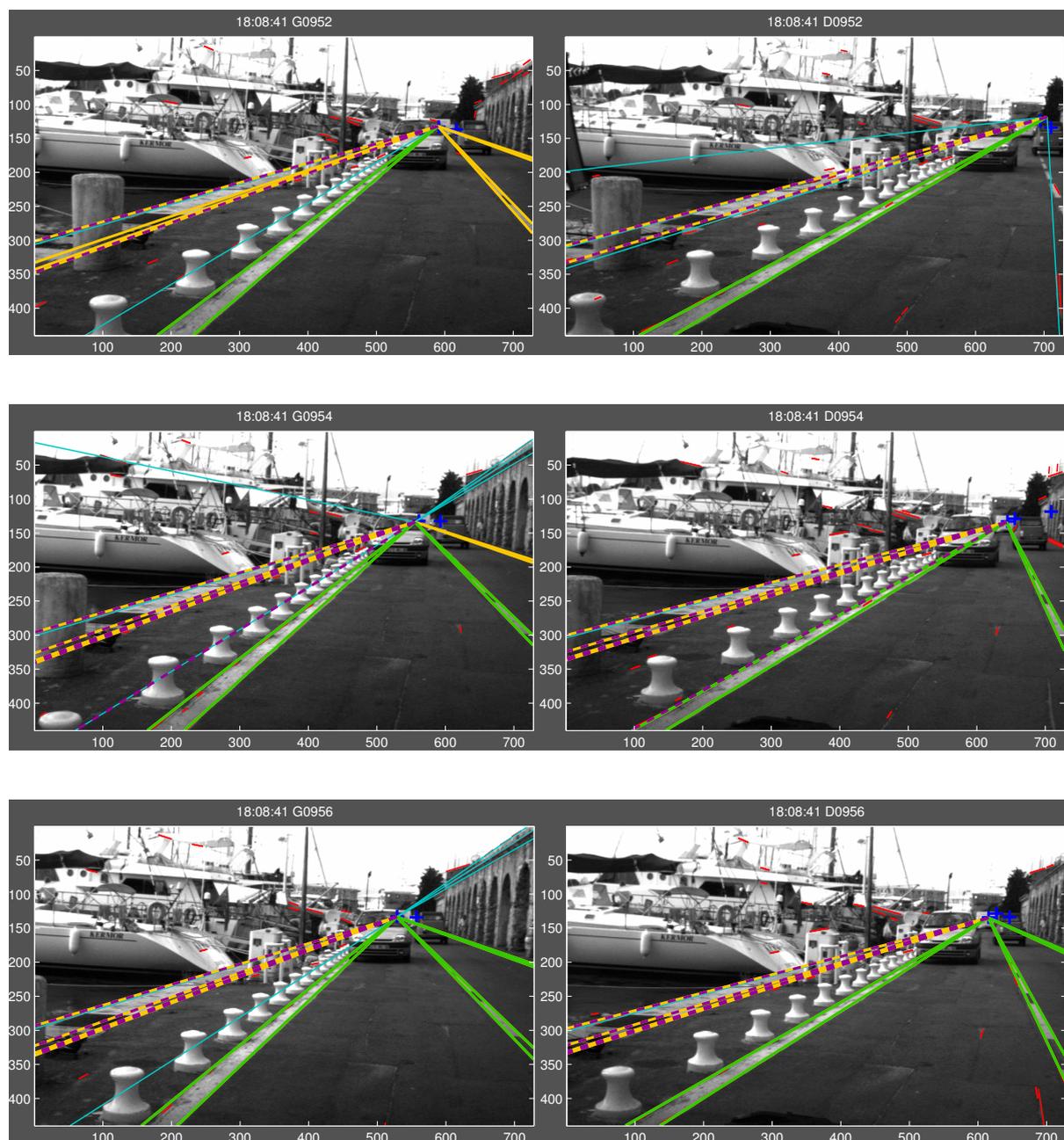


FIG. 5.12 – Identification des bandes principales dès leur réapparition dans le couple d'images. Durant la période pendant laquelle les bandes principales centrales et droites ne sont pas visibles, le mouvement de lacet du véhicule est important. La précision des prédictions fournies par le filtre de Kalman sur les abscisses des bandes principales autorise toutefois d'identifier correctement chacune d'elles dès l'identification de chacune d'elles qu'elles apparaissent à nouveau dans le champ de vision.

Nous commençons par étudier les transformations que subissent les FPs coplanaires entre deux images consécutives à partir de deux types de mouvements : une translation en ligne droite et une rotation. Nous en déduisons des nouveaux critères de mise en correspondance, qui, une fois vérifiés, permettent d'initier le processus itératif qui alterne les étapes de mise en correspondance et d'estimation de l'homographie. Puis, nous détaillerons les opérations indispensables à la transposition de l'estimation de l'homographie d'un plan entre deux images d'un couple stéréo à deux images consécutives.

5.2.1 Mouvement relatif des points d'intérêt coplanaires entre deux poses de la caméra

Le mouvement relatif des FPs coplanaires entre deux images doit répondre à deux contraintes antagonistes de manière à robustifier l'estimation de l'homographie. Ainsi, le mouvement apparent entre les projections d'un même FP doit être :

- plus grand que l'incertitude sur la localisation de ses coordonnées. Des déplacements trop faibles sont à l'origine de bruits de mesure difficilement surmontables : les contraintes de déplacements imposées par les couples de FPs en correspondance ne contraignent pas de manière suffisante l'homographie. Le système formé par les coordonnées des couples de FPs est alors mal conditionné dû à une perte d'observabilité : plusieurs compositions de mouvements de translation et rotation sont alors solutions,
- d'amplitude limitée de façon à ce que les propriétés de colorimétrie et de gradient qui caractérisent les primitives soient suffisamment proches pour faciliter la future mise en correspondance et maintenir la cohérence de la distribution spatiale des primitives entre deux poses de la caméra.

5.2.1.1 Observations

Nous présentons en Fig. 5.13 le résultat de l'étape de suivi de FPs à partir de deux mouvements de caméras : une translation et une rotation. Le champ de vision dans la ROI est dégagé ce qui permet de détecter de nombreux FPs, représentés à l'origine par des '+' jaunes. Parmi ces FPs, ceux qui vérifient l'homographie H_{st}^{n-1} (en bas) et H_{st}^n (en haut) ont été coloriés en rose : ils sont supposés coplanaires. Les FPs mis en correspondance lors de l'estimation de $H_{g(d)}^n$ sont représentés par un 'o' vert dans l'image $\mathcal{I}_{g(d)}^{n-1}$ et un segment vert représente le mouvement apparent dans l'image $\mathcal{I}_{g(d)}^n$.

Le mouvement apparent du DVP '**' selon l'axe vertical est quasiment nul dans les deux cas : aucun mouvement d'inclinaison n'est enregistré. A gauche, dans le cas du déplacement en ligne droite, nous vérifions que la distance qui sépare deux FPs en correspondance dépend essentiellement de l'ordonnée du FP dans l'image de référence. Les FPs situés "à l'infini" (proches de l'horizon) ont des mouvements verticaux nuls ou très petits alors que ceux du premier plan (bas de la ROI) sont les plus grands. A droite, l'amplitude des mouvements apparents des FPs dans le cas de la rotation est plus complexe. Nous retrouvons toutefois une dépendance en fonction de l'ordonnée mais la composante due au mouvement de rotation dominant est elle inversement proportionnelle à l'ordonnée du FP considéré. La caméra étant au centre de la rotation, les FPs les plus éloignés présentent les mouvements latéraux les plus importants. Parmi ces FPs éloignés figure le DVP qui se situe à l'infini et fournit donc un majorant de la composante du mouvement relatif dû aux rotations de la caméra entre deux poses.

La disparité des distances entre projections de FPs coplanaires qui subissent le même

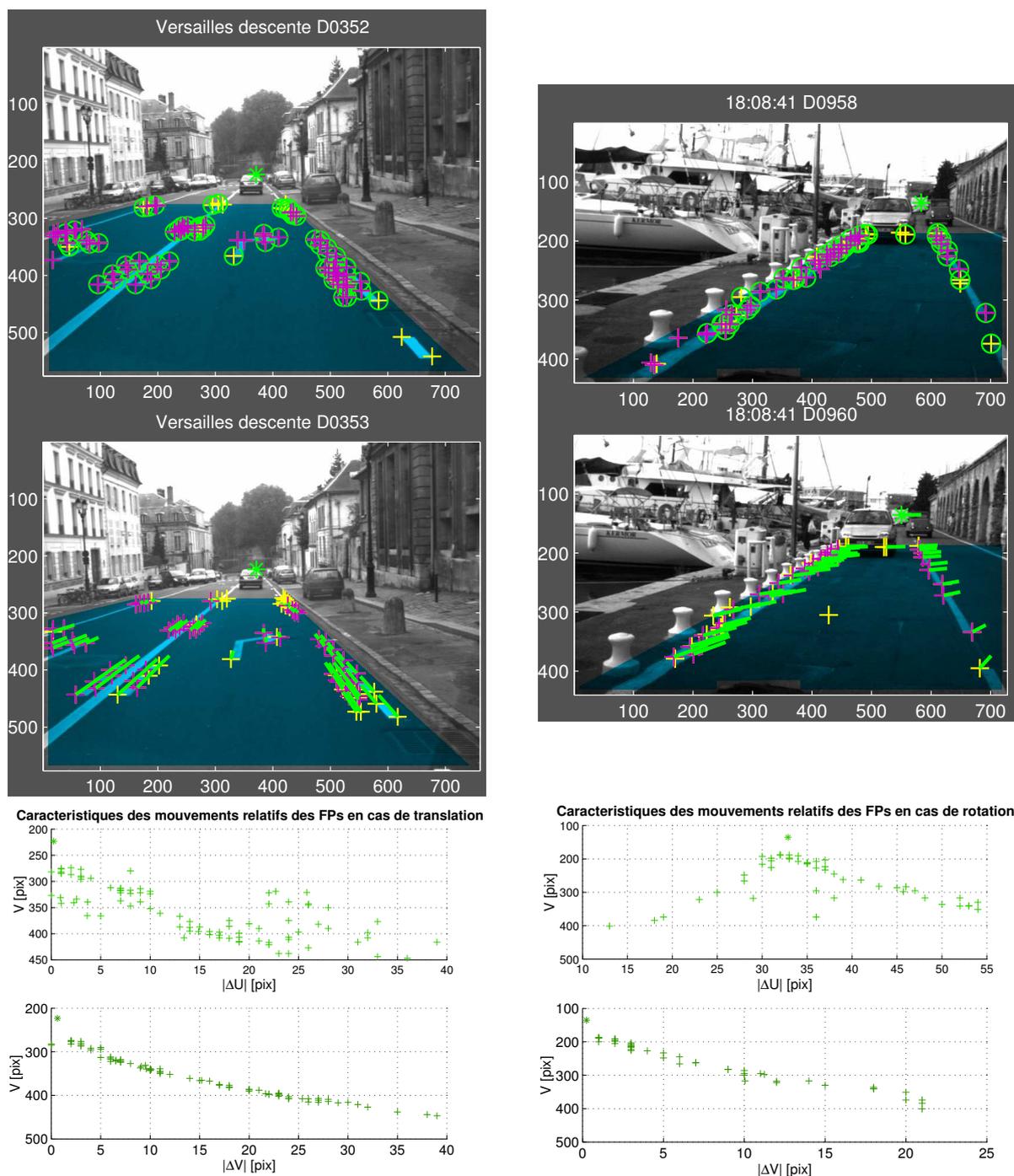


FIG. 5.13 – Résultats de la mise en correspondance de FPs entre images consécutives suite à un mouvement de translation pure (à gauche) et un mouvement de rotation dominant (à droite). En haut, les FPs détectés dans la ROI sont représentés par des '+' jaunes, les roses sont supposés coplanaires et les segments verts représentent le mouvement apparent entre les deux images. En bas, la décomposition des mouvements apparents entre les FPs en correspondance est présentée selon les ordonnées des FPs, mesurées dans l'image de référence. La composante verticale des mouvements apparents dépend exclusivement de de l'ordonnée des FPs contrairement à la composante horizontale qui dépend aussi du mouvement de rotation de la caméra selon l'angle de lacet.

déplacement complique la phase de mise en correspondance car si les FPs dont les écarts de projection les plus réduits ne contraignent pas assez le déplacement entre deux images, ils sont généralement les plus nombreux. Afin d'assurer la convergence de la recherche vers les seuls FPs coplanaires, il nous faut prendre en compte le plus tôt possible les FPs situés dans la partie basse de l'image qui présentent les déplacements les plus importants car ils contraignent de manière forte le mouvement de l'intégralité du plan.

Suivant la vitesse du véhicule et la fréquence d'acquisition des images, le mouvement apparent des FPs extraits au premier plan peut varier de quelques *pixels* à plusieurs dizaines. La prise en compte de tels écarts dans l'étape de mise en correspondance des FPs n'est pas envisageable pour deux raisons : l'écart maximal dépend du mouvement et de la disposition des FPs dans la ROI et l'algorithme de Pilu ne peut aboutir car, par principe, seuls les FPs dont les mouvements relatifs les plus faibles seront retenus. C'est pourquoi, nous devons disposer, tout comme pour l'estimation de l'homographie stéréo, d'une prédiction de l'homographie recherchée de façon à réduire les écarts entre les projections des couples en correspondance.

5.2.1.2 Linéarisation des mouvements de rotation

L'unique information a priori dont nous disposons sur le déplacement du véhicule est qu'il est en train d'avancer ou qu'il est à l'arrêt. Nous savons d'autre part que le mouvement relatif des FPs entre deux images successives est essentiellement contraint par le FOE (foyer d'expansion) qui est l'intersection commune des droites qui portent les vecteurs déplacements des FPs observés entre deux poses.

Or, l'expérience que nous avons acquise pour estimer le DVP à partir de droites-supports, dont la représentativité était de l'ordre de la centaine de *pixels*, nous permet d'affirmer que l'estimation du FOE à partir de quelques segments de quelques dizaines de *pixels* ne peut aboutir. De plus, le FOE n'est caractéristique que des mouvements de translation de la caméra : lorsque le mouvement inclut des mouvements de rotation, les vecteurs déplacements des primitives comprennent une composante de rotation qui les écartent de la direction du FOE.

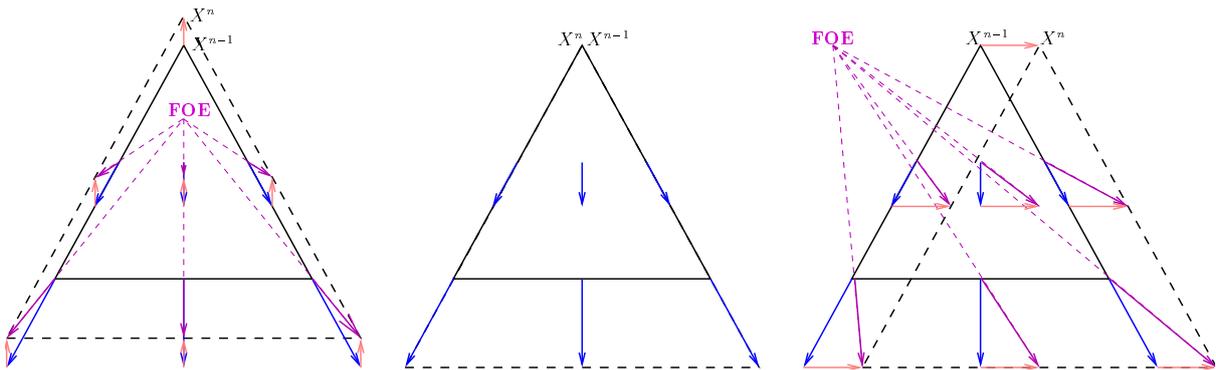


FIG. 5.14 – Champ de composition des vitesses de FPs coplanaires entre deux poses de la caméra. Les vecteurs bleus représentent le mouvement apparent des FPs dû au mouvement de translation selon l'axe optique de la caméra. Leur amplitude augmente linéairement en fonction de leur ordonnée. Les vecteurs roses correspondent, eux, au mouvement de rotation selon l'angle de lacet. L'introduction du mouvement de rotation a écarté le FOE de sa position initiale, localisée en X^{n-1} .

De manière simplifiée, le mouvement d'un véhicule correspond principalement à un mou-

vement de translation selon son axe de déplacement auquel se superpose des mouvements de rotation selon les angles de lacet et d'inclinaison. Notre objectif étant de calculer une prédiction de chaque primitive coplanaire suivie, nous proposons d'assimiler le champ de vitesse complexe induit par les mouvements de translation et rotation de la caméra en deux mouvements de translation selon les axes horizontal et vertical. Cette hypothèse simplificatrice signifie que le mouvement du véhicule peut être décomposé par un mouvement de translation vertical dû à l'avancée du véhicule et un éventuel mouvement d'inclinaison tandis que le mouvement de rotation selon l'angle de lacet peut être modélisé au premier ordre par une translation latérale.

Nous illustrons cette hypothèse couramment employée par de nombreux auteurs ([Branca et al., 1997; Pears and Bojian, 2001]) avec la Fig. 5.14. Le cas particulier d'un déplacement correspondant à une translation pure à vitesse constante est représenté au centre. Lorsque la caméra subit un mouvement de rotation selon l'angle d'inclinaison (à gauche) ou de lacet (à droite), une approximation du mouvement de rotation peut être fournie par le mouvement apparent du DVP qui, rappelons-le, ne dépend que des mouvements de rotation car il appartient au plan à l'infini. Une prédiction des coordonnées des primitives coplanaires peut donc être calculée par composition des mouvements de translations réels (bleus) et approchés (oranges).

5.2.2 Recherche d'une prédiction de l'homographie

Contrairement à la transformation qui lie les deux repères caméras du banc stéréo, le déplacement d'une caméra entre deux poses est inconnu. Cependant, la connaissance a priori du mouvement plan en marche avant du véhicule et les contraintes de continuité de mouvements au cours du temps nous permettent cependant d'imposer que l'évolution du déplacement se fera de manière continue.

5.2.2.1 Connaissance a priori du mouvement apparent des points d'intérêt

L'introduction de l'ensemble des connaissances a priori sur le déplacement plan vers l'avant du véhicule et de la disposition du banc stéréo par rapport à la route nous permettent de déterminer pour chaque FP coplanaire de $\mathcal{I}_{g(d)}^{n-1}$ une région grossière de l'image $\mathcal{I}_{g(d)}^n$ dans laquelle se situe sa projection courante. En réduisant l'espace de projection de chaque FP coplanaire de $\mathcal{I}_{g(d)}^{n-1}$, nous pouvons présélectionner un jeu de FPs de $\mathcal{I}_{g(d)}^n$ parmi lesquels se trouve la représentation courante du FP considéré. L'algorithme de Pîlu est alors employé de façon à identifier parmi les FPs présélectionnés quel est le FP de $\mathcal{I}_{g(d)}^n$ correspondant.

La région de $\mathcal{I}_{g(d)}^n$ dans laquelle se situe la projection courante d'un FP est calculée en utilisant la composition simplifiée du champ des vitesses présentée au paragraphe précédent. Nous disposons pour cela des coordonnées du DVP dans les deux images ($\mathbf{x}_{g(d)}^{n-1}$ et $\mathbf{x}_{g(d)}^n$), ainsi que d'une estimation de la composante verticale maximale du mouvement apparent des FPs coplanaires mesurée à l'itération précédente (voir Fig. 5.15). En supposant que la continuité du déplacement du véhicule soit respectée, nous pouvons calculer une prédiction de la localisation de chaque primitive suivie. La prédiction est calculée en estimant un mouvement radial proportionnel à l'avancée du véhicule, auquel se rajoute les composantes dues aux mouvements de rotations.

Les prédictions ainsi calculées sont peu précises en raison du bruit important qui demeure sur l'estimée du déplacement en translation et de l'approximation des mouvements de rotation. La région de $\mathcal{I}_{g(d)}^n$ dans laquelle sont présélectionnés les FPs susceptibles d'être en correspondance est par conséquent élargit de façon à ne pas retenir le FP correspondant en employant des critères trop stricts. Nous associons donc à chaque FP de $\mathcal{I}_{g(d)}^{n-1}$ une région de prospection de

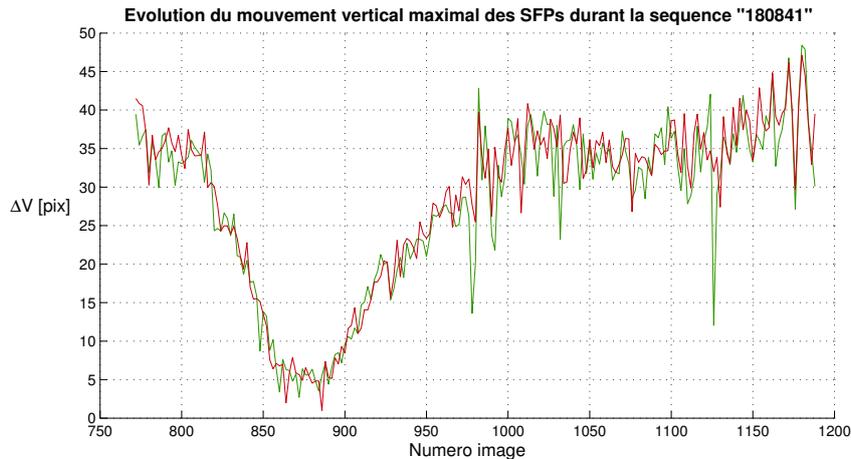


FIG. 5.15 – Chronogramme de l'estimation du mouvement vertical maximal entre deux images consécutives. Les estimations correspondent à la valeur de la composante verticale d'un FP situé au bas de l'image : elles sont calculées en supposant que la composante verticale des FPs suit une loi linéaire en fonction de l'ordonnée des FPs. Les itérations pour lesquelles de fortes discontinuités apparaissent contiennent peu de FPs au premier plan. Un filtrage temporel de type médian sur les 3 derniers échantillons est utilisé pour réduire ces discontinuités.

$\mathcal{I}_{g(d)}^n$ qui s'apparente au demi-disque inférieur, centré sur les coordonnées \mathbf{p}_k^{n-1} et de rayon le double du maximum des composantes latérales et verticales du mouvement apparent prédit.

Le résultat d'une telle segmentation est fourni en Fig. 5.16. Les rayons des régions de projection évoluent essentiellement avec l'ordonnée des FPs dans le cas qui est présenté car le déplacement en translation est important et les mouvements de rotation imperceptibles. Le cas présenté correspond aux deux premières images d'une séquence, l'estimation de la composante verticale maximale a été arbitrairement fixée à 100 *pixels*.

A l'initialisation du processus, nous pouvons seulement supposer que les mouvements de rotation du véhicule seront limités. En fixant que les mouvements relatifs entre les deux premières images n'excèdent pas un seuil limite *ecart_max*, nous pouvons délimiter dans $\mathcal{I}_{g(d)}^2$ des régions circulaires centrées aux coordonnées des FPs coplanaires détectés dans $\mathcal{I}_{g(d)}^1$ dans lesquelles se situent leur nouvelle projection. Le mouvement du véhicule étant soit nul, soit vers l'avant, nous pouvons tronquer la moitié supérieure de chaque disque.

La troncature du demi-disque supérieur s'avère trop sévère dans le cas de mouvement de rotation d'inclinaison non négligeable lorsque le véhicule se déplace à petite vitesse. L'exemple de la Fig. 5.17 montre le cas particulier un certain nombre de FPs correctement mis en correspondance "remontent" dans l'image. Nous nous contenterons donc de tronquer la région du disque au-delà de $dist_max/2$, après la prise en compte du mouvement apparent du DV, au-dessus de l'ordonnée de chaque FP.

5.2.2.2 Sélection des FPs potentiellement en correspondance

Tous les FPs détectés dans la région de projection de l'image $\mathcal{I}_{g(d)}^n$ ne sont pas susceptibles de représenter la projection courante d'un FP coplanaire de $\mathcal{I}_{g(d)}^{n-1}$. Ainsi, le calcul du coefficient de corrélation entre ce dernier et les FPs détectés dans la région de projection permet de rejeter sans ambiguïté des coins qui ne peuvent être la représentation courante du FP considéré. La

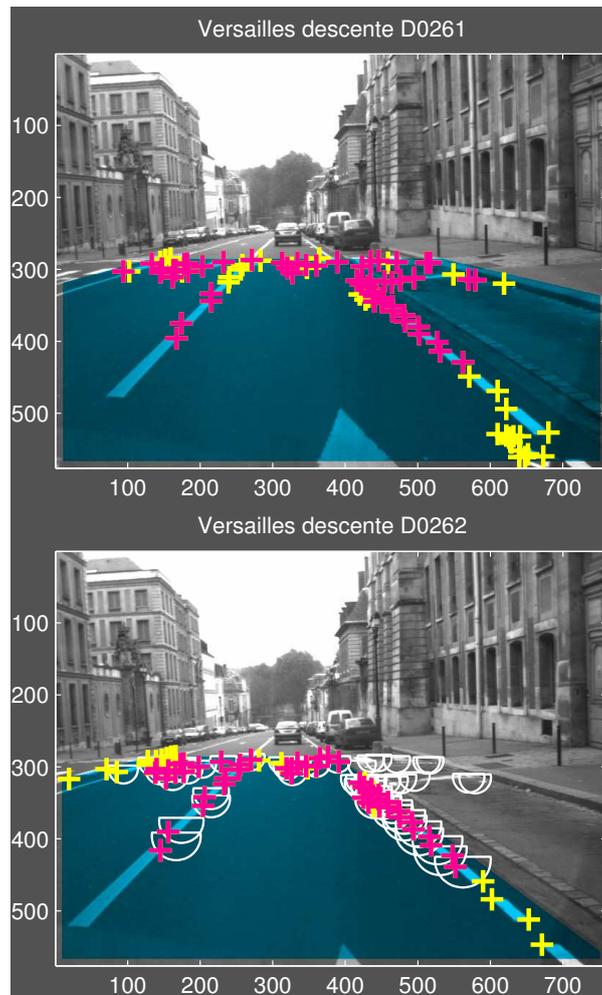


FIG. 5.16 – Présélection des FPs de l'image courante grâce à la seule connaissance a priori du mouvement entre deux images successives. L'amplitude du mouvement relatif des FPs est supposée proportionnelle à l'ordonnée relative du FP considéré par rapport à celle du DVP, toutes deux exprimées dans l'image précédente. Les FPs de $\mathcal{I}_{g(d)}^n$ susceptibles d'être en correspondance avec un FP coplanaire de $\mathcal{I}_{g(d)}^{n-1}$ sont recherchés dans les disques tronqués blancs.

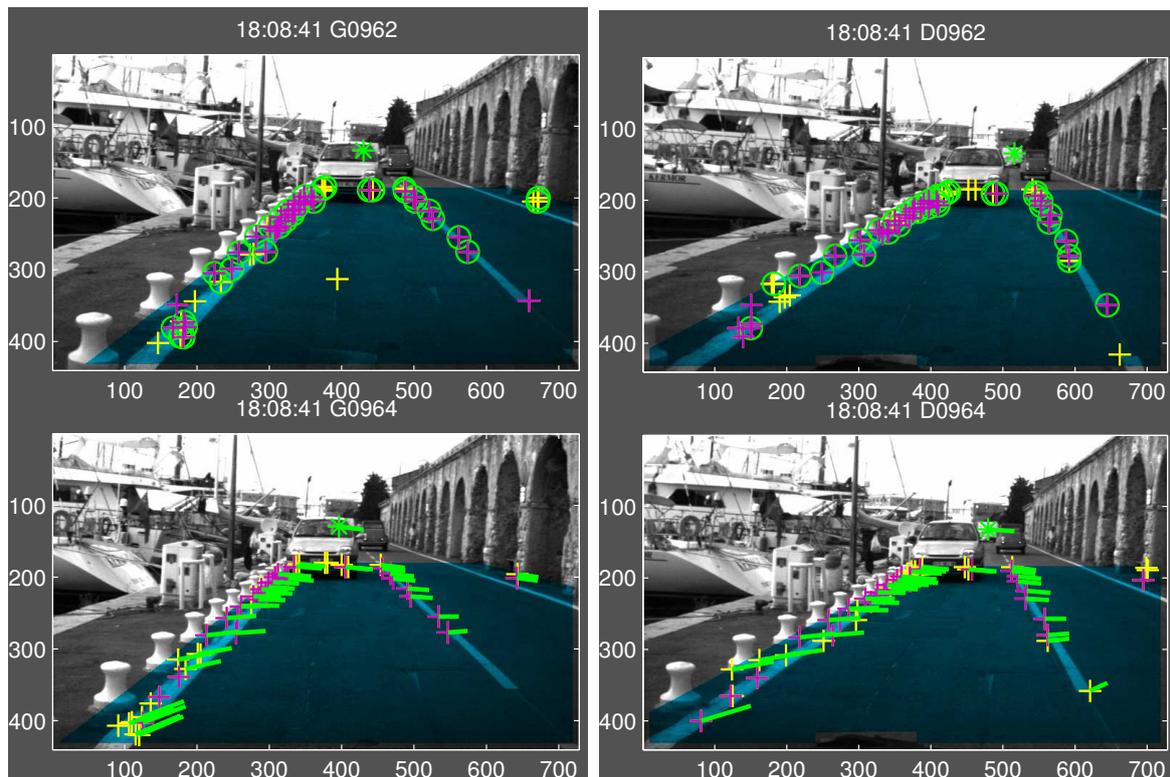


FIG. 5.17 – Cas particulier où le fort mouvement d'inclinaison du véhicule entre deux poses de la caméra induit un mouvement relatif vers le haut de certains FPs coplanaires dans les deux images du couple stéréo.

recherche de la projection courante dans $\mathcal{T}_{g(d)}^n$ est donc opérée selon un nombre restreint de FPs présélectionnés en fonction de leur appartenance à la région de projection et de leurs niveaux de gris.

Le mouvement que subissent les régions autour des FPs entre deux images successives est essentiellement un "glissement" vers le bas qui peut s'apparenter à un effet de zoom avant. La déformation engendrée par ce type de mouvement est moins importante que celle due à un mouvement de translation latérale, observée entre les deux images de la paire stéréo. Lorsqu'un mouvement de rotation se superpose à la translation, son action est appliquée essentiellement aux FPs les plus éloignés de la caméra et par conséquent les mieux caractérisés. Ainsi, la transformation due à un mouvement de rotation de la caméra sera sans répercussion notable sur son coefficient de corrélation. La sélection des FPs potentiellement en correspondance s'effectue donc en comparant le coefficient de corrélation des FPs présélectionnés avec le seuil $seuil_corr = 0,5$, utilisé lors des mises en correspondance entre images du couple stéréo.

Fort des constats précédents, nous utilisons dans un premier temps un seuil du coefficient de corrélation spécifique pour les FPs appartenant à deux images consécutives : $seuil_corr_gd = 0,7$. Ce seuil permet une discrimination beaucoup plus sévère et robustifie la tâche de mise en correspondance. Nous avons néanmoins été contraint de l'abandonner car dans de très rares cas, ce seuil écarte des couples de FPs réellement en correspondance. Nous utilisons par conséquent le même seuil que celui utilisé entre les deux images du couple stéréo $seuil_corr$.

En supposant que la caractérisation de chacun des couples de FPs réellement en correspondance soit maximale, l'algorithme de Pilu mettra en correspondance des couples de FPs dont le coefficient de corrélation est supérieur au $seuil_corr$ mais dont l'écart avec les coordonnées du FP considéré de $\mathcal{T}_{g(d)}^{n-1}$ sera le plus réduit.

Or, dans le cas d'un motif dupliqué, par exemple petits pointillés blancs délimitant des places de stationnement (voir Fig. 5.18), deux à plusieurs FPs peuvent être sélectionnés comme potentiellement en correspondance avec un FP car le déplacement de la caméra est plus important que la distance qui sépare ces coins de la scène. Afin de contraindre la recherche du meilleur FP parmi les sélectionnés, il suffit de substituer à chaque distance entre couples de FPs potentiellement en correspondance $|\mathbf{p}_k^{n-1} - \mathbf{p}_q^n|$, une distance d_0 égale à 1 *pixel*. La matrice de proximité *Dist*, utilisée par l'algorithme de Pilu, sera remplie de d_0 .

Ainsi, seuls les couples dont le coefficient de corrélation est maximal de manière croisée sont retenus. Des mises en correspondance erronées peuvent apparaître mais elles sont généralement écartées au fur de l'affinement de l'estimation de $\mathbf{H}_{g(d)}^n$. L'inconvénient de cette méthode est le nombre relativement réduit de couples de FPs identifiés lorsque plusieurs FPs sélectionnés sont susceptibles de représenter une même projection.

La prise en compte erronée de couples de FPs potentiellement en correspondance se traduit généralement par un nombre réduit de couples de FPs identifiés et bien évidemment une estimation de l'homographie $\mathbf{H}_{g(d)}^n$ fautive. Nous proposons dans les deux prochains paragraphes deux hypothèses que nous avons développées afin d'obtenir une précision plus fine des prédictions des primitives coplanaires.

5.2.2.3 Hypothèse d'un déplacement du véhicule à vitesse constante

Dans l'hypothèse où de nombreux FPs sont détectés sur le plan de la route de manière uniformément répartie dans la ROI, la mise en correspondance des FPs coplanaires entre deux images consécutives est relativement aisée en supposant que le véhicule se déplace à vitesse constante. Le mouvement de la caméra ayant peu le temps d'évoluer entre deux instants d'ac-

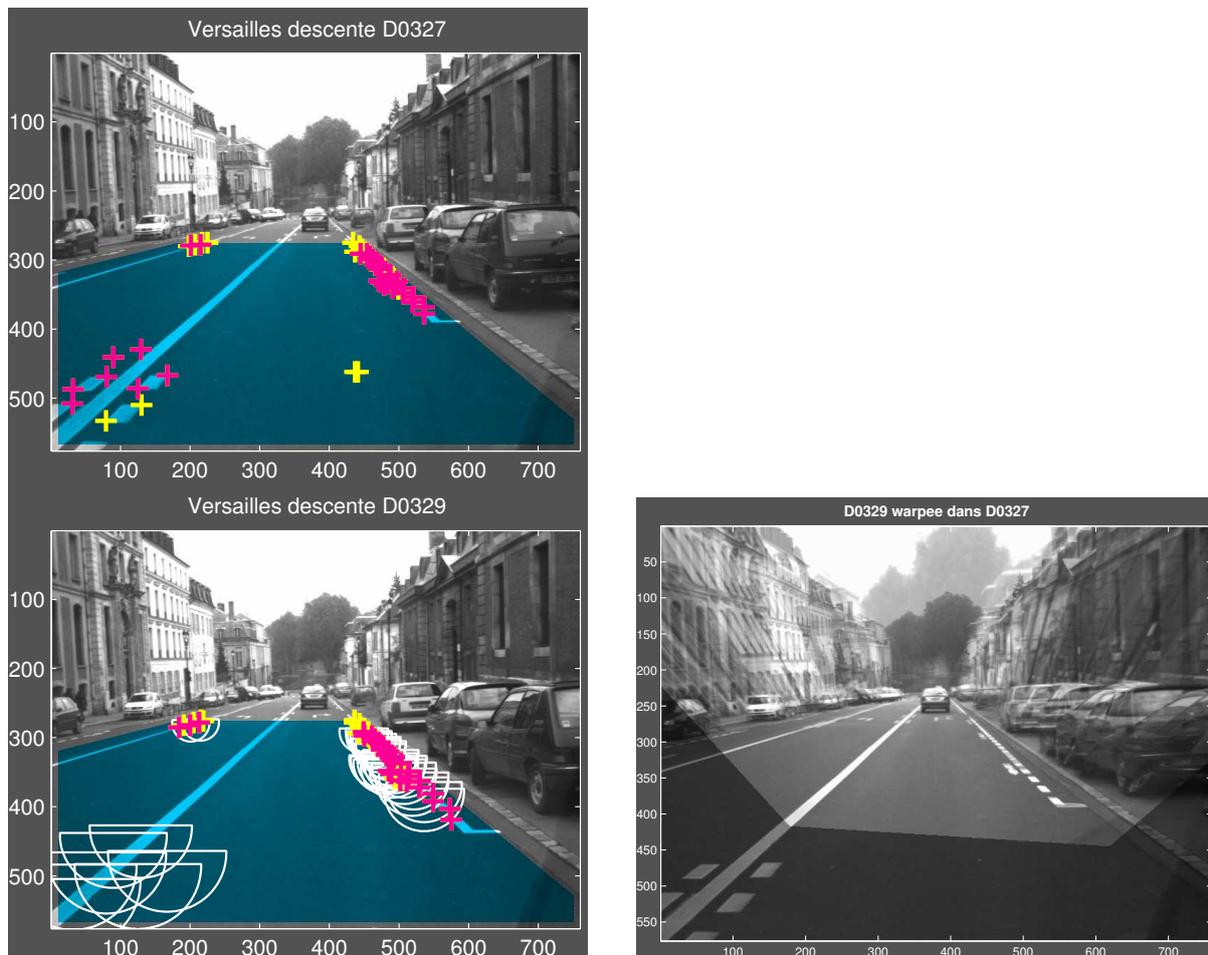


FIG. 5.18 – La présélection de FPs en correspondance est complexifiée lorsque certains FPs sont détectés le long d'un motif répétitif. En raison de la vitesse importante du véhicule, certaines régions de projections contiennent plus d'un FP de $\mathcal{I}_{g(d)}^n$ susceptibles d'être en correspondance avec un FP de $\mathcal{I}_{g(d)}^{n-1}$. Le résultat de la mise en correspondance erronée est présenté grâce au warping de l'image $\mathcal{I}_{g(d)}^n$ dans $\mathcal{I}_{g(d)}^{n-1}$. La superposition des pointillés blancs des deux images qui limitent l'aire de stationnement n'est pas effective, des FPs ont été mis en correspondance de manière erronée.

quisition, une prédiction de l'homographie recherchée est fournie par l'estimation obtenue à l'itération précédente :

$$\widetilde{H_{g(d)}^n} = H_{g(d)}^{n-1}. \quad (5.24)$$

L'intérêt d'une telle méthode est de déterminer une région de projection des FPs de l'image précédente dont la dimension est constante quelle que soit l'ordonnée du FP considéré. Une prédiction des coordonnées des FPs coplanaires de $\mathcal{I}_{g(d)}^{n-1}$ est calculée, la région de projection dans laquelle se situe la projection courante du FP considéré est alors un disque, centré sur les coordonnées de la prédiction, de rayon constant. Comme pour la mise en correspondance entre FPs du couple stéréo, le rayon de la région est pondéré par le facteur pds , relatif à l'ordonnée du FP considéré. Une tolérance de $dist_max$ permet de présélectionner des FPs du premier plan dont le mouvement relatif mesuré dépasse 40 *pixels*.

Dès qu'une prédiction $\widetilde{H_{g(d)}^n}$ de l'homographie est disponible, l'estimation de $H_{g(d)}^n$ correspond exactement à la méthode que nous avons développée pour estimer H_{st}^n . Le calcul est effectué de manière itérative en alternant les phases de mises en correspondance et d'estimations de l'homographie recherchée.

Cependant, lorsque le nombre de FPs coplanaires est réduit et/ou que leur distribution spatiale dans la ROI est limitée, l'emploi de $H_{g(d)}^{n-1}$ en tant que prédiction de $H_{g(d)}^n$ est délicat car la dérive aidant, de moins en moins de couples de FPs sont correctement mis en correspondance. Une difficulté inhérente à la méthode proposée est alors d'estimer la qualité de l'homographie estimée. En effet, les couples de FPs mis en correspondance et retenus pour estimer une homographie vérifient un même critère de distance de projection $seuil_dist = 1,63 \text{ pixels}$. Les moments statistiques (moyenne, variance) des écarts mesurés ne permettent donc pas de qualifier la fiabilité de l'homographie obtenue. Seule la décomposition de l'homographie a posteriori peut valider l'estimation obtenue en vérifiant la continuité des estimations de déplacements et de la décomposition de la normale au plan selon les trois axes. Cette vérification ne permet hélas pas d'identifier des couples de FPs mis en correspondance de manière erronée.

Nous avons déjà souligné le rôle important que jouent les FPs détectés au premier plan : leurs déplacements importants contraignent fortement le mouvement relatif du plan observé entre deux images. Par conséquent, si une homographie permet de mettre en correspondance suffisamment de FPs du premier plan, nous pourrions considérer que son estimation est correcte. Cependant, nous ne sommes pas assurés de détecter des FPs au premier plan car la caractérisation des coins détectés diminue avec la distance qui les sépare de la caméra. L'ultime possibilité de quantifier l'exactitude de l'estimation d'une homographie est alors de vérifier que la distribution spatiale des FPs en correspondance est uniforme sur la ROI. D'après nos expérimentations, si parmi les FPs en correspondance au moins deux d'entre eux sont séparés de quelques centaines de *pixels* en abscisse et en ordonnée et que la majorité des couples ne sont pas alignés le long d'un même contour, il est possible d'utiliser l'estimation $H_{g(d)}^{n-1}$ comme prédiction de $H_{g(d)}^n$.

5.2.2.4 Hypothèse de mouvement relatif des points d'intérêt à vitesse constante

La distribution non-uniforme des FPs dans la ROI étant une configuration récurrente quelle que soit la séquence considérée, nous nous sommes efforcés de déterminer une autre méthode plus générique qui permette de raffiner le calcul de la prédiction des coordonnées des FPs. Dans l'éventualité où les coordonnées d'un FP \mathbf{p}_k sont connues aux itérations $(n-2)$ et $(n-1)$, il est possible de calculer une prédiction $\widetilde{\mathbf{p}_k}^n$ de la position du FP à l'itération n en supposant que le

mouvement apparent du FP dans l'image soit régi par un modèle à vitesse constante :

$$\widetilde{p}_k^n = p_k^{n-1} + (p_k^{n-1} - p_k^{n-2}). \quad (5.25)$$

La précision d'une telle prédiction est relative car nous avons vu plus haut que l'amplitude des mouvements apparents des FPs entre deux images est proportionnelle à l'ordonnée du FP dans l'image de référence, ici la précédente. Les FPs qui se situent dans la partie basse de l'image ayant les mouvements relatifs les plus importants auront une prédiction de leur prochaine position moins précise que d'autres FPs situés plus dans la ROI. Cependant, cette prédiction s'avère plus proche de la position courante recherchée que celle qui définit une région grossière dans laquelle plusieurs correspondants potentiels sont détectés.

Pour les FPs dont les coordonnées sont connues aux itérations $(n-2)$ et $(n-1)$, la recherche du correspondant dans l'image $\mathcal{I}_{g^{(d)}}^n$ consiste donc à rechercher parmi les FPs préalablement sélectionnés ceux qui vérifient l'algorithme de Pilu en utilisant cette fois-ci l'écart de projection qui sépare la prédiction de ses correspondants. Lorsque le nombre et la répartition spatiale des couples de FPs identifiés le permet, une prédiction de l'homographie $\widetilde{H}_{g^{(d)}}^n$ est calculée et initie par conséquent le processus itératif qui alterne l'étape de mise en correspondance des FPs et l'estimation de l'homographie. Dans le cas inverse, la mise en correspondance de tous les FPs de la ROI est opérée à partir des couples de FPs potentiellement en correspondance.

Au cours de l'étape de mise en correspondance, Les couples erronés, formés à partir de FPs non en correspondance, sont rejetés en vérifiant que le mouvement apparent entre les deux images a pour support une droite censée converger vers le FOE. Les coordonnées du FOE sont alors calculées de manière très approximative. Nous supposons en effet que son ordonnée est celle du DVP et que son abscisse peut être calculée comme la moyenne des droites qui supportent les vecteurs déplacements des couples de FPs identifiés. Les couples de FPs dont la droite qui supporte leur mouvement apparent se situe à plus de 200 *pixels* de cette approximation grossière des coordonnées du DVP sont rejetés.

L'intérêt de rechercher une prédiction de l'homographie à partir des couples de FPs extraits grâce à une prédiction de leur position courante plutôt que d'utiliser l'estimée à l'itération précédente se justifie à partir du cas présenté en Fig. 5.19. Nous présentons le résultat de la mise en correspondance des FPs détectés sur le plan de la route lorsque leur répartition spatiale n'est pas uniforme. Dans la Fig. 5.19(a), la prédiction de l'homographie a été initialisée grâce à l'estimée de l'itération précédente. Pour information, cette estimée permettait de mettre en correspondance toutes les projections coplanaires ('+' roses de l'image du haut) qui vérifiait l'homographie stéréo H_{st}^{942} . Le résultat de la mise en correspondance des FPs entre les images \mathcal{I}_g^{942} et \mathcal{I}_g^{944} est correct mais délaisse de nombreuses projections pourtant identifiées coplanaires à l'aide des estimations de H_{st}^{942} et H_{st}^{944} .

Le résultat de la même opération initialisée à l'aide d'une prédiction \widetilde{H}_g^{944} calculée à partir des couples de FPs identifiés grâce à la prédiction de leur position courante dans \mathcal{I}_g^{944} est présenté en Fig. 5.19(b). Un FP proche de la bande principale centrale est mis en correspondance et permet d'obtenir un système mieux conditionné. Le nombre final de couples correctement mis en correspondance est supérieur. Le warping de \mathcal{I}_g^{942} dans \mathcal{I}_g^{944} selon les estimées de H_g^{944} calculées à partir des deux prédictions différentes sont respectivement présentées en Fig. 5.19(c) et Fig. 5.19(d).

L'estimation de l'homographie est entreprise en incluant les caractéristiques des bandes principales des deux images qui contraignent le mouvement de la route en dehors de la région où se situent les FPs. Malgré cela, l'estimée calculée à partir de $\widetilde{H}_g^{944} = H_g^{942}$ est valide seulement dans

le voisinage des couples de FPs en correspondance, contrairement à celle dont la prédiction a été calculée à partir des couples de FPs détectés grâce à la prédiction de leur nouvelle coordonnée.

D'autre part, nous pouvons remarquer que l'estimation d'une homographie qui ne permet de projeter correctement qu'une fraction de la ROI met correctement en correspondance des couples de FPs coplanaires. Cela conforte notre proposition de calculer la prédiction de l'homographie à partir de la prédiction des positions courantes des FPs suivis depuis plus d'une image car si l'estimée $H_{g(d)}^{n-1}$ est perfectible, les coordonnées des FPs, elles, ne le sont pas. Ainsi, la prédiction de l'homographie sera calculée à partir des couples de FPs identifiés grâce à la prédiction des coordonnées des FPs dans l'image courante car cette méthode est indépendante du résultat de l'homographie estimée à l'itération précédente.

5.2.3 Suivi de points d'intérêt mal caractérisés

Nous avons vu que l'extraction de FPs sur le plan de la route permet de détecter une majorité de coins à partir du marquage de la signalisation horizontale en raison du fort contraste qui généralement apparaît entre les bandes "blanches" et la quasi-uniformité du revêtement. Or, la présence de bandes du marquage de la signalisation horizontale n'est pas assurée de manière continue sur une même séquence et dépend du niveau de matérialisation de la route suivie. Afin d'extraire suffisamment de primitives coplanaires pour entreprendre une estimation d'homographie, il peut être tentant de baisser le seuil du score de Harris de manière à détecter plus de FPs.

Cette option est à proscrire pour deux raisons. La robustesse des FPs extraits est étroitement liée à leur score de Harris, abaisser le seuil du score de Harris aura pour conséquence d'introduire des FPs peu robustes dont la précision de la localisation sera d'autant moins grande. De plus, l'augmentation du nombre de FPs extraits complexifie la tâche de mise en correspondance des FPs les moins bien caractérisés car plusieurs FPs dans un voisinage proche sont susceptibles de représenter la projection courante d'un même FP. Des erreurs de mise en correspondance sont alors inévitables et corrompent de manière significative les estimations d'homographies.

Cependant, le détail des projections d'une flèche peinte au sol (voir Fig.5.20), observée dans deux couples d'images séparés de plusieurs dizaines d'itérations, montrent que suivant la saillance d'un coin ou de sa situation dans l'image, l'évolution de son score de Harris peut varier dans de proportions telles que certains d'entre eux ne seront pas pris en compte si nous appliquons de manière stricte la sélection des FPs en fonction du seuil critique *seuil_harris*. La majorité des FPs généralement extraits sur le plan de la route dans les diverses séquences à notre disposition sont moins bien caractérisés que les extrémités de la flèche présentée. Nous devons assouplir notre critère de sélection si nous souhaitons pouvoir utiliser les FPs les moins bien caractérisés lorsqu'ils sont observés au premier plan.

Dans l'objectif d'assurer une répartition quasi-uniforme des FPs dans la ROI, nous avons pris le parti de ne travailler qu'à partir d'un nombre limité de FPs dans chaque région de l'image. Dans le cas de route matérialisée, les coins les mieux caractérisés sont détectés aux extrémités saillantes du marquage au sol. Bien que celui-ci soit formé de régions peintes de manière uniformément, il est toutefois possible d'extraire dans les régions correspondantes de nombreux FPs en réduisant le critère de sélection *seuil_harris*. Ces FPs sont généralement très mal caractérisés et très mal localisés. Il est par conséquent inutile de rechercher des FPs en dehors des extrémités du marquage d'autant plus que le mouvement latéral de la projection du plan de la route est déjà contraint par le mouvement des bandes principales.

Nous pouvons alors supposer que la ROI est composée de deux types de régions distinctes :

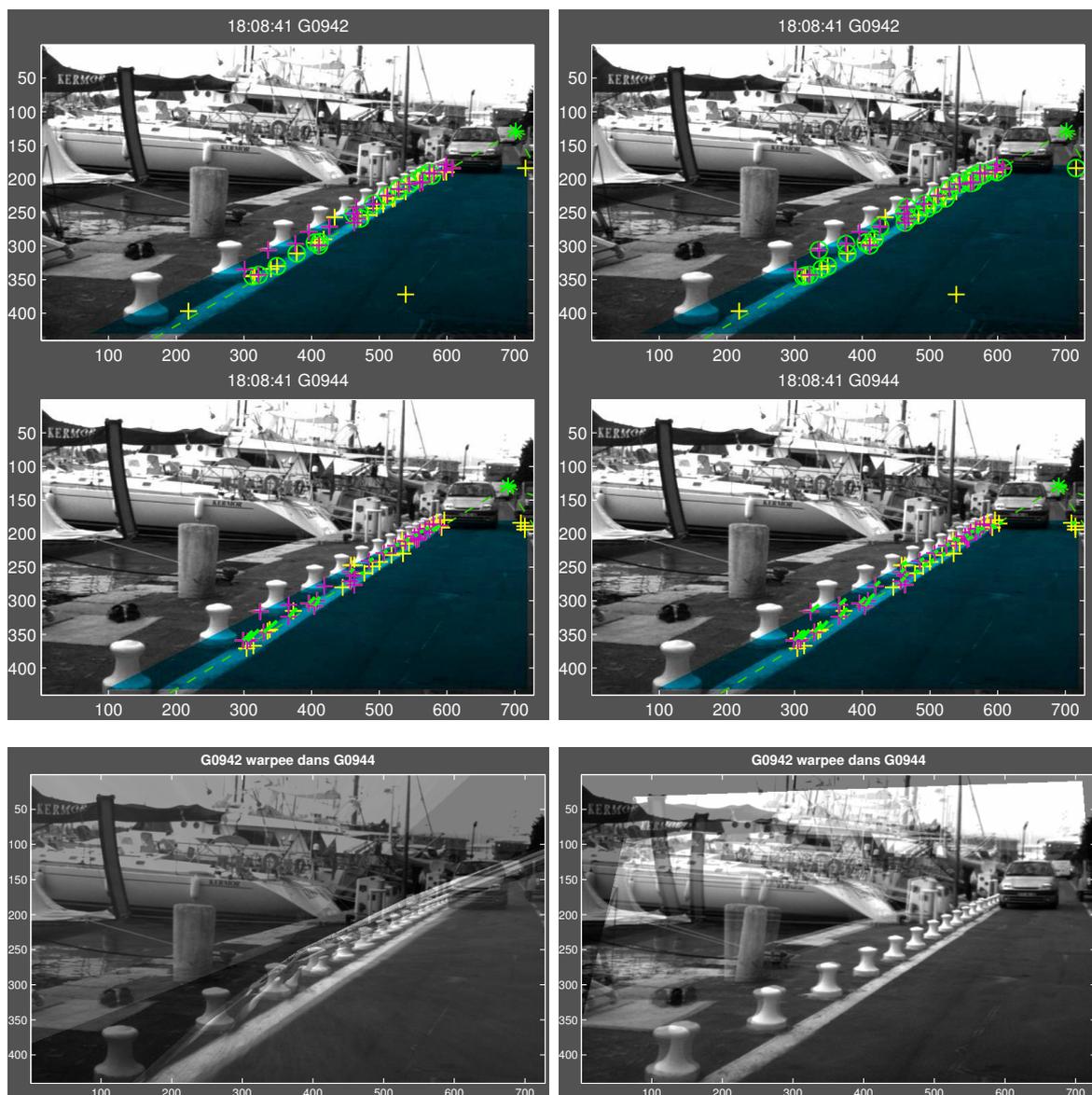


FIG. 5.19 – Confrontation du choix de la prédiction d'une homographie lorsque la distribution des primitives n'est pas uniforme dans la ROI. A gauche, la prédiction de l'homographie correspond à l'estimée obtenue à l'itération précédente. A droite, la prédiction des positions courantes des FPs coplanaires suivis sur plus d'une itération permet de mettre en correspondance la majorité d'entre eux. L'homographie calculée à partir de ces couples de FPs coplanaires est utilisée comme prédiction de l'homographie recherchée et permet d'identifier plus de couples de FPs coplanaires en correspondance que la méthode précédente. En dessous, le warping de l'image courante dans la précédente en fonction des homographies estimées confirme la pertinence d'utiliser la deuxième option puisque le plan de la route est correctement projeté sur la majorité de la ROI.

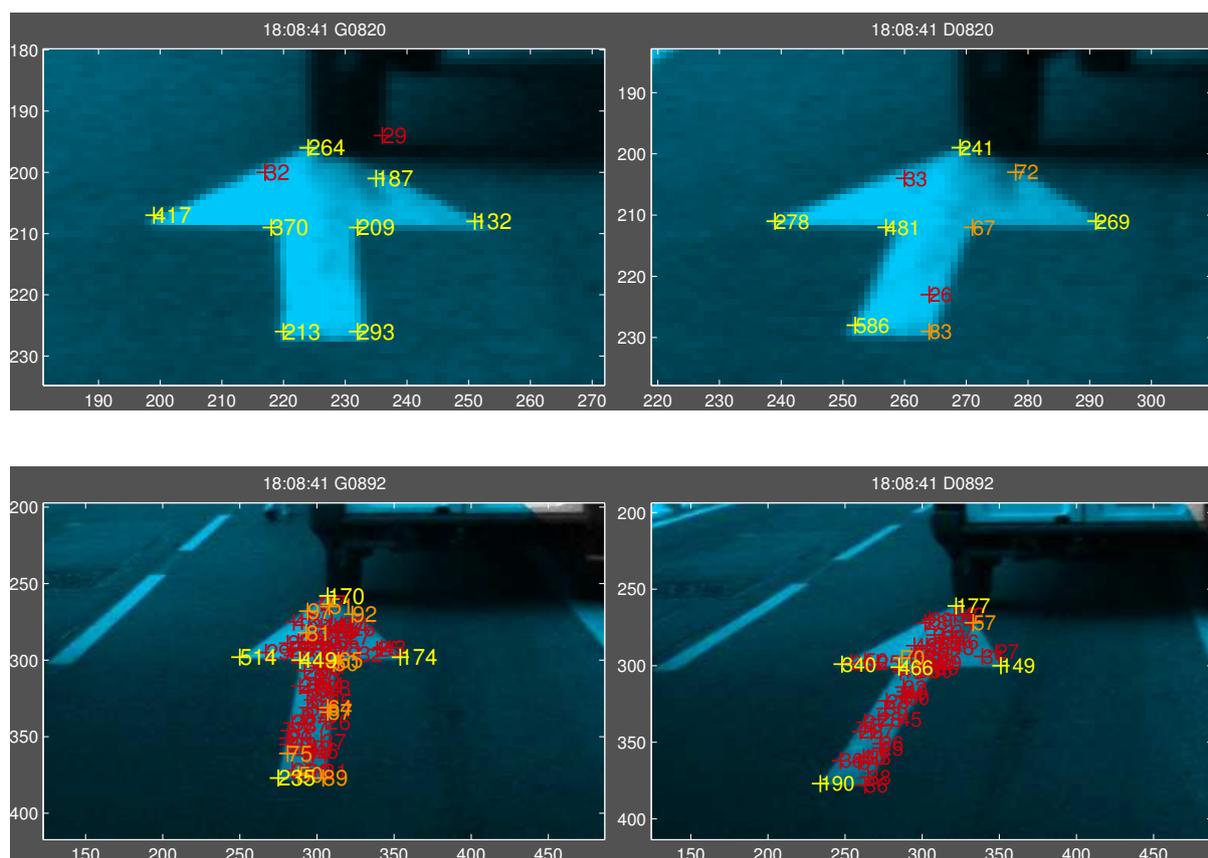


FIG. 5.20 – La détection de points d'intérêt selon l'angle de vue et la proximité de la caméra. Dans un premier temps, seuls les FPs jaunes sont considérés car leur score de Harris est supérieur à `seuil_harris`. Lors de la phase de présélection des FPs de l'itération courante, la recherche de FPs susceptibles de représenter la projection courante de FPs suivis à l'itération précédente s'effectue parmi tous les FPs détectés. Les FPs en orange ont un score de Harris compris entre $[\text{seuil_harris}/2; \text{seuil_harris}]$ alors que les rouges sont en dessous de $\text{seuil_harris}/2$. Les extrémités de la flèche représentées par des angles obtus ont des scores de Harris très faibles, la précision de leur localisation est peu fiable.

les bandes principales où le nombre de FPs à extraire doit être limité et le reste. Dans les régions délimitées par les bandes principales, nous supposons que le nombre de FPs pouvant être extraits est restreint (un ou deux dans le cas de bandes discontinues) par tranche d'ordonnées de 50 *pixels* de hauteur. A l'opposé, dans les régions de la ROI en dehors des bandes principales, tous les FPs dont le seuil est supérieur à *seuil_harris* sont retenus (voir Fig 5.21). Si ce nombre est inférieur à cinq, nous complétons le groupe de FPs détectés avec d'autres FPs dont le score de Harris est inférieur à *seuil_harris*. Cette sectorisation de l'image permet de s'affranchir d'amas de FPs qui réduisent la robustesse des tâches de mise en correspondance et favorise l'estimation de l'homographie en réduisant le nombre de cas où le système est mal conditionné.

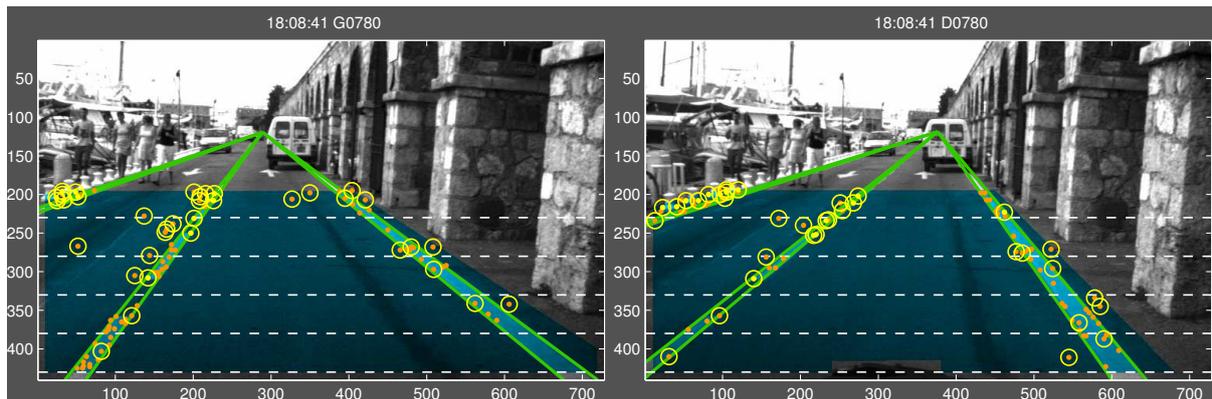


FIG. 5.21 – Sectorisation et découpage de la ROI. Les FPs initialement détectés sont représentés par un '.' orange. Dans un premier temps, seuls ceux dont le score de Harris est supérieur à *seuil_harris* sont retenus. Suite à l'étape d'uniformisation spatiale des FPs détectés dans la ROI, de nombreux FPs oranges sont finalement sélectionnés, essentiellement dans les régions peu texturées de la route.

5.2.4 Sélection des points d'intérêt coplanaires

Nous présentons sur la Fig. 5.22 le résultat de la mise en correspondance des FPs entre deux images consécutives en présence d'obstacles. Nous pouvons d'ores et déjà remarquer que les FPs supposés coplanaires suite aux estimations des homographies stéréo ('+' roses) le sont effectivement. L'emploi d'un banc stéréo se justifie alors pleinement.

Ce résultat permet de valider l'hypothèse de coplanarité des FPs retenus par H_{st}^n . Cinq FPs détectés sur le véhicule en stationnement sont néanmoins mis en correspondance entre les images I^{n-1} et I^n du fait que leur coefficient de corrélation et les écarts croisés de leurs prédictions respectives via H_{st}^n et H_{st}^{n-1} vérifient les critères imposés. En revanche, il est satisfaisant d'observer qu'aucun FP détecté n'est retenu sur le véhicule en mouvement en train de dépasser le véhicule d'expérimentation.

La mise en correspondance de FPs entre deux images ne permet donc pas de vérifier leur appartenance au plan de la route. Il faut donc contraindre plus la sélection : nous considérerons désormais un FP sur le plan de la route à la condition que parmi ses quatre représentations dans les couples d'images précédentes et courantes, au moins trois d'entre elles soient liées grâce aux estimations d'homographies H_{st}^{n-1} , H_{st}^n , H_g^n et H_d^n . Dans le cas présenté précédemment, les FPs détectés sur le véhicule utilitaire seront encore considérés comme coplanaires. Il nous faut par conséquent affiner notre propos et distinguer deux cas :

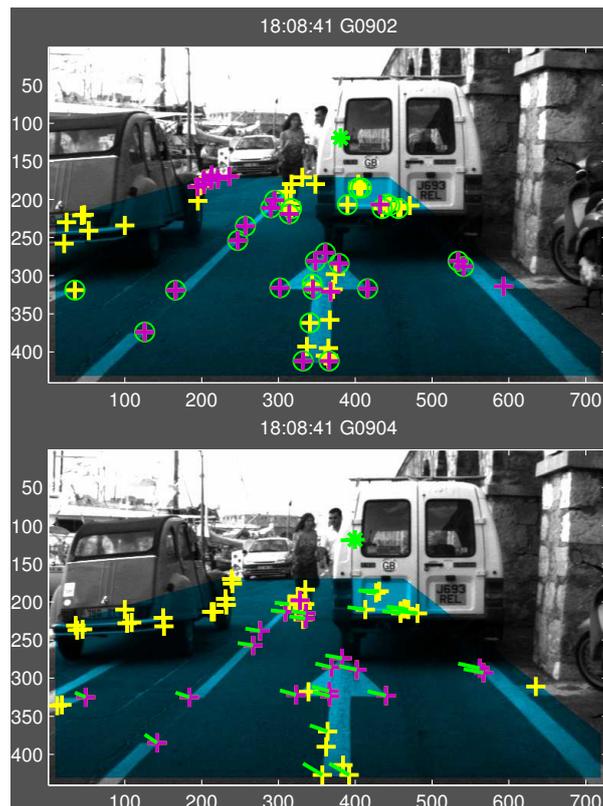


FIG. 5.22 – Détection de FPs coplanaires en présence d'obstacles. Les FPs extraits de la ROI sont représentés par des '+' jaunes. Les FPs '+' roses sont supposés coplanaires car ils ont permis l'estimation de l'homographie stéréo H_{st}^{n-1} (en haut) et H_{st}^n (en bas). La mise en correspondance des FPs entre les deux images est représenté dans I^{n-1} par un 'o' vert et un segment vert dans I^n qui représente le mouvement du FP entre les deux images. Les DVPs '*' sont colorés en vert car leurs coordonnées vérifient le critère de distance suite à la projection des FPs de I^{n-1} dans I^n grâce à l'homographie estimée $H_{g(d)}^n$. Des FPs détectés sur le véhicule utilitaire ont été mis en correspondance bien qu'ils n'appartiennent pas au plan de la route.

- soit les projections du FP dans le couple d'images courantes sont en correspondance par l'intermédiaire de H_{st}^n : il suffit alors que l'une d'entre elle soit liée à sa projection dans l'image qui précède pour que les projections courantes du FP soient considérées comme coplanaires,
- soit l'algorithme ne parvient pas à mettre en correspondance un FP détecté dans une image courante suivant l'homographie stéréo H_{st}^n : le FP manquant n'est pas détecté ou n'est pas mis en correspondance car l'estimation de H_{st}^n n'est pas assez précise. Si la projection courante est liée à la précédente et que cette dernière est liée à sa projection dans l'autre image stéréo via H_{st}^{n-1} alors la projection courante sera considérée comme coplanaire.

L'application d'une telle sélection permet d'éliminer trois des quatre FPs détectés sur le véhicule utilitaire et mis en correspondance suite à l'estimation de $H_{g(d)}^n$. Le FP conservé par erreur est celui situé au centre de la portière gauche car il vérifie aussi H_{st}^n . L'élimination de FPs dont les projections précédentes et courantes ne vérifient pas au moins deux homographies dont une stéréo, nous impose d'affiner l'estimation des homographies courantes de manière à vérifier qu'aucune primitive coplanaire détectée ne soit sélectionnée par erreur.

Nous n'avons pas implémenté ces nouvelles recherches car non seulement il est très rare qu'une nouvelle primitive coplanaire soit détectée ou rejetée et d'autre part le gain en terme de précision d'une telle opération est difficilement quantifiable. En effet, seule la décomposition des matrices d'homographies permet d'observer les effets de l'élimination de FPs supposés coplanaires jusqu'alors. D'après nos expérimentations, dans les pires des cas, certains paramètres évoluent de moins de 5%.

5.2.5 Résultats

La mise en oeuvre de ce seuil adaptatif pour détecter les projections des FPs tant qu'ils sont présents à l'image permet de suivre toutes les extrémités de la flèche peinte au sol sur plusieurs dizaines d'itérations même si certaines d'entre elles sont mal caractérisées.

Nous présentons en Fig. 5.23(a), le résultat du suivi des FPs coplanaires détectés aux extrémités de la flèche peinte au sol. La trajectoire du véhicule durant cette période se décompose en deux parties : d'abord il ralentit en serrant sur sa droite pour se laisser dépasser par la 2CV puis déboîte vers la gauche pour contourner le véhicule utilitaire.

Les FPs détectés aux extrémités de l'empenne de la flèche sont fortement caractérisés car ils sont formés d'une région très claire généralement saillante située dans une région sombre et uniforme. Cependant les coins aux extrémités de la hampe sont représentés par un angle obtus à partir du second tiers du suivi. Ces coins mal détectés par le filtre de Harris sont mal localisés et non retenus dans un premier temps du à leur faible score.

La recherche des FPs potentiellement en correspondance lors du calcul de la prédiction de l'homographie $H_{g(d)}^n$ permet de retrouver parmi les FPs écartés les projections de la plupart d'entre eux. Leur localisation peu précise conduit aux irrégularités des trajectoires des extrémités droites présentées en Fig. 5.23(b). Aucun FP ne correspondait pour les projections représentées par un '+', leurs coordonnées ont été calculées grâce à un procédé que nous développerons dans la partie 6.3.1.2. Nous désirons souligner que les coordonnées de ces FPs estimées vérifient, elles, la continuité des trajectoires.

Le nombre d'images sur lequel les FPs sont suivis n'est en aucun cas représentatif de la qualité du suivi. Par contre, le fait de détecter sans exception toutes les extrémités de la flèche est satisfaisant d'autant plus que la trajectoire de chacun d'eux correspond à un mouvement

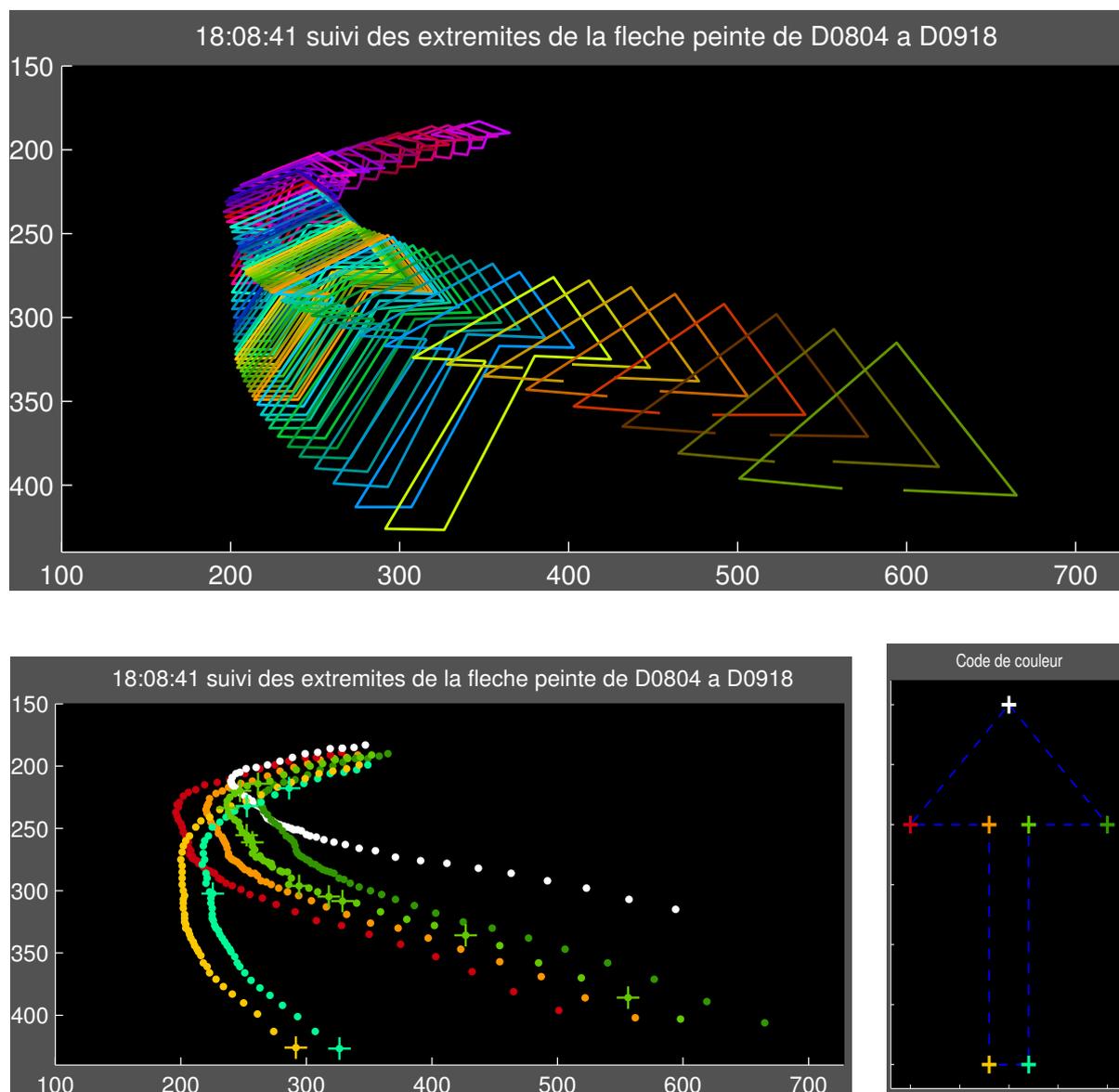


FIG. 5.23 – Suivi dans les images droites des FPs qui délimitent les extrémités de la flèche peinte située devant le véhicule en stationnement. En haut, le suivi de toutes les extrémités de la flèche est effectif malgré une trajectoire de la caméra complexe. En bas à gauche, trajectographie des extrémités suivies avec des coordonnées calculées pour les FPs représentés par une '+'. En bas à droite, code de couleurs utilisé pour représenter la trajectographie des différentes extrémités.

complexe de la caméra.

5.3 Discussion et conclusion

Dans la première partie du chapitre, nous présentons comment segmenter de manière fiable la région de l'image apparentée à la route grâce au suivi temporel des coordonnées du DVP et des abscisses des bandes principales. Nous rappelons que ces estimations interviennent en début d'itération, avant les estimations des homographies stéréo et successives. Certains résultats s'appuient sur des données fausses : les coordonnées des DVPs sont parfois incorrectes lorsque l'observabilité du système est critique, ce qui influence l'extraction des bandes principales. Malgré des erreurs résiduelles, nous parvenons à estimer de manière précise la position des bandes principales dans chaque image lorsqu'elles sont masquées ou en dehors du cadre de l'image. Nous sommes ainsi capables de déterminer à tout instant les limites gauche et droite de la ROI, représentation du plan de la route dans chacune des images.

Les estimations des homographies stéréo et successives permettent de valider le résultat la segmentation de l'image en vérifiant que les coordonnées du DVP et des bandes principales se projettent correctement d'une image vers une autre. Nous n'avons pas implémenté de telles vérifications pour deux raisons. Les estimations d'homographies dépendent de la distribution spatiale et du nombre de primitives coplanaires pouvant être détectées. Or, nous avons pu vérifier qu'une estimation d'homographie incorrecte a tendance à invalider les projections des DVPs et des bandes principales alors qu'elles sont correctes. De plus, la méthode présentée parvient à s'affranchir d'erreurs d'extraction : il est inutile de reprendre l'étape de segmentation de l'image lorsque les estimations d'homographies sont déjà réalisées.

Les résultats cohérents du suivi de coordonnées du DVP et de l'abscisse des bandes principales sur la séquence du port d'Antibes ne permettent certes pas de valider les modèles d'évolution de ces deux variables. Cependant, la séquence test s'avère beaucoup plus compliquée qu'elle ne le paraît pour l'extraction du DVP et le suivi des bandes principales. En effet, le changement de voie de circulation est une manoeuvre qui implique une modification conjointe de la position du DVP et des bandes principales. Lorsque le véhicule suit une voie de circulation donnée, seul le DVP est susceptible d'évoluer en cas de virage mais son déplacement est contraint par l'intersection des bandes principales qui encadrent le véhicule de manière continue ce qui induit des variations des abscisses réduites par rapport à celles que nous suivies.

Nous avons appliqué dans la seconde partie de ce chapitre la méthodologie développée au chapitre précédent pour estimer l'homographie stéréo à l'homographie relative au déplacement d'une caméra entre deux poses. Parmi les différentes méthodes proposées pour le calcul de la prédiction de l'homographie recherchée, celle effectuée à partir des couples de FPs mis en correspondance permet d'obtenir des résultats satisfaisants même lorsque la distribution spatiale des primitives coplanaires observées n'est pas suffisamment contraignante.

La mise en correspondance des primitives coplanaires permet d'identifier des primitives coplanaires que les homographies stéréo n'ont pas détectées en tant que telles. A l'inverse, des primitives mal caractérisées et détectées sur des obstacles statiques peuvent vérifier l'homographie entre deux images successives durant de nombreuses itérations. L'estimation de l'homographie est par conséquent biaisée car seules les primitives coplanaires doivent être en correspondance. Nous verrons au chapitre suivant comment la redondance des informations collectées sur les coordonnées des primitives coplanaires observées dans plusieurs couples d'images permet d'obtenir des résultats robustes et d'une précision sub-pixellique.

Chapitre 6

Trajectographie du véhicule

6.1 Position du problème

Dans les chapitres précédents, nous avons présenté comment segmenter puis suivre les éléments structurants d'un environnement urbain considéré comme planaire par morceaux. La paire stéréo étant embarquée à bord d'un véhicule, le suivi des primitives coplanaires est rendu complexe par l'ensemble des aléas dus à l'emploi de caméras en environnement d'extérieur mais aussi en raison du caractère dynamique de l'environnement urbain : le champ de vision est réduit par la présence d'obstacles statiques ou en mouvement qui occultent les plans segmentés.

La nature de l'environnement urbain ainsi que la supposition de mouvement plan du véhicule autorisent à modéliser le mouvement des primitives coplanaires observées dans l'image par un mouvement à vitesse constante. Un processus de filtrage fournit une prédiction des projections des primitives détectées à l'itération précédente dans les images courantes. Cette prédiction facilite la tâche de mise en correspondance et améliore l'estimation de l'homographie induite.

Lorsque la précision des homographies calculées est insatisfaisante, une solution consiste à prendre en compte l'image du plan observé en une troisième pose. Le tenseur trifocal a été largement étudié dans la littérature car il joue un rôle analogue à celui de la matrice fondamentale pour trois vues. Son intérêt est, d'une part, de propager les contraintes épipolaires entre des primitives observées à l'origine dans deux images et, d'autre part, de lever l'ambiguïté liée au facteur d'échelle des homographies par triangulation. Le tenseur trifocal est généralement employé pour déterminer la géométrie d'un système trinoculaire paire stéréo et/ou l'auto-calibration de caméra(s) mobile(s). Hélas, comme toutes les méthodes fondées sur l'estimation de la matrice fondamentale, la mise en oeuvre du tenseur trifocal est délicate car la résolution du système imposé par la matrice fondamentale est très sensible aux bruits. Nous n'avons donc pas retenu cette approche qui ne nous semblait pas exploiter au mieux les propriétés induites par un système de vision stéréo se déplaçant dans une scène planaire.

L'estimation des trois homographies courantes H_{st}^n , H_g^n et H_d^n a été calculée jusqu'alors à partir de primitives coplanaires mises en correspondance de façon croisée entre chaque couple d'images retenu. Cependant, comme le montre la Fig. 6.1, la mise en correspondance des primitives coplanaires entre deux couples d'images stéréo ne sera totale que si les correspondances sont vérifiées entre toutes les paires d'images disponibles.

L'introduction simultanée des contraintes de géométrie épipolaire liées à l'observation d'un plan à partir de plusieurs poses peut être pris en compte par le biais de la super-homographie, à la quelle nous consacrerons la deuxième partie de ce chapitre. Avant cela, nous rappellerons dans une première partie comment décomposer les homographies pour estimer les mouvements

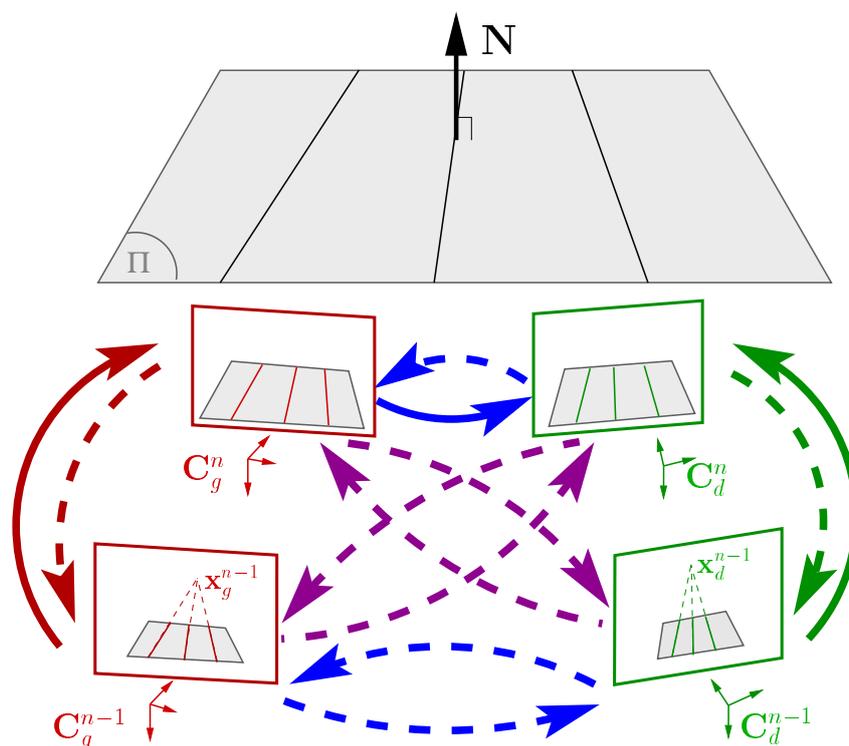


FIG. 6.1 – Détail des homographies induites entre 2 couples de paires d'images. L'estimation des homographies courantes H_{st}^n , H_g^n et H_d^n , représentées en trait plein, sera améliorée si les primitives coplanaires utilisées sont mises en correspondance dans toutes les images disponibles. La mise en correspondance des primitives entre images duales de paire stéréo différente (en pointillés violets) introduit de la redondance dans les contraintes imposées par la géométrie épipolaire.

de translation et rotation induits par le déplacement d'une caméra entre deux poses.

6.2 Décomposition des homographies

La matrice d'homographie exprimée dans l'espace euclidien $\mathbf{H}_e = \mathbf{K}^{-1}\mathbf{H}\mathbf{K}$ (avec \mathbf{K} matrice de paramètres intrinsèques de la caméra) contient les mouvements de translation \mathbf{t} et rotation \mathbf{R} de la caméra caractérisant son déplacement entre deux poses. L'homographie capture à la fois la structure géométrique et le mouvement apparent d'un plan observé dans une image. L'extraction des paramètres de déplacements 3D est étroitement liée à la normale \mathbf{n} et à la distance d du plan considéré par rapport au repère de la caméra dans sa pose initiale :

$$\mathbf{H}_e = \mathbf{R} + \frac{\mathbf{t} \cdot \mathbf{n}^t}{d}. \quad (6.1)$$

L'homographie euclidienne est définie sans facteur d'échelle ($\mathbf{p}_2 = \mathbf{H}_e \mathbf{p}_1$), la décomposition de l'homographie \mathbf{H}_e permet d'identifier les 6 paramètres du déplacement (3 angles de rotation \mathbf{R} , 3 composantes de translation normalisées \mathbf{t}), 2 composantes de la normale au plan \mathbf{n} car le vecteur est normalisé et la distance d qui sépare la pose initiale de la caméra du plan.

6.2.1 Résolution

Nous rappelons ici les résultats établis par Faugeras [Faugeras, 1993]. Du fait de sa non-linéarité, la résolution de \mathbf{H}_e admet quatre solutions différentes dans le cas général. Nous proposons dans un premier temps de les énumérer puis de restreindre le choix aux seules solutions qui ont une existence physique.

6.2.1.1 Rappel

La matrice \mathbf{H}_e est diagonalisable par décomposition de la forme SVD : $\mathbf{H}_e = \mathbf{U}\mathbf{D}\mathbf{V}^t$ avec \mathbf{D} , matrice diagonale contenant les trois valeurs propres ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) et \mathbf{U} et \mathbf{V} des matrices orthogonales telles que : $\mathbf{U}^t\mathbf{U} = \mathbf{V}^t\mathbf{V} = \mathbf{I}_3$. La matrice diagonale \mathbf{D} est composée de nouveaux paramètres de déplacements (\mathbf{R}^* , \mathbf{t}^* , \mathbf{n}^* , d^*) que nous pouvons associer aux paramètres recherchés :

$$\mathbf{D} = \mathbf{R}^* + \frac{\mathbf{t}^* \cdot \mathbf{n}^{*t}}{d^*} \iff \begin{cases} \mathbf{R} &= s\mathbf{U}\mathbf{R}^*\mathbf{V}^t \\ \mathbf{t} &= \mathbf{U}\mathbf{t}^* \\ \mathbf{n} &= \mathbf{V}\mathbf{n}^* \\ d &= sd^* \\ s &= \det(\mathbf{U})\det(\mathbf{V}) \end{cases} \quad (6.2)$$

La décomposition de la normale \mathbf{n}^* dans une base canonique formée des vecteurs $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ ($\mathbf{n}^* = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \alpha_3\mathbf{e}_3$) fournit un jeu de trois équations :

$$\lambda_i \mathbf{e}_i = d^* \mathbf{R}^* \mathbf{e}_i + \mathbf{t}^* \alpha_i \quad \forall i \in \{1, 2, 3\}. \quad (6.3)$$

Le vecteur \mathbf{n} étant unitaire et \mathbf{V} orthogonale, \mathbf{n}^* est aussi unitaire, ce qui s'écrit $\sum_{i=1}^3 \alpha_i^2 = 1$. La substitution du vecteur \mathbf{t}^* d'une des équations du système 6.3 dans les deux autres mène à :

$$d^* \mathbf{R}^* (\alpha_j \mathbf{e}_i - \alpha_i \mathbf{e}_j) = \lambda_i \alpha_i \mathbf{e}_i - \lambda_j \alpha_j \mathbf{e}_j \quad \forall i \neq j. \quad (6.4)$$

La matrice \mathbf{R}^* représentant une matrice de rotation ($\det(\mathbf{R}^*) = 1$), la norme du vecteur \mathbf{n}^* est conservée. En élevant au carré l'égalité précédente, une nouvelle formulation de l'écriture précédente est :

$$\begin{cases} (d^{*2} - \lambda_2^2)\alpha_1^2 + (d^{*2} - \lambda_1^2)\alpha_2^2 = 0 \\ (d^{*2} - \lambda_3^2)\alpha_2^2 + (d^{*2} - \lambda_2^2)\alpha_3^2 = 0 \\ (d^{*2} - \lambda_1^2)\alpha_3^2 + (d^{*2} - \lambda_3^2)\alpha_1^2 = 0 \end{cases} \quad (6.5)$$

Le système précédent peut être considéré comme un système linéaire en fonction des variables $\alpha_1^2, \alpha_2^2, \alpha_3^2$. Afin de rejeter la solution évidente, le déterminant du système doit être nul :

$$(d^{*2} - \lambda_1^2)(d^{*2} - \lambda_2^2)(d^{*2} - \lambda_3^2) = 0. \quad (6.6)$$

La résolution de la décomposition de l'homographie \mathbf{H}_e dépend donc de l'ordre de multiplicité de ses valeurs propres. Trois cas sont à distinguer : les valeurs propres sont toutes distinctes, deux d'entre elles sont égales, toutes sont égales. Dans les trois cas, nous supposons que $d^{*2} = \lambda_2^2$. Sachant que la norme du vecteur \mathbf{n}^* est unitaire, le système à résoudre est :

$$\begin{cases} (d^{*2} - \lambda_1^2)\alpha_2^2 = 0 \\ (d^{*2} - \lambda_3^2)\alpha_2^2 = 0 \\ (d^{*2} - \lambda_1^2)\alpha_3^2 + (d^{*2} - \lambda_3^2)\alpha_1^2 = 0 \\ \alpha_3^2 + \alpha_2^2 + \alpha_1^2 = 1 \end{cases} \quad (6.7)$$

Nous traitons désormais du cas général où les trois valeurs propres sont distinctes et précisons au fur du développement des cas particulier où la valeur propre est de multiplicité double ou triple. Les deux premières lignes imposent $\alpha_2 = 0$, tandis que la substitution de de la quatrième dans la troisième conduit à :

$$\begin{cases} \alpha_1 = \epsilon_1 \sqrt{\frac{\lambda_1^2 - \lambda_2^2}{\lambda_1^2 - \lambda_3^2}} \\ \alpha_2 = 0 \\ \alpha_3 = \epsilon_3 \sqrt{\frac{\lambda_3^2 - \lambda_2^2}{\lambda_3^2 - \lambda_1^2}} \end{cases} \quad (6.8)$$

avec ϵ_1 et ϵ_3 égal à $(+/-1)$.

6.2.1.2 Décomposition du mouvement selon l'ordre de multiplicité

Nous traiterons seulement du cas $d^* > 0$, le cas $d^* < 0$ correspondant à une réécriture du système qui conduit à des résultats différents à partir de développements similaires. Nous considérons donc $d^* = \lambda_2$. L'application de l'Equ. 6.3 au vecteur de la base canonique \mathbf{e}_2 montre que \mathbf{R}^* est une rotation autour de cet axe ($\mathbf{R}^* \mathbf{e}_2 = \mathbf{e}_2$). La matrice \mathbf{R}^* peut s'écrire sous la forme :

$$\mathbf{R}^* = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \quad (6.9)$$

Les éléments de cette matrice sont obtenues en vérifiant simultanément les Equ. 6.4 avec $\{i, j\}$ respectivement égaux à $\{1, 3\}$ et Equ. 6.7 :

$$\begin{cases} \cos(\theta) = \frac{\lambda_1 \alpha_3^2 + \lambda_3 \alpha_1^2}{\lambda_2} \\ \sin(\theta) = (\lambda_1 - \lambda_3) \frac{\alpha_1 \alpha_3}{\lambda_2} \end{cases} \quad (6.10)$$

ce qui permet de déterminer les composantes du vecteur \mathbf{t}^* :

$$\mathbf{t}^* = (\lambda_1 - \lambda_2) \begin{vmatrix} \alpha_1 \\ 0 \\ -\alpha_3 \end{vmatrix} \quad (6.11)$$

Les déplacements \mathbf{t}^* et \mathbf{R}^* dépendent des paramètres α_1 et α_3 : ils admettent donc quatre solutions différentes suivant les valeurs de ϵ_1 et ϵ_3 . Dans le cas traité jusqu'alors, le mouvement entre les deux poses de la caméra correspond à une translation \mathbf{t} et une rotation \mathbf{R} .

Lorsque la valeur propre est de multiplicité deux ($\lambda_1 = \lambda_2$), les composantes du vecteur \mathbf{n}^* sont telles que $\alpha_1 = \alpha_2 = 0$, soit $\alpha_3 = \epsilon_3$. Le cas $\lambda_3 = \lambda_2$ est identique au précédent par symétrie. La rotation \mathbf{R}^* est telle que $\cos(\theta) = \sin(\theta) = 0$ et la translation \mathbf{t}^* s'exprime alors en fonction de α_3 :

$$\begin{cases} \mathbf{R}^* &= \mathbf{I}_3 \\ \mathbf{t}^* &= (\lambda_1 - \lambda_3)\mathbf{n}^* \end{cases} \quad (6.12)$$

Le déplacement correspond à un mouvement de translation ($\mathbf{R} = \mathbf{0}$) qui ne dépend que de ϵ_3 . Deux solutions sont donc possibles.

Une valeur propre ayant une multiplicité d'ordre 3 est à l'origine d'un système indéterminé : $\lambda_1 = \lambda_2 = \lambda_3$ ce qui suppose que le vecteur \mathbf{n}^* est non-défini ($\alpha_1, \alpha_2, \alpha_3$ sont quelconques). La décomposition des mouvements est telle que :

$$\begin{cases} \mathbf{R}^* &= \mathbf{I}_3 \\ \mathbf{t}^* &= \mathbf{0} \end{cases} \quad (6.13)$$

ce qui induit un déplacement correspondant à une rotation pure ($\mathbf{t} = \mathbf{0}$). La solution est, dans ce cas, unique mais la normale au plan \mathbf{n} est inconnue.

6.2.1.3 La prise en compte de la contrainte d'observabilité

Dans le cas général, quatre solutions sont envisageables selon les signes de α_1 et α_3 . Leur valeur numérique dépend du signe de $d^* = (+/-)\lambda_2$, d'après le système Equ. 6.7. Parmi ces quatre solutions, seules deux d'entre elles ont une réalité physique, imposée par la contrainte d'observabilité : les objets observés par une caméra se situent dans le demi-espace au-delà de l'écran, leur ordonnée selon l'axe optique z est positive. Ainsi, en considérant les coordonnées projectives \mathbf{p}_1 et \mathbf{p}_2 d'un point de l'espace observé à partir de deux poses d'une caméra sont telles que :

$$\begin{cases} \mathbf{p}_2 &= \mathbf{R}\mathbf{p}_1 + \mathbf{t} \\ \mathbf{n}^t \mathbf{p}_2 &= 0 \end{cases} \quad (6.14)$$

En introduisant ces égalités dans la décomposition Equ. 6.1, on obtient :

$$d\mathbf{p}_2 = d\mathbf{R}\mathbf{p}_1 + d\mathbf{t} = (\mathbf{H}\mathbf{e} - \mathbf{t}\mathbf{n}^t)\mathbf{p}_1 + d\mathbf{t} = \mathbf{H}\mathbf{e}\mathbf{p}_1 - \mathbf{t}(\mathbf{n}^t\mathbf{p}_1 - d) = \mathbf{H}\mathbf{e}\mathbf{p}_1 \quad (6.15)$$

Les ordonnées selon l'axe optique de ces deux projections seront de même signe si et seulement si :

$$\frac{z_2}{z_1} = \frac{\mathbf{H}\mathbf{e}_3^t \cdot \mathbf{p}_2}{d} > 0 \quad \text{avec} \quad \mathbf{H}\mathbf{e} = \begin{vmatrix} \mathbf{H}\mathbf{e}_1^t \\ \mathbf{H}\mathbf{e}_2^t \\ \mathbf{H}\mathbf{e}_3^t \end{vmatrix} \quad (6.16)$$

Si l'homographie $\mathbf{H}\mathbf{e}$ et les coordonnées de \mathbf{P} sont connues dans les deux images, nous pouvons en déduire le signe de d , puis celui de d^* ($d^* = d/s$). Le paramètre s étant calculé lors de la

décomposition de H_e , seules deux solutions sur les quatre énumérées ont une vérité physique. Parallèlement, dans le cas où la valeur propre d^* est de multiplicité 2, la contrainte d'observabilité est vérifiée par une seule des deux solutions.

6.2.2 Résultats

La décomposition de l'homographie induite par le déplacement relatif d'un plan correspond au cas où l'ordre de multiplicité de la valeur propre d^* est égal à 2. Parmi les deux solutions possibles, la détermination de la solution qui vérifie la contrainte d'observabilité nécessite de connaître la projection dans l'image de l'orientation de la normale au plan \mathbf{n} qui induit l'homographie.

Dans le cas particulier de la route supposée plane par morceaux, la normale sera proche de la verticale et orientée vers le ciel, ce qui se traduira dans l'espace image par une projection quasi-verticale si le plan de l'image peut être assimilé à un plan vertical. Une étude au cas par cas devra être menée dans l'éventualité du suivi de nouveaux plans.

Les chronogrammes de la Fig. 6.2 représentent la décomposition des homographies H_g^n et H_d^n estimées à partir de la séquence du port d'Antibes. Nous rappelons que sur cette séquence, la géométrie des caméras est inconnue, nous avons été contraint d'introduire une matrice générique de paramètres intrinsèques K_0 que nous supposons identique pour les deux caméras ($K_g = K_d = K_0$).

De manière générale, les déplacements estimés à partir de la décomposition des homographies estimées par chacune des caméras sont cohérents avec la trajectoire du véhicule. Le mouvement de translation normalisé s'effectue principalement selon l'axe optique, proche de la direction de déplacement du véhicule. Le véhicule ralentit, s'arrête quasiment, puis repart pour circuler à vitesse constante. Le mouvement de rotation mesuré est, lui, essentiellement porté par l'axe des ordonnées, proche de la verticale. Nous retrouvons chacune des étapes qui jalonnent le déplacement :

- la décélération en ligne droite jusqu'à l'image 850,
- le déportement sur la droite à très faible vitesse jusqu'à l'image 892,
- le déboîtement vers la voie de gauche pour contourner le véhicule utilitaire jusqu'à l'image 922,
- le rabattement vers la voie de droite jusqu'à l'image 968,
- le réalignement du véhicule sur sa voie jusqu'à l'image 1004,
- le recentrage du véhicule sur sa voie jusqu'à l'image 1040,
- la poursuite du trajet en ligne droite à vitesse constante.

Les erreurs d'estimation des mouvements de translation et rotation, observées entre les images [936;948], sont essentiellement dues à une distribution linéaire des primitives coplanaires détectées. La région de la route observée dans ces images se résume en effet à l'extrémité de la voie de gauche et son bas coté, soit une bande plane d'environ 4m de large centrée autour de la bande principale gauche. La ROI, représentant l'espace navigable, ne tient pas compte de la région extérieure à la bande principale. Les FPs extraits sont pour une moitié détectés sur la bande blanche peinte au sol tandis que l'autre moitié correspond aux coins des plots qui délimitent la région de la route réservée à la circulation des véhicules. Ce deuxième ensemble de FPs se situe généralement au-dessus du plan de la route. Le système formé par les primitives mises en correspondance n'étant pas observable, l'estimation de l'homographie est incorrecte.

Un second passage difficile correspond au moment où le véhicule test s'apprête à croiser le deuxième véhicule en stationnement. Les FPs détectés sur la route sont principalement localisés

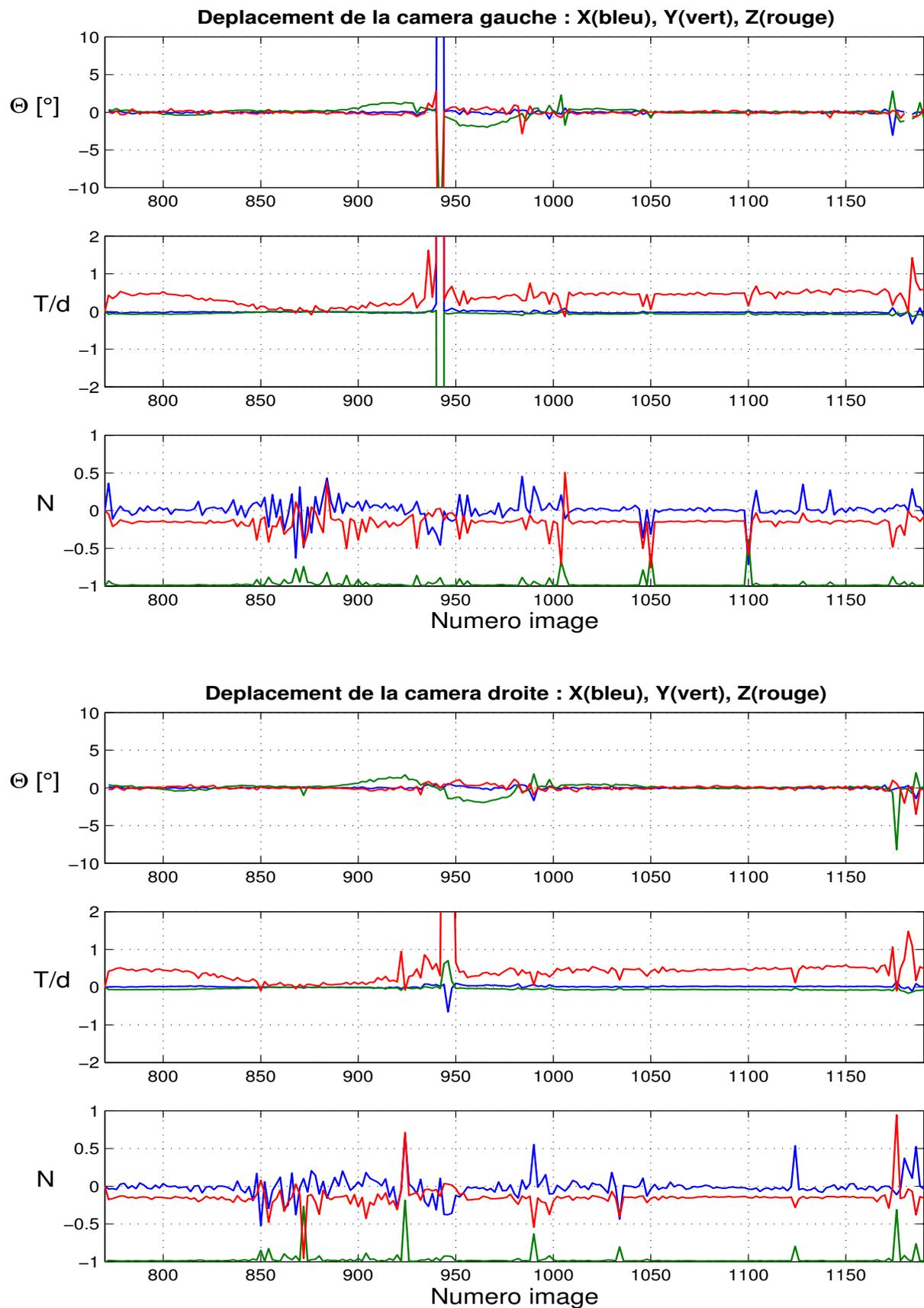


FIG. 6.2 – Estimation des mouvements de translation et rotation de chacune des caméras entre chaque pose. La normale au plan intervenant dans la composition de l'homographie, son orientation discontinue montre que certaines estimations des homographies ne sont pas cohérentes.

autour des pointillés qui forment la bande principale centrale. La projection de la bande principale centrale se trouvant proche des axes optiques des 2 caméras, l'orientation des pointillés dépend alors de la position relative de la caméra par rapport à la bande. Les FPs détectés aux extrémités saillantes de chaque parallélogramme sont complémentaires et ne peuvent donc pas être mis en correspondance (voir Fig. 6.3).

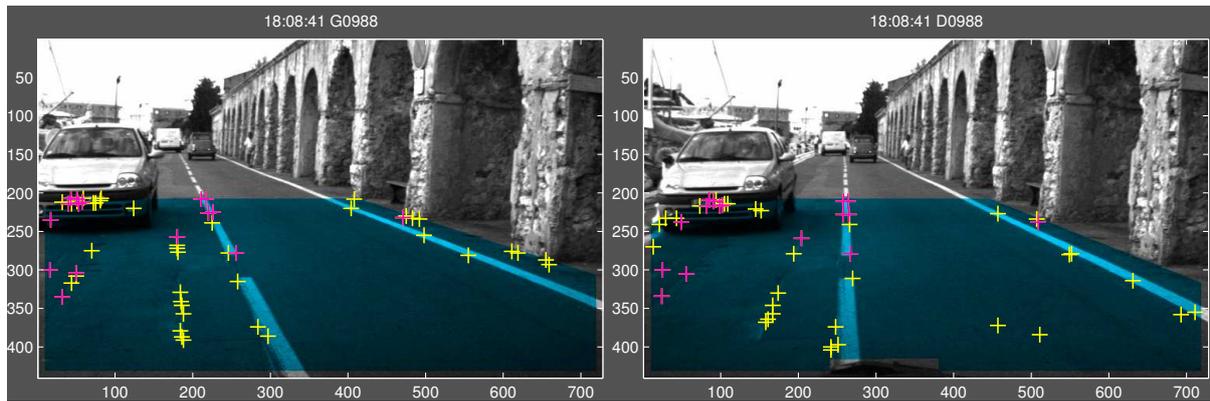


FIG. 6.3 – Les FPs détectés aux coins saillants des parallélogrammes formés par le marquage horizontal sont différents suivant la position relative de la caméra par rapport à la bande principale centrale. Le calcul de l'homographie entre ces deux images est alors difficile vu le peu de primitives coplanaires pouvant être mises en correspondance.

Les discontinuités mesurées lors du mouvement de translation du véhicule aux alentours des images $\{1000; 1050; 1100\}$ sont de nouveau dues à une concentration non-uniforme des FPs détectés dans la ROI. L'abaissement du seuil de Harris dans les régions les moins texturées ne permet pas d'extraire des FPs, même à faibles scores. Ainsi, la région de la route autour de la bande principale droite présente une asphalté très homogène où aucun FP n'est détecté. La majorité des FPs extraits et mis en correspondance se situent sur les coins des pointillés qui forment la bande principale centrale et sur le second véhicule en stationnement. Le peu de primitives coplanaires détectées ne permet en effet pas de rejeter les FPs situés sur le véhicule correctement mis en correspondance.

Enfin, les fluctuations importantes des composantes de la normale au plan \mathbf{N} estimées entre les images $[850; 900]$ sont dues au faible déplacement du véhicule. En effet, la composante verticale maximale du mouvement apparent des nombreux FPs équi-répartis de manière uniforme dans la ROI durant cette période est inférieur à 10 pixels (voir Fig. 5.15). Cela signifie que les homographies sont calculées à partir de couples de FPs dont l'amplitude des mouvements apparents sont de l'ordre de la précision des FPs extraits. Plusieurs familles de déplacements (translations, rotations et normale au plan) sont susceptibles d'être solutions, ce qui justifie la grande variabilité de la normale au plan.

Nous disposons toutefois d'une autre estimation de la décomposition de la normale, exprimée dans le même repère, grâce à la décomposition de l'homographie stéréo H_{st}^n . Nous pouvons vérifier grâce à la Fig. 4.25 que la décomposition de la normale est durant cet intervalle beaucoup plus précise entre les deux images de la paire stéréo du fait que la distance qui sépare les deux caméras est plus importante que le déplacement du véhicule entre deux poses.

6.3 Définition et propriété de la super-homographie

Dans la section précédente, nous avons considéré chaque homographie H_{st}^n, H_g^n et H_d^n de façon indépendante. Il est clair que cette manière de procéder n'exploite pas au mieux la spécificité de notre application, à savoir, une paire stéréo rigide en mouvement.

Notre objectif va être désormais d'estimer l'ensemble des homographies liées aux déplacements de la paire stéréo en un unique calcul. Nous allons montrer que la prise en compte des positions des primitives coplanaires observées dans différentes vues permet d'améliorer la robustesse et la qualité des estimations en utilisant la cohérence des mouvements relatifs des primitives entre images. Pour cela, nous allons utiliser la notion de super-homographie introduite par Malis et Cipolla [Malis and Cipolla, 2000]. La super-matrice d'homographies permet une formulation implicite des contraintes qui lient les homographies de trois vues différentes $\mathcal{I}_a, \mathcal{I}_b$ et \mathcal{I}_c d'une même scène coplanaire :

$$H_{ca} \propto H_{cb} \cdot H_{ba}. \quad (6.17)$$

A notre connaissance, nous sommes les premiers à mettre en oeuvre cette méthode pour ce type d'application.

La prise en compte des connaissances a priori de la scène n'est exploitée actuellement que lors de la phase de mise en correspondance des primitives que l'algorithme parvient à détecter et que nous espérons coplanaires. L'algorithme utilisé introduit jusqu'alors des contraintes spatio-temporelles seulement entre deux couples d'images. De plus, la qualité des estimations des homographies de l'itération courante est directement liée à la vérification de la relation $H_{st}^n \propto H_d^{n-1} \cdot H_{st}^{n-1} \cdot H_g^n$ qui dépend elle-même de la qualité de l'homographie stéréo à l'itération précédente.

La méthode que nous allons détailler permet d'estimer les homographies croisées entre un ou plusieurs plans observés à partir de plusieurs poses et/ou caméras. Dans notre cas, nous nous restreignons, dans un premier temps, à l'observation du plan de la route à partir de positions successives de la paire stéréo au cours du temps. La prise en compte de plusieurs couples d'images a l'intérêt de minimiser l'incertitude due aux erreurs d'extraction des primitives dans les images et de réduire l'instabilité numérique lorsque les déplacements entre deux vues sont faibles.

6.3.1 Propriétés de la matrice de super-homographie

6.3.1.1 Rappel

Soient m le nombre d'images disponibles d'une scène planaire. La matrice de super-homographie $\mathcal{H} [3m, 3m]$ contient toutes les homographies $[3,3]$ croisées qui lient les $m(m-1)$ couples d'images pouvant être formés. Les matrices de la diagonale principale \mathcal{H}_{ii} représentent l'identité I_3 alors que les matrices du bloc triangulaire supérieur \mathcal{H}_{ij} avec $i \geq j$ représentent les homographies inverses de celles du bloc triangulaire inférieur :

$$\mathcal{H} = \begin{bmatrix} \mathcal{H}_{11} & \cdots & \mathcal{H}_{1m} \\ \vdots & \ddots & \vdots \\ \mathcal{H}_{m1} & \cdots & \mathcal{H}_{mm} \end{bmatrix} = \begin{bmatrix} I_3 & \cdots & H_{m1}^{-1} \\ \vdots & \ddots & \vdots \\ H_{m1} & \cdots & I_3 \end{bmatrix} \quad (6.18)$$

La transposition de la contrainte Equ. 6.17 permet de développer l'élevation au carré de la matrice \mathcal{H} . Ainsi chaque élément de la matrice vérifie :

$$(\mathcal{H}_{ji})^2 = \sum_{r=1}^m \mathcal{H}_{jr} \cdot \mathcal{H}_{ri}$$

$$(6.19)$$

Les contraintes spatio-temporelles entre les homographies estimées entre les m vues seront donc respectées si et seulement si :

$$(\mathcal{H})^2 = m.\mathcal{H} \quad (6.20)$$

La diagonale principale de la matrice \mathcal{H} étant formée de matrices identités \mathbf{l}_3 , son rang est supérieur ou égal à 3. A l'opposé, chaque ligne matricielle (groupe de trois lignes consécutives) de \mathcal{H} est une combinaison linéaire de trois autres, ce qui implique que le rang de \mathcal{H} ne peut dépasser 3 :

$$\text{rang}(\mathcal{H}) = 3 \quad \forall m \geq 3 \quad (6.21)$$

6.3.1.2 Super-primitives

Nous associons à la matrice de super-homographie des super-primitives qui ne sont autres que le regroupement sous forme de vecteur $[3m, 1]$ des coordonnées homogènes des primitives observées dans les m images. Nous introduisons alors deux nouveaux types d'éléments les super-points d'intérêt (SFPs) et les super-lignes de fuite (SVLs). En généralisant, la relation entre l'égalité projective de deux projections du k^{eme} FP et de la s^{eme} VL dans les images \mathcal{I}_i et \mathcal{I}_j (conformément à 4.32), nous obtenons :

$$\mathbf{P}_k^j \propto \mathcal{H}_{ji}.\mathbf{P}_k^i \quad \text{et} \quad \mathbf{L}_s^j \propto \mathcal{H}_{ji}^{-t}.\mathbf{L}_s^i \quad (6.22)$$

Les SFPs et SVLs vérifiant des systèmes linéaires transposés inverses, nous traiterons dans la suite de cette section que du cas des SFPs sachant qu'une réécriture des coordonnées qui forment les SVLs permet d'exploiter le système linéaire \mathcal{H} vérifié par les SFPs.

La matrice \mathcal{H} étant de rang 3, cela signifie que parmi ces $3m$ valeurs propres, seules trois sont non-nulles ($\lambda_1 = \lambda_2 = \lambda_3 = m$). La généralisation de l'égalité projective précédente permet alors d'écrire pour les représentations du k^{eme} SFP $\mathbf{P}_k = [\mathbf{p}_k^{1t}, \mathbf{p}_k^{2t}, \dots, \mathbf{p}_k^{mt}]^t$ dans les m images :

$$\Gamma_k \mathbf{P}_k = \mathcal{H}.\mathbf{P}_k \quad (6.23)$$

où $\Gamma_k = \text{diag}(\gamma_1.\mathbf{l}_3, \gamma_2.\mathbf{l}_3, \dots, \gamma_m.\mathbf{l}_3)$ est une matrice diagonale dont les γ_k représentent les facteurs de mise à l'unité des 3^{eme} coordonnées homogènes de chacun des vecteurs \mathbf{p}_k^r avec $r \in [1; m]$ qui constituent le SFP \mathbf{P}_k . La multiplication de l'égalité précédente par \mathcal{H} conduit à l'égalité suivante :

$$\mathcal{H}.\Gamma_k.\mathbf{P}_k = (\mathcal{H})^2.\mathbf{P}_k \quad (6.24)$$

$$= m.\mathcal{H}.\mathbf{P}_k \quad (6.25)$$

$$= m.\Gamma_k.\mathbf{P}_k \quad (6.26)$$

soit en posant $\widetilde{\mathbf{P}}_k = \Gamma_k.\mathbf{P}_k$, la réécriture de l'égalité précédente signifie que $\widetilde{\mathbf{P}}_k$ est un vecteur propre de la matrice \mathcal{H} associé à la valeur propre $\lambda = m$:

$$\mathcal{H}.\widetilde{\mathbf{P}}_k = m.\widetilde{\mathbf{P}}_k \quad (6.27)$$

La matrice \mathcal{H} étant diagonalisable (son rang étant égal à la dimension de l'espace engendré), il existe toujours une matrice de passage \mathbf{Q} telle que : $\mathcal{H} = \mathbf{Q}^{-1}.\mathbf{D}.\mathbf{Q}$ avec $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3m})$. La matrice \mathbf{Q} étant non-singulière, les vecteurs propres qui constituent ses colonnes sont linéairement indépendants les uns des autres et forment une base de \mathcal{R}^{3m} .

Les SFPs peuvent être représentés par une combinaison linéaire de la base formée par les vecteurs propres $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3$ associés à la valeur propre triple $\lambda = m$:

$$\widetilde{\mathbf{P}}_{\mathbf{k}} = \alpha_1 \cdot \mathbf{Q}_{\lambda_1} + \alpha_2 \cdot \mathbf{Q}_{\lambda_2} + \alpha_3 \cdot \mathbf{Q}_{\lambda_3} \quad (6.28)$$

La relation précédente admet une interprétation géométrique : soit trois points non-colinéaires situés sur un plan, ils forment une base à partir de laquelle tout point coplanaire peut être exprimé par une combinaison linéaire. L'homographie étant une transformation affine, la composition de vecteurs est conservée. Ainsi la combinaison linéaire qui lie un point coplanaire à sa base est invariante : les coordonnées des projections d'un point coplanaire dans les m vues vérifient la même combinaison linéaire des projections des vecteurs de base.

6.3.1.3 Décomposition de la matrice de super-homographie

De l'espace projectif à l'espace euclidien Chacune des homographies H_{ji} qui forment la matrice \mathcal{H} est connue à un facteur d'échelle près dans l'espace projectif, les coordonnées homogènes des primitives étant normalisées dans chacune des images \mathcal{I}_i et \mathcal{I}_j considérées. Or, l'égalité projective qui lie les coordonnées homogènes des primitives en correspondance via \mathcal{H} devient une égalité au sens strict lorsque ces mêmes primitives sont exprimées dans l'espace euclidien. L'homographie euclidienne à l'origine de l'homographie exprimée dans l'espace projectif est égale à :

$$H_{eji} = K_j^{-1} \cdot H_{ji} \cdot K_i \quad (6.29)$$

où K_i, K_j représentent respectivement les paramètres intrinsèques des caméras en positions \mathbf{C}_i et \mathbf{C}_j .

La matrice \mathcal{H} est estimée grâce aux coordonnées des super-primitives, normalisées via la matrice Γ_k (6.23). L'expression des homographies H_{eji} à partir de leur duales exprimées dans l'espace projectif exige donc d'identifier chacun des facteurs d'échelle γ_k de l'ensemble des homographies qui constituent \mathcal{H} . Les trois valeurs propres d'une homographie normalisée sont telles que $\lambda_1 \geq \lambda_2 > \lambda_3$ avec $\lambda_2 = 1$. Il est par conséquent possible de déterminer le facteur d'échelle de toute homographie exprimée dans l'espace euclidien : il s'agit de la seconde valeur propre lorsque ces dernières sont rangées par ordre (dé-)croissant.

Estimation de la normale au plan Chacune des homographies euclidiennes qui forment \mathcal{H}_e contient la transformation rigide qui permet de passer du repère lié à la caméra \mathbf{C}_i à celui de la caméra \mathbf{C}_j . La décomposition de l'homographie en mouvement de rotation \mathbf{R}_{ji} et translation \mathbf{t}_{ji} nécessite la connaissance de la représentation de la normale au plan \mathbf{n}_i qui induit l'homographie et de la distance d_i qui le sépare de \mathbf{C}_i :

$$H_{eji} = \mathbf{R}_{ji} + \frac{\mathbf{t}_{ji} \cdot \mathbf{n}_i^t}{d_i} \quad (6.30)$$

A partir de la décomposition précédente, nous pouvons vérifier l'égalité suivante quel que soit $k > 0$:

$$[\mathbf{n}_j]_x^k \cdot H_{ij}^t = H_{ji} \cdot [\mathbf{n}_i]_x^k \quad (6.31)$$

où $[\mathbf{n}_j]_x \cdot H_{ij}^t = [\mathbf{n}_j]_x \cdot \mathbf{R}_{ji}$ est comparable à une matrice essentielle ($[\mathbf{t}]_x \cdot \mathbf{R}$). La matrice ainsi formée a donc une valeur propre double et une troisième nulle ce qui permet d'introduire deux contraintes fortes sur le produit ainsi formé.

Sachant que pour tout vecteur \mathbf{v} et toute matrice A , $\det(A) \cdot A \cdot [\mathbf{v}]_x \cdot A^t = [A^{-t}\mathbf{v}]_x$, l'Equ. 6.31 peut se mettre sous la forme suivante dans le cas particulier où $k = 1$:

$$[\mathbf{n}_j]_x^k \cdot H_{ij}^t = H_{ji} \cdot [\mathbf{n}_i]_x^k \cdot H_{ij}^t \quad (6.32)$$

soit,

$$\mathbf{n}_j = B_{ji} \cdot \mathbf{n}_i \quad (6.33)$$

avec $B_{ji} = \det(H_{ji}) \cdot H_{ji}^{-t}$. Il est donc possible d'estimer l'orientation de la normale au plan observé à partir de la pose \mathbf{C}_i en connaissant celle observée en \mathbf{C}_j .

Estimation des mouvements de rotation et translation Dans l'éventualité où les paramètres intrinsèques des caméras sont connus, il est alors envisageable d'estimer une matrice de super-homographie $\mathcal{H}e$ exprimée dans l'espace euclidien par analogie avec l'expression 6.29. Il nous faut pour cela définir une super-matrice de paramètres intrinsèques telle que $\mathcal{K} = \text{diag}(K_1, K_2, \dots, K_m)$. Nous pouvons alors estimer une matrice de super-homographie dans l'espace euclidien :

$$\mathcal{H}e = \mathcal{K}^{-1} \cdot \mathcal{H} \cdot \mathcal{K} \quad (6.34)$$

Cette matrice de super-homographie exprimée dans l'espace euclidien est représentative des déplacements rigides entre les poses des diverses caméras. Chaque déplacement correspond au mouvement de translation et de rotation entre deux caméras, compte-tenu de la distance qui sépare la caméra qui sert de référence du plan observé. Ainsi, la super-homographie $\mathcal{H}e$ est représentative de tous les déplacements entre les m poses de la (des) caméra(s) et se décompose en :

$$\mathcal{H}e = \mathcal{R} + T \cdot \mathcal{N}^t \quad (6.35)$$

avec :

$$\mathcal{R} = \begin{bmatrix} R_{11} & \cdots & R_{1m} \\ \vdots & \ddots & \vdots \\ R_{m1} & \cdots & R_{mm} \end{bmatrix}, \quad T = \begin{bmatrix} \frac{t_{11}}{d_1} & \cdots & \frac{t_{1m}}{d_m} \\ \vdots & \ddots & \vdots \\ \frac{t_{m1}}{d_1} & \cdots & \frac{t_{mm}}{d_m} \end{bmatrix}, \quad \mathcal{N} = \begin{bmatrix} \mathbf{n}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{n}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{n}_3 \end{bmatrix} \quad (6.36)$$

où $\dim(\mathcal{R}) = (3m; 3m)$, $\text{rang}(\mathcal{R}) = 3$ et $\dim(T) = \dim(\mathcal{N}) = (3m; m)$. La matrice \mathcal{R} est symétrique et vérifie la relation 6.20, ses valeurs propres sont identiques à celles de \mathcal{H} .

6.3.1.4 Estimation itérative de la super-homographie

Considérons dans un premier temps que la matrice de super-homographie \mathcal{H} soit connue. Sa diagonalisation fournit une matrice de passage \mathbf{Q} dont les colonnes \mathbf{Q}_r représentent les vecteurs propres λ_r associés aux valeurs propres de la matrice \mathcal{H} :

$$\mathcal{H} = \mathbf{Q}^{-1} \cdot \mathbf{D} \cdot \mathbf{Q} \quad (6.37)$$

avec $\mathbf{D} = \text{diag}(m, m, m, 0, \dots, 0)$. La matrice \mathbf{Q} étant non-singulière, les vecteurs propres forment une base de \mathcal{R}^{3m} . Soient 0P_k , les coordonnées du k^{eme} SFP, initiées en concaténant les coordonnées \mathbf{p}_k dans les m vues. Ce vecteur de \mathcal{R}^{3m} , qui n'est autre qu'une représentation bruitée de $\widetilde{\mathbf{P}}_k$, peut alors être décomposé dans la base \mathbf{Q} :

$${}^0\mathbf{P}_k = \alpha_1 \mathbf{Q}_1 + \alpha_2 \mathbf{Q}_2 + \alpha_3 \mathbf{Q}_3 + \cdots + \alpha_{3m} \mathbf{Q}_{3m} \quad (6.38)$$

La base des super-primitives solutions étant constituée des trois vecteurs propres associés à $\lambda = m$, nous devons calculer les coordonnées théoriques de chacune des super-primitives de façon à supprimer les composantes selon les $(3m - 3)$ autres vecteurs propres de la combinaison linéaire précédente. Nous appliquons alors la relation 6.27 à 6.38 :

$$\begin{aligned} {}^1\mathbf{P}_k &= \frac{1}{m} \mathcal{H} \cdot {}^0\mathbf{P}_k \\ &= \frac{1}{m} (\alpha_1 \mathcal{H} \cdot \mathbf{Q}_1 + \alpha_2 \mathcal{H} \cdot \mathbf{Q}_2 + \alpha_3 \mathcal{H} \cdot \mathbf{Q}_3 + \cdots + \alpha_{3m} \mathcal{H} \cdot \mathbf{Q}_{3m}) \\ &= \frac{\lambda_1}{m} (\alpha_1 \mathbf{Q}_1 + \alpha_2 \frac{\lambda_2}{\lambda_1} \mathbf{Q}_2 + \alpha_3 \frac{\lambda_3}{\lambda_1} \mathbf{Q}_3 + \cdots + \alpha_{3m} \frac{\lambda_{3m}}{\lambda_1} \mathbf{Q}_{3m}) \end{aligned} \quad (6.39)$$

L'estimée des coordonnées de la super-primitive après w itérations est alors égale à :

$${}^w\mathbf{P}_k = \left(\frac{\lambda_1}{m}\right)^w \left(\alpha_1 \mathbf{Q}_1 + \alpha_2 \left(\frac{\lambda_2}{\lambda_1}\right)^w \mathbf{Q}_2 + \alpha_3 \left(\frac{\lambda_3}{\lambda_1}\right)^w \mathbf{Q}_3 + \cdots + \alpha_{3m} \left(\frac{\lambda_{3m}}{\lambda_1}\right)^w \mathbf{Q}_{3m} \right) \quad (6.40)$$

Nous retrouvons ainsi la relation 6.28 lorsque nous nous replaçons dans le cas idéal où seules les trois premières valeurs propres sont non-nulles et égales à m . Les coordonnées estimées de ${}^w\mathbf{P}_k$ sont obtenues par un moyennage de celles obtenues à l'itération précédente. Conformément à [Robert, 1993], le bruit de mesure relevé sur les coordonnées des diverses primitives détectées est considéré comme gaussien, l'opération de moyennage revient donc à éliminer les composantes bruitées des coordonnées de chaque super-primitive.

Nous nous replaçons désormais dans le cadre général où la matrice \mathcal{H} est inconnue. L'estimée initiale ${}^0\mathcal{H}$, obtenue à partir des super-primitives ${}^0\mathbf{P}_k$, est imparfaite dans le sens où ses valeurs propres ne correspondent pas exactement à celles de \mathcal{H} . Les vecteurs propres vérifient néanmoins ${}^w\mathcal{H} \cdot {}^w\mathbf{Q}_r = {}^w\lambda_r \cdot {}^w\mathbf{Q}_r$. A chaque nouvelle itération w , l'estimation ${}^w\mathcal{H}$ est affinée ce qui engendre une nouvelle matrice de passage ${}^w\mathbf{Q}$. L'expression 6.40 est alors modifiée de la sorte :

$${}^w\mathbf{P}_k = \frac{\prod_{r=1}^w \lambda_1}{m^w} \left(\alpha_1 {}^w\mathbf{Q}_1 + \alpha_2 \prod_{r=1}^w \left(\frac{\lambda_2}{\lambda_1}\right)^r {}^w\mathbf{Q}_2 + \alpha_3 \prod_{r=1}^w \left(\frac{\lambda_3}{\lambda_1}\right)^r {}^w\mathbf{Q}_3 + \cdots + \alpha_{3m} \prod_{r=1}^w \left(\frac{\lambda_{3m}}{\lambda_1}\right)^r {}^w\mathbf{Q}_{3m} \right) \quad (6.41)$$

L'estimation de la super-homographie ${}^w\mathcal{H}$ consiste donc pour chaque itération w à l'établissement alternatif des super-primitives ${}^w\mathbf{P}_k^n$ à partir des projections \mathbf{p}_k des primitives en correspondance et à l'estimation à proprement parler de la super-homographie ${}^w\mathcal{H}$ correspondante. Lorsque le processus se termine, les coordonnées des super-primitives sont toutes liées entre elles, le $\text{rang}({}^w\mathcal{H}) = 3$.

La convergence de l'algorithme n'est pour l'instant pas démontrée. Les auteurs parviennent néanmoins à observer une convergence à l'aide de simulations numériques où la variance du bruit introduit atteignait $\sigma^2 = 1000$ pixels. Il va de soi que la rapidité de convergence de l'algorithme dépend de la qualité de l'estimation initiale ${}^0\mathcal{H}$.

6.3.2 Implémentation de la méthode

L'estimation du calcul de la super-homographie dans notre cas nécessite une adaptation de la méthode proposée puisque nous disposons à chaque itération de deux images contraintes par la géométrie épipolaire du banc stéréo. De même, les faisceaux de VLs détectées sur le plan de la route n'est pas introduit de manière directe dans le calcul car le système linéaire vérifié

par les VLs est différent de celui vérifié par les FPs. C'est pourquoi, nous avons été contraint d'adapter le calcul de la super-homographie à notre configuration.

6.3.2.1 Adaptation à notre configuration

L'emploi d'une paire stéréo modifie essentiellement la notation des super-primitives puisque nous traitons ici le cas de paires d'images. Ainsi, afin de respecter la convention précédente, les m vues sont désormais réparties en $m/2$ couples d'images stéréo. L'estimation du calcul de \mathcal{H}^n à l'itération courante n s'effectue donc à partir des projections des super-primitives détectées durant les $\frac{m}{2}$ dernières itérations :

$$\mathcal{H}^n = \begin{bmatrix} \mathbf{l}_3 & (\mathbf{H}_{\text{st}}^{n-\frac{m}{2}+1})^{-1} & \dots \\ \mathbf{H}_{\text{st}}^{n-\frac{m}{2}+1} & \mathbf{l}_3 & \dots \\ \mathbf{H}_{\text{g}}^{n-\frac{m}{2}+1} & \dots & \mathbf{l}_3 & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \ddots & \dots & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & \dots & \dots & \dots & \ddots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \mathbf{H}_{\text{g}}^n & \dots & \mathbf{l}_3 & (\mathbf{H}_{\text{st}}^n)^{-1} & \dots \\ \dots & \dots & \dots & \dots & \dots & \mathbf{H}_{\text{d}}^n & \mathbf{H}_{\text{st}}^n & \mathbf{l}_3 & \dots \end{bmatrix} \quad \text{avec} \quad \mathbf{P}_{\mathbf{k}}^n = \begin{bmatrix} \mathbf{p}_{\text{gk}}^{n-\frac{m}{2}+1} \\ \mathbf{p}_{\text{dk}}^{n-\frac{m}{2}+1} \\ \mathbf{p}_{\text{gk}}^{n-\frac{m}{2}+2} \\ \vdots \\ \vdots \\ \mathbf{p}_{\text{gk}}^{n-1} \\ \mathbf{p}_{\text{dk}}^{n-1} \\ \mathbf{p}_{\text{gk}}^n \\ \mathbf{p}_{\text{dk}}^n \end{bmatrix} \quad (6.42)$$

L'emploi de la super-homographie permet de prendre en compte des super-primitives dont toutes les projections ne sont pas connues dans les m vues. La première estimation ${}^0\mathcal{H}^n$ est en effet construite à partir de primitives coplanaires en correspondance dans des couples d'images. Les primitives pour lesquelles la projection est inconnue dans une image ne sont donc pas utilisées. L'estimation de ${}^0\mathcal{H}^n$ réalisée, le calcul des coordonnées des super-primitives ${}^1\mathbf{P}_{\mathbf{k}}$ est alors mené en prenant garde d'initialiser les r coordonnées $\mathbf{p}_{\mathbf{k}}^r$ des projections inconnues de ${}^0\mathbf{P}_{\mathbf{k}}^n$ par des zéros ; l'expression 6.40 devient alors :

$${}^1\mathbf{P}_{\mathbf{k}}^n = \frac{1}{m^*} \cdot {}^0\mathcal{H}^n \cdot {}^0\mathbf{P}_{\mathbf{k}}^n \quad (6.43)$$

avec $m^* \leq m$ nombre de projections mises en correspondance dans les différentes vues. La sélection des super-primitives retenues est réalisée en fixant $m^* \geq m/2$ de façon à conserver suffisamment de contraintes entre les projections des différentes vues et entreprendre l'estimation de la super-homographie à partir d'un nombre suffisant de super-primitives.

6.3.2.2 Prise en compte des lignes de fuite

Dans cette partie, la terminologie lignes de fuite s'applique au faisceau de VLs mis en correspondance au cours des estimations des homographies \mathbf{H}_{st}^n , \mathbf{H}_{g}^n et \mathbf{H}_{d}^n , soit les projections des bandes principales et plus précisément leurs médianes dont les caractéristiques sont filtrées.

La représentation du faisceau de droites dans plusieurs images Le nombre de SFPs existants dont les projections dans les images courantes commun à m images est réduit en comparaison des temporaires dont la coplanarité des projections qui les constituent n'a pas encore été vérifiée. L'introduction des SVLs dans le calcul de super-homographie s'avère donc

nécessaire de façon à augmenter la proportion de super-primitives coplanaires mais aussi à contraindre le mouvement du plan de la route dans des régions de l'image où peu ou pas de FPs sont mis en correspondance.

L'introduction directe des coordonnées des SVLs complexifie la résolution du système linéaire initialement vérifié par les coordonnées des SFPs. Afin de s'affranchir de problèmes délicats liés au conditionnement d'un système hybride, formé par des points et des droites, nous proposons de substituer aux faisceaux de VLs en correspondance leurs intersections avec une droite sécante coplanaire dont les coordonnées sont connues dans les m images. La probabilité qu'une telle droite existe dans l'image à chaque itération est minime voire nulle. C'est pourquoi, nous proposons de créer une droite virtuelle permettant de prendre en compte dans le calcul de la super-homographie l'information contenue dans les VLs par le biais de points d'intérêt virtuels.

Les points d'intérêt virtuels La droite virtuelle qui intercepte le faisceau de SVLs pour créer les points d'intérêt virtuels (*Virtual Feature Points*) est construite à partir d'un couple de SFPs dont les projections sont connues dans les m images. Bien que le choix parmi les SFPs soit restreint, il apparaît néanmoins délicat de déterminer quelle est la meilleure droite virtuelle qui permettra de localiser des VFPs là où la route manque de texture.

Apport de l'introduction d'une droite virtuelle Une solution consiste à rechercher le couple de SFPs permettant de construire une droite virtuelle proche de l'horizontale, située dans la partie basse de la ROI. Les VFPs ainsi créés auront des mouvements relatifs parmi les plus importants et seront répartis sur la largeur de la ROI. Un exemple de l'intérêt d'introduire une telle droite virtuelle horizontale est illustré en Fig. 6.4 où les images présentées sont composées par le warping de l'image gauche courante \mathcal{I}_g^n dans l'image droite courante \mathcal{I}_d^n via l'homographie stéréo H_{st}^n , extraite de \mathcal{H}^n . La prise en compte des VFPs dans la représentation permet de considérablement améliorer la projection du plan de la route de l'image gauche vers l'image droite. En effet, nous pouvons remarquer qu'aucun SFP n'est extrait le long de la bande principale droite représentée par une bande blanche continue. Les contraintes de projection sont toutes localisées sur la moitié gauche de la route. Sans l'introduction des VFPs, il est normal que la bande principale droite soit mal projetée.

D'autre part, nous avons reporté sur les images résultats les coordonnées des primitives de \mathcal{I}_d^n extraites ('.' cyans et '+' rouges foncés) et théoriques obtenues grâce à l'estimation de \mathcal{H}^n ('o' bleus, 'o' violets). Les écarts statistiques mesurés entre les coordonnées extraites et théoriques des SFPs varient peu suivant la prise en compte ou pas des VFPs : la moyenne et la variance sont respectivement de 1,032 et 1,056 pixels sans l'introduction des VFPs et 1,068 et 0,834 pixels avec.

Cette comparaison montre que caractériser la justesse d'une estimation d'homographie par les erreurs de projection statistiques entre primitives est trompeur. L'homographie estimée grâce à l'introduction des VFPs est nettement meilleure que celle obtenue sans, bien que les erreurs de projection des primitives soit comparable. L'introduction des VFPs accroît l'uniformité de la répartition des super-primitives. Il apparaît donc plus opportun de qualifier la justesse d'une homographie en mesurant l'erreur de projection de la région planaire dans sa globalité plutôt que d'utiliser des primitives non-uniformément réparties.

Introduction de plusieurs droites virtuelles La recherche d'une droite virtuelle horizontale située dans la partie basse de la ROI s'avère infructueuse dans la majorité des cas. Dès lors, la recherche de plusieurs droites virtuelles à partir des SFPs extraits permet d'introduire

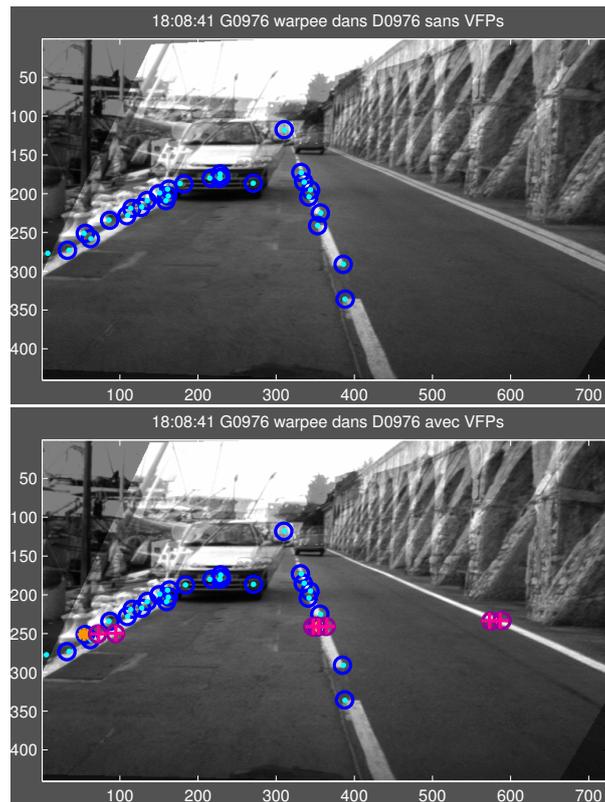


FIG. 6.4 – Comparaison de la qualité de l’homographie stéréo H_{st}^n extraite de \mathcal{H}^n , estimée à partir des seuls FPs en commun aux m images (‘.’ cyans) ou en tenant compte des VFPs (‘+’ rouges foncés). L’introduction des VFPs permet de mieux caractériser la région apparentée à la route : la projection du plan de la route de l’image gauche dans l’image droite est plus fidèle autour de la bande principale droite. Les coordonnées théoriques des FPs (‘o’ bleus) et des VFPs (‘o’ violets) sont aussi proches dans les deux cas des coordonnées extraites.

s'impose car elles devraient permettre d'introduire des VFPs dans chaque région de la ROI sans texture ou cachée par un obstacle. Nous pouvons alors contraindre le mouvement apparent de la projection du plan observé à souhait car tout couple de points formé par l'association de SFP(s) et/ou de VFP(s) sera à l'origine d'une nouvelle droite virtuelle dont les coordonnées seront connues dans les m images.

La contre-partie de l'emploi de plusieurs droites virtuelles est la détermination du jeu optimal de droites virtuelles nécessaire pour rendre le système linéaire sur-conditionné. Deux options sont envisageables :

- la première s'inspire des travaux de [Aufrière et al., 2001] que nous avons développé en 2.3.2.2. Chaque bande principale est alors tronçonnée en région de hauteur qui augmente avec un pas logarithmique avec l'ordonnée. Dans l'hypothèse où aucun FP n'est détecté dans un tronçon, il nous faut déterminer si il existe le couple de SFPs dont la droite virtuelle interceptera la SVL qui passe à proximité, entre les limites haute et basse du tronçon considéré,
- la seconde consiste à déterminer un jeu de droites virtuelles minimum qui permettra de localiser des VFPs le long des bandes principales.

La première option correspond au cas idéal vers lequel nous devons tendre mais la recherche des couples de FPs risque de nécessiter la mise en oeuvre de plusieurs droites virtuelles de manière à satisfaire toutes les contraintes souhaitées. La deuxième option, bien que rudimentaire permet l'introduction de VFPs qui contraignent le mouvement apparent du plan considéré à partir d'un nombre limité de droites virtuelles.

Les médianes des bandes principales forment un jeu de SVLs réduit mais fiable car leurs caractéristiques ont l'intérêt d'être filtrées. Elles délimitent de surcroît la projection du plan de la route à l'image. La recherche de VFPs sur ce jeu de droites particulier contraint donc de manière significative le mouvement apparent du plan de la route.

La précision des coordonnées des VFPs dépend étroitement des couples de SFPs sélectionnés pour initier les droites virtuelles. Il va de soi que l'incertitude sur les coordonnées des droites virtuelles est inversement proportionnelle à la distance qui sépare les couples de SFPs. Par contre, si l'intersection d'une droite virtuelle avec une SVL a lieu en dehors du cadre de l'image, la fiabilité de ses coordonnées sera relative. Or, dans le cas où la distribution spatiale des super-primitives ne contraint pas suffisamment le déplacement de la projection du plan, prendre en compte des VFPs mal localisés ne peut que fausser l'estimation de \mathcal{H} . C'est pourquoi nous ne tenons compte des seuls VFPs dont les coordonnées se situent dans les limites de la ROI en tolérant cependant que les abscisses soient inférieures à un dépassement correspondant au quart de la largeur de l'image.

Trois types de distributions spatiales de SFPs posent problème : il s'agit de cas où les SFPs sont soit alignés le long d'un contour (Fig. 6.5(a)), soit situés dans la partie haute de la ROI (Fig. 6.5(b)) ou en présence d'obstacles (Fig. 6.5(c)). Dans les trois cas, une proportion importante de la ROI est dépourvue de SFP et par conséquent de contrainte de déplacement entre les m images.

Actuellement, la méthodologie retenue pour la recherche de droites virtuelles est la suivante. Les droites virtuelles sont recherchées autour de l'horizontale en tentant d'optimiser le compromis entre la position dans la ROI la plus basse possible et une distance minimale de 200 *pixels* entre les couples de SFPs. Les droites virtuelles sont recherchées autour d'orientations privilégiées $[-25^\circ; 25^\circ]$ avec un pas de 5° . Les droites virtuelles ayant les orientations les plus importantes en valeur absolue permettent de créer des VFPs dans des régions de la ROI opposées de celles où se situent les FPs coplanaires dans les cas où le système est mal conditionné.

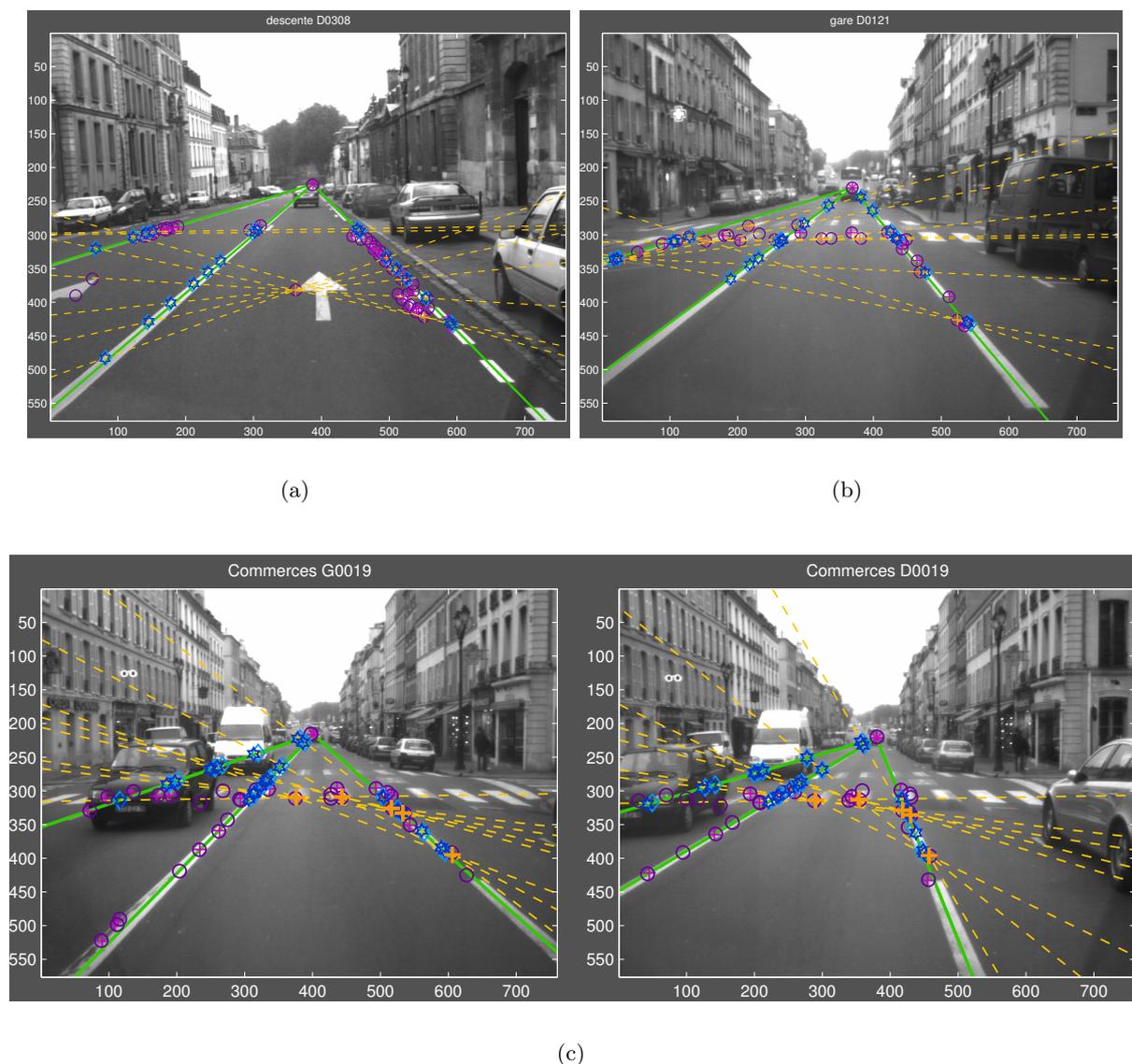


FIG. 6.5 – Cas où la distribution spatiale des SFPs ('+' mauves) n'est pas suffisamment uniforme pour contraindre l'intégralité de la projection du plan de la route. Les droites virtuelles (pointillées oranges) sont formées à partir de SFPs existants dont les projections dans les images courantes ont été identifiées ('+' oranges). L'interception des droites virtuelles avec les médianes des bandes principales (vertes) est à l'origine de VFPs ('◇' bleu-ciels) qui permettent de rigidifier le mouvement apparent de la ROI entre les m images lors du calcul de \mathcal{H} . Les FPs et VFPs dont les projections vérifient la super-homographie sont respectivement représentés par un 'o violet' et un '◻ bleu foncé'.

6.3.2.3 Initialisation de la super-homographie

Dans cette section, nous expliquons comment mettre en oeuvre l'estimation de la super-homographie à partir des primitives des images courantes, mises en correspondance lors des estimations des homographies (H_{st}^n, H_g^n et H_d^n). La liaison entre les primitives en correspondance permettra d'identifier, d'une part, les projections courantes des super-primitives existantes et, d'autre part, de nouvelles super-primitives potentielles si l'estimation de \mathcal{H}^n valide les mises en correspondance de projections jusqu'alors isolées. La contribution de chaque nouvelle itération se résumant à substituer les projections des super-primitives les plus anciennes par les plus récentes, nous verrons dans le troisième paragraphe l'influence d'une telle manipulation sur l'initialisation de la matrice de super-homographie ${}^0\mathcal{H}^n$.

La sélection des super-primitives La composition des super-primitives à partir des couples de primitives en correspondance dans les deux derniers couples d'images stéréo est une étape importante dans l'estimation de la super-homographie. Chacune des homographies estimée durant l'itération courante fournit une liste de couples de FPs en correspondance. Il nous faut établir les liens entre ces primitives de façon à identifier les nouvelles projections des SFPs qui ont permis l'estimation de \mathcal{H}^{n-1} à l'itération précédente.

Nous vérifions dans un premier temps que les correspondances qui lient les projections courantes d'un même coin de la scène sont cohérentes. Il nous faut identifier parmi les couples de primitives mis en correspondance par H_{st}^n quelles sont les représentations courantes des couples de primitives identifiés à l'itération précédente. Il suffit pour cela de vérifier que les projections courantes d'une super-primitive via H_g^n et H_d^n sont en correspondance, ce qui peut se résumer pour les projections précédentes et courantes de la super-primitive \mathbf{P}_k par :

$$\mathbf{P}_{kd}^n = [H_{st}^n \cdot H_g^n] \cdot \mathbf{P}_{kg}^{n-1} = [H_d^n \cdot H_{st}^{n-1}] \cdot \mathbf{P}_{kg}^{n-1} \quad (6.44)$$

La mise en oeuvre est immédiate : un tableau de correspondance *Liaisons* avec le numéro des primitives en correspondance dans chaque image. Chaque ligne représente une image ($\mathcal{I}_g^{n-1}, \mathcal{I}_d^{n-1}, \mathcal{I}_g^n, \mathcal{I}_d^n$), chaque colonne correspond à une super-primitive identifiée à l'itération précédente. Le tableau comprend 6 lignes :

- la première référence correspond au numéro du SFP dans la séquence : il sera non-nul si le SFP a permis l'estimation de \mathcal{H}^{n-1} à l'itération précédente,
- la deuxième et la troisième représentent les couples de primitives en correspondance grâce à H_{st}^{n-1} ,
- les deux suivantes contiennent respectivement les numéros des primitives couplées aux deux premières via H_g^n et H_d^n ,
- la dernière contient le résultat de la mise en correspondance par H_{st}^n entre les primitives de la 3^{ème} ligne (\mathcal{I}_g^n) et celle de l'image droite courante (\mathcal{I}_d^n).

La cohérence des mises en correspondance entre les deux derniers couples d'images est vérifiée si les éléments de la quatrième et cinquième ligne sont identiques. Lorsque ce n'est pas le cas, la liaison la moins certaine est rejetée, ce qui revient à ne plus tenir compte de la projection courante.

Un exemple de cette procédure est illustré en Fig. 6.6 où nous avons reporté les FPs détectés dans les deux derniers couples d'images. Seules les correspondances entre les couples d'images stéréo sont mis en évidence par soucis de clarté. Nous avons reporté dans le Tab. 6.1 quelques exemples de liaisons établies à partir des couples de FPs mis en correspondance lors du calcul des homographies H_{st}^n, H_g^n et H_d^n . Si le véhicule-test est quasiment à l'arrêt dans le couple d'images

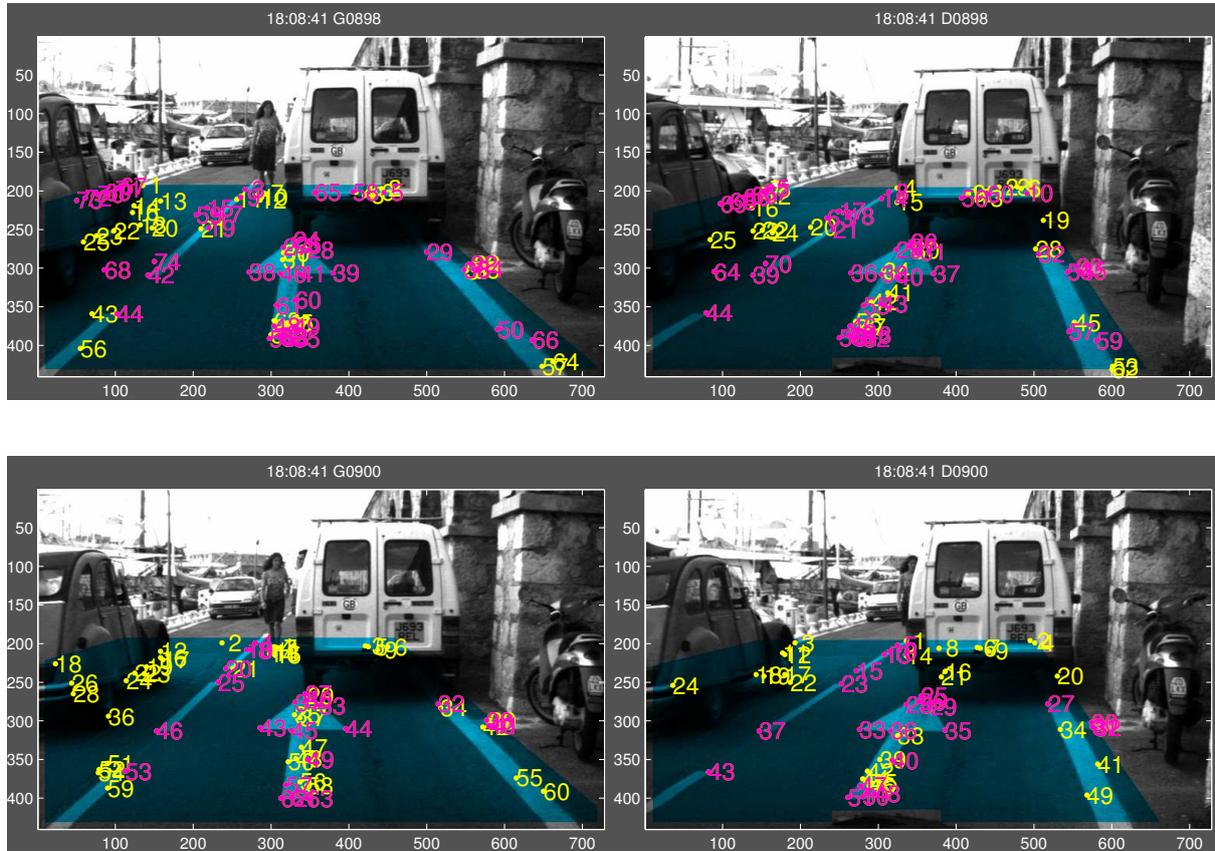


FIG. 6.6 – Synthèse des couples de primitives mis en correspondance grâce aux estimations des homographies courantes. Les FPs extraits sont représentés en jaune, les roses vérifient respectivement H_{st}^{n-1} et H_{st}^n dans les couples d'images supérieur et inférieur. Les FPs roses sur le véhicule en déplacement qui vérifient H_{st}^{n-1} n'étaient pas fournis par le détecteur de Harris mais correspondent à la propagation par \mathcal{H}^{n-1} de SFPs coplanaires actuellement masqués par l'obstacle.

origine	BP centrale						2CV			flèche			utilitaire		route			
n°	64	77	97	110	126	140	131	148	171	78	79	83	149	163	0	0	0	0
\mathcal{I}_g^{n-1}	44	42	19	17	9	3	69	71	67	39	38	40	65	58	48	49	50	35
\mathcal{I}_d^{n-1}	44	39	21	18	14	8	65	67	5	37	36	38	56	60	46	48	57	0
\mathcal{I}_g^n	53	46	25	20	8	1	0	0	0	44	43	45	0	3	57	58	0	40
\mathcal{I}_d^n	43	37	23	15	10	5	0	0	0	35	33	36	0	0	47	0	49	31
\mathcal{I}_d^n	43	37	23	15	10	5	0	0	0	35	33	36	0	0	0	0	0	31

TAB. 6.1 – Exemples de liaisons établies suite aux estimations des homographies courantes à partir des 2 couples d'images précédents. Chaque colonne représente un SFP actif au début de l'itération courante si son numéro (deuxième ligne) est non-nul ou qui sera à l'origine d'un SFP temporaire autrement.

précédentes, le faible déplacement qui se produit entre les deux itérations permet de rejeter les FPs détectés sur le véhicule en stationnement qui étaient mis en correspondance de manière erronée jusqu'à l'itération précédente. De même, nous pouvons observer que les FPs les mieux caractérisés (coins des marquages au sol) sont tous correctement suivis sans ambiguïté (valeurs des lignes 5 et 6 identiques) contrairement à d'autres (sur la route) où la mise en correspondance dans les 4 vues n'est pas effective.

Le calcul de la super-homographie est entrepris à partir de tous les couples de FPs identifiés grâce aux calculs des trois homographies courantes. Deux groupes de SFPs se distinguent alors :

- les existants correspondent aux SFPs qui ont permis l'estimation de \mathcal{H}^{n-1} (première ligne du tableau de liaisons non-nulle),
- les nouveaux temporaires sont formés par les associations des couples de FPs identifiés lors de l'établissement des liaisons entre les couples d'images précédents et courants.

Les nouveaux SFPs peuvent être formés à partir de seulement un couple de FPs : les projections dans les $(m - 2)$ autres images sont alors supposées inconnues.

L'identification des super-primitives coplanaires L'estimation de \mathcal{H}^n permet de détecter les couples de primitives non-coplanaires. Il s'agit des SFPs qui présentent des écarts importants entre les coordonnées estimées par \mathcal{H}^n et celles extraites par le filtre de Harris et ce, dans l'une des m images. Deux cas doivent alors être distingués :

- le SFP est existant : il est suivi depuis plusieurs itérations, la probabilité qu'il ne corresponde pas à une primitive coplanaire est faible. L'écart relevé entre les coordonnées extraites et estimées est dû soit à une erreur d'extraction, soit à une estimation de \mathcal{H}^n mal conditionnée, soit à une combinaison des deux possibilités. Le SFP est alors marqué comme SFP potentiellement non-coplanaire. Le SFP est éliminé lorsque le SFP est marqué durant $m/2$ itérations consécutives, ce qui permet d'assurer la continuité du plan sur lequel les FPs sont censés être détectés.
- le SFP est temporaire : au moins une des liaisons entre les FPs qui le constituent est erronée, le SFP est composé de projections qui ne sont pas toutes coplanaires : le SFP est rejeté.

Quel que soit le type du SFP, si l'une de ses projections dans l'une des images courantes était inconnu avant le calcul de \mathcal{H}^n , il nous faut retrouver le FP correspondant. Le FP dont nous connaissons les coordonnées est recherché parmi tous les FPs extraits par le détecteur de Harris comme celui étant le plus proche. Si plusieurs possibilités sont possibles, le calcul du coefficient de corrélation est entrepris avec la projection dans l'image précédente. Le coefficient de corrélation est en effet plus élevé que celui calculé avec la projection dans l'autre image du couple stéréo : le résultat du calcul est par conséquent plus discriminant. Le FP qui présente le meilleur score sera considéré comme la projection courante du SFP dans l'image courante considérée.

Dans l'éventualité où le coefficient de corrélation s'avère inférieur à *seuil_corr* ou qu'aucun FP ne corresponde aux coordonnées estimées, cela signifie que soit le FP est caché par un obstacle, soit que son extraction est devenue difficile. Dans les deux cas, le SFP est marqué comme précédemment de façon à ce que les SFPs existants puissent justifier qu'au moins une des projections dans les $m/2$ derniers couples d'images corresponde à un FP détecté dans la scène. Nous nous affranchissons de la sorte à la propagation de SFPs masqués par des obstacles.

La Fig. 6.7 illustre les différents marqueurs que nous utilisons pour caractériser les projections courantes des FPs qui constituent les SFPs pour une itération donnée. Les FPs extraits grâce au détecteur sont représentés en cyan. Les FPs '+' en bleu foncé vérifie l'homographie

stéréo courante H_{st}^n . Les SFPs qui vérifient \mathcal{H}^n sont représentés en vert, les roses présentent un écart entre coordonnées estimées et extraites dans l'une des images. L'une des projections dans le couple d'images courantes est manquante pour les SFPs oranges alors que les deux manquent pour les SFPs rouges.

Plusieurs points sont à relever suite à cette figure. Nous pouvons vérifier qu'à l'exception des FPs masqués par le véhicule en déplacement tous les SFPs sont détectés sur le plan de la route. L'estimation de H_{st}^n a permis de rejeter des couples de FPs correctement mis en correspondance par les homographies courantes mais situés au-dessus de la route. Le SFP rose est effectivement mal localisé : il représente à l'origine l'intersection de l'empennage de la flèche avec sa base. Le coin correspondant n'est pas détecté dans l'image de droite car la saillance est obtuse. Un FP détecté dans le voisinage ayant un coefficient de corrélation qui satisfait *seuil_corr* a été associé de manière malencontreuse.

Enfin, l'identification de nouvelles super-primitives s'effectue à partir des SFPs temporaires qui ont vérifié H_{st}^n . Les précautions que nous avons établies pour s'assurer que les SFPs représentent les projections de coins de la scène effectivement détectés dans plusieurs images se justifient encore plus lors de leur initialisation. Seuls les SFPs temporaires dont au moins trois des quatre projections dans les deux derniers couples d'images correspondent à des FPs extraits sont considérés comme de nouveaux SFPs. Nous imposons ainsi au minimum une mise en correspondance stéréo et une autre entre deux images consécutives tout en vérifiant les coefficients de corrélation entre les différentes projections.

Initialisation de la super-homographie La matrice super-homographie se décomposant en trois parties conformément à la relation 6.42 : m matrices identités I_3 , un bloc triangulaire inférieur et son dual le supérieur tel que $\mathcal{H}_{ji} \cdot \mathcal{H}_{ji} = I_3$ quelles que soient les images \mathcal{I}_i et \mathcal{I}_j . Le calcul des homographies qui forment \mathcal{H} n'est donc fait que pour les homographies telles que $i < j$ quels que soient $i, j \in [2, m]$. Seules les deux dernières lignes du bloc triangulaire inférieur de ${}^0\mathcal{H}^n$ font référence aux images courantes \mathcal{I}_g^n et \mathcal{I}_d^n , conformément à la Fig. 6.8.

L'initialisation des $(m - 2)$ premières lignes du bloc triangulaire inférieur peut donc être réalisée à partir des $(m - 2)$ dernières lignes du bloc triangulaire inférieur de \mathcal{H}^{n-1} . Le calcul des $(2m - 3)$ homographies relatives aux images courantes est alors mené en composant les homographies croisées telles qu'elles sont décrites dans la figure. La composition des homographies entre les images précédentes et courantes revient à multiplier les homographies qui forment les 2 dernières lignes de \mathcal{H}^{n-1} par respectivement H_g^n, H_d^n .

L'introduction du bloc constitué des $(m - 2)$ dernières lignes et colonnes de \mathcal{H}^{n-1} permet non seulement de réduire le nombre d'opérations nécessaires à l'initialisation de \mathcal{H}^n mais contraint fortement le conditionnement du système puisque le rang du bloc est égal à 3. Ainsi, dans le cas où $m > 3$, le rang de la matrice initialisée de la sorte est égal à $m + 3$ alors qu'il est égal à $3m$ quand l'initialisation est opérée par concaténation des homographies calculées entre chacune des images.

6.3.2.4 Estimation de la super-homographie

Durée d'observation des super-primitives La séquence "descente" enregistrée à Versailles correspond à un déplacement en pente sur un large boulevard à 5 voies dont deux de stationnement où la signalisation horizontale présente un fort contraste avec le revêtement de la route. La trajectoire du véhicule est rectiligne sur les 4/5 du parcours puis se déporte sur la gauche pour quitter la voie de droite en direction de la centrale. Le champ de vision est complètement

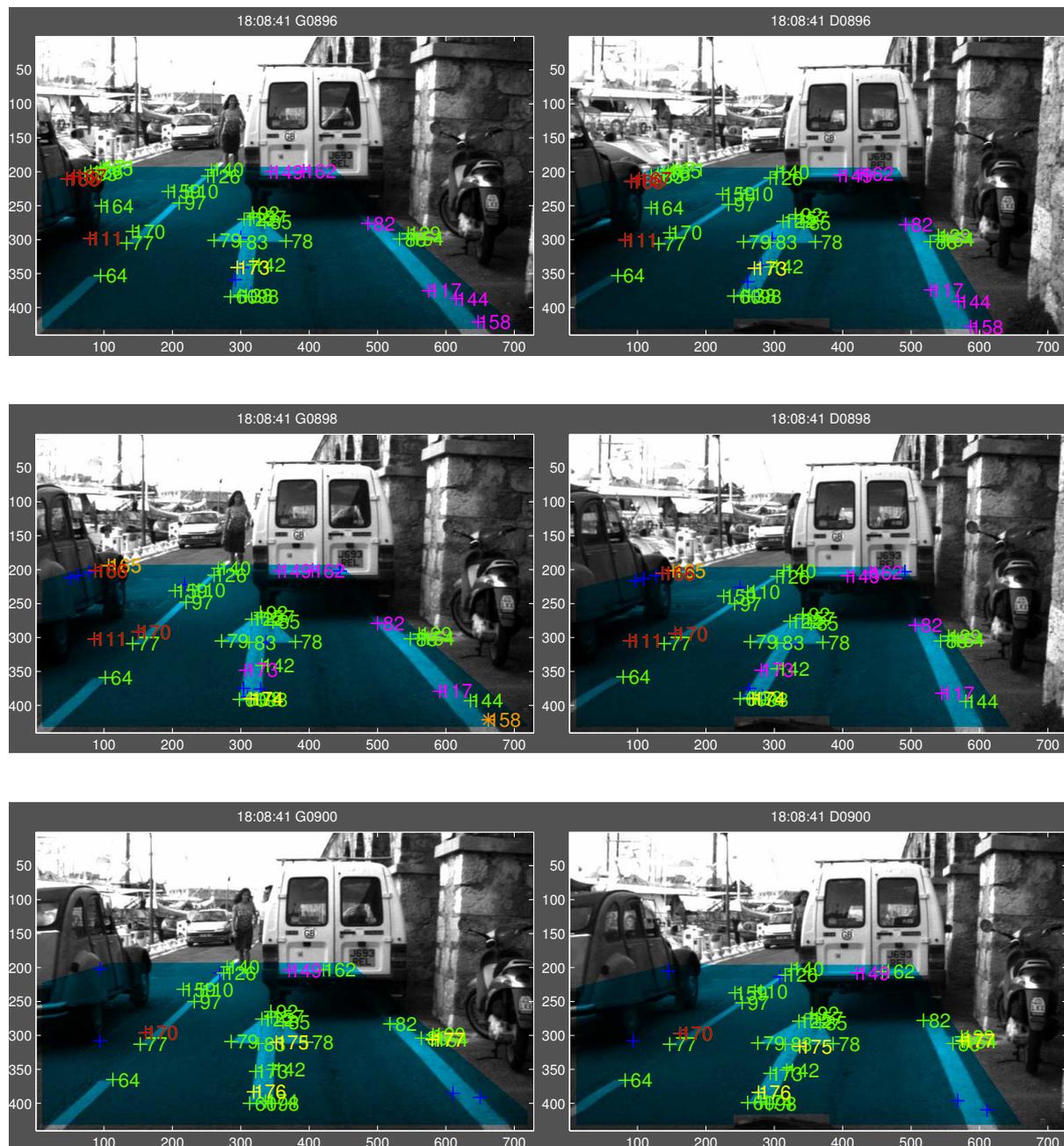


FIG. 6.7 – Caractérisation des super-primitives qui ont permis de calcul de la super-homographie. La présence du véhicule en déplacement permet de vérifier la robustesse de la méthode qui permet de localiser les FPs coplanaires bien qu'ils soient momentanément masqués. Les SFPs verts sont ceux qui vérifient la super-homographie, les roses présentent un écart au-delà du seuil toléré entre les coordonnées estimées par \mathcal{H}^n et les coordonnées détectées. Les SFPs jaunes sont nouveaux tandis que les oranges et rouges ont respectivement une et deux de leurs projections non identifiées dans le couple d'images courantes.

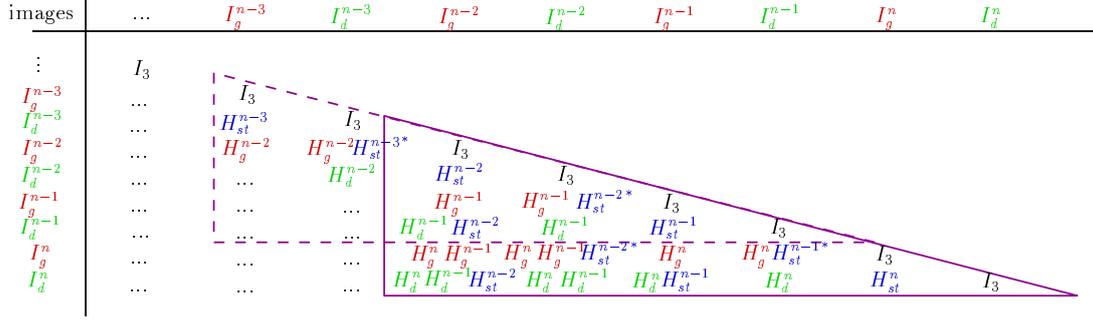


FIG. 6.8 – Représentation de la composition du bloc triangulaire inférieur des matrices \mathcal{H}^{n-1} (pointillés) et \mathcal{H}^n (trait plein) à partir des homographies qui les constituent. L'initialisation de ${}^0\mathcal{H}^n$ est calculée pour partie à partir de l'estimée de \mathcal{H}^{n-1} . Les homographies relatives aux images \mathcal{I}_g^n et \mathcal{I}_d^n sont calculées par composition des homographies courantes H_{st}^n, H_g^n et H_d^n . La notation H_{st}^{n*} correspond à l'inverse H_{st}^{n-1} a été introduite par soucis de lisibilité.

dégagé : tous les FPs détectés dans la ROI sont coplanaires.

Le relevé du nombre de FP détectés dans les images du couple stéréo (en haut de la Fig.6.9 montre que nos efforts pour détecter des FPs en plus de ceux des coins de la signalisation horizontale sont effectifs. Les populations de FPs détectés dans l'image gauche (en rouge) et droite (en vert) sont comparables sur l'intégralité de la séquence. L'évolution conjointe au cours du temps du nombre de FPs extraits est essentiellement dû à l'observation temporaire d'éléments de la signalisation horizontale comme un passage piéton à deux reprises et d'une flèche peinte sur chacune des deux voies observées à quatre reprises.

Nous retrouvons cette même similarité dans le second chronogramme en comparant les quantités de FPs mis en correspondance lors des estimations des homographies H_g^n, H_d^n . Ces résultats sont néanmoins trompeurs : le champ de vision commun aux 2 caméras est restreint au centre de la ROI comme le montre le nombre moins important de FPs mis en correspondance grâce à l'estimation de H_{st}^n (en bleu). Cela signifie que les couples de FPs supplémentaires mis en correspondance lors des estimations des homographies H_g^n, H_d^n appartiennent à des champs de vision propres à chaque caméra.

Le nombre de SFPs utilisé pour estimer \mathcal{H} étant commun à trois couples d'images stéréo, il est normal qu'il soit inférieur au nombre de FPs minimal en correspondance pour une itération donnée. Néanmoins, il apparaît clairement que le nombre de SFPs rejetés lors de l'estimation de la super-homographie est insignifiant durant cette séquence car le nombre de SFPs est légèrement inférieur ou égal au nombre de FPs mis en correspondance entre les deux seules images courantes de la paires stéréo. D'autre part, l'évolution lente du nombre de SFPs signifie que le nombre de SFPs qui sortent du cadre de l'image, est compensé par l'apparition et la détection de nouveaux SFPs.

L'histogramme de la Fig. 6.9(b) permet de vérifier que la présente configuration des caméras n'est pas adaptée à notre étude. La durée pendant laquelle les SFPs sont observées est trop courte : seuls les FPs qui ont des projections situées au centre de chaque image sont observés sur quelques dizaines d'itérations. Les FPs détectés autour des bandes principales gauche ou droite sortent rapidement du champ de vision d'une des deux caméras. A titre de comparaison, nous rappelons que toutes les extrémités d'une flèche peinte au sol ont été suivies durant une cinquantaine d'itérations (soit une centaine d'images) à partir de la séquence du port d'Antibes. Cependant, la vitesse plus élevée du véhicule (50 km/h) et le plan incliné de la route justifie le

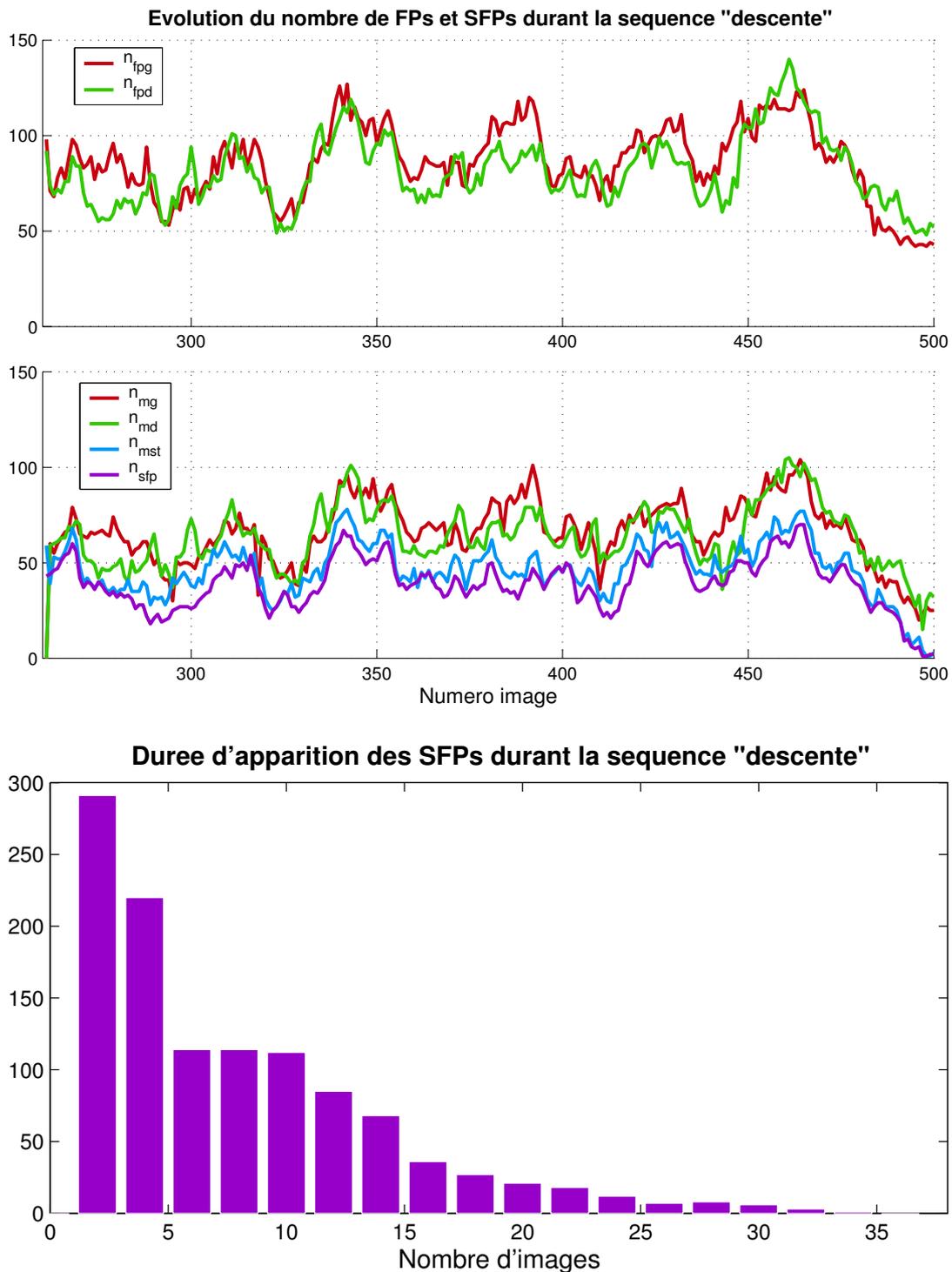


FIG. 6.9 – Evolution du nombre de points d'intérêt au cours de la séquence "descente". En haut sont reportés respectivement le nombre de FPs détectés dans chaque image, le nombre de FPs mis en correspondance lors des estimations des homographies H_g^n , H_d^n , H_{st}^n et le nombre de super-primitives qui vérifient l'estimation de la super-homographie. En bas, l'histogramme représente le nombre d'images durant lesquelles les SFPs sont détectées.

fait que la détection des flèches peintes sur cette séquence ne procurent pas des SFPs observés durant de longues périodes.

Choix du nombre de vues m à considérer La recherche de nouvelles super-primitives impose de fixer le nombre de vues m à considérer. En effet, le nombre de projections détectées d'une nouvelle super-primitive lors de la prochaine itération ($n + 1$) devra être suffisamment important pour dépasser le seuil $m/2$ de façon à justifier sa propagation par \mathcal{H}^{n+1} . Dans le pire des cas, aucune correspondance ne sera détectée à partir des primitives détectées dans les images \mathcal{I}_g^{n+1} et \mathcal{I}_d^{n+1} .

Ainsi seules les primitives détectées durant les itérations ($n - 1$) et n permettent de valider le seuil $m/2$. Or, nous avons imposé que le nombre minimal de primitives détectées en correspondance entre les itérations précédentes et courantes était égal à trois pour justifier la création d'une nouvelle super-primitive, soit $m/2 \geq 3$. Ainsi, le calcul de la super-homographie à partir d'au moins trois couples d'images stéréo permet de calculer les coordonnées dans les images courantes de super-primitives et répond aux contraintes de cohérence que nous imposons lors de la création des super-primitives.

Il va de soi que deux couples stéréo suffisent à calculer une super-homographie : son rôle sera réduit à l'affinement des estimations des homographies courantes. La définition des contraintes imposées pour la création de nouvelles super-primitives à partir des primitives mises en correspondance entre les itérations précédentes et courantes est adaptée au cas $m/2 = 3$ que nous avons retenu. Les affinages des estimations des homographies calculées durant l'itération courante (H_{st}^n, H_g^n et H_d^n) sont calculées grâce aux coordonnées des super-primitives des itérations ($n - 2$), ($n - 1$) et n .

Une modification de ce paramètre est toutefois envisageable mais il nous faut noter que le choix du nombre d'images à considérer dépend de deux facteurs antagonistes si nous faisons abstraction du temps de calcul : le nombre de contraintes imposées augmente intuitivement avec m ce qui réduit inversement le nombre de primitives observées dans la région commune à toutes les images.

Algorithme d'estimation de la super-homographie Les coordonnées des super-primitives des itérations précédentes sont supposées comme coplanaires. L'estimation de la super-homographie \mathcal{H}^n s'effectue en deux passes. La première écarte les super-primitives dont les projections courantes détectées \mathbf{p}_k^n sont trop éloignées de leurs estimées ${}^1\mathbf{p}_k^n$. Les coordonnées de ces super-primitives sont alors considérées comme inconnues. L'estimation de \mathcal{H}^n est relancée une seconde fois si au moins une liaison de SFP a été invalidée.

Lors de la première itération de l'estimation ($w = 1$) de chaque passe, une vérification de la cohérence des VFPs est effectuée en mesurant la distance $|\mathbf{}^1\mathbf{p}_k - \mathbf{}^0\mathbf{p}_k|$ qui sépare les coordonnées calculées de celle fournies par ${}^1\mathcal{H}^n$. Ces dernières étant construites à partir de droites virtuelles, les intersections obtenues peuvent être imprécises et par conséquent s'opposer à une convergence rapide de l'estimation de \mathcal{H}^n .

Ainsi, le rang de la matrice ${}^w\mathcal{H}$ décroît rapidement de $3m$ à 3. Il suffit dans la majorité des cas de $w = 2$ pour obtenir la convergence. Les cas qui nécessitent plus d'itérations contiennent généralement des primitives dont les projections entre images sont erronées. Lorsque le nombre d'itérations dépasse 4, l'une des homographies courantes doit être remise en cause. Cela arrive dans le cas de systèmes très mal conditionnés où le nombre de primitives mise en correspondance est réduit et mal réparti spatialement dans la ROI.

A l’opposé, la convergence du calcul n’assure aucunement que les SFPs sélectionnées soient détectées sur le plan considéré. En effet, la méthode telle qu’elle est implémentée a tendance à privilégier la solution qui satisfait le plus grand nombre de super-primitives. Si leur répartition dans la ROI n’est pas uniforme ou qu’un certain nombre d’entre elles sont détectées en dehors du plan considéré, la convergence du calcul peut être obtenue à partir d’un plan moyen qui minimise la distance entre les primitives 3D considérées.

Lorsque la convergence n’est pas atteinte ($\text{rang}(\mathcal{H}^n) \neq 3$), l’estimation de la matrice de super-homographie ainsi que les coordonnées estimées des super-primitives sont incorrectes. L’algorithme conserve pour les estimations courantes des homographies H_{st}^n, H_g^n, H_d^n celles extraites directement à partir des couples d’images. De même, les jeux de SFPs et SVLs sont seulement actualisés, les coordonnées courantes de leurs projections sont enregistrées lorsqu’elles sont connues.

Raisons de la non-convergence du calcul de la super-homographie L’estimation de la super-homographie impose par principe que les coordonnées des primitives aux itérations précédentes soient connues et supposées correctes car elles ont été validées par le calcul de \mathcal{H}^{n-1} . Lorsque l’estimation de \mathcal{H} nécessite plus de deux itérations, cela indique que le système est mal conditionné. Plusieurs cas sont à envisager :

- des mises en correspondance entre projections de super-primitives sont erronées. L’identification des projections incriminées est alors facilement obtenue en comparant les coordonnées estimées et détectées des super-primitives. Etant donné que certaines projections courantes peuvent être inconnues, nous avons alors tout intérêt à réaliser la comparaison à partir des coordonnées les plus anciennes. Les super-primitives ayant des projections dont les coordonnées dans les différentes images sont liées, si les coordonnées aux estimations précédentes évolues, cela signifie que la mise en correspondance est erronée. Une nouvelle estimation de \mathcal{H} doit être effectuée en considérant les coordonnées courantes des super-primitives comme inconnues,
- les projections des droites virtuelles peuvent ne pas correspondre dans les images courantes lorsque les coordonnées des SFPs dans les images courantes sont fausses. Une nouvelle tentative de l’estimation de \mathcal{H} sans les VFPs peut rapidement converger dans de tels cas,
- lorsque la distribution spatiale des primitives dans les images courantes est mauvaise, les estimations des homographies courantes sont peu fiables contrairement aux primitives mises en correspondance comme nous l’avons vu dans la partie 5.2. Lors de l’initialisation de ${}^0\mathcal{H}^n$, la substitution des homographies H_{st}^n, H_g^n, H_d^n par des homographies calculées à partir des SFPs et des VFPs permet dans la majorité des cas d’obtenir des estimations des homographies courantes de meilleure facture car contraintes sur des régions de la ROI plus importantes,
- les primitives filtrées (DVPs et bandes principales) s’écartent parfois du modèle qui régit leur déplacement ou plus simplement elles ne sont plus observables. Or, toutes les estimations d’homographies sont calculées en introduisant systématiquement leurs coordonnées. La mise en oeuvre de la recherche des projections erronées ne peut être envisagée à partir d’un calcul d’écart de projection étant donné que les homographies sont biaisées.

La mise en oeuvre de tests qui permettent de définir quelle est parmi la liste précédente la cause de la non-convergence de l’algorithme nous permet désormais d’obtenir une convergence en un maximum de 4 passes.

6.3.2.5 Récapitulatif

L'intégralité des opérations successives nécessaires à l'estimation de la super-homographie \mathcal{H}^n peut se résumer de la sorte :

1. $r=0$,
2. mise en correspondance des couples de primitives identifiés grâce aux estimations des homographies H_{st}^n, H_g^n et H_d^n ,
3. création du tableau de vecteurs ${}^0P_k^n$,
4. recherche parmi ${}^0P_k^n$ des SFPs dont les projections sont connues dans les m vues pour créer les VFPs ${}^0VP_k^n$,
5. initialisation de ${}^0\mathcal{H}^n$ à partir de ${}^0\mathcal{H}^{n-1}$ et H_{st}^n, H_g^n, H_d^n ,
6. **tant que** $\text{rang}({}^r\mathcal{H}^n) > 3$,
 - (a) $r = r+1$,
 - (b) calcul des coordonnées estimées des SFPs ${}^rP_k^n$ en appliquant l'Equ. 6.41,
 - (c) estimation par bloc de matrice $[3;3]$ de ${}^r\mathcal{H}^n$ à partir des SFPs ${}^rP_k^n$,
 - (d) **si** $r == 1$,
 - i. recherche puis élimination des VFPs dont les coordonnées estimées ${}^1VP_k^n$ sont distantes des coordonnées extraites ${}^0VP_k^n$ dans l'une des m images,
 - (e) **fin si**
7. **fin tant que**
8. recherche des SFPs dont les coordonnées estimées sont distantes des coordonnées extraites,
9. **si** des mises en correspondance sont erronées,
 - (a) rejet des projections courantes des SFPs dont les liaisons ne sont pas vérifiées en annulant les coordonnées courantes dans ${}^0P_k^n$,
 - (b) nouvelle estimation de \mathcal{H}^n à partir de ${}^0\mathcal{H}^n$ et des ${}^0P_k^n$,
10. **fin si**
11. recherche parmi les FPs détectés dans les images courantes de FPs susceptibles de représenter les projections inconnues dont les coordonnées sont fournies par \mathcal{H}^n ,
12. mise à jour des SFPs existants,
13. création de nouveaux SFPs.

6.3.3 Reconstruction de la route au long d'une séquence

Nous présentons dans cette dernière partie des résultats perfectibles mais encourageant sur les perspectives qu'ils ouvrent. Nous avons superposé les projections du plan de la route extraits des images enregistrées par une même caméra tout au long d'une même séquence. La calibration de la paire stéréo étant inconnue, nous avons redressé "à la main" la première image

de la séquence puis avons appliqué cette même transformation aux images suivantes tout en tenant compte du déplacement de la caméra entre chacune d'elles.

Le redressement consiste à déterminer l'homographie H_V^1 qui transforme le parallélogramme qui représente la ROI de la première image de la séquence en un rectangle où les projections des bandes principales apparaissent parallèles conformément à la Fig. 6.10. L'homographie H_{v_n} qui permet de positionner la n^{eme} image par rapport à la première n'est autre que la composition des homographies $H_{g(d)}^n$, estimées durant la séquence :

$$H_V^n = H_v^1 \cdot H_{g(d)}^2 \cdot H_{g(d)}^3 \dots H_{g(d)}^{n-1} \cdot H_{g(d)}^n. \quad (6.45)$$



FIG. 6.10 – Choix de la première image pour reconstituer le plan de la route au long de la séquence. L'homographie qui permet de passer du repère caméra à un repère virtuel en vue de dessus nécessite que la ROI contienne des informations relatives à la direction orthogonale à celle de la route de façon à compenser l'orientation du véhicule. L'estimation de l'homographie qui permet une telle transformation est obtenue par composition de l'homographie qui transforme le parallélogramme de la ROI en un rectangle puis par une rectification qui aligne les bandes du passage piéton.

Sur la séquence du port d'Antibes, le mauvais conditionnement du système formé par le peu de primitives coplanaires détectées à partir de l'itération 940 (fin du déboîtement pour contourner le véhicule stationné) ne permettent pas d'obtenir une reconstruction sur la totalité de la séquence (voir Fig. 6.11). Nous pouvons néanmoins vérifier sur les 85 premières images que la reconstitution du plan de la route est correcte : l'alignement des bandes est respecté, les détails du revêtement sont continus malgré les mouvements de rotation du véhicule. Quelques imperfections apparaissent au niveau de l'empenne de la flèche peinte sur la voie de droite lorsque le véhicule est quasiment à l'arrêt. Les homographies estimées sont très peu fiables (5 pixels de déplacement vertical maximal pour un FP situé au bas de l'image) ce qui justifie la dégradation de la reconstitution.

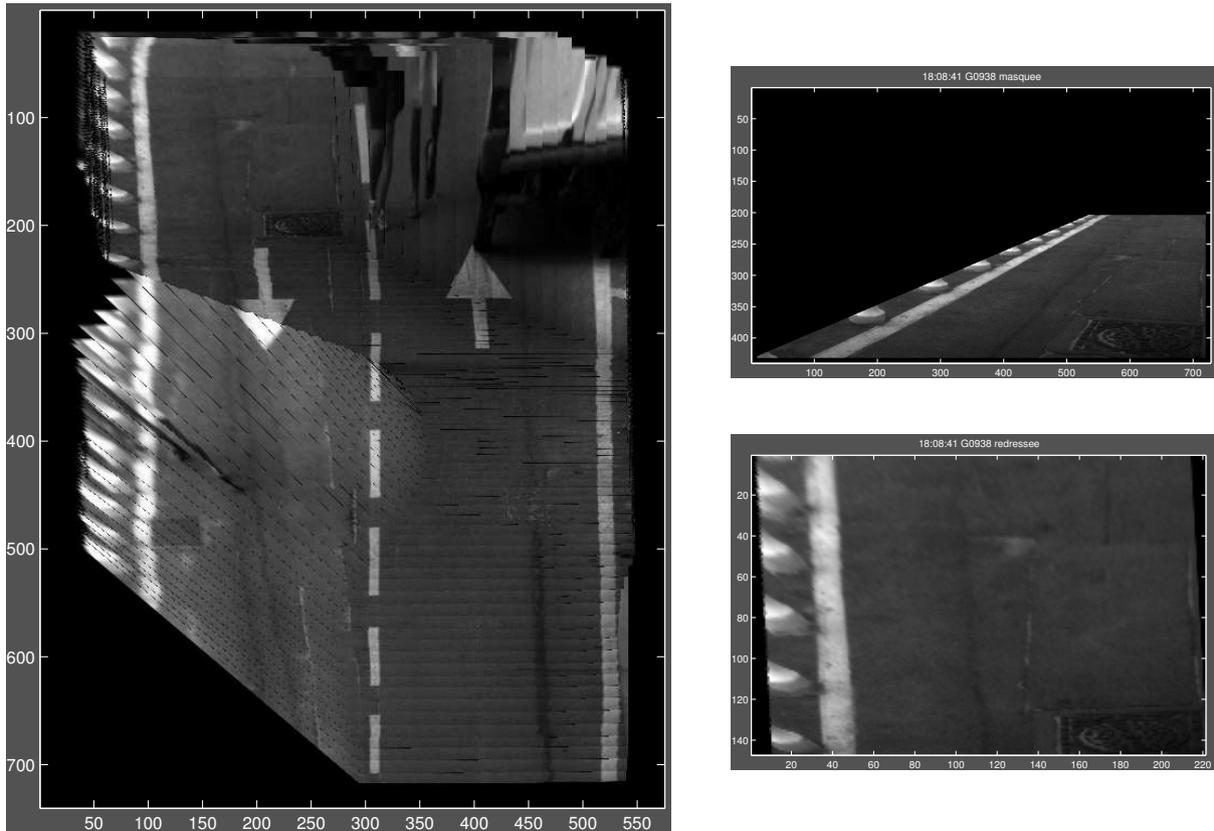


FIG. 6.11 – *Reconstruction du plan de la route sur la séquence du port d'Antibes. La reconstruction est partielle en raison d'une erreur d'estimation du déplacement en fin de déboîtement alors que la région de la route observée est réduite. A droite sont présentées la dernière image masquée valide et sa projection en vue de dessus, exprimée dans le repère de la première image (770).*

Les chronogrammes de la Fig. 6.12 représentent les estimations de déplacements de la caméra droite au cours de la séquence descente, enregistrée le long du château de Versailles. Malgré l'absence d'obstacles et un déplacement en ligne droite à vitesse quasi-constante sur les 4/5 de la séquence puis un changement de voie sur la dernière partie, l'algorithme n'est néanmoins pas parvenu à estimer correctement certaines homographies. Cela s'explique par deux raisons :

- l'homogénéité du revêtement ne permet pas de détecter des FPs en dehors des coins de la signalisation horizontale qui s'avère de surcroît surabondante au niveau de la bande prin-

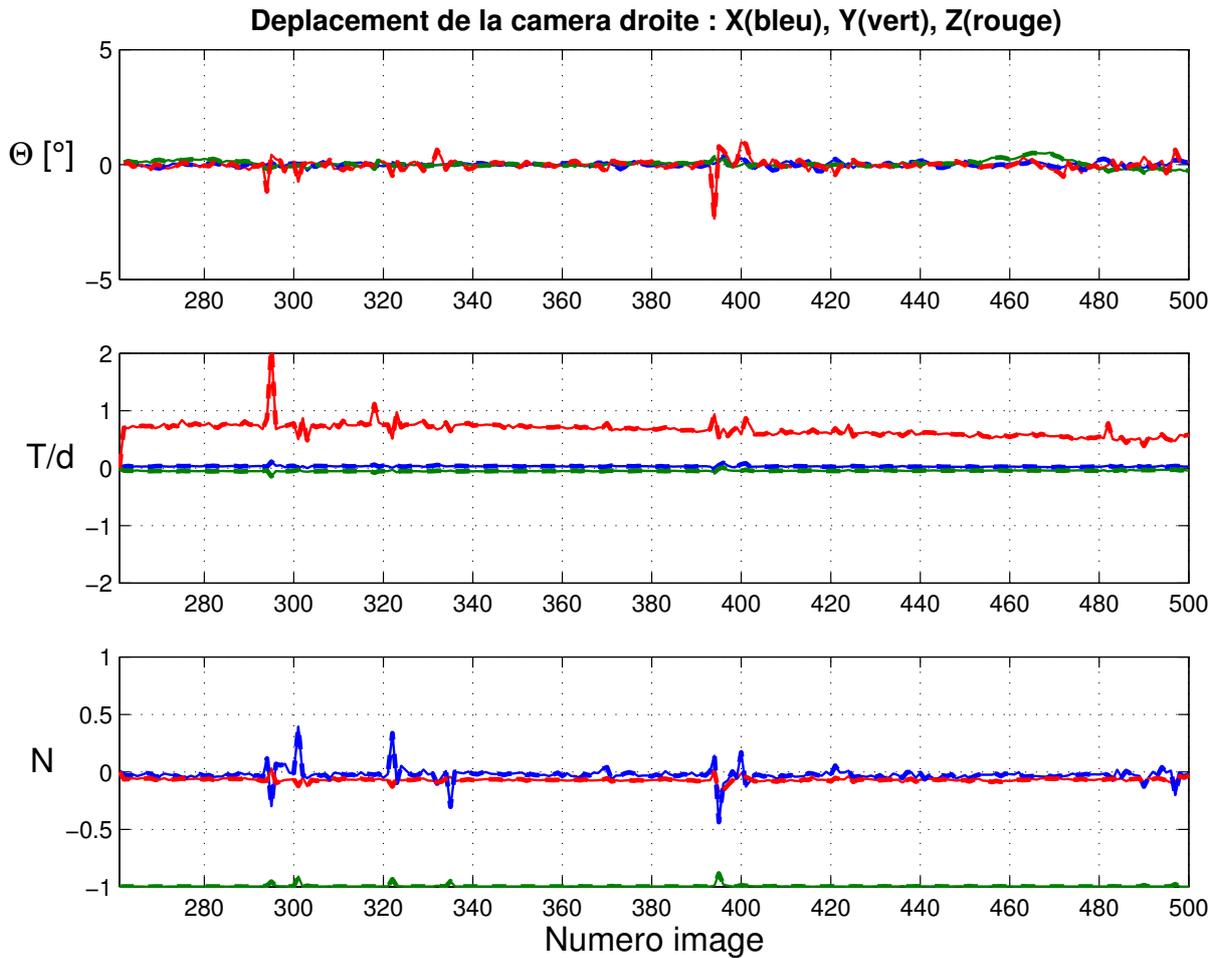


FIG. 6.12 – Décomposition des mouvements de translation et rotation de la caméra droite au cours de la séquence "descente". Le véhicule se déplace à vitesse constante en ligne droite sur les 4/5 de la séquence pour finalement changer de voie sur la dernière partie en délaissant la voie de droite pour se diriger vers la centrale.

principale droite et quasiment inexistante autour de la bande principale gauche essentiellement représentée par une ligne blanche continue,

- la configuration de la paire stéréo s'oppose à la mise en correspondance de la majorité des FPs détectés au premier plan et la vitesse du véhicule est telle que généralement la mise en correspondance des FPs détectés autour du motif des pointillés est de manière quasi-générale ambiguë entre deux images consécutives. En effet, le mouvement apparent de ces FPs est plus important que la distance qui sépare les projections d'un même motif.

La séquence comprend dans son intégralité 239 images ([261;500]). La distance parcourue est estimée à partir d'un plan de la ville à 200 m qui ont été parcourus à une vitesse d'environ 50 km/h. Les reconstitutions du plan de la route présentées en Fig. 6.13 illustrent l'influence des erreurs d'estimations des homographies dans ce processus en boucle ouverte. La première, à gauche, correspond au déplacement en ligne droite de la trajectoire (images [270;468]). Le cumul des erreurs, essentiellement selon l'angle de lacet du véhicule, s'oppose à la reconstruction du plan de la route à l'abord de l'action de déboîtement.

A l’opposé, la reconstitution de droite se distingue de la précédente car elle débute à partir de l’itération 330, c’est à dire, après les imprécisions mesurées en début de séquence. La route est correctement recomposée : les passages piétons sont perpendiculaires à la direction de la route, les détails du revêtement sont continus. De plus, nous pouvons observer la manoeuvre de déboîtement car contrairement à la séquence antiboise, le mouvement de rotation du véhicule est moins important ce qui permet d’observer une région suffisamment importante de la route et d’y détecter des FPs mieux répartis.

Dans les deux exemples, le défaut de reconstruction observé à la mi-séquence (brusque discontinuité de la bande principale centrale) est imputable à la période relativement longue qui précède le décalage durant laquelle les FPs extraits le sont de manière quasi-exclusive le long de la bande principale droite formée des petits pointillés. Une fois de plus, le système insuffisamment contraint par une mauvaise répartition spatiale des primitives coplanaires détectées est à l’origine des erreurs d’estimations de l’homographie et par conséquent de la reconstruction.

La plupart des discontinuités à l’origine des erreurs de reconstruction ont de grandes amplitudes, ce qui devrait permettre leurs détections et faciliter la mise en oeuvre d’une correction adaptée. Le principal facteur qui s’oppose à la reconstitution de l’intégralité de la séquence est le fait que nous ne disposons pas d’informations sur le déplacement du véhicule autres que celles extraites des images. Une des possibilités dont nous disposons pour corriger les erreurs d’estimation des homographies serait d’entreprendre un filtrage temporel sur les estimations de mouvements de translation et de rotation de façon à s’affranchir de telles discontinuités et par conséquent robustifier les résultats obtenus sur de longues distances parcourues.

6.4 Conclusion

La décomposition d’une homographie planaire nécessite de connaître a priori l’orientation du plan observé afin de choisir la solution correcte parmi les deux solutions admissibles. Dans le cas d’un véhicule en déplacement en milieu urbain, les principaux plans susceptibles d’être observés sont la route et d’éventuelles façades verticales. Les normales de ces plans sont respectivement portées par la verticale et l’horizontale. En supposant une configuration des caméras où l’axe optique des caméras appartient à un plan horizontal, les projections des normales associées aux plans suivis sont alors connues.

La matrice de super-homographie permet d’estimer en un unique calcul toutes les homographies qui lient un nombre d’images m supérieur ou égal à trois d’une même scène planaire. Les homographies ainsi estimées vérifient la contrainte de composition $H_{ca} = H_{cb}H_{ba}$ pour tous les triplets d’images $\mathcal{I}_a, \mathcal{I}_b, \mathcal{I}_c$. La super-homographie est calculée à partir de super-primitives qui contiennent les coordonnées des primitives coplanaires dans les m vues. L’actualisation des super-primitives est réalisé à partir des primitives mises en correspondance lors des estimations des homographies courantes H_{st}^n, H_g^n et H_d^n .

La redondance des contraintes spatio-temporelles introduite grâce à la prise en compte des projections des primitives coplanaires détectées dans plusieurs couples d’images a deux intérêts significatifs. D’abord, les coordonnées des super-primitives peuvent être qualifiées de ”théoriques” car la précision sub-pixellique des projections correspond aux mesures obtenues en s’affranchissant des différents bruits de mesure de la chaîne d’acquisition. Ensuite, la mise en correspondance des primitives sur des couples de vues stéréo et consécutives permet de rejeter des couples de primitives non coplanaires qui peuvent vérifier un type d’homographies durant quelques itérations mais en aucun cas les deux.

Les résultats de la super-homographie sont toutefois dépendants de la quantité et de la

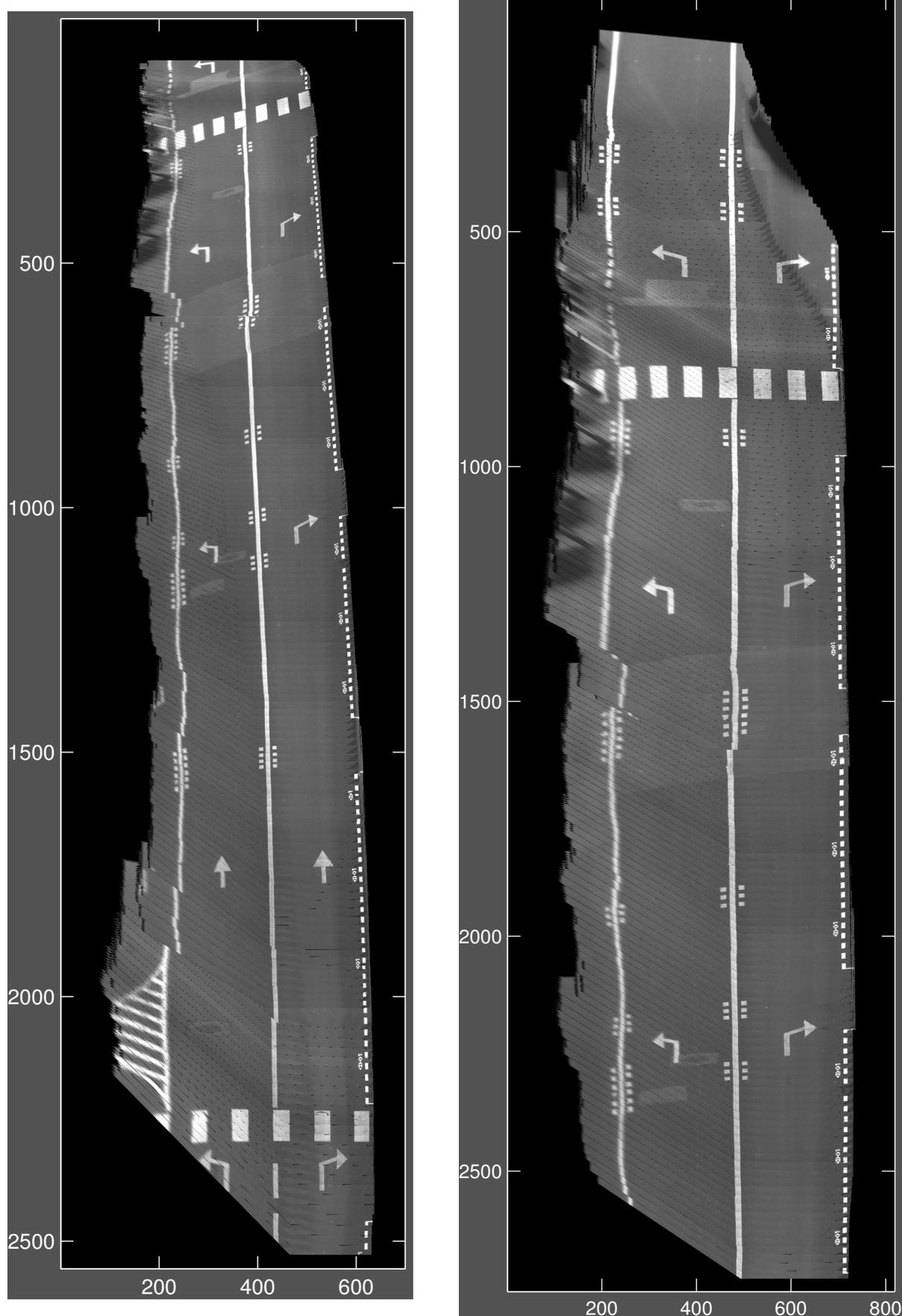


FIG. 6.13 – Reconstitution du plan de la route à partir de deux instants différents. A gauche, la reconstitution débute avec la séquence mais une erreur d'estimation de l'homographie à l'itération de 295 est à l'origine d'une légère erreur sur l'estimation de l'azimut de la caméra qui s'avère fatal lorsque le véhicule commence son déboîtement. A droite, la reconstitution commence à l'itération 200, ce qui permet d'éviter les erreurs de déboîtement.

qualité des couples de primitives mis en correspondance lors des estimations des homographies de l'itération courante. Lorsque au moins un des systèmes est mal conditionné, la pertinence des homographies extraites de la super-homographies peuvent être dégradées. Pour pallier à cela, nous avons introduit des primitives coplanaires virtuelles qui nous permettent de contraindre le mouvement apparent de la projection d'un plan dans des régions dépourvues originellement de primitives. Ces primitives virtuelles ne sont autres que des combinaisons linéaires des primitives identifiées comme coplanaires et connues dans les m vues. Les points d'intérêt virtuels utilisés correspondent à l'intersection du faisceau de bandes principales, mis en correspondance et filtré tout au long de la séquence, avec des droites virtuelles, formées de super-primitives dont les projections sont connues dans les m images.

L'introduction de ces nouvelles primitives améliorent considérablement la répartition spatiale des primitives dans chaque image et robuste l'estimation de la super-homographie. La super-homographie fournit ainsi des estimations cohérentes d'homographies dont les systèmes mal conditionnés s'opposent généralement à une estimation à partir des seules primitives extraites dans l'image.

Enfin, nous présentons dans la dernière partie le résultat de la reconstitution du plan de la route observée en vue de dessus par une caméra virtuelle en réalisant un mosaicing de la projection du plan de la route au cours du déplacement du véhicule. Les images ainsi composées signifient que nous sommes parvenus à estimer fidèlement le déplacement du véhicule se déplaçant à 50 km/h sur une distance de plusieurs dizaines de mètres par la seule extraction de primitives sur une route très matérialisée sur le coté droit mais au revêtement homogène alors que la configuration du banc stéréo est peu adaptée à notre algorithme.

Ces résultats encourageants méritent toutefois de nombreuses améliorations. En effet, malgré tous nos efforts, nous ne sommes pas parvenus à présenter une reconstitution du plan de la route à partir de la séquence enregistrée sur le port à Antibes. La région de la route observée dans les images du couple stéréo durant une dizaine d'itérations est tellement réduite que nous nous félicitons déjà d'être parvenus à estimer durant la majorité des itérations qui caractérisent cette période le fort mouvement de rotation du véhicule avec des configurations de primitives aussi dégénérées.

Chapitre 7

Conclusion et perspectives

Au mois d'octobre 2004, date à laquelle ont été enregistrées les dernières séquences vidéo à notre disposition, les ressources informatiques nécessaires à l'acquisition des images synchronisées en pleine résolution de la paire stéréo interdisait encore à l'enregistrement des données de ne serait-ce qu'un seul des nombreux autres capteurs qui équipaient le véhicule d'expérimentation. Ce dernier étant certainement l'un des mieux instrumentés actuellement en France, cette première remarque permet de situer le contexte scientifique de cette étude. Il nous faut cependant souligner que cette contrainte matérielle a non seulement fortement influencé la méthode que nous avons développée mais s'est aussi opposée à la démarche rigoureuse de confrontation de nos résultats à une vérité terrain ou capteur dont nous ne disposons pas.

Notre objectif a donc été d'extraire le maximum d'informations sur le déplacement d'un véhicule en milieu urbain à partir des seules données fournies par un banc stéréo non-calibré. Dans un premier temps, nous avons segmenté dans le couple d'images les principaux plans susceptibles d'être observés. A partir d'un modèle de projection de route simple, nous parvenons à détecter les limites du plan de la route en supposant que le marquage au sol qui délimitent les voies de circulation est formé de bandes équidistantes parallèles à la direction de la route, tout au moins au premier plan de l'image.

La région d'intérêt apparentée à la projection de la route dans les images peut contenir des obstacles. L'estimation de l'homographie stéréo induite par le plan de route entre les deux images du couple permet d'identifier quelles sont parmi les primitives extraites par un détecteur de Harris celles qui sont détectées sur le plan de la route. Cette opération réalisée, le déplacement de chaque caméra entre deux poses consécutives est calculé en estimant l'homographie dans les séquences temporelles des caméras gauche et droite qui met en correspondance un maximum des primitives coplanaires identifiées lors de l'estimation des homographies stéréo.

Le nombre ainsi que le niveau de caractérisation peu élevé des primitives pouvant être détectées sur une région aussi peu texturée que la route a conduit la plupart des auteurs à se détourner d'une méthode basée sur l'extraction de primitives pour estimer le déplacement d'un véhicule sur route. De plus, en raison des nombreuses incertitudes sur les coordonnées des primitives, la majorité des seuils d'acceptation utilisés pour vérifier que deux primitives sont en correspondance ou coplanaires sont choisis relativement lâches. La principale répercussion de ces deux remarques est que la mise en correspondance de primitives à partir de l'estimation d'une homographie s'avère peu discriminante : suivant la distribution spatiale des primitives réellement coplanaires, certains couples détectés proches du plan peuvent être considérés de manière erronée coplanaires.

Dès lors, l'introduction simultanée de contraintes spatio-temporelles sur les projections des

primitives coplanaires permet de discriminer parmi les couples de primitives ceux qui sont effectivement coplanaires. La super-homographie généralise à plusieurs vues la règle de composition des homographies et fournit en un unique calcul les coordonnées théoriques des primitives coplanaires avec une précision sub-pixelique. Les cas où la répartition spatiale des primitives n'est pas uniforme dans les différentes images est compensé par l'introduction de primitives virtuelles, qui sont des combinaisons linéaires des primitives existantes.

La reconstitution du plan de la route sur plusieurs dizaines de mètres démontre qu'il est possible d'estimer le déplacement d'un véhicule sur route à partir d'un seul système de stéréovision embarqué. Autrement dit, un capteur de stéréovision peut être considéré comme un système palliatif aux pertes d'informations des systèmes de géo-référencement dans le cas d'un déplacement en milieu urbain.

Les bâtiments qui structurent l'environnement urbain sont généralement perçus comme des obstacles à la réception de signaux satellitaires empêchant l'utilisation en continu d'un système de géo-référencement. L'intérêt de la méthode proposée réside dans le fait que les éléments structurants de l'environnement urbain sont considérés comme autant de sources de surfaces planaires texturées sur lesquelles une estimation du déplacement du véhicule peut être effectuée. La transposition de la méthodologie mise en oeuvre pour la route est directe pour toute surface plane de la scène statique.

Les estimations de déplacement obtenues grâce au suivi de façades verticales n'ont toutefois pas encore été validées. Le nombre important de primitives pouvant être extraites ainsi que leur caractérisation sont sans commune mesure avec les primitives extraites sur la route. Cependant, il est probable que la distance qui sépare les façades des caméras ne permettra pas d'obtenir des estimations de déplacements aussi précises que celles fournies par les primitives détectées au premier plan sur la route.

Par contre, la redondance des informations recueillies grâce au suivi de plusieurs plans devrait permettre d'extraire des informations relativement fiables qui pourraient être utilisées comme une estimation grossière du déplacement du banc stéréo entre deux poses. Inséré dans un processus de filtrage temporel, cette estimation permettrait de disposer d'une prédiction du déplacement du banc stéréo et robustifierait l'étape de mise en correspondance des primitives coplanaires détectées sur la route entre deux images consécutives.

Le fait de ne pas disposer d'une prédiction du déplacement entre deux poses est la difficulté majeure à laquelle nous avons été confronté au cours de cette thèse. Nous avons mis en évidence que l'estimation du déplacement à partir d'un échantillonnage temporel de période fixe n'est pas adapté à la gamme de vitesse d'un véhicule en milieu urbain :

- trop lent, le mouvement apparent des primitives n'est pas suffisamment contraignant, plusieurs combinaisons de déplacements sont susceptibles de correspondre,
- trop rapide, des primitives sortent rapidement du champ de vision, l'estimation du déplacement est effectué à partir d'un jeu de primitives réduit et mal réparti spatialement.

Dans les deux cas, l'estimation de l'homographie induite par le plan observé est délicat en raison d'un mauvais conditionnement du système formé par les coordonnées des primitives mises en correspondance.

Nous atteignons alors sans surprise les limites de la méthode : l'emploi d'un unique capteur est à ce niveau un réel handicap auquel il sera toutefois facile de trouver une solution. L'introduction des estimations de déplacement fournies par un odomètre peuvent, par exemple, être utilisées de façon à déclencher l'acquisition des images en fonction de la distance parcourue par le véhicule. L'adaptation du pas temporel d'échantillonnage se justifie d'autant plus qu'il s'avère inutile d'estimer le déplacement du véhicule lorsque celui-ci est immobile.

Il nous paraît opportun de prochainement valider la segmentation du plan de la route sur des séquences moins rectilignes. A notre décharge, la majorité des séquences urbaines à notre disposition sont essentiellement constituées de portions de lignes droites entrecoupées de jonctions courbes et de croisements en fourche ou en croix. Le modèle linéaire de route devra certainement être amélioré. D'autre part, nous ne disposons pas de séquences urbaines où le trafic est très dense. De nouvelles expérimentations doivent être menées de façon à vérifier qu'un champ de vision restreint aux premiers mètres durant de longues périodes suffit à estimer son déplacement entre deux poses.

Nous avons en effet l'intime conviction que la méthode que nous avons développée est réellement adaptée au flux de circulation en milieu urbain. Le suivi d'une dizaine de points équi-répartis autour des bandes pointillées qui délimitent la voie de circulation d'un véhicule sur une région de la route comprise entre 5 m et 10 m de profondeur suffit à correctement estimer le déplacement du véhicule sur quelques centaines de mètres.

Une partie moins évidente consiste à calibrer le système de stéréo-vision en ligne à partir du suivi de plusieurs plans. Différents éléments de bibliographie ont d'ores et déjà été regroupés sur le sujet. Dans l'hypothèse du déplacement du véhicule dans un environnement de type canyon urbain, l'introduction de contraintes géométriques fortes telles que le parallélisme et l'orthogonalité des façades verticales par rapport au plan de la route devrait favoriser la mise en oeuvre d'une solution dédiée.

D'autre part, la configuration du banc stéréo doit être adaptée à nos ambitions. Au cours de cette étude, nous avons eu maintes fois l'occasion de mesurer à quel point l'emploi de caméras à large champ était indispensable à la mise en oeuvre de la méthode proposée. En effet, si nous faisons l'analogie avec un conducteur, le fait d'aborder une trajectoire en courbe ne restreint pas pour autant le champ de vision de la route. Le regard du conducteur se détourne et si besoin est, le conducteur a la possibilité de tourner la tête de façon à toujours conserver une vision globale de la route. C'est évidemment loin d'être le cas avec les séquences vidéo à notre disposition.

Une alternative, essentiellement coûteuse en développement, à ce problème serait de motoriser la paire stéréo de manière à opérer une étape de suivi sur le principal plan segmenté, en l'occurrence la route. L'objectif de la motorisation serait de compenser les mouvements de rotation du véhicule afin que la région de la route observée soit toujours identique. La mise en oeuvre de cette tâche consisterait à réaliser un asservissement sur les coordonnées du DVP dans une des deux images. Cette seule opération garantirait une observation optimale du plan de la route et fournirait par la même occasion l'azimut du véhicule par rapport à la direction de la route.

Une perspective intéressante, qui n'a été abordée que de manière indirecte dans cette thèse, est la détection d'obstacles. En effet, la segmentation des principaux plans qui structurent la scène statique devrait nous permettre d'élaborer une reconstruction d'un modèle 2,5D de la scène en ligne. Dès lors, tous les éléments de la scène qui ne vérifient pas le mouvement apparent de ces plans principaux peuvent être considérés comme des obstacles qu'il sera possible de discriminer par la suite entre obstacles statiques et obstacles dynamiques.

Annexe A

Détail des séquences utilisées

A.1 Séquence du port d'Antibes



FIG. A.1 – *Echantillons de la séquence du port d'Antibes enregistrés par la caméra gauche entre les images [770;1050] avec un pas de d'échantillonnage de 20.*

A.2 Séquence descente enregistrée à Versailles

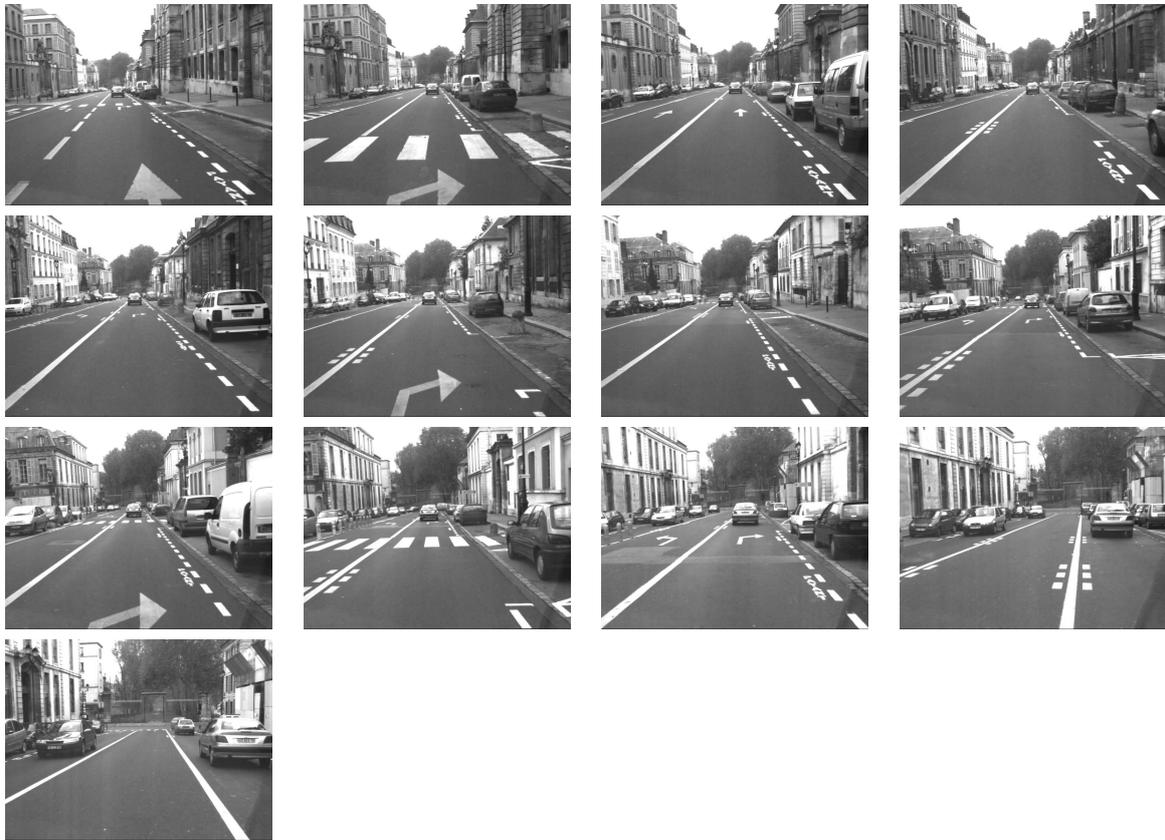


FIG. A.2 – Echantillons de la séquence descente enregistrés à Versailles par la caméra droite entre les images [260 ;500] avec un pas de d'échantillonnage de 20.

Annexe B

Algorithme de RANSAC

Jusqu' alors nous avons supposé que l'estimation de l'homographie entre deux jeux de FPs était perturbée uniquement par la présence de bruit de mesure sur les coordonnées des primitives en correspondance. Or, la prise en compte accidentelle de couples de FPs non-coplanaires est à l'origine d'erreurs d'estimation, difficilement identifiables, et souvent irrémédiables. L'objectif est maintenant de traiter le cas de ces mesures aberrantes. Pour cela, nous utilisons la méthode robuste de RANSAC.

B.1 Principe

L'estimateur robuste RANSAC (RANdom SAmple Consensus), élaboré par [Fischler and Bolles, 1981], permet de déterminer de manière statistique quels sont les échantillons qui vérifient un modèle donné parmi une population donnée. L'algorithme de RANSAC est largement employé car le résultat fourni est correct même pour une large proportion de mesures aberrantes dans la population. Trois paramètres sont à définir a priori :

- quelle est la tolérance t à partir de laquelle un échantillon ne vérifie plus le modèle ?
- quelle est la proportion d'échantillons de la population qui ne vérifient pas le modèle (outliers) ?
- quelle est la probabilité d'écarter un échantillon qui vérifie le modèle en considérant une tolérance t fixée ?

La méthode consiste à tirer aléatoirement parmi la population le nombre minimal d'éléments qui permettent d'estimer les paramètres du modèle et de comptabiliser le nombre d'éléments qui vérifient à la tolérance fixée près chacune des estimations aléatoires. Le tirage solution est celui qui regroupe le plus grand nombre d'éléments sous le même modèle. L'estimation du modèle est alors affinée en prenant en compte l'ensemble des éléments qui le vérifie.

La tolérance t est généralement imposée de manière empirique en fonction du modèle et du type de données à traiter. Par contre, il est d'usage de fixer les deux derniers paramètres qui permettent de fixer le nombre de tirages aléatoires à opérer pour obtenir un résultat fiable. Ainsi, la probabilité p qu'un échantillon sélectionné parmi les N tirages aléatoires soit dépourvu d'outlier est fixée à 99% ($p = 0,99$) et la probabilité α qu'un élément retenu vérifie effectivement le modèle est égale à 95% ($\alpha = 0,95$). Autrement dit, la probabilité de retenir de manière erronée un outlier comme élément qui vérifie le modèle est de 5%.

Le nombre de tirages aléatoires N doit donc être suffisamment élevé de façon à vérifier les deux probabilités précédentes. Supposons δ la probabilité qu'un élément vérifie le modèle, la

probabilité qu'un élément soit un outlier est donc $(1 - \delta)$. Soit s le nombre minimal d'éléments pour former un échantillon, la probabilité d'obtenir un échantillon qui contienne un outlier est donc $\epsilon = (1 - \delta^s)$. La probabilité $1 - p$ de tirer un échantillon avec un outlier après N tirages aléatoires est égale à $(1 - \delta^s)^N$, nous en déduisons la valeur de N :

$$N = \frac{\log(1 - p)}{\log(1 - (1 - \epsilon)^s)} \quad (\text{B.1})$$

La détermination du seuil d'acceptation d'un élément à un modèle donné dépend de l'erreur qui lui est associée. L'erreur de mesure est supposée gaussienne avec une moyenne nulle et une déviation standard σ . Dans le cas d'un point censé vérifier une homographie, l'écart entre sa position initiale et la position idéale varie en fonction des deux axes : la codimension de l'erreur est donc égale à 2. L'erreur de mesure $\|\mathbf{p}_k - \widehat{\mathbf{p}}_k\|^2$ suit donc une loi de type χ_2^2 avec deux degrés de liberté. La probabilité que la valeur d'une variable aléatoire de type χ_2^2 soit inférieure à k^2 est fournie par la loi de distribution cumulée $F_2(k^2) = \int_0^{k^2} \chi_2^2(\xi) d\xi$. La tolérance t est alors choisie de manière à vérifier l'égalité suivante :

$$t^2 = F_2^{-1}(\alpha) \cdot \sigma^2 \quad (\text{B.2})$$

soit pour $\alpha = 0,95$, $t = \sqrt{5,99}\sigma$.

B.2 Implémentation

Dans notre cas, l'emploi d'une telle méthode se justifie essentiellement lors de l'initialisation de l'algorithme. Nous rappelons que tant que toutes les bandes principales ne sont pas identifiées dans le couple d'images stéréo, la région d'intérêt se résume à la partie basse de l'image délimitée par les droites-supports sous la ligne d'horizon. De nombreuses primitives sont susceptibles d'être détectées en dehors de la région de la route supposée coplanaire. De plus, la présence éventuelle d'obstacles dans les régions ainsi segmentées sont à l'origine de nouvelles primitives non-coplanaire.

Ainsi, la proportion de primitives effectivement coplanaires détectées dans les régions segmentées est dans un premier temps totalement inconnu. C'est pourquoi l'emploi d'un estimateur robuste se justifie pleinement lors de l'initialisation de l'algorithme où l'étape de mise en correspondance est de surcroît initialisée à partir d'une prédiction grossière de l'homographie stéréo H_{st}^0 .

En supposant qu'une majorité de couples de primitives soient détectés sur la route, l'algorithme de RANSAC nous assure de trouver une fraction des couples supposés sur la route. L'estimation itérative qui suit cette étape de pré-sélection doit permettre d'identifier d'éventuels couples de primitives qui ne vérifiaient pas l'homographie formée par les quatre couples tirés aléatoirement, mais qui formeront à la fin de la première estimation H_{st}^1 les premières primitives coplanaires. L'emploi d'un tel algorithme lors des itérations suivantes ($n > 1$) est inutile : les contraintes imposées par la connaissance de l'environnement ainsi que le suivi des primitives coplanaires permettent de s'affranchir de cette coûteuse étape en temps de calcul en raison du nombre de tirages aléatoires.

Bibliographie

- G. Adiv. Automatic recovery of relative camera rotations for urban scenes. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume PAMI-7, pages 384–401, Head Island, SC, USA, July 1985.
- D. Aubert, K. Kluge, and C. Thorpe. Autonomous navigation of structured city roads. In *The International Society for Optical Engineering 1388*, pages 141–151, Boston, MA, USA, Nov. 8-9 1990.
- R. Aufrère, R. Chapuis, and F. Chausse. A model-driven approach for real-time road recognition. *Machine Vision and Applications*, 2(13) :95–107, 2001.
- H. Bechetoille. Localisation absolue d’un mobile à l’aide d’un radar continu à modulation de fréquence. Rapport de dea, Université Pierre et Marie Curie, Paris VI, 1994.
- S. Benhimane and E. Malis. Mise en correspondance d’images à différentes résolutions à l’aide d’invariants aux paramètres intrinsèques. In *14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA’04)*, Toulouse, France, Jan. 27-30 2004.
- S. Benhimane and E. Malis. Vision-based control for car platooning using homography decomposition. In *IEEE International Conference on Robotics and Automation (ICRA’05)*, Barcelona, Spain, 2005.
- A. Benschraï, A. Bertozzi, A. Broggi, A. Fascioli, S. Mousset, and G. Toulminet. Stereo vision-based feature extraction for vehicle detection. In *IEEE Intelligent Vehicle Symposium (IV’02)*, pages 58–63, Versailles, France, June 17-21 2002.
- M. Bertozzi, A. Broggi, M. Cellario, A. Fascioli, P. Lombardi, and M. Porta. Artificial vision in road vehicles. *IEEE*, 90(7) :1258–1271, 2002.
- P. Bonnifait. *Localisation précise en position et attitude des robots mobiles d’extérieur évolution lente*. Thèse, Ecole Centrale de Nantes, Nantes, France, Nov. 1997.
- J. Borenstein, H.R. Everett, and L. Feng. *Where am I? Sensors and methods for mobile robot positioning*. University of Michigan, <http://www-personal.umich.edu/~johannb/shared/pos96rep.pdf>, Apr. 1996.
- S. Botton, Y. Duquenne, Y. Egels, M. Evens, and P. Willis. *GPS, localisation et navigation*. Editions Hermès, Paris, 1997.

- A. Branca, E Stella, and A Distante. Mobile robot navigation using egomotion estimates. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'97)*, Grenoble, France, Sep. 8-12 1997.
- A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi. Shape-based pedestrian detection. In *IEEE Intelligent Vehicles Symposium*, Dearbon (MI), USA, Oct. 2000.
- J.F. Canny. A computational approach to edge detection. In *IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 679–698, 1986.
- T. Chen. Development of a vision-based positioning system for high density area. In *Asian Conference on Remote Sensing (ACRS'99)*, Hong Kong, China, Nov 22-25 1999.
- G. Dudeck and D. Jugessur. Robust place recognition using local appearance based methods. In *IEEE Transactions on Robotics and Automation (ICRA'00)*, pages 1030–1035, San-Francisco, CA, Apr. 24-28 2000.
- F.-X. Espiau. *Métrologie 3D par vision active sur des objets naturels sous-marins*. Thèse, INRIA, Sophia-Antipolis, France, Feb. 2002.
- O. Faugeras. *Three-Dimensional Computer Vision : a Geometric Viewpoint*. MIT Press, 1993.
- M.A. Fischler and R.C. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. In *ACM*, pages 381–395, June 1981.
- U. Franke, D. Gavrilla, S. Goerzig, F. Linder, F. Paetzold, and C. Woehler. Autonomous driving approaches downtown. *IEE Intelligent Systems*, 13(6), 1999.
- J.-M. Gaudin. *Conception d'un système hybride GPS-Estime pour la localisation en condition de visibilité satellitaire réduite*. Thèse, Université de Montpellier II, Montpellier, France, Dec. 2000.
- D. M. Gavrila, J. Giebel, and S. Munder. Vision-based pedestrian detection : the protector system. In *IEEE Intelligent Vehicle Symposium (IV'04)*, Parma, Italy, June 2004.
- A. Georgiev and P.K. Allen. Localization methods for a mobile robot in urban environments. In *IEEE Transactions on Robotics and Automation (ICRA'04)*, 2004.
- J.J. Gibson. *The perception of visual world*. Greenwood Press, Boston, Ma, 1950.
- T. Gotoh, M. Kudo, J. Toyama, and M. Shimbo. Geometry reconstruction of urban scenes by tracking vertical edges. In *3rd International Conference on Knowledge-based intelligent information engineering systems*, pages 455–458, Adelaide, Australia, Aug. 31-Sep.1 1999.
- J.J. Guerrero and C. Sagues. *Uncalibrated vision-based on lines for robot navigation*, volume 6, pages 759–777. Elsevier, mechatronics 11 edition, Sept 2001.
- R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- J.-B. Hayet. *Contribution à la navigation d'un robot mobile sur amers visuels texturés dans un environnement structuré*. Thèse, Université Paul Sabatier, Toulouse, France, Jan. 2003.

- B. Heisele. Motion-based object detection and tracking in color images sequences. In *Fourth Asian Conference on Computer Vision (ACCV'00)*, pages 1028–1033, Taipei, China, Jan 8-11 2000.
- C. Hilario, J.M. Collado, J.M. Armingol, and de la Escalera A. Driver assistance system based on computer vision for vehicle detection. In *5th IFAC Symposium on Intelligent Autonomous Vehicles (IAV'04)*, Lisbon, Portugal, July 5-7 2004.
- C. T. Hong, T. H. Rasmussen, and C. E. Shneier. Road detection and tracking for autonomous mobile robots. In *PIE Aerosense Conference*, Orlando, FL, Apr. 2002.
- R. Horaud. Vision 3d projective, affine et euclidienne. In *Cours de DEA*, 1998.
- B.K.P. Horn and E.J. Weldon. Direct methods for recovering motion. *IJCV*, 2(1) :51–76, June 1988.
- Z. Hu and K. Uchimura. Dynamical road modeling and matching for direct visual navigation. In *Sixth IEEE Workshop on Applications of Computer Vision (WACV 2002)*, pages 237–241, Orlando, FL, USA, Dec. 3-4 2002.
- F. Jurie and M. Dhome. A simple and efficient template matching algorithm. In *8th IEEE International Conference on Computer Vision*, pages 544–549, Vancouver, BC, Canada, July 7-14 2001.
- M. Kais, S. Morin, A. de la Fortelle, and C. Laugier. Geometrical model to drive vision systems with error propagation. In *8th International Conference on Control, Automation, Robotics and Vision (ICARCV'04)*, Kunming, China, Dec. 3-9 2004.
- W. Kasprzak, H. Niemann, and D. Wetzel. Adaptive road parameter estimation in monocular image sequences. In *5th British Machine Vision Conference (BMVC'94)*, pages 691–700, York, U.K., Sept. 13-16 1994.
- H. Katsura, J. Miura, M. Hild, and Y. Shirai. A view-based outdoor navigation using object recognition robust to changes of weather and seasons. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'03)*, pages 2974–2979, Las Vegas, Nev., USA, Oct. 27-31 2003.
- Qifa Ke and Takeo Kanade. Transforming camera geometry to a virtual downward-looking camera : Robust ego-motion estimation and ground-layer detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, pages 390–7, Madison, USA, June 2003.
- Jeong-Hun Kim, Jong-Soo CHOI, Kyung-Hyun Yoon, and Young-Ho Chai. Vision-based augmented reality by geometric constrained camera calibration. In *Eighth Korea-Japan Joint Workshop on Frontiers of Computer Vision*, Sapporo, Japan, 2002.
- K. Kluge and S. Lakshmanan. Lane boundary detection using deformable templates : effects of image subsampling on detected lane edges. In *Second Asian Conference on Computer Vision (ACCV'95)*, pages 329–339, Singapore, Dec. 5-8 1995.
- L. Kosecka and W. Zhang. Efficient computation of vanishing points. In *IEEE International Conference on Robotics and Automation ICRA'02*, pages 3321–3327, Washington DC, USA, May 2002.

- R. Labayrade and D. Aubert. In-vehicle obstacles detection and characterization by stereovision. In *In-Vehicle Cognitive Computer Vision Systems (IVC2VS)*, pages 13–19, Graz, Austria, Apr. 03 2003.
- X. Lebegue and J.K. Aggarwal. Generation of architectural cad models using a mobile robot. In *IEEE International Conference on Robotics and Automation (ICRA'94)*, pages 711–717, San-Diego, CA, USA, May 8-13 1994.
- J.C.H. Leung and G.F. McLean. Vanishing point matching. In *IEEE International Conference on Image Processing (ICIP'96)*, Lausanne, Switzerland, Sept. 16-19 1996.
- J. Lobo and J. Dias. Vision and inertial sensor cooperation ausing gravity as a vertical reference. *IEEE Transactions on pattern analysis and machine intelligence*, 25(12) :1597–1607, Dec. 2003.
- Jianye Lu, Ming Yang, Hong Wang, and Bo Zhang. Vision-based real-time road detection in urban traffic. In *The International Society for Optical Engineering (SPIE'02)*, pages 75–82, San-Jose, CA, USA, 2002.
- E. Malis and R. Cipolla. Multi-view constraints between collineations : application to self-calibration from unknown planar structures. In *European Conference on Computer Vision (ECCV'00)*, volume 2, pages 610–624, Dublin, EIRE, June 2000.
- A. Mallet. *Localisation d'un robot mobile autonome en environnements naturels*. Thèse, Institut National Polytechnique de Toulouse, Toulouse, France, Mai 2001.
- R. Mandelbaum, G. Salgian, H. Sawhney, and M. Hansen. Terrain reconstruction for ground and underwater robots. In *IEEE Transactions on Robotics and Automation (ICRA '00)*, pages 879–884, San-Francisco, CA, Apr. 24-28 2000.
- M. Mata, J.M. Armingol, A. de la Escalera, and M.A. Salichs. A visual landmark recognition system for topological navigation of mobile robots. In *IEEE Transactions on Robotics and Automation (ICRA '01)*, pages 1124–1129, Seoul, South Korea, May 21-26 2001.
- R. Mester and M. Muehlich. Improving motion and orientation estimation using an equilibrated total least squares approach. In *International Conference on Image Processing*, volume 2, pages 929–932, Thessaloniki, Greece, Oct. 7-10 2001.
- Hans Moravec. Towards automatic visual obstacle avoidance. In *The 5th International Joint Conference on Artificial Intelligence*, pages 584–586, Cambridge, U.K., 1977.
- R. Okada, Y. Taniguchi, K. Furukawa, and K. Onoguch. Obstacle detection using projective invariant and vanishing lines. In *9th IEEE International Conference on Computer Vision (ICCV'03)*, Nice, France, Oct. 13-16 2003.
- M. Okutomi, K. Nakano, J. Maruyama, and T. Hara. Robust estimation of planar regions for visual navigation using sequential stereo images. In *IEEE International Conference on Robotics and Automation (ICRA '02)*, pages 3321–3327, Washington DC, USA, May 2002.
- J.-M. Vézien P.-L. Bazin. Shape and motion estimation from geometry and motion modeling. In *Machine Vision Applications*, pages 1612–1619, Tokyo, Japan, Dec. 2001.

- N. Pears and L. Bojian. Ground plane segmentation for mobile robot visual navigation. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'01)*, pages 1513–1518, Maui, HI, USA, Oct. 29–Nov 3 2001.
- M. Pilu. A direct method for stereo correspondence based on singular value decomposition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pages 261–266, San Juan, Puerto Rico, June 17–19 1997.
- T. Rabie, G. Auda, A. El-Rabbany, A. Shalaby, and B. Abdulhai. Active-vision based traffic surveillance and control. In *Vision Interface Annual Conference (VI'01)*, Toronto, Canada, June 7–9 2001.
- C. Rasmussen. Grouping dominant orientations for ill-structured road following. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04)*, Washington, DC, USA, 2004.
- P. Reisman, O. Mano, Avidan S., and Shashua A. Crowd detection in video sequences. In *5th IFAC Symposium on Intelligent Autonomous Vehicles (IAV'04)*, Lisbon, Portugal, July 5–7 2004.
- L. Robert. *Perception stéréoscopique de courbes et de surfaces tridimensionnelles. Applications à la robotique mobile*. Thèse, Ecole Polytechnique, 1993.
- C. Rother. A new approach to vanishing point detection in architectural environments. In *11th British Machine Vision Conference (BMVC'00)*, pages 647–655, Bristol, UK, Sept. 11–14 2000.
- E Royer, M Lhuilier, M. Dhome, and T. Chateau. Towards an alternative gps sensor in dense urban environment from visual memory. In *15th British Machine Vision Conference (BMVC'04)*, London, U.K., Sept. 7–9 2004.
- C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris*. Thèse, Institut National Polytechnique de Grenoble, Grenoble, France, Jul.2nd 1996.
- G. Scott and H. Longuet-Higgins. An algorithm for associating the features of two patterns. In *Royal Society of London*, volume B244, pages 21–26, 1991.
- S. Se and M. Brady. Road feature detection and estimation. *Machine Vision and Applications*, 14(3) :157–165, 2003.
- A. Shashua, Y. Gdalyahu, and G. Hayon. Pedestrian detection for driving assistance systems : Single-frame classification and system level performance. In *IEEE Intelligent Vehicles Symposium (IV'04)*, Parma, Italy, June 2004.
- G. Silveira, J. Azinheira, P. Rives, and S. Bueno. Line following visual servoing for aerial robots combined with complementary sensors. In *IEEE International Conference on Advanced Robotics (ICRA'03)*, Coimbra, Portugal, July 2003.
- M. Snaith, D. Lee, and P. Probert. A low-cost system using sparse vision for navigation in the urban environment. *Image and Vision Computing*, 16 :225–233, 1997.

- M.A. Sotelo, F.J. Rodriguez, and L. Magdalena. Vision-based navigation system for autonomous urban transport vehicles in outdoor environments. In *IEEE Intelligent Vehicles Symposium (IV'02)*, pages 52–57, Versailles, France, June 17-21 2002.
- G.P. Stein, O. Mano, and A. Shashua. A robust method for computing vehicle ego-motion. In *IEEE Intelligent Vehicle Symposium (IV'00)*, pages 362–368, Piscataway, NJ, USA, Oct. 3-5 2000.
- A. Takahashi, Y. Ninomiya, M. Ohta, M. Nishida, and M. Takayama. Rear view lane detection by wide angle camera. In *IEEE Intelligent Vehicle Symposium (IV'02)*, pages 148–153, Versailles, France, June 17-21 2002.
- A. Talukder, S. Goldberg, L. Matthies, and A. Ansar. Real-time detection of moving objects in a dynamic scene from moving robotic vehicles. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'03)*, pages 1308–1313, Las Vegas, Nev., USA, Oct. 27-31 2003.
- C.J. Taylor, J. Malik, and J. Weber. A real-time approach to stereopsis and lane-finding. In *IEEE Intelligent Vehicles Symposium (IV'96)*, pages 207–212, New-York, USA, 1996.
- Q. M. Tian, Y. P. Luo, and D. C. Hu. Pedestrian detection in nighttime driving. In *Third International Conference on Image and Graphics (ICIG'04)*, Hong Kong, China, Dec. 2004.
- R. Turchetto and R. Manduchi. Visual curb localization for autonomous navigation. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'03)*, pages 1336–1342, Las Vegas, Nev., USA, Oct. 27-31 2003.
- Alessandro Victorino. *La commande référencée capteur : une approche robuste au problème de navigation, localisation et cartographie simultanées pour un robot d'intérieur*. Thèse, Université de Nice Sophia-Antipolis, 2002.
- R. Wang, Y. Xu, and Y. Zhao. A vision-based road edge detection algorithm. In *Sixth IEEE Workshop on Applications of Computer Vision (WACV'02)*, pages 237–241, Orlando, Florida, USA, Dec. 3-4 2002.
- C. Watman, D. Austin, N. Barnes, G. Overett, and S. Thmpson. Fast sum of absolute differences visual landmark detector. In *5th IFAC Symposium on Intelligent Autonomous Vehicles (IAV'04)*, Lisbon, Portugal, July 5-7 2004.
- M. Werman, S. Banerjee, S.D. Roy, and Maolin Qiu. Robot localization using uncalibrated camera invariants. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (ICVPR'99)*, pages 353–359, Fort Collins, CO, USA, June 23-25 1999.
- Yong-In Yoon, Jang-Hwam Im, Dae-Hyun Kim, and Jong-Soo Choi. 3d reconstruction using three vanishing points from a single image. In *The 2002 International Conference on Circuit Systems, Computers and Communication*, Phuket, Thailand, July 2002.
- Q. Yu, H. Araujo, and H. Wang. Stereo-vision based real time obstacle detection for urban environments. In *11th International Conference on Advanced Robotics (ICAR'03)*, pages 1671–1676, Coimbra, Portugal, June 30-July 3 2003.
- L. Zhao and C. Thorpe. Stereo and neural network-based pedestrian detection. In *IEEE Transactions on Intelligent Transportation Systems*, volume 1, pages 148–154, Sept. 2000.