



HAL
open science

Contribution à la modélisation d'environnements par vision monoculaire dans un contexte de robotique aéro-terrestre

Sébastien Bosch

► **To cite this version:**

Sébastien Bosch. Contribution à la modélisation d'environnements par vision monoculaire dans un contexte de robotique aéro-terrestre. Automatique / Robotique. Université Paul Sabatier - Toulouse III, 2007. Français. NNT: . tel-00181794

HAL Id: tel-00181794

<https://theses.hal.science/tel-00181794>

Submitted on 24 Oct 2007

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

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

Thèse

préparée au

Laboratoire d'Analyse et d'Architecture des Systèmes du CNRS

en vue de l'obtention du

Doctorat de l'Université de Toulouse

délivrée par l'Université Toulouse III - Paul Sabatier

École Doctorale : Système

Spécialité : Informatique

par

Sébastien BOSCH

Contribution à la modélisation d'environnements par vision monoculaire dans un contexte de robotique aéro-terrestre

Soutenue le 1 octobre 2007 devant le jury composé de :

Ryad Benosman	Rapporteur
William Puech	Rapporteur
Olivier Strauss	Examineur
Patrice Dalle	Examineur
Guy Le Besnerais	Invité
Rachid Alami	Examineur, Président
Simon Lacroix	Examineur, Directeur de thèse

LAAS-CNRS

7, Avenue du Colonel Roche

31077 Toulouse Cedex 4

*A Cécile, qui a illuminé ces travaux et
à son mémorable hachis parmentier*

Remerciements

Je tiens à remercier Malik Ghallab, directeur du LAAS du CNRS, pour m'avoir accueilli dans ce laboratoire. Je remercie également Raja Chatila, directeur à mon arrivée du groupe de Robotique et Intelligence Artificielle, de m'avoir permis d'y travailler lors de mon arrivée. Je remercie de même Rachid Alami, directeur du groupe Robotique et Interactions, de m'avoir laissé le temps d'achever ma thèse.

J'exprime ma sincère reconnaissance à mes deux rapporteurs Ryad Benosman de l'ISIF et Oliver Strauss du département robotique du LIRMM, pour avoir accepté d'être rapporteurs de mes travaux de thèse ainsi que pour leur lecture pertinente du manuscrit. Merci à Guy Le Besnerais et Patrice Dalle, pour avoir accepté d'être membre de mon jury de thèse. Un grand merci également à William Puech, sans qui je n'aurais soutenu à temps.

Je voudrais adresser mes plus chaleureux remerciements à mon directeur de thèse Simon Lacroix, qui a su me témoigner toute sa confiance et son soutien au cours de ces quatre années de travail et qui a su me guider lorsque j'en avais le plus besoin ou me laisser du temps quand c'était nécessaire.

Je tiens aussi à exprimer ma reconnaissance toute particulière à toute la COMETS team : Jeremi Gancet, Matthieu Herb, Anthony Mallet et Leonardo Solaque, avec qui j'ai vécu tant de très bons moments et tellement appris.

Merci à tous les TAS : Sylvain, Gauthier, Panos, Thierry en vous souhaitant le meilleur pour la suite.

Mes remerciements ne pourraient oublier mes inestimables co-locataires du bureau SLAM : Abedallatif, Maxime et Thomas. Merci à vous pour votre patience exemplaire à écouter mes chantonades ou mes tapotements agaçés. Un grand merci à tous les amateurs de café que je n'ai déjà cité : Alexandre, Fred, Leonard, Guillaume et Nicolas. Merci pour tous nos échanges et merci pour votre amitié.

Many thanks to Fernando for our exchange of views and for his "english of the mountains". Many wishes for you and Vicky.

Enfin un grand merci aux ours polaires qu'ils soient de Toulouse ou d'ailleurs, et bien évidemment toutes mes pensées vont à ma famille et à mes chers parents.

Résumé

Les travaux de cette thèse s'articulent principalement autour de la détection de surfaces planes à partir de séquences d'images faiblement calibrées. La détection de telles surfaces offre des possibilités pour l'atterrissage autonome de drones, la coopération air/sol par la fusion de modèles de traversabilité ou encore la cartographie aérienne.

Considérant d'abord le contexte d'images aériennes faiblement localisées (GPS métrique, centrale inertielle bas coût...), nous exploitons les propriétés des homographies, qui définissent le déplacement entre deux images de points appartenant à un même plan du monde. Nous avons utilisé cette propriété pour segmenter les régions des images contenant la projection d'un plan. Un soin particulier a été donné à l'évaluation de différents algorithmes d'estimation linéaire robuste dans le cadre de l'estimation d'homographie à partir de points appariés. Des données de synthèse ont été utilisées afin de mettre en évidence l'influence des paramètres de l'environnement sur la qualité des estimées. Ces résultats ont été par la suite confrontés à des données réelles.

Ces techniques permettent de segmenter les points appariés selon qu'ils appartiennent ou non à une zone plane. Afin de produire une description dense (continue) des zones détectées, nous avons proposé une amélioration du score de corrélation normalisée croisée. En considérant un modèle adaptatif de l'influence des variances des niveaux de gris sur les scores de corrélation, nous avons introduit une méthode de segmentation automatique de zones planes offrant de bons résultats.

L'introduction de modèles probabilistes nous a permis de fusionner les observations au fur et à mesure du déplacement de la caméra, et de construire des grilles locales qui représentent la probabilité de planéité sur des zones.

Enfin dans l'optique de la coopération entre robots aériens et terrestres, nous avons étendu ces travaux au contexte de la robotique terrestre et de la fusion de modèles aéro-terrestres.

Table des matières

1	Introduction	1
1.1	Contexte : la robotique aéro-terrestre	1
1.2	Problématique considérée	2
1.3	Organisation du manuscrit	4
2	Géométrie et estimation en vision	5
2.1	Éléments de base	5
2.1.1	Notations	5
2.1.2	Espaces métrique, euclidien, affine, et projectif	6
2.1.3	Normes	7
2.1.4	La décomposition en valeurs singulières	7
2.2	Géométrie projective	8
2.2.1	Notions de géométrie projective	8
2.2.2	Reconstruction 3D	13
2.3	Estimation d'homographie à partir de correspondances de points	16
2.3.1	Considérations sur le système linéaire	17
2.3.2	Estimation robuste : approches itératives	19
2.3.3	Estimation robuste : méthodes de vote	21
2.3.4	Le cas de l'affinité	23
2.3.5	Estimation robuste d'homographie, un bilan	24
3	Détection de zones traversables	25
3.1	Objectifs, présentation de l'approche	25
3.2	Estimation d'homographie	27
3.2.1	Considérations sur l'estimation d'homographie	27
3.2.2	Analyse des estimateurs robustes avec des données de synthèse	29
3.2.3	Analyse des estimateurs robustes avec des données réelles	37
3.2.4	Algorithme d'estimation mis en oeuvre	41
3.3	Etude quantitative des possibilités de détection	41
3.4	Segmentation des zones planes	46
3.4.1	Score de corrélation ZNCC (Zero Normalized Cross Correlation)	46
3.4.2	Limitations du score ZNCC	47

3.4.3	Correction du score de corrélation ZNCC en ligne	49
3.4.4	Correction du score de corrélation ZNCC hors-ligne	52
3.5	Résultats	55
3.5.1	Robustesse	55
3.5.2	Facteur d'échelle	57
3.5.3	Influence du recouvrement	58
3.5.4	Comparaison entre les deux méthodes de correction	59
3.5.5	Problèmes en suspens	59
4	Mise à jour d'un modèle probabiliste de traversabilité	61
4.1	Introduction	61
4.2	Fusion des informations	62
4.2.1	Construction de la grille locale G_L	62
4.2.2	Cas d'observations binaires	62
4.2.3	Cas d'observations continues	65
4.3	Géoréférencement	67
4.3.1	Fusion utilisant les capteurs	68
4.3.2	Fusion par homographie	72
5	Vers une modélisation intégrant données terrestres et aériennes	77
5.1	Introduction	77
5.2	Modèles numériques de terrain	78
5.3	Modèles de l'environnement basés sur les segments	79
5.3.1	Modélisation par segments 3D pour un robot terrestre	80
5.3.2	Modélisation par segments 3D pour un robot aérien	81
5.4	Modèles de l'environnement basés sur les plans	84
5.5	Bilan	87
6	Conclusion et perspectives	89
A	Optimisation du calcul de score de corrélation ZNCC	93
B	Relation entre parallaxe et score de corrélation ZNCC	95
C	Extraction du mouvement à partir des homographies	97
D	Algorithmes de mise en correspondance exploités	101

Table des figures

1.1	Différents modèles nécessaires à la coopération entre robots terrestres et aériens	3
2.1	Modélisation d'un système avec deux caméras	9
2.2	Modélisation d'un système avec deux caméras dans le cas de points sur un plan	10
2.3	Illustration de la géométrie épipolaire	11
2.4	Topologie des techniques d'estimation par méthodes de vote	24
3.1	Carte de traversabilité établie par un robot terrestre	25
3.2	Algorithme de détection de plan	26
3.3	Illustration des étapes de la segmentation de zones planes	28
3.4	Illustration du phénomène de parallaxe relative à un plan	29
3.5	Illustration des données de synthèse	30
3.6	Représentation du pourcentage d' <i>outliers</i> détecté Ω_O et du résidu sur les <i>inliers</i> ϵ_I en fonction de d_I et l_O	32
3.7	Lignes de niveau de la surface Ω_O , associées à différents vecteurs normaux au plan \mathbf{n}_I , en fonction de d_I et l_O pour $\Omega_O = 80\%$	33
3.8	Lignes de niveau de la surface Ω_O , associées à différents angles de rotation ϕ entre les caméras, en fonction de d_I et l_O pour $\Omega_O = 80\%$	34
3.9	Lignes de niveau des surfaces Ω_O de chaque estimateur robuste, en fonction de d_I et l_O pour $\Omega_O = 80\%$	35
3.10	Lignes de niveau des surfaces Ω_O de chaque estimateur robuste, en fonction de d_I et l_O pour $\Omega_O = 80\%$	36
3.11	Comparaison des résidus sur les <i>inliers</i> ϵ_I avec ou sans normalisation des coordonnées	37
3.12	Résidus ϵ_I des estimateurs robustes sur des paires d'images aériennes . . .	38
3.13	Résidus ϵ_I des estimateurs robustes sur des paires d'images terrestres . . .	39
3.14	Géométrie d'une paire d'images dans un cas simplifié de translation pure .	42
3.15	Illustration de l'influence de d_O et d_I sur le résidu	44
3.16	Illustration des types d'artefacts de corrélation	48

3.17 Répartitions des scores de corrélation ZNCC en fonction de différentes classes de variance \mathcal{C}_l	49
3.18 Fonction de répartition discrète $F_l(s)$ des scores de corrélation ZNCC en fonction de différentes classes de variance \mathcal{C}_l pour deux séries d'images distinctes	50
3.19 Scores de corrélation ZNCC en fonction de différentes classes de variance pour une seule paire d'image	51
3.20 Comparaison entre intégrales de répartition et densités de probabilités expérimentales et modélisées	53
3.21 Comparaison entre intégrales de répartition et densités de probabilités expérimentales et modélisées pour une seule paire d'image	54
3.22 Mise en relation de la médiane S_M et de la variance σ^2	55
3.23 Illustration de la segmentation de zones planes	56
3.24 Influence du facteur d'échelle sur la segmentation de zones planes	57
3.25 Influence de la baseline sur la segmentation de zones planes	58
3.26 Comparaison entre correction des scores de corrélation en ligne et hors ligne	60
4.1 Illustration de fusions successives d'observations binaires dans une grille locale G_L	64
4.2 Illustration d'une grille locale construite à partir d'observations binaires . .	65
4.3 Représentation de la densité de probabilité $p(S_k \bar{c})$ déterminée expérimentalement à partir des résultats de la segmentation	66
4.4 Comparaison des méthodes de fusion binaire et continue	67
4.5 Illustration d'une grille locale construite à partir d'observations continues .	68
4.6 Géométrie de la projection de la grille locale G_L sur la grille globale G_G . .	69
4.7 Illustration de la projection d'une grille locale G_L vers une grille globale G_G	70
4.8 Illustration des problèmes de mises à jour de la grille globale G_G	71
4.9 Comparaison des mosaïques avec et sans zones planes	73
4.10 Illustration de la fusion par homographie de grilles locales G_L	74
4.11 Illustration de la fusion par homographie des probabilités de grilles locales G_L (probabilités)	75
5.1 Modèle numérique de terrain	79
5.2 Résultat de SLAM monoculaire avec segments	80
5.3 Illustration de l'estimation d'homographie à partir de segments	83
5.4 Illustration de la mise en correspondance de segments par homographie . .	84
5.5 Illustration du déplacement pour un robot terrestre	85
5.6 Illustration d'une grille locale construite à partir d'images terrestres	88
A.1 Schéma de principe de l'optimisation du calcul ZNCC	94
B.1 Illustration de la probabilité d'obtenir une parallaxe selon le score de corrélation	95

C.1	Extraction du mouvement à partir des homographies pour des images aériennes	98
C.2	Extraction du mouvement à partir des homographies pour des images terrestres	99
D.1	Résultat de l'algorithme de mise en correspondance de points	102
D.2	Résultat de l'algorithme de mise en correspondance de segments	103

Liste des tableaux

2.1	Caractérisation des transformations de l'espace 3D	6
2.2	Influence du pourcentage d' <i>outliers</i> r_o et de la probabilité de réaliser un tirage sans <i>outliers</i> P sur le nombre de tirages aléatoires m nécessaires . . .	23
3.1	Moyenne des résidus ϵ_I des estimateurs robustes sur des paires d'images aériennes	40
3.2	Moyenne des résidus ϵ_I des estimateurs robustes sur des paires d'images terrestres	40
3.3	Correspondance entre des résidus ϵ_O en pixels et des outliers à une distance l_O du plan de référence $\mathbf{\Pi_I}$	45
3.4	Définition des distributions de Weibull et Gumbel (aussi appelée distribution de Fisher-Tippett)	52

Liste des Algorithmes

1	Algorithme générique d'estimation d'homographie par méthode de vote . . .	22
2	Calcul automatique des seuils τ_{S_i} des scores de corrélation ZNCC	53
3	Optimisation du calcul du score de corrélation ZNCC	94

Chapitre 1

Introduction

1.1 Contexte : la robotique aéro-terrestre

La mise en œuvre de véhicules sans pilote en environnements extérieurs se justifie principalement pour des missions d'intervention en milieu hostile et pour des missions de surveillance à caractère répétitif. Les contextes applicatifs sont nombreux, essentiellement dans les domaines de l'environnement et de la sécurité civile (missions d'exploration, de surveillance ou d'observation), de la défense (missions de logistique, de reconnaissance, voire de feu), et même de l'exploration scientifique – planétaire notamment.

Dans ces contextes, l'autonomie des engins, même si elle n'est pas totale, permet de libérer l'opérateur des tâches de télé-pilotage, des tâches routinières et des fonctions de sécurité du système exigeant une forte réactivité. Depuis le début des années 1990, de nombreux travaux portant sur l'autonomie des véhicules terrestres ont mobilisé les chercheurs en robotique, et si il n'existe aujourd'hui pas encore de systèmes opérationnels capable d'effectuer des missions autonomes dans des environnements mal connus, des grands progrès ont été accomplis – et particulièrement illustrés par le Darpa Grand Challenge en 2005 [JFR 06].

Dans la majorité des contextes considérés, il y a un grand intérêt à faire coopérer plusieurs engins pour réaliser les missions : une approche multi-engins offre bien sûr une plus grande redondance et une complémentarité des moyens d'action et de recueil d'informations. En particulier, la coopération entre engins terrestres et aériens offre des possibilités très intéressantes, notamment grâce à leurs complémentarités de perception et de déplacement. Différents schémas de coopération peuvent être envisagés entre robots terrestres et aériens : il vont de l'assistance mutuelle (les robots terrestres exploitent des informations collectées par les robots aériens pour définir leurs déplacements, ou les robots terrestres fournissent un modèle précis de zones d'atterrissage aux robots aériens par exemple), à la réalisation de tâches conjointes, telle l'exploration coordonnée d'une région donnée de l'environnement.

La mise en œuvre de tels systèmes requiert bien entendu le développement de travaux

couvrant un très large spectre de la robotique. Parmi ces travaux, la *modélisation de l'environnement intégrant données terrestres et aériennes* joue un rôle essentiel. En effet, la construction de modèles de l'environnement peut être la finalité de la mission des robots (cas de la robotique d'exploration), mais elle est surtout nécessaire à la prise des différentes décisions pour mener à bien les missions impliquant les robots : c'est par l'intermédiaire des représentations de l'environnement construites que les robots pourront établir des schémas de coopération.

Projets existants. Ce n'est que récemment que des chercheurs en robotique considèrent les problèmes posés par la coopération entre robots terrestres et aériens – le plus souvent dans des contextes militaires. Ainsi, au SPAWAR de San Diego [Harbour 05, Mullens 06], différents scénarios ont été définies : les robots aériens détectent des cibles au sol et transmettent leurs coordonnées aux robots terrestres qui assurent leur suivi, ou bien des missions d'exploration consistent à fusionner données aériennes et terrestres pour assurer une surveillance de l'environnement. Malheureusement, peu d'informations sur l'état d'avancement de ces scénarios et sur les fonctions développées pour les mettre en œuvre sont publiées. D'autres travaux ont été menés dans le laboratoire GRASP de l'université de Pennsylvanie [Mullens 04], où en particulier des cibles posées au sol sont localisées grâce à des données perçues par les engins terrestres et aériens [Chaimowicz 05].

Remarquons que parmi ces travaux réalisés par des équipes de chercheurs en robotique, aucun ne concernent la fusion d'informations terrestres et aériennes pour la construction de modèles de l'environnement. C'est plutôt dans le communauté des Systèmes d'Information Géographiques que de tels travaux sont actuellement menés, dans l'objectif de construire des modèles 3D réalistes de zones urbaines [Frueh 01].

Au LAAS, la robotique aéro-terrestre constitue le contexte dans lesquels sont menés des travaux depuis quelques années [Lacroix 05]. Un projet du PIR Robea du CNRS en collaboration avec l'Onera et le Centre d'Expertise Parisien de la DGA a notamment permis de mieux préciser les problématiques de recherche sur la modélisation de l'environnement et la coopération multi-robots [Lacroix 06]. Début 2007, un projet d'ampleur en collaboration avec l'Onera (financé par la DGA) a été initié [PEA 07].

1.2 Problématique considérée

Pour pouvoir établir des schémas de coopération entre robots terrestres et aériens, différents types de modèles de l'environnement doivent être construits, selon le contexte de la mission et les besoins au niveau décisionnel. Par exemple, un modèle tridimensionnel est nécessaire à la planification des placements pour permettre des observations, un modèle de traversabilité est nécessaire à la planification des déplacements des robots terrestres, et des informations particulières sont nécessaires au contrôle de l'exécution des déplacements (figure 1.1). Le maintien de la cohérence spatiale de ces différents modèles

implique bien entendu que les robots soient localisés les uns par rapport aux autres, problème qui est fortement lié aux processus de modélisation (techniques de cartographie et localisation simultanées). La fusion des informations provenant des différentes sources et le développement de capacités d'interprétation des scènes modélisées sont aussi des thèmes importants, particulièrement pour passer de l'autonomie des déplacements à une réelle autonomie décisionnelle.

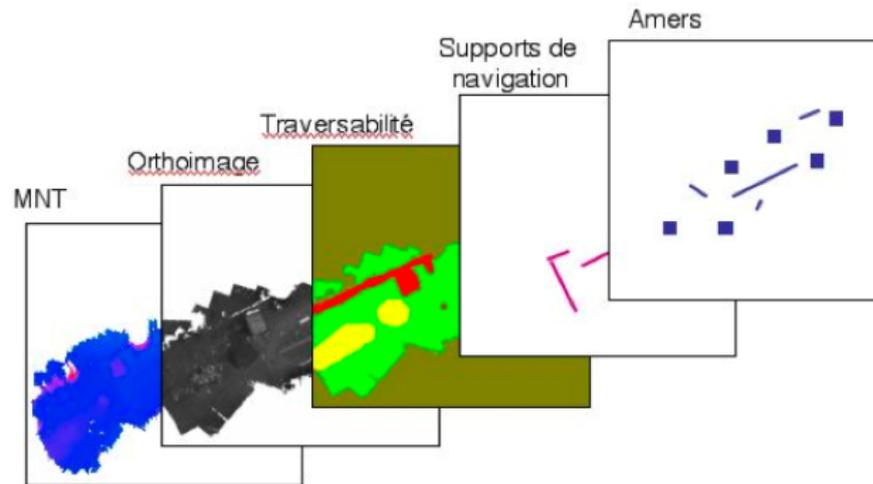


FIG. 1.1: *Différents modèles nécessaires à la coopération entre robots terrestres et aériens. De manière analogue à un système d'information géographique, différentes "couches" constituent le modèle de l'environnement construit par les robots (certaines peuvent intégrer des informations initiales). C'est grâce à la localisation que la cohérence spatiale de chaque représentation et entre les différentes représentations est maintenue*

Les travaux proposés dans ce manuscrit s'inscrivent dans ce contexte de développement de fonctionnalités de perception pour favoriser la collaboration entre robots terrestres et aériens. Ils exploitent exclusivement la vision monoculaire, et se basent sur des fonctions d'extraction de zones planaires présentes dans l'environnement. La volonté d'exploiter la vision monoculaire tient au fait que nous souhaitons développer des algorithmes qui puissent être exploités à bord de drones bas-coût, voire de micro-drones : ces contraintes nous ont aussi conduit à développer des approches qui ne dépendent pas d'une localisation précise du drone. Nos travaux visent d'abord à construire un modèle de traversabilité pour un robot terrestre sur la base de séquences d'images aériennes monoculaires. Puis, différentes fonctions de perception basées sur des séquences monoculaires acquises par un robot terrestre ou aérien sont proposées, dans le but de permettre la construction de modèles de l'environnement qui intègrent les données acquises par les différents types de robots.

Contributions Les principales contributions de cette thèse sont les suivantes :

- Estimation robuste d’homographie sur la base de points mis en correspondance dans des séquences d’images. Cette méthode permet de segmenter les points appariés, selon qu’ils appartiennent à des zones traversables ou non.
- Densification des informations de traversabilité : sur la base des appariements de points qui correspondent à des zones traversables, nous proposons une méthode qui permet de déterminer l’ensemble des pixels qui correspondent à des zones traversables. Cette méthode exploite un score de corrélation entre les pixels, qui ne requiert aucun seuil préalablement fixé.
- Construction d’un modèle probabiliste de traversabilité : les zones planes détectées au fur et à mesure de l’acquisition des images sont fusionnées en un modèle global qui exprime la probabilité que les zones perçues sont traversables.
- Construction d’un modèle plus complet basé sur des primitives planes : différentes approches visant à extraire les primitives planes présentes dans l’environnement ont été développées et évaluées, sur la base de données terrestres et aériennes.

1.3 Organisation du manuscrit

Le manuscrit est organisé comme suit :

- Le chapitre 2 rappelle les principaux résultats connus dans l’état de l’art qui concernent la géométrie de la vision projective, en insistant plus particulièrement sur les techniques d’estimation d’homographies à partir de points mis en correspondance dans deux images.
- Le chapitre 3 présente notre approche de détection de zones traversables sur la base de deux images non positionnées. Une analyse empirique des techniques d’estimation d’homographies est menée, puis notre algorithme d’estimation est présenté. À l’issue de l’estimation de l’homographie, les zones traversables sont décrites par des points épars : le moyen de compléter cette description sur la base d’une mise en correspondance dense entre les images est proposée.
- Le chapitre 4 présente une approche pour intégrer les détections de zones traversables effectuées lors des déplacements en un modèle global qui exhibe la probabilité de traversabilité pour un robot terrestre. Les informations déduites de chaque paire d’images consécutives sont fusionnées selon un formalisme bayésien.
- Le chapitre 5 présente différentes fonctions qui visent à construire un modèle plus complet sur la base de séquences monoculaires d’images aériennes et terrestres.

Le manuscrit se conclut enfin par une discussion des travaux menés, et quelques annexes précisent des aspects techniques et des extensions de nos travaux.

Chapitre 2

Géométrie et estimation en vision

2.1 Éléments de base

2.1.1 Notations

Nous présentons ici les notations qui sont employées au long de cette thèse. Par soucis de consistance avec les publications en vision par ordinateur nous avons employé majoritairement celles de [Hartley 04] qui ont tendance à se généraliser.

Différentes entités géométriques sont utilisées par la suite : points, droites et plans. Concernant les points et les droites, ils sont notés en italique et majuscule lorsqu'il est fait référence à des éléments 3D et en minuscule pour ceux en 2D. On aura ainsi M , L et m , l . Les plans sont notés Π . Sauf indication contraire, ces entités géométriques sont toujours exprimées en coordonnées homogènes ainsi qu'en vecteur colonne. Comme tout vecteur en général, ils sont écrit en caractères gras. On respecte enfin la casse entre les éléments géométriques et leur coordonnées donnant ainsi : $m \rightarrow \mathbf{m} = (x, y, 1)^\top$ ou $M \rightarrow \mathbf{M} = (X, Y, Z, 1)^\top$. L'expression des coordonnées d'un point \mathbf{m} dans le repère R_C est notée $\mathbf{m}|_{R_C}$.

Les matrices sont désignées par des caractères dans une police distincte telle que \mathbf{H} . Leur dimension, lorsqu'elle est spécifiée, est notée par les indices $n \times m$. Leurs lignes sont désignées en caractères minuscules. On a ainsi :

$$\mathbf{H}_{3 \times 3}^\top = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3]$$

Concernant les écritures matricielles de base, on note $^{-T}$ l'inverse de la transposée. Le produit vectoriel est noté \times . Rappelons que la notation $\mathbf{x} \times \mathbf{y}$ est équivalente à $[\mathbf{x}]_\times \mathbf{y}$, où $[\mathbf{x}]_\times$ est la matrice antisymétrique :

$$[\mathbf{x}]_\times = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}$$

L'égalité à un facteur multiplicatif près est notée \sim . La norme euclidienne L_2 des vecteurs ou entre des points est notée $\|\mathbf{m}\|^2 = \mathbf{m}^\top \mathbf{m}$. La même notation est utilisée pour la distance de Frobenius entre des matrices.

Enfin, les entités qui seront estimées, telles que par exemple les points, seront notées $\hat{\mathbf{m}}$ et celles qui seront normalisées $\bar{\mathbf{m}}$. La représentation euclidienne extraite de coordonnées homogènes sera noté $\tilde{\mathbf{m}}$ de tel sorte que $\mathbf{m} = (\tilde{\mathbf{m}}, 1)^\top$.

2.1.2 Espaces métrique, euclidien, affine, et projectif

Il est nécessaire de définir les espaces géométriques dans lesquels nous allons situer nos travaux. ¹

L'espace Euclidien tout d'abord, est l'espace physique "naturel" tel que nous le percevons usuellement. Les grandeurs physiques y sont mesurables avec une unité et les transformations sont rigides. L'espace métrique (anglicisme de *metric*) décrit le même espace à un facteur d'échelle près. Une grande part de nos travaux se place dans cet espace, jusqu'à ce qu'une mesure de distance euclidienne permettent de lever l'ambiguïté du facteur d'échelle. L'espace affine, ne conserve plus les rapports de longueur ou les angles, mais le parallélisme y reste un invariant. Enfin, l'espace projectif ne conserve plus le parallélisme. Le tableau 2.1 résume les différentes caractéristiques de ces espaces dont les invariants qui ne sont plus conservés dans les espaces plus généraux.

Espace	Transformation	ddl	Modélisation	Invariants
Euclidien	Transformation Rigide	6	$\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}$	longueurs, aires, volumes
Métrique	Similitude	7	$\begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}$	angles, rapports de longueurs points circulaires, conique absolue
Affine	Affinité	12	$\begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}$	parallélisme, centre de gravité, plan à l'infini, rapports d'aire
Projectif	Homographie Projection perspective	15	$\begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^\top_{1 \times 3} & v \end{bmatrix}$	birapport, colinéarité, intersection

TAB. 2.1: *Caractérisation des transformations de l'espace 3D. La matrice \mathbf{A} est inversible, \mathbf{R} est une matrice de rotation 3D, \mathbf{t} un vecteur de translation 3D et v un scalaire. "ddl" signifie nombre de degré de liberté de la transformation. Les invariants des espaces les plus généraux sont valables dans espaces les moins généraux.*

¹Le lecteur nous excusera de *franciser* les termes anglo-saxons en vigueur, ceci afin d'être cohérent avec la littérature et d'éviter les contres-sens.

2.1.3 Normes

Distance euclidienne entre deux points

Considérons \mathbf{x} et \mathbf{y} deux points de l'espace euclidien de dimension p , notés respectivement en coordonnées homogènes $\mathbf{x}_{(p+1) \times 1} = (\tilde{\mathbf{x}}, x)^\top$ et $\mathbf{y}_{(p+1) \times 1} = (\tilde{\mathbf{y}}, y)^\top$. La distance euclidienne entre ces deux points est alors :

$$d_E(\mathbf{x}, \mathbf{y}) = \left\| \frac{\tilde{\mathbf{y}}}{y} - \frac{\tilde{\mathbf{x}}}{x} \right\|^2$$

Distance algébrique entre deux points

La distance géométrique entre ces deux points est alors :

$$d_A(\mathbf{x}, \mathbf{y}) = \|x\tilde{\mathbf{y}} - y\tilde{\mathbf{x}}\|^2$$

Elle est égale à la distance euclidienne si x et y sont égal à 1.

Facteur de biais entre ces deux normes

On observe très simplement que d_E et d_A sont égaux à un facteur près que l'on appellera *facteur de biais* w_{AE} tel que :

$$d_E(\mathbf{x}, \mathbf{y}) = w_{AE} d_A(\mathbf{x}, \mathbf{y}) \text{ avec } w_{AE} = \frac{1}{x^2 y^2}$$

La norme de Frobenius

La norme de Frobenius est l'application d'une norme 2 sur les éléments d'une matrice. On a ainsi :

$$\|\mathbf{A}\|^2 = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 = \text{trace}(\mathbf{A}\mathbf{A}^\top) = \sum_{i=1}^{\min\{m,n\}} \sigma_i^2$$

avec σ_i les valeurs singulières de \mathbf{A} telles que définies dans ce qui suit.

2.1.4 La décomposition en valeurs singulières

La décomposition en valeurs singulières (*Singular Value Decomposition* ou SVD) est une décomposition aux propriétés très intéressantes dont il est souvent fait usage dans nos travaux. Elle permet de résoudre un système linéaire exact ou d'obtenir la solution au sens des moindres carrés d'un système linéaire surdimensionné. Cette décomposition s'appuie sur le théorème d'algèbre suivant :

Toute matrice $\mathbf{A}_{N \times M}$ telle que $N > M$ peut être décomposée comme le produit suivant :

$$\mathbf{A}_{N \times M} = \mathbf{U}_{N \times M} \mathbf{W}_{M \times M} \mathbf{V}_{M \times M}^\top \text{ avec } \mathbf{W} = \text{diag}_{M \times M}(\sigma_1, \dots, \sigma_M)$$

U est colonne-orthogonale dans le sens où ses colonnes sont orthonormales ($U^T \cdot U = 1$) et V est orthonormale ($V^T \cdot V = V \cdot V^T = 1$). W contient les valeurs singulières de A et V ses vecteurs singuliers.

Propriété additionnelle (1) : Minimisation de distance de Frobenius

Considérons une matrice A de dimension 3 et de rang plein et la matrice A' de dimension 3 et rang 2, telle qu'elle minimise la norme de Frobenius $\|A - A'\|^2$. Il a été montré par [Tsai 84] que la matrice A' peut s'obtenir à partir de A par la méthode suivante. Si la décomposition par SVD de A est $A = U \text{diag}(r, s, t) V^T$ avec $r \geq s \geq t$, alors $A' = U \text{diag}(r, s, 0) V^T$.

Une autre propriété du même ordre existe si l'on recherche A' telle que 2 de ses valeurs propres soient égales. On a alors $A' = U \text{diag}(\frac{r+s}{2}, \frac{r+s}{2}, 0) V^T$.

Propriété additionnelle (2) : Solution de système linéaire sous contrainte

Lorsque la matrice A est sur-dimensionnée, la solution \hat{h} , au sens des moindres carrés, du système $Ah = 0$ sous la contrainte $\|h\| = 1$ est le vecteur propre correspondant à la plus petite valeur propre de $A^T A$. Elle s'obtient donc aisément via la SVD comme étant la colonne de V correspondant à la plus petite valeur singulière de A . Ce résultat est d'un grand intérêt en présence de bruit sur les valeurs de A .

2.2 Géométrie projective

Nous ne pouvons faire une étude complète de la géométrie projective en quelques pages, mais nous avons souhaité présenter une synthèse des différentes notions et grandeurs que l'on y trouve. Même s'il n'est pas fait usage dans nos travaux de toutes ces notions, cela permet de mieux situer nos travaux.

2.2.1 Notions de géométrie projective

Caméras

La figure 2.1 illustre la géométrie d'un système à deux caméras. Le modèle le plus simple permettant de décrire la perception du monde par la caméra est le modèle sténopé qui représente une projection centrale du monde sur le plan image. Le centre de la projection est le centre optique de la caméra, le plan image appelé aussi plan focal est le plan tel que $Z = f$ dans le repère euclidien R_{cam} associé au centre optique. Dans ce repère, l'axe des Z passe par le centre optique C et est appelée axe principal de la caméra. La projection d'un point de l'espace $M_{|R_{cam}} = (X, Y, Z, 1)^T$ sur ce plan image est donc un

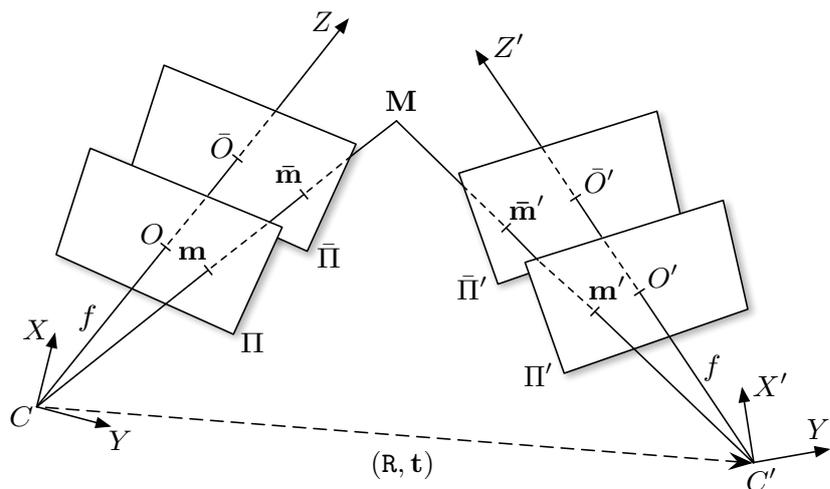


FIG. 2.1: Modélisation d'un système avec deux caméras

point $\mathbf{m}_{|R_{cam}} = (f \frac{X}{Z}, f \frac{Y}{Z}, f)^\top$. Cette relation peut être noté sous la forme :

$$\mathbf{m}_{|R_{cam}} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{M}_{|R_{cam}} = \text{diag}(f, f, 1) [\mathbf{I} | \mathbf{0}] \mathbf{M}_{|R_{cam}}$$

Pour exprimer la projection de $\mathbf{M}_{|R_{cam}}$ dans un repère R_{im} associé au plan image il faut prendre en compte les paramètres intrinsèques de la caméra CCD. Si ces pixels ne sont pas carrés la projection ne sera pas la même sur les axes x et y de l'image. Ces facteurs d'échelles sont pris en compte par l'ajout des paramètres α_u et α_v au modèle. Ils représentent la longueur focale en pixels dans chacune des directions x et y .

En outre, en plaçant l'origine du repère R_{im} en haut à gauche des images on se doit alors d'effectuer une translation entre cette origine et le point principal, intersection de l'axe principal et le plan image. Nous obtenons alors la matrice de calibration ² suivante :

$$\mathbf{C} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Elle nous permet de définir la projection $\mathbf{m}_{|R_{im}} = (u, v, 1)^\top$ d'un point du monde dans le repère image comme étant :

$$\mathbf{m}_{|R_{im}} = \mathbf{C} [\mathbf{I} | \mathbf{0}] \mathbf{M}_{|R_{cam}}$$

Cette relation étant définie à un facteur près, on notera que tout point sur le même rayon a le même projeté dans le plan image. Corrolairement, on ne peut retrouver à partir d'un

² *calibration* est un anglicisme qui tends à supplanter dans la communauté vision le terme correct qui est *étalonnage*.

point $\mathbf{m}_{|R_{im}}$ les coordonnées de $\mathbf{M}_{|R_{cam}}$ qu'à un facteur d'échelle près. Ceci nous amène au plan focal normalisé, défini comme le plan parallèle au plan focal et tel que $Z = 1m$. Les projetés d'un même point $\mathbf{M}_{|R_{cam}}$ sur ces deux plans sont reliés par l'expression suivante :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \bar{x} \\ \bar{y} \\ 1 \end{pmatrix} \quad \text{soit} \quad \mathbf{m}_{|R_{im}} = C\bar{\mathbf{m}}_{|R_{cam}} \quad (2.1)$$

Maintenant, considérons que la caméra s'est déplacée de la position C à C' , nous pouvons modéliser ce déplacement par une translation et une rotation (\mathbf{R}, \mathbf{t}) . La projection d'un point $\mathbf{M}_{|R_{cam}}$ exprimée dans le repère associé à la première caméra dans le plan image de la deuxième caméra peut alors s'écrire :

$$\mathbf{m}_{|R_{im_2}} = C [\mathbf{R}|\mathbf{t}] \mathbf{M}_{|R_{cam}}$$

Homographie

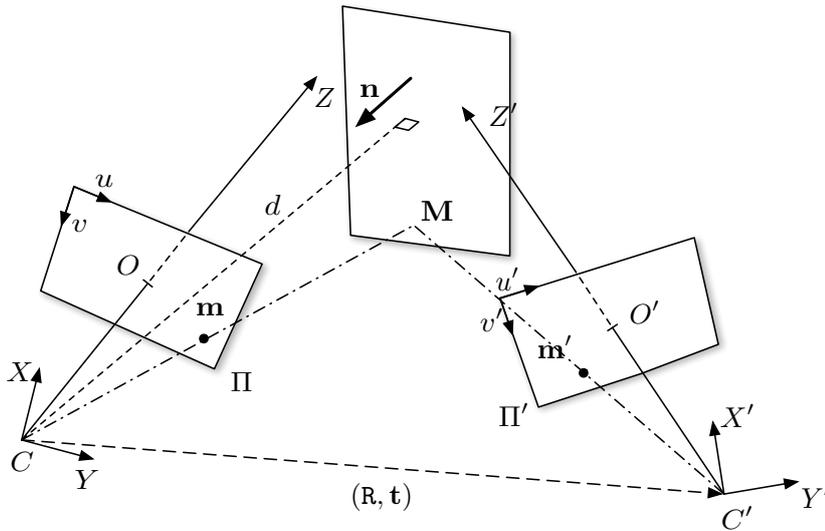


FIG. 2.2: Modélisation d'un système avec deux caméras dans le cas de points sur un plan.

La figure 2.2 illustre la géométrie relative à un système à deux vues lorsque les points perçus dans les images sont sur un plan du monde. Cette propriété permet de définir une homographie 2D H : la transformation reliant les projetés $(\mathbf{m}, \mathbf{m}')$ sur deux plans images d'un même point \mathbf{M} sur un plan Π . En fait, on peut dire que le plan Π induit une homographie H entre les images.

En considérant que le plan Π a une normale \mathbf{n} et que sa distance au centre optique C de la première caméra est d , on peut alors considérer que les points \mathbf{M} de ce plan

$\Pi = (\mathbf{n}^\top, d)^\top$ vérifie l'équation :

$$\mathbf{n}^\top \mathbf{M} + d = 0$$

En associant ce résultat aux matrices de projection $\mathbf{P} = \mathbf{C}[\mathbf{I}|\mathbf{0}]$ et $\mathbf{P}' = \mathbf{C}[\mathbf{R}|\mathbf{t}]$ on peut alors exprimer l'homographie dans le plan focal normalisé sous la forme :

$$\bar{\mathbf{H}} = \left(\mathbf{R} - \frac{\mathbf{t}\mathbf{n}^\top}{d} \right)$$

Dans le plan image, on obtient alors les relations $\mathbf{m}' = \mathbf{H}\mathbf{m}$ et $\mathbf{l} = \mathbf{H}^\top \mathbf{l}'$ où \mathbf{H} est de la forme :

$$\mathbf{H} = \mathbf{C} \left(\mathbf{R} - \frac{\mathbf{t}\mathbf{n}^\top}{d} \right) \mathbf{C}^{-1} \quad (2.2)$$

L'homographie \mathbf{H} étant une grandeur épipolaire elle est définie à un facteur près et a 8 degrés de liberté.

Géométrie épipolaire

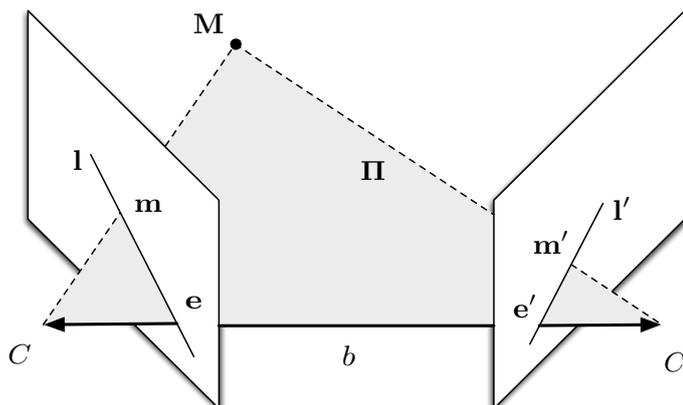


FIG. 2.3: Illustration de la géométrie épipolaire.

La géométrie épipolaire est la géométrie entre deux vues étudiant les propriétés liées à l'intersection des plans images avec l'ensemble des plans, dits *épipolaires*, contenant l'axe entre les deux centres optiques. Cet axe est appelé *baseline*. La figure 2.3 illustre cette géométrie. Comme il y est montré, à un point \mathbf{M} est associé deux projections \mathbf{m} et \mathbf{m}' dans chaque plan image. Ces trois points, les centres optiques \mathbf{C} , \mathbf{C}' et la baseline sont tous coplanaires.

En considérant les intersections de la baseline avec chaque plan image en les points que l'on appelle *épipoles*, on peut constater qu'à un point \mathbf{m} est associée une droite \mathbf{l}' dite *épipolaire* passant par son correspondant \mathbf{m}' et l'épipôle \mathbf{e}' de la deuxième caméra.

En remarquant maintenant que si la position du point \mathbf{M} varie sur le même rayon de la première caméra, alors le projeté \mathbf{m}' variera, mais pas \mathbf{m} , ni le plan épipolaire Π , ni les

épipoles, ni les droites épipolaires. *Il existe donc une relation entre les points d'une image et les primitives épipolaires qui est indépendante de la structure de la scène et ne dépend que des paramètres des caméras et de mouvements.*

Cette relation a été formalisée ([Faugeras 93b], [Hartley 04]) en modélisant la relation point à ligne existant entre un point \mathbf{m} et sa ligne épipolaire associée \mathbf{l}' par la matrice fondamentale \mathbf{F} . Un grand nombre de paramétrisations ont été proposées pour la définir, nous évoquerons simplement les suivantes :

$$\mathbf{l}' = \mathbf{F}\mathbf{m} = \mathbf{C}'^{-\top} [\mathbf{t}]_{\times} \mathbf{R} \mathbf{C}^{-\top} \mathbf{m} = [\mathbf{e}']_{\times} \mathbf{H}_{\Pi} \mathbf{m} \quad (2.3)$$

où $[\mathbf{t}]_{\times}$ est la matrice antisymétrique (ou de produit vectoriel) de \mathbf{t} , $[\mathbf{e}']_{\times}$ celle de \mathbf{e} et \mathbf{H}_{Π} l'homographie associée au plan Π et à (\mathbf{R}, \mathbf{t}) .

Comme la matrice $[\mathbf{e}]_{\times}$ est de rang 2 et \mathbf{H}_{Π} de rang 3, la matrice fondamentale \mathbf{F} n'est que de rang 2. Étant une matrice homogène de déterminant nul elle a 7 degrés de liberté. Ces caractéristiques soulèvent de nombreux points pour ce qui est de sa paramétrisation et son estimation. Celle-ci repose sur l'équation directement dérivable de (2.3) :

$$\mathbf{m}'^{\top} \mathbf{F} \mathbf{m} = 0 \quad (2.4)$$

Les propriétés de la matrice fondamentale et ses relations avec différentes entités géométriques ont été amplement étudiées. Le lecteur est invité à se reporter par exemple à [Hartley 04] qui en décrit un grand nombre.

A la matrice fondamentale est associée la matrice essentielle qui décrit la contrainte épipolaire dans le plan focal normalisé. On obtient ainsi simplement :

$$\begin{aligned} \bar{\mathbf{m}}'_{cam}{}^{\top} \mathbf{E} \bar{\mathbf{m}}_{cam} &= 0 \\ \text{avec } \mathbf{F} &= \mathbf{C}'^{-\top} \mathbf{E} \mathbf{C}^{-1} \text{ et } \mathbf{E} = \mathbf{C}^{\top} \mathbf{F} \mathbf{C}' \end{aligned} \quad (2.5)$$

Relations entre l'homographie et la géométrie épipolaire

Il est intéressant de déterminer si les homographies sont nécessairement compatibles avec la géométrie épipolaire. Par construction, une homographie induite par un plan l'est nécessairement, mais ce n'est pas le cas de toutes les matrices 3×3 à 8 degrés de liberté.

En effet, si on considère des points $(\mathbf{m}, \mathbf{m}')$ validant une homographie et la géométrie épipolaire, on peut alors écrire :

$$\mathbf{m}'^{\top} \mathbf{F} \mathbf{m} = (\mathbf{H}\mathbf{m})^{\top} \mathbf{F} \mathbf{m} = \mathbf{m}^{\top} (\mathbf{H}^{\top} \mathbf{F}) \mathbf{m} = 0$$

Ceci implique donc que la matrice $\mathbf{H}^{\top} \mathbf{F}$ soit anti-symétrique et vérifie alors l'équation suivante :

$$\mathbf{H}^{\top} \mathbf{F} + \mathbf{F}^{\top} \mathbf{H} = \mathbf{0} \quad (2.6)$$

Cette anti-symétrie, induite par la géométrie épipolaire, n'apporte que 5 contraintes sur les 8 degrés de liberté de \mathbf{H} . La validation de la géométrie épipolaire par \mathbf{H} ne permet

donc pas la déterminer uniquement. Un espace à 3 paramètres décrit donc l'ensemble des homographies associées à une matrice fondamentale F .

De nombreuses autres propriétés entre F et H peuvent être déduites. Par exemple si H est une homographie induite par un plan, alors il existe une décomposition de F telle que $F = [e']_{\times} H = H^{-T} [e]_{\times}$.

Primitives géométrique à l'infini

La géométrie projective permet de modéliser des primitives à l'infini comme par exemple l'intersection de droites parallèles. Ces primitives possèdent des invariances et des liens avec des grandeurs épipolaires permettant leur estimation.

Ainsi, si l'on se place dans l'espace projectif \mathbb{P}^2 et que l'on considère deux droites parallèles $\mathbf{l} = (a, b, c)^T$ et $\mathbf{l}' = (a, b, c')^T$ on peut calculer leur intersection comme le point $\mathbf{m} = \mathbf{l} \times \mathbf{l}' = (b, -a, 0)^T$ qui se trouve à l'infini. Tous les points à l'infini sont sur la même ligne à l'infini $\mathbf{l}_{\infty} = (0, 0, 1)^T$. En effet, ces points peuvent s'écrire $\mathbf{m} = (x, y, 0)^T$ et ils vérifient donc tous $\mathbf{m}^T \mathbf{l}_{\infty} = 0$.

Dans l'espace projectif \mathbb{P}^3 le plan à l'infini $\mathbf{\Pi}_{\infty}$ et la conique absolue $\mathbf{\Omega}_{\infty}$ sont définis. Le plan $\mathbf{\Pi}_{\infty} = (0, 0, 0, 1)^T$ a les propriétés suivantes :

- Tous les plans parallèles ont pour intersection une droite à l'infini contenu dans $\mathbf{\Pi}_{\infty}$.
- L'intersection d'une droite avec une autre ou un plan qui lui parallèle est un point de $\mathbf{\Pi}_{\infty}$.
- Le plan $\mathbf{\Pi}_{\infty}$ est invariant à une transformation projective si et seulement si cette transformation est une affinité.

La conique absolue $\mathbf{\Omega}_{\infty}$ est une conique sur le plan à l'infini $\mathbf{\Pi}_{\infty}$ tel que les coordonnées des points $\mathbf{M} = (X, Y, Z, W)^T$ de la conique vérifient :

$$\left. \begin{array}{l} X^2 + Y^2 + Z^2 \\ W^2 \end{array} \right\} = 0$$

Elle a pour propriété principale d'être invariante à une transformation projective si et seulement si cette transformation est une similitude.

Un autre résultat d'importance est que l'image de la conique absolue (IAC) dans le plan image est la conique $\omega = (CC^T)^{-1}$. Il s'agit d'un résultat très important pour l'auto-calibration car ω est estimable en combinant des connaissances sur l'orthogonalité d'éléments de la scène, la matrice de calibration et l'homographie à l'infini \mathbf{H}_{∞} .

Enfin on peut s'intéresser à l'homographie à l'infini \mathbf{H}_{∞} induite par le plan à l'infini $\mathbf{\Pi}_{\infty}$. En reprenant l'équation (2.2) et en faisant tendre d vers l'infini, on obtient :

$$\mathbf{H}_{\infty} = \mathbf{C}\mathbf{R}\mathbf{C}^{-1}$$

2.2.2 Reconstruction 3D

Afin d'estimer une reconstruction 3D de l'environnement il est nécessaire d'estimer les matrices de projections des caméras que l'on peut noter $\mathbf{P} = \mathbf{C}[\mathbf{I}|\mathbf{0}]$ et $\mathbf{P}' = \mathbf{C}[\mathbf{R}|\mathbf{t}]$ où \mathbf{P}' est

exprimé dans le même repère que P et contient donc les paramètres du mouvement entre les deux prises de vue.

Extraction de caméras projectives

La matrice fondamentale, comme il est montré par Hartley, correspond à une infinité de paires de caméra. *Il y a une ambiguïté projective.* En revanche pour chaque paire de caméra il existe une seule matrice fondamentale. En effet, si la matrice de projection de la première caméra $P = [I|0]$ alors $P' = [SF|e']$ où e' est l'épipole tel que $e'F = 0$ et S n'importe quelle matrice symétrique. une forme canonique de cette décomposition est donnée par Hartley [Hartley 04] (p.238 - résultat 8.15). Cette ambiguïté projective est le fondement de la méthode de reconstruction stratifiée.

Extraction de caméras métriques par approche stratifiée

La connaissance de la matrice fondamentale permet une reconstruction à une transformation projective près. En apportant une connaissance supplémentaire comme l'hypothèse d'un mouvement en translation pure, la connaissance du plan à l'infini Π_∞ , de l'homographie à l'infini H_∞ , ou un ratio connu sur une ligne il est alors possible de contraindre la transformation projective en une transformation affine. Une reconstruction affine est alors possible mais par définition garantie la paralléité mais pas les angles.

On peut remarquer que $H_\infty = C'RC^{-1}$ où C et C' sont les matrices de calibration et R la matrice de rotation entre les 2 positions de la caméra. Une autre méthode pour obtenir une reconstruction affine est de faire l'hypothèse d'une caméra affine.

Afin d'obtenir une reconstruction métrique, il reste alors à estimer la conique absolue Ω_∞ reposant sur Π_∞ . En pratique, on s'intéresse à l'image de la conique absolue ω (IAC Image of Absolute Conic) estimable à partir d'hypothèses sur les matrices de calibration et l'orthogonalité d'éléments de la scène.

Extraction de caméras métriques via la matrice essentielle

Il est possible d'extraire les paramètres des modèles de caméras de la matrice essentielle à un facteur d'échelle près et 4 solutions possibles en décomposant la matrice essentielle par une SVD . L'ambiguïté de ces quatres solutions peut-être aisément levée par une connaissance approximative de la position relative des caméras. Hartley donne ces quatre solutions [Hartley 04] (p.240 - result 8.19) en considérant la matrice de projection de la première caméra tel que $P = [I|0]$.

D'autres solutions pour extraire la matrice de projection ont été proposées. Faugeras [Faugeras 93b] et [Horn 90] montrent ainsi comment extraire les paramètres (R, t) de E à partir des égalités suivantes :

$$\begin{aligned} \mathbf{t}\mathbf{t}^T &= \frac{1}{2} \text{Trace}(\mathbf{E}\mathbf{E}^T)\mathbf{I} - \mathbf{E}\mathbf{E}^T \\ (\mathbf{t}^T\mathbf{t})\mathbf{R} &= \mathbf{E}^{*\top} - [\mathbf{t}]_{\times}\mathbf{E} \end{aligned} \tag{2.7}$$

Nous pouvons alors déterminer t à un signe près. Cette ambiguïté peut se lever en introduisant une connaissance à priori sur le signe d'une des composantes.

Extraction de caméras à partir d'homographie par approche stratifiée

En combinant homographie H et propriétés de la géométrie épipolaire il est possible d'estimer la matrice fondamentale F et donc d'obtenir une reconstruction perspective.

Ainsi, en sélectionnant deux points mis en correspondance $(\mathbf{m}_1, \mathbf{m}_2) \rightarrow (\mathbf{m}'_1, \mathbf{m}'_2)$ qui ne sont pas sur un plan, il suffit de calculer l'intersection des lignes $(H\mathbf{m}_1)_{\times} \mathbf{m}'_1$ et $(H\mathbf{m}_2)_{\times} \mathbf{m}'_2$ pour trouver l'épipôle \mathbf{e}' . La matrice fondamentale est alors $F = [\mathbf{e}']_{\times} H$.

Il est ensuite possible de la même manière que précédemment d'obtenir une reconstruction métrique en utilisant les propriétés de l'homographie et de la conique à l'infini ou la matrice essentielle si les matrices de calibration sont connues.

On peut aussi obtenir la matrice fondamentale dans le cas multi-planaire en vérifiant l'équation (2.6) pour les homographies associées à chacun des plans [Szeliski 98].

Extraction des caméras à partir d'homographie en connaissant la calibration

Tsai et Huang [Tsai 81] proposent une méthode pour extraire deux solutions de paramètres $(R, \mathbf{t}, \mathbf{n}, d)$ à un facteur près à partir d'une homographie dans le plan focal normalisé. Pour cela, ils font l'hypothèse de petites rotations. Sous les mêmes hypothèses, ils proposent une autre méthode s'appuyant sur la décomposition en valeurs singulières [Tsai 82] par utilisation de la SVD .

[Triggs 98] propose une autre méthode reposant aussi sur la décomposition en valeurs singulières mais qui ne fait pas l'hypothèse de petites rotations.

La géométrie projective dans nos travaux

Comme nous souhaitons extraire des plans de nos séquences d'images, l'estimation et l'utilisation des homographies est toute indiquée. Toutefois, il est important de noter que notre perspective de faible coût et de faible encombrement nous contraint en aérien à exploiter des capteurs de positionnement peu précis (GPS métrique et centrale inertielle bas coût), ce qui nous oblige à estimer les déplacements via les images si nous souhaitons en avoir une estimation précise.

On notera aussi qu'en imagerie aérienne, la caméra évolue le plus souvent dans un plan horizontal, et l'environnement survolé est principalement monoplaire. Ces contraintes empêchent toute estimation de la matrice fondamentale et par là même la reconstruction stratifiée qui en découle. En effet, nous cumulons alors mouvement critique et estimation dégénérée de F [Sturm 96, Torr 98, Kanatani 98, Gurdjos 03]. L'estimation des épipôles pour estimer F à partir de H est impossible [Lourakis 02, Vincent 01], ainsi que l'utilisation de contraintes multi-planaires [Bartoli 01, Bartoli 03].

Sans connaissance de l'environnement, sans hypothèse de caméra affine, avec un plan perçu quasi fronto-parallèle aux caméras empêchant l'estimation des lignes et points de fuite, il n'est pas non plus possible d'estimer les primitives à l'infini comme H_∞ ou ω . Nous sommes donc réduits à exploiter une portion congrue des outils de la géométrie projective avec seulement l'estimation d'homographie et l'estimation des paramètres $(\mathbf{R}, \mathbf{t}, \mathbf{n}, d)$ par [Triggs 98].

Remarquons enfin l'existence d'autres approches d'extraction de plans ne s'appuyant pas explicitement sur la géométrie projective. Ainsi [Garcia-Padro 01] utilise un descripteur de texture pour trouver des zones planes. Une telle approche nous semble assez pauvre, dans le sens où elle nécessite la satisfaction d'hypothèses peu réalistes en environnements extérieurs (une texture d'herbe n'implique pas que le terrain couvert d'herbe soit plat).

On retrouve aussi des approches basées sur estimation de déplacement dense, telle que [Gelgon 97, Odobez 97], qui détectent la parallaxe à partir de DLP (Displaced Frame Difference), ou [Besnerais 05] qui utilisent le flux optique. Ce sont des approches intéressantes et adaptées à notre contexte, mais que nous n'avons pas souhaité approfondir pour des raisons de coût en termes de temps de calcul et puissance nécessaires.

2.3 Estimation d'homographie à partir de correspondances de points

L'estimation de transformation projective 2D appelée communément homographie est au coeur de notre approche de détection de plans. Comme nous l'avons introduit en § 2.2.1, une homographie est définie comme étant la transformation H supportée par les correspondances \mathbf{m}_i et \mathbf{m}'_i , projection de points \mathbf{M}_i positionnés sur un même plan 3D.

Ces correspondances valident donc l'équation $\mathbf{m}'_i = H\mathbf{m}_i$, ce qui permet avec $m = (x, y, z)^\top$ d'obtenir trivialement le système suivant :

$$\begin{bmatrix} z'_i \mathbf{m}_i^\top & \mathbf{0} & -x'_i \mathbf{m}_i^\top \\ \mathbf{0} & -z'_i \mathbf{m}_i^\top & y'_i \mathbf{m}_i^\top \end{bmatrix} \begin{bmatrix} \mathbf{h}_1^\top \\ \mathbf{h}_2^\top \\ \mathbf{h}_3^\top \end{bmatrix} = \mathbf{0}$$

qui sera écrit par la suite $\mathbf{A}_i \mathbf{h} = \mathbf{0}$ où \mathbf{A}_i est une matrice 2×9 . On peut donc obtenir cette relation pour chacun des N couples $\{\mathbf{m}_i, \mathbf{m}'_i\}$ ce qui nous permet d'obtenir le système :

$$\mathbf{A} \mathbf{h} = \mathbf{0} \quad \text{avec } \mathbf{A} \text{ matrice } 2N \times 9 \quad (2.8)$$

Nous allons procéder dans ce qui suit à quelques rappels concernant la recherche de solution de ce type de système linéaire en insistant sur les méthodes conduisant à une solution robuste.

2.3.1 Considérations sur le système linéaire

Avant d'aborder plus spécifiquement les méthodes de résolution du système linéaire (2.8), nous allons en préciser certaines particularités.

Propriétés algébriques

Tout d'abord il est intéressant de remarquer que \mathbf{A} est de rang 8 et que son noyau est de dimension 1. Il suffit donc de 4 correspondances pour trouver une solution \mathbf{h} . \mathbf{H} est définie à un facteur d'échelle près, il est donc nécessaire d'ajouter une contrainte supplémentaire tel que $\|\mathbf{h}\| = 1$. Il est aussi possible de fixer l'un des paramètres de \mathbf{H} : $H_{33} = 1$ a pour conséquence de créer un nouveau système de la forme :

$$\begin{bmatrix} x_i z'_i & y_i z'_i & z_i z'_i & 0 & 0 & 0 & -x_i x'_i & -y_i x'_i \\ 0 & 0 & 0 & -x_i z'_i & -y_i z'_i & -z_i z'_i & x_i y'_i & y_i y'_i \end{bmatrix} \mathbf{h} = \begin{bmatrix} z_i x'_i \\ -z_i y'_i \end{bmatrix}$$

h est de dimension 8, et en concaténant comme précédemment la relation précédente pour toutes les correspondances on obtient :

$$\mathbf{A}\mathbf{h} = \mathbf{b} \quad \text{avec } \mathbf{A} \text{ matrice } 2N \times 9 \text{ et } \mathbf{b} \text{ vecteur } 2N \times 1 \quad (2.9)$$

Cette expression a l'avantage de réduire l'espace de recherche pour h ce qui peut ainsi en faciliter l'estimation, comme on le détaillera par la suite. Toutefois, comme il est suggéré par [Hartley 04], ce système peut s'avérer instable quand $H_{33} = 0$ ou du moins tends vers 0. Ceci peut intervenir lorsque la ligne de fuite du plan estimé coïncide avec le centre optique, comme dans le cas du plan du sol et de l'horizon. Un tel cas de figure est fort peu probable en aérien mais peut arriver pour un robot terrestre.

Notions de distances

La résolution du système (2.8) au sens des moindres carrés est équivalente à la minimisation non pas de la distance euclidienne mais de la distance algébrique. En effet, la fonction de coût minimisée est la suivante :

$$\epsilon_A(\mathbf{h}) = \sum_{i=1}^N \|\mathbf{A}_i \mathbf{h}\|^2 = \sum_{i=1}^N \left\| \begin{bmatrix} z'_i(\mathbf{h}_1 \mathbf{m}_i)^\top - x'_i(\mathbf{h}_3 \mathbf{m}_i)^\top \\ -z'_i(\mathbf{h}_2 \mathbf{m}_i)^\top + y'_i(\mathbf{h}_3 \mathbf{m}_i)^\top \end{bmatrix} \right\|^2$$

ce qui est équivalent à :

$$\epsilon_A(\mathbf{h}) = \sum_{i=1}^N \left\| \begin{bmatrix} z'_i \hat{x}'_i - \hat{z}'_i x'_i \\ y'_i \hat{z}'_i - \hat{y}'_i z'_i \end{bmatrix} \right\|^2 = \sum_{i=1}^N d_A(\mathbf{m}'_i, \mathbf{H} \mathbf{m}_i)^2 = \sum_{i=1}^N r_{A_i}^2(\mathbf{h}) \quad (2.10)$$

L'erreur ainsi calculée n'est pas satisfaisante par deux aspects : (1) c'est la minimisation d'une distance algébrique qui n'a pas de sens physique dans l'espace euclidien. Celle ci peut ainsi nous écarter de solutions plus naturelles. (2) la distance ne prend en compte que

l'erreur de reprojection dans la deuxième image : il s'agit donc d'une *erreur de transfert non-symétrique*.

Toutefois, nous utiliserons tout de même ce système pour chercher \mathbf{H} . En effet, le coût réduit des algorithmes de minimisation de système linéaire permet d'obtenir une première approximation acceptable de \mathbf{H} , qui peut ensuite servir d'initialisation à une recherche itérative.

En outre, deux solutions rapides permettent de minimiser l'influence de l'utilisation d'une distance algébrique :

La normalisation des coordonnées : introduite par Hartley dans [Hartley 97] dans le cadre de l'estimation de matrice fondamentale, elle améliore considérablement la recherche d'une solution par l'algorithme des 8 points lors de l'estimation de matrice fondamentale. Il en a aussi démontré l'intérêt dans le cadre de l'estimation d'homographie [Hartley 04]. Il a été démontré par la suite que cette normalisation faisait tendre l'erreur algébrique vers l'erreur de reprojection. Cette normalisation peut être conduite de deux manières distinctes :

- Mise à l'échelle isotropique : on translate le centre du nuage de points à l'origine et on met à l'échelle leur distance à l'origine de telle sorte que la distance moyenne soit $\sqrt{2}$.
- Mise à l'échelle non-isotropique : le centroïde des points est aussi placé à l'origine puis leurs coordonnées sont mises à l'échelle de telle sorte que les deux moments principaux des points soient égaux à 1.

Précisons que Hartley effectue la mise à l'échelle non-isotropique à l'aide d'une factorisation de Cholesky mais il a montré que cette méthode n'apportait pas d'amélioration notable par rapport à la version isotropique.

L'optimisation quasi-linéaire : Cette méthode proposée par Bartoli [Bartoli 03] permet de s'affranchir aussi en partie de l'erreur d'estimation introduite par la minimisation de la distance algébrique. Il définit tout d'abord la fonction de coût duale de (2.10) en utilisant la distance euclidienne :

$$\epsilon_E(\mathbf{h}) = \sum_{i=1}^N d_E(\mathbf{m}'_i, \mathbf{H}\mathbf{m}_i)^2 = \sum_{i=1}^N r_{E_i}^2(\mathbf{h})$$

Comme mentionné en (§ 2.1.3) le facteur de biais entre les deux fonctions de coûts ϵ_E et ϵ_A peut s'écrire $w_{AE_i} = \frac{1}{z_i^2 (\mathbf{h}_3 \mathbf{m}_i)^2}$. Il est alors possible de réaliser une minimisation itérative par moindres carrés repondérés, en créant une nouvelle fonction de coût $\epsilon_{A,k+1}$ à minimiser par moindres carrés qui prend en compte les facteurs de biais. Cet algorithme est dit *quasi-linéaire* et atteint une convergence au bout de quelques itérations. Il a l'avantage d'être peu coûteux, simple à mettre en oeuvre, et de proposer une minimisation de la distance euclidienne.

Notions de robustesse

Il sera, dans la suite de ce mémoire, fréquemment abordé la notion d'algorithme *robuste*. Ce qualificatif un peu équivoque signifie ici une tolérance des algorithmes par rapport à des données d'entrées erronées. Il sera notamment employé dans le cadre de l'estimation d'homographie (§ 2.3). Pour de tels algorithmes, estimant une transformation à partir de correspondances de points, deux causes d'erreurs sont possibles sur les données d'entrées :

1. Les correspondances de points sont fausses. Avec l'algorithme de mise en correspondance de point que nous utilisons [Jung 01], ce cas de figure peut être considéré comme rare.
2. Les correspondances de points ne satisfont pas le type de transformation que l'on cherche à estimer : autrement dit le type de transformation que l'on essaie d'estimer à partir des correspondances ne décrit pas de manière suffisamment complète le déplacement des correspondances.

Ce second cas de figure sera prépondérant dans nos travaux. Nous utiliserons par la suite les appellations anglophones suivantes :

- Les *inliers* (points cohérents) sont les correspondances de points satisfaisant la transformation recherchée.
- Les *outliers* (points aberrants) sont les correspondances de points ne satisfaisant pas la transformation recherchée.

2.3.2 Estimation robuste : approches itératives

[Madsen 04] est une référence complète et claire pour les problèmes de moindres carrés non-linéaires. Le lecteur intéressé pourra aussi étudier la synthèse d'Hartley [Hartley 04] (annexe §A.6) et son adaptation de l'algorithme de Levenberg-Marquardt au cas de l'estimation d'homographie. La lecture de [Triggs 00] et son application aux techniques d'ajustement de faisceaux s'avère aussi profitable.

Enfin, nous citerons [Benhimane 04] qui a proposé une intéressante optimisation de ces techniques appliquée à l'estimation d'homographies. Nous reportons encore une fois le lecteur intéressé par une estimation itérative de H vers [Malis 05] qui réalise une synthèse de ces fonctions dans ce cadre et un contexte de vision robotique avec contraintes en temps de calcul fortes.

Si nous ne nous détaillons pas les spécificités de ces méthodes d'estimation nous allons décrire les fonctions de coûts qui peuvent être employées afin d'obtenir une estimée robuste de H .

Least Median of Squares - LMS

Le Least Median of Squares (LMS ou LMEDS) [Rousseeuw 84][Rousseeuw 87] est une reformulation du problème de régression linéaire standard. En deux dimensions ceci

revient à identifier la bande la plus étroite, définie par 2 lignes parallèles, qui contient la moitié des points plus un. L'expression de sa fonction de coût est la suivante :

$$\epsilon(\mathbf{h}) = \text{median}(r_i^2(\mathbf{h})) \quad (2.11)$$

En trouvant le \mathbf{h} minimisant cette fonction de coût, on ne prends pas en compte les $N/2$ résidus les plus importants rendant l'estimation robuste à 50% de correspondances erronées.

Least Trimmed Squares - LTS

Le Least Trimmed Squares minimise la somme des carrés des k premiers résidus rangés par ordre croissant :

$$\epsilon(\mathbf{h}) = \sum_{i=1}^k (r_i^2(\mathbf{h})) \quad (2.12)$$

Il a comme propriété d'accélérer la vitesse de convergence du LMS [Malis 05][Rousseeuw 87] et de lisser la fonction de coût. Contrairement au LMS, il est ici possible de fixer k et donc de s'adapter à une approximation sur le pourcentage d'outliers r_O présents dans les correspondances.

M-estimateurs

Suivant le même principe que la fonction de coût RANSAC (§ 2.3.3) ces estimateurs cherchent à minimiser l'influence des résidus les plus grands en proposant toutefois une fonction de pondération des résidus différentiable [Zhang 95b]. On cherche alors à minimiser la fonction de coût :

$$\epsilon(\mathbf{h}) = \sum_{i=1}^N \rho(r_i(\mathbf{h})) \quad (2.13)$$

Huber [Huber 81] propose la fonction ρ suivante :

$$\rho(r_i(\mathbf{h})) = \begin{cases} \frac{1}{2}r_i^2(\mathbf{h}) & \text{si } r_i(\mathbf{h}) \leq t \\ t(|r_i(\mathbf{h})| - \frac{t}{2}) & \text{si } r_i(\mathbf{h}) > t \end{cases} \quad \text{avec } t = 1.345\hat{\sigma}$$

La minimisation de (2.13) conduit à chercher une solution à :

$$\sum_{i=1}^N \psi(r_i(\mathbf{h})) \frac{\partial r_i(\mathbf{h})}{h_j} = 0 \quad \text{pour } j = 1, \dots, 9$$

où $\psi(x) = \frac{\partial \rho(x)}{\partial x}$ est appelée fonction d'influence. En définissant la fonction de pondération : $w(x) = \frac{\psi(x)}{x}$, on peut reformuler l'équation précédente sous la forme :

$$\sum_{i=1}^N w(r_i(\mathbf{h})) r_i(\mathbf{h}) \frac{\partial r_i(\mathbf{h})}{h_j} = 0 \quad \text{pour } j = 1, \dots, 9$$

Comme décrit par [Zhang 95b] cette formulation reprend celle de l'estimation itérative par la méthode des moindres carrés repondérés (Iterative Reweighted Least Squares, IRLS) :

$$\sum_{i=1}^N w(r_i^{(k-1)}(\mathbf{h})) r_i^2(k)(\mathbf{h}) = 0$$

Il suffit pour le résoudre d'itérer les minimisations d'un système du type $\mathbf{D}_k \mathbf{A} \mathbf{h} = \mathbf{D}_k \mathbf{b}$ jusqu'à atteindre une convergence. \mathbf{D}_k est une matrice diagonale contenant les poids $w(r_i^{(k-1)}(\mathbf{h}))$ (que l'on notera w_i^{k-1} par la suite).

Dans le cas de la fonction de pondération de Huber les poids w_i^{k-1} sont calculés par :

$$w_i^{k-1} = \begin{cases} 1 & \text{si } |r_i^{k-1}| \leq t \\ \frac{t}{|r_i^{k-1}|} & \text{si } |r_i^{k-1}| > t \end{cases} \quad \text{et } t = 1.345\hat{\sigma}$$

$\hat{\sigma}$ est une estimation robuste de l'écart type dans le sens où il quantifie l'écart type du bruit sur les résidus correspondant aux *outliers*. On l'estime donc par une statistique robuste, le MAD (Median Absolute Deviation) :

$$\hat{\sigma} = 1.48 \operatorname{median}(\mathbf{r} - \operatorname{median}(\mathbf{r}))$$

En lieu et place de la fonction de pondération de Huber on pourra préférer celle de Fair [Rey 83] qui, comme Huber, garantit une solution unique et a, en outre, comme intérêt de ne pas nécessiter l'estimation d'une statistique supplémentaire comme le MAD pour pondérer le système linéaire. Les poids w_i^{k-1} sont alors obtenus par :

$$w_i^{k-1} = \frac{1}{1 + |r_i^{k-1}|/c} \quad \text{avec } c = 1.3998$$

Il est enfin à noter que les M-estimateurs peuvent être utilisés postérieurement à une pré-étape de filtrage des outliers, comme un moyen d'affiner l'estimation de \mathbf{H} . Une telle étape de préfiltrage peut être menée par les méthodes de votes décrites dans ce qui suit.

2.3.3 Estimation robuste : méthodes de vote

Les estimateurs que nous mentionnons, à l'exception de la transformée de Hough, s'appuient sur une technique de tirages aléatoires de type Monte-Carlo. Ils visent à réaliser suffisamment d'estimations de \mathbf{H} à partir d'autant d'ensembles de correspondances distincts pour que l'un de ses ensembles ne contienne que des *inliers*. L'estimée $\hat{\mathbf{H}}$ obtenue sera donc exempte de l'influence d'*outliers* et normalement optimale.

Les estimées de \mathbf{H} sont obtenues usuellement par minimisation de l'équation (2.9) et donc d'une distance algébrique non symétrique. L'algorithme (1) résume ces approches.

Algorithme 1 : *Algorithme générique d'estimation d'homographie par méthode de vote.*

Considérons le système initial $\|\mathbf{A}\mathbf{h} = \mathbf{0}\|$ avec \mathbf{A} matrice $2N \times 9$.

1. Effectuer m tirages aléatoires de loi uniforme de $p = 4$ entrées de \mathbf{A}
 2. Pour chaque tirage indicé i :
 - (a) Estimer $\hat{\mathbf{h}}_i$ au sens des moindres carrés pour les p entrées tirées.
 - (b) Calculer $\epsilon(\hat{\mathbf{h}}_i)$ selon la fonction de coût retenue (RANSAC, LMS, LTS...)
 3. Conserver $\hat{\mathbf{h}}$, le $\hat{\mathbf{h}}_i$ dont le résidu est minimal
 4. Selon la méthode, réitérer une estimation de \mathbf{h} après pondération des \mathbf{A}_i en fonction des résidus obtenus pour $\hat{\mathbf{h}}$.
-

Transformée de Hough

Nous citons pour mémoire la transformée de Hough [Hough 59] qui dans le contexte d'estimation d'homographies à 8 ou 9 paramètres est une méthode trop coûteuse. Il s'agit d'une méthode de vote robuste qui travaille dans l'espace des paramètres en discrétisant ceux-ci. On obtient alors des hypercubes dans l'espace d'état auxquels on associe des accumulateurs. À partir d'un jeu de données minimal, on estime les paramètres et on incrémente l'accumulateur associé. Une étape de recherche de maxima dans les accumulateurs est alors conduite et produit le jeu de paramètres correspondant à la meilleure estimation.

RANSAC

La méthode de vote RANSAC [Fischler 87] pour *Random Sample Consensus* est une approche probabiliste offrant des performances plus rapides que les méthodes de votes classiques. Cette démarche repose sur une fonction de coût non linéaire annulant les résidus les plus grands et donc des *outliers* si l'estimée considérée est juste. On peut écrire cette fonction de coût sous la forme :

$$\epsilon(\mathbf{h}) = \sum_{i=1}^N \rho(r_i(\mathbf{h}))$$

avec :

$$\rho(r_i(\mathbf{h})) = \begin{cases} 0 & \text{si } r_i^2(\mathbf{h}) \leq t \\ 1 & \text{si } r_i^2(\mathbf{h}) > t \end{cases} \quad \text{et } t = 2.5\hat{\sigma}^2$$

Le principe de cette méthode est de réitérer le calcul de cette fonction de coût, estimée à partir de p correspondances tirées aléatoirement sur les N existantes, jusqu'à être sûr d'avoir effectué au moins un tirage ne contenant pas d'*outliers* avec une probabilité notée

P . Le nombre de tirages aléatoires de p correspondances parmi N (et conséquemment de $\epsilon(\mathbf{h})$) dépend du pourcentage d'outliers r_o :

$$m = \frac{\log(1 - P)}{\log(1 - (1 - r_o)^p)}$$

On remarquera que le nombre de correspondances N n'influe pas sur le nombre de tirages aléatoires nécessaires. Le tableau 2.2 illustre l'influence de r_o et P sur m .

	$r_o = 0.1$	$r_o = 0.2$	$r_o = 0.3$	$r_o = 0.4$	$r_o = 0.5$	$r_o = 0.6$	$r_o = 0.7$	$r_o = 0.8$
$P = 0.95\%$	3	6	11	22	47	115	368	1870
$P = 0.99\%$	4	9	17	33	71	178	566	2875

TAB. 2.2: Influence du pourcentage d'outliers r_o et de la probabilité de réaliser un tirage sans outliers P sur le nombre de tirages aléatoires m nécessaires. Le nombre p de correspondances par tirage est fixé à 4 ce qui nous place dans le cas d'une estimation d'homographie.

Least Median of Squares - LMS

Cette approche consiste à utiliser la méthode probabiliste du RANSAC en remplaçant la fonction de coût par celle de l'équation (2.11). L'avantage principal est de ne pas avoir à estimer la distribution du bruit sur les correspondances. Toutefois [Rousseeuw 87] puis [Hettmansperger 92] illustrent l'instabilité numérique de cet estimateur en présence de bruit blanc sur les correspondances. Afin de compenser ceci il est proposé de pratiquer une nouvelle estimation du système en pondérant les entrées. L'écart type robuste (*robust standard deviation*) est ainsi introduit :

$$\hat{\sigma} = 1.4826(1 + 5/(n - p))\sqrt{r_{med}}$$

avec $\sqrt{r_{med}}$ la plus petite médiane des résidus. Les poids utilisés pour la réestimation du système sont alors :

$$w_i^{k-1} = \begin{cases} 1 & \text{si } w_i^{k-1} \leq 2.5\hat{\sigma} \\ 0 & \text{si } w_i^{k-1} > 2.5\hat{\sigma} \end{cases}$$

Least Trimmed of Squares - LTS

De même que pour § 2.3.3 on peut combiner approche probabiliste de type RANSAC et la fonction de coût LTS (2.12).

2.3.4 Le cas de l'affinité

Dans l'hypothèse d'un mouvement planaire parallèle au plan Π 3D étudié, c'est à dire un déplacement de la caméra composé d'une translation parallèle au plan Π et une

rotation orthogonale à la translation, les points mis en correspondance $(\mathbf{m}, \mathbf{m}')$ vérifient une affinité :

$$\mathbf{H}_A = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad (2.14)$$

Il est possible de l'estimer par des méthodes similaires à celles décrites précédemment pour une homographie. Comme les paramètres h_{31} et h_{32} de \mathbf{H}_A sont nuls, avec des coordonnées homogènes du type $\mathbf{m} = (x, y, 1)^\top$, le facteur de biais $w_{AE} = \frac{1}{\mathbf{h}_{A3}\mathbf{m}}$ est égal à 1. Les distances algébrique $d_A(\mathbf{m}', \mathbf{H}_A\mathbf{m})^2$ et euclidienne $d_E(\mathbf{m}', \mathbf{H}_A\mathbf{m})^2$ étant alors égales, il n'y a plus lieu de réaliser d'optimisation quasi-linéaire.

L'affinité estimée par SVD est alors la meilleure au sens des moindres carrés de la distance euclidienne.

2.3.5 Estimation robuste d'homographie, un bilan

Comme on n'a pu le voir au § 2.3 les techniques d'estimation robustes d'homographie sont nombreuses et variées.

La figure 2.4 dresse une topologie des techniques possibles d'estimation par méthodes de vote. Les M-estimateurs et optimisation quasi-linéaire pouvant être implémentés par des algorithmes IWLS utilisant les estimées de $\hat{\mathbf{H}}$ et $\hat{\mathbf{A}}$ issues des estimateurs par méthodes de vote.

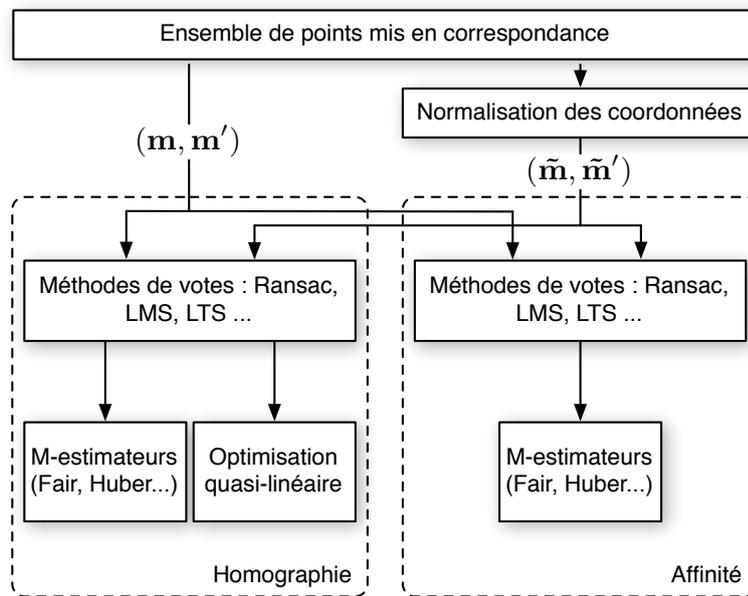


FIG. 2.4: Topologie des techniques d'estimation par méthodes de vote.

Chapitre 3

Détection de zones traversables

3.1 Objectifs, présentation de l'approche

La détection de zones planaires à partir d'imagerie aérienne s'inscrit dans notre contexte de robotique aéro-terrestre, où les zones planaires détectées peuvent tout d'abord initier un modèle de traversabilité de l'environnement utilisé par un robot terrestre qui cherche à évoluer dans l'environnement survolé. La figure 3.1 illustre ce type de modèle, construit par un robot terrestre sur la base de données stéréoscopique. Il est aussi possible d'exploiter le plan détecté comme une orthoimage, sur laquelle le robot terrestre peut se localiser.



FIG. 3.1: *Carte probabiliste de traversabilité établie par un robot terrestre à partir de données 3D fournies par stéréovision. La probabilité de traversabilité est codée en couleurs, selon une échelle traversable/vert à obstacle/rouge.*

La détection de zones planaires dans une séquence d'images monoculaires peut être menée de façons distinctes : en aval, à la suite d'une étape de reconstruction d'un modèle de

l'environnement à base de points ou lignes [Baillard 99] ou encore en amont en travaillant directement sur les images. Dans le premier cas, il sera d'abord requis de procéder à une reconstruction 3D de l'environnement. Dans le second, on pourra détecter les zones planes dans les images, les segmenter et estimer leur géométrie 3D sans procéder à une reconstruction totale de la scène.

Bien que la littérature soit très abondante en ce domaine, le contexte d'imagerie aérienne dans lequel nous sommes placé apporte deux contraintes majeures :

- L'environnement est très souvent majoritairement constitué d'un même plan (le plan du sol)
- Les zones planes ou non-planes ne sont pas connues *a-priori*.

Ces deux contraintes rendent caduques nombre de méthodes utilisant la géométrie épipolaire afin de reconstruire l'environnement à partir de séquences monoculaires. Nous avons donc opté pour une approche visant à estimer les zones planes au sein même de la séquence d'image avant d'en retrouver les caractéristiques 3D.

La méthode que nous proposons repose sur l'utilisation des propriétés des homographies et sur une étude approfondie du comportement du score de corrélation ZNCC. Elle est résumée dans la figure 3.2.

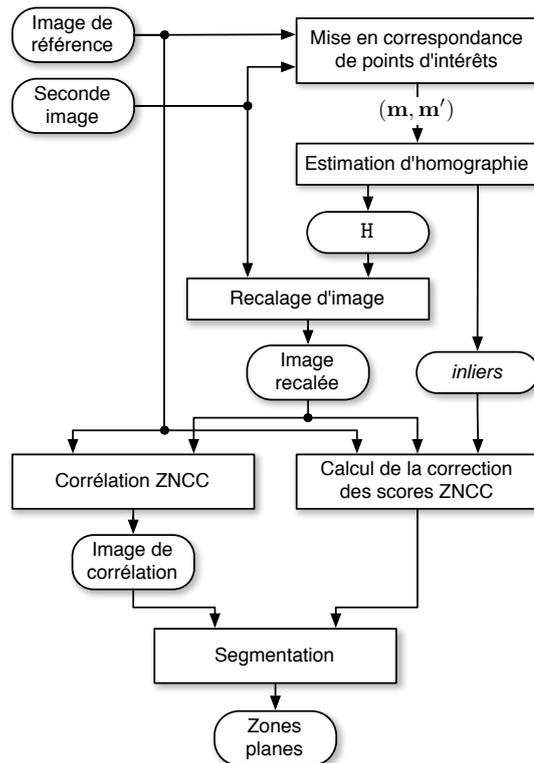


FIG. 3.2: Algorithme de détection de plan.

Elle repose sur l'idée très simple, que l'ensemble des points $\{\mathbf{m}, \mathbf{m}'\}$ mis en correspondance dans une paire d'images, s'ils sont la projection de points \mathbf{M} coplanaires, peuvent

être mis en relation au travers d'une et une seule homographie H . Corrolairement, estimer une homographie H pour un ensemble de points $\{\mathbf{m}, \mathbf{m}'\}$ équivaut à établir la coplanarité des points \mathbf{M} dont ils sont la projection. L'estimation de H sera abordée en détail au § 3.2.

Si un triplet $\{\mathbf{m}, \mathbf{m}', H\}$ définit une relation de coplanarité entre points, la segmentation de la zone planaire ainsi définie n'est pas pour autant immédiate. La méthode que nous proposons pour augmenter (densifier) la description de la zone détectée sera détaillée au § 3.4.

Enfin, dans le but de maximiser l'utilisation de l'information que nous pouvons extraire des images, nous avons proposé un modèle probabiliste de traversabilité décrit au § 4. Il permet de ne pas se limiter à une paire d'image et de fusionner les informations de traversabilité acquises le long d'une ou plusieurs séquences.

3.2 Estimation d'homographie

Nous souhaitons déterminer l'ensemble des points mis en correspondance $\mathcal{E}_I : \{\mathbf{m}, \mathbf{m}'\}_I$ tel qu'il contienne les projections de points sur le plan 3D Π_I . Incidemment nous trouverons aussi l'ensemble des correspondances \mathcal{E}_O , projections de points qui n'y sont pas. La recherche de \mathcal{E}_I nous permet ainsi de définir une homographie H_I qui est la transformation entre les points \mathbf{m} et \mathbf{m}' de \mathcal{E}_I . La recherche de l'homographie H_I , définie par les paramètres du mouvement (\mathbf{R}, \mathbf{t}) et ceux de l'environnement (\mathbf{n}_I, d_I) étant notre finalité, une estimée de \mathcal{E}_I incomplète ne nous gêne pas, l'homographie H_I associée étant la même. En revanche, nous souhaitons minimiser l'intégration de points de \mathcal{E}_O dans \mathcal{E}_I , ces points altérant l'estimation de H_I .

Comme nous l'avons précédemment abordé au § 2.3, un certain nombre de méthodes robustes permettent de s'affranchir de ce que nous appellerons les *outliers* afin d'obtenir une estimée de H_I fiable. Nous allons présenter dans ce qui suit quelques considérations sur les paramètres influençant ces estimateurs, puis présenter nos analyses de leurs performances dans un certain nombre de situations de synthèse et réelles.

3.2.1 Considérations sur l'estimation d'homographie

La robustesse dans le cadre de l'estimation d'homographie :

Nous pouvons préciser la notion de robustesse telle que nous l'avons décrite au § 2.3.1. On cherche ici à estimer par des algorithmes robustes des triplets $\{\mathbf{m}, \mathbf{m}', H\}$ tel que H définisse pleinement la transformation entre les correspondances $\{\mathbf{m}, \mathbf{m}'\}$. On peut donc définir :

- les *inliers* sont les correspondances de points qui satisfont l'homographie H estimée et qui appartiennent donc à un plan Π_I .
- les *outliers* sont les correspondances de points qui ne satisfont pas H et ne sont donc pas sur le plan Π_I associé.

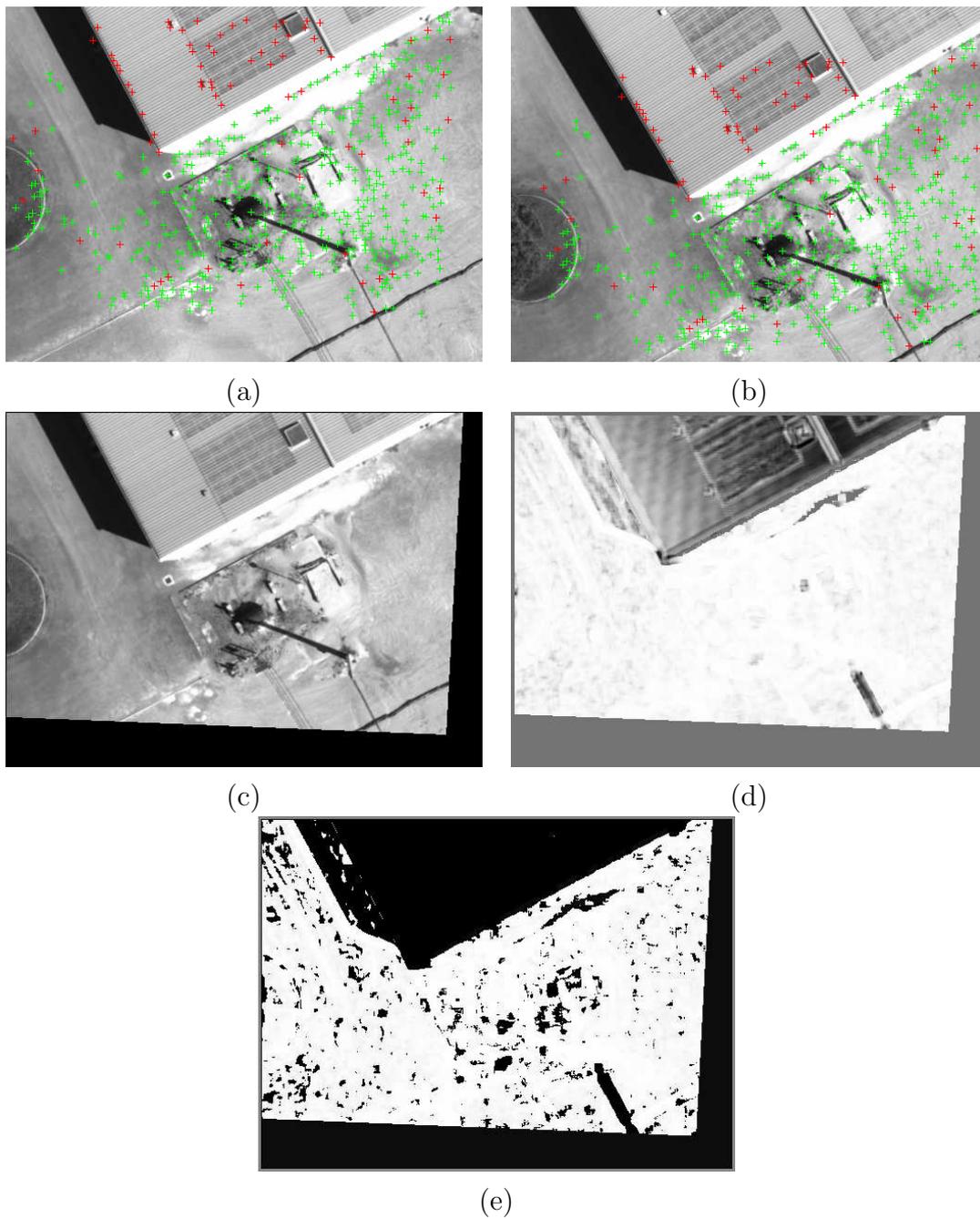


FIG. 3.3: *Illustration des étapes de la segmentation de zones planes. Les images (a) et (b) sont respectivement les images de référence et l'image suivante. Les points verts ont été segmentés par une estimation d'homographie robuste comme des inliers et les rouges comme des outliers . L'image (c) est l'image (b) recalée par l'homographie estimée. (d) est l'image de corrélation ZNCC entre les images (a) et (c), L'image (e) est l'image (d) seuillée afin de mettre en évidence les zones planes.*

Influence de la parallaxe La parallaxe est l'incidence du changement de position de l'observateur sur l'observation d'un objet. Dans notre problème nous nous intéressons plus particulièrement à la parallaxe relative au plan Π_I telle que décrite par la formule suivante :

$$\mathbf{m}' = H_I \mathbf{m} + \rho \mathbf{e}' \quad (3.1)$$

où \mathbf{e}' est l'épipôle dans la seconde image, avec $F^T \mathbf{e}' = 0$. \mathbf{e}' ne dépend que des paramètres (\mathbf{R}, \mathbf{t}) et est donc indépendant de la structure de l'environnement observé. ρ décrit donc cette structure. Afin d'essayer de la quantifier nous avons choisi de l'exprimer relativement au plan Π_I en optant pour les paramètres (\mathbf{n}_I, d_I, l_O) où l_O est la distance au plan Π_I des points M_O .

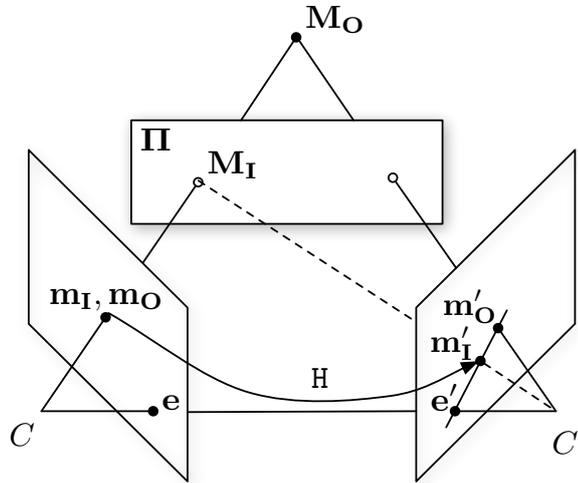


FIG. 3.4: Illustration du phénomène de parallaxe relative à un plan. \mathbf{m}_I et \mathbf{m}_O sont confondus car ils sont les projections respectives de M_I et M_O , deux points sur la même raie.

3.2.2 Analyse des estimateurs robustes avec des données de synthèse

Méthodologie :

Afin d'évaluer les différents algorithmes d'estimation nous avons choisi dans un premier temps d'utiliser des données de synthèse. Pour cela, nous avons calculé des ensembles de correspondances $\{\mathbf{m}, \mathbf{m}'\}$ à partir de paramètres du mouvement (\mathbf{R}, \mathbf{t}) et du plan Π_I recherché (\mathbf{n}_I, d_I) connus. On a alors :

$$s\mathbf{m} = C[\mathbf{I}_{3 \times 3} | \mathbf{0}_{3 \times 1}] \mathbf{M} \quad \text{et} \quad s\mathbf{m}' = CR[\mathbf{I}_{3 \times 3} | \mathbf{t}_{3 \times 1}] \mathbf{M}$$

Et on peut calculer l'homographie H_I :

$$H_I = C \left(R - \frac{\mathbf{t}\mathbf{n}_I^T}{d_I} \right) C^{-1}$$

En calculant $\hat{\mathbf{m}}'$ la reprojection de \mathbf{m} selon H_I on peut aussi obtenir une valeur théorique du résidu :

$$\epsilon_{O_{VT}} = d_A(\mathbf{m}', \hat{\mathbf{m}}')^2 \quad \text{avec } s\hat{\mathbf{m}}' = H_I \mathbf{m}$$

En créant un ensemble des points \mathbf{M}_I appartenant à Π_I et un autre ensemble de points \mathbf{M}_O qui en sont distincts on peut alors recréer des jeux de données simulés et mettre à l'épreuve les algorithmes d'estimation robuste et évaluer l'influence de (R, \mathbf{t}) et la parallaxe sur leurs performances. Nous avons ainsi choisi d'utiliser des jeux de données composés de N_I correspondances sur le plan Π_I et N_O outliers à une distance de ce plan défini par une loi uniforme d'écart type l_O . Cette configuration est illustrée sur la figure 3.5.

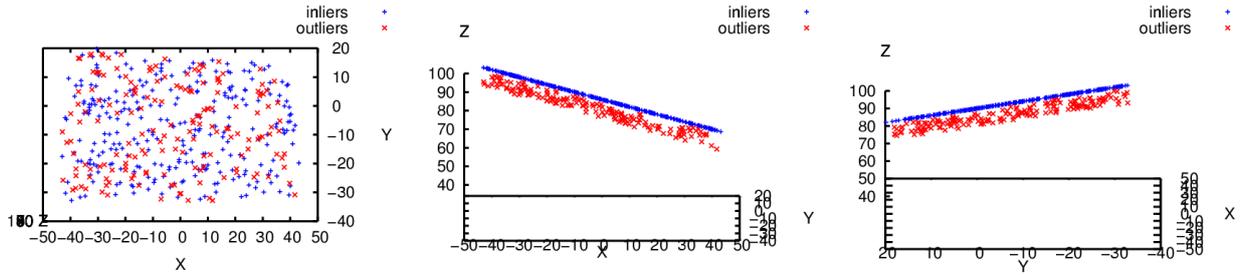


FIG. 3.5: Représentation des points \mathbf{M} dans le repère de la première caméra. (a) est la vue tel que projeté sur le plan image, (b) et (c) sont des vues de coupe permettant d'apprécier l'écart entre inliers et outliers.

L'illustration de la figure 3.5 ne montre pas la vraie distribution des points \mathbf{M}_I mais les points de référence pour lesquels on détermine leurs projetés \mathbf{m} et \mathbf{m}' avant de leur appliquer un bruit blanc de loi normale centrée réduite $\mathcal{N}(0, 0.5)$. Le même bruit est appliqué sur les projetés de l'ensemble \mathcal{E}_O . L'erreur induite par ce bruit est celle qui correspond à la détection des points d'intérêts.

Nous avons choisi un pourcentage d'outliers de 40% : au delà de 50% les estimateurs ne sont par construction plus capables de fournir une estimation fiable.

Enfin, les résultats présentés par la suite pour chaque jeu de paramètres $R, \mathbf{t}, \mathbf{n}_I, d_I, l_O$ ont été calculés pour 50 générations aléatoires de points vérifiant ces paramètres puis moyennés.

Estimateurs testés :

Nous avons opté pour des estimateurs robustes associés à des méthodes de vote minimisant les fonctions de coût LMS (§ 2.3.3) et LTS (§ 2.3.3). Ces deux méthodes ont

été employées afin de minimiser une homographie \mathbf{H}_I ainsi qu'une affinité \mathbf{A}_I . Nous avons en outre utilisé les solutions ainsi calculées comme solutions initiales pour des méthodes de moindres carrés repondérés utilisant le M-estimateur proposé par Fair et présenté au § 2.3.2. Nous avons aussi mis en pratique l'optimisation quasi-linéaire tel que décrite en § 2.3.1 sur les estimées de \mathbf{H}_I . Enfin, toutes ces méthodes ont été testées avec et sans la normalisation des coordonnées introduite par Hartley et décrite au § 2.3.1. Au total 20 estimateurs ont donc été évalués.

Il est à noter que nous ne cherchons pas en premier lieu à quantifier exactement les performances des estimateurs mais à déterminer l'influence des certains paramètres tels que la normale ou la rotation. Comme pour ces deux études les autres estimateurs présentent les mêmes types de comportement, nous ne présentons que des résultats basés sur l'estimation d'homographie et d'affinité par la minimisation de la fonction de coût LMS avec ou sans normalisation des coordonnées.

Résultats généraux :

Afin de quantifier les estimateurs et l'influence des différents paramètres, nous avons choisi d'utiliser le résidu $\epsilon_A(\mathbf{h})$ comme premier indicateur naturel de qualité des estimées. Mais notre objectif étant avant tout d'obtenir l'estimée de l'ensemble \mathcal{E}_I la plus correcte possible, nous avons aussi introduit Ω_O comme le ratio entre le nombre d'*outliers* détectés et le vrai nombre d'*outliers* de nos jeux de données. Il aurait semblé plus logique d'utiliser un ratio sur les *inliers* Ω_I mais Ω_O est symétrique et plus contrasté ce qui rend la lecture des résultats plus aisée.

Une première conclusion générale et indépendante de tout paramètre de mouvement est l'influence attendue de d_I et l_O sur les performances des estimateurs. Comme la figure 3.6 le présente, les performances décroissent fort logiquement si le plan $\mathbf{\Pi}_I$ est éloigné des caméras (augmentation de d_I) ou si les outliers sont distribués près de $\mathbf{\Pi}_I$. Il est intéressant de noter que $\epsilon_A(\mathbf{h})$ s'avère être un piètre indicateur de la qualité de nos estimées de \mathbf{H}_I puisque qu'il présente des minimas pour les meilleures et les plus mauvaises estimations de \mathcal{E}_I .

Influence de la normale

La normale au plan \mathbf{n}_I influe sur l'angle de vue de l'environnement observé et donc sur la parallaxe et les résidus des points mis en correspondance. Afin de quantifier de quelle manière cette influence agit sur les performances des estimateurs nous avons généré des correspondances en faisant varier d_I et l_O pour différentes normales \mathbf{n}_I . La figure 3.7 illustre les performances pour les différents jeux de données de nos quatre estimateurs représentatifs en représentant les lignes de niveaux $\Omega(d_I, l_O) = 80\%$.

Pour bien lire ces courbes, il est à remarquer par exemple que pour chacune de ces courbes la meilleure performance est celle représentée par la courbe rouge. En effet, le jeu

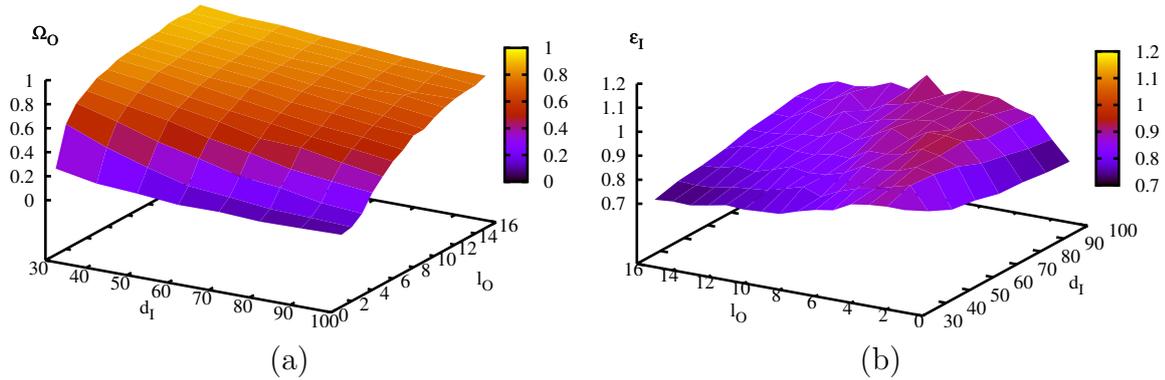


FIG. 3.6: Représentation du pourcentage d'outliers détecté Ω_O et du résidu sur les inliers ϵ_I en fonction de d_I et l_O .

de données associé est alors celui pour lequel la limite de bonne estimation l_O est la plus basse en tout d_I .

Ceci considéré, on peut alors noter sur les figures (b) et (d) que les estimateurs affines produisent de meilleurs résultats en estimant l'ensemble \mathcal{E}_O plus précisément que les estimateurs homographiques des figures (a) et (c). On peut constater que ces deux types d'estimateurs produisent des résultats similaires pour des d_I faibles, puis que l'écart se creuse à mesure que d_I augmente. Ce résultat est très prévisible du fait des paramètres rigides (\mathbf{R}, \mathbf{t}) choisis simplifiant ainsi l'expression de l'homographie considérée \mathbf{H}_I . Ses paramètres h_{31} et h_{32} sont alors égaux à 0 ce qui la rend donc affine. L'intérêt de ce résultat est donc qu'il est préférable d'utiliser des estimateurs affines aux estimateurs homographiques lorsque la transformation est affine.

On remarquera enfin que la normalisation des coordonnées utilisée par les estimateurs des figures (c) et (d) n'apporte ici pas d'amélioration de performance par rapport aux estimateurs ne l'utilisant pas ((a) et (b)).

Influence de la rotation

L'influence de la rotation \mathbf{R} sur les résidus et la parallaxe est tout aussi naturelle que celle de la normale \mathbf{n}_I . Comme celle-ci nous l'avons quantifiée en séparant les estimateurs affines (b) et (d) des estimateurs homographiques (a) et (c) et les estimateurs avec normalisation de coordonnées (c) et (d) des estimateurs sans normalisation (a) et (b). La figure figure 3.8 présente donc les lignes de niveaux $\Omega(d_I, l_O) = 80\%$ pour différentes valeurs de ϕ , ces valeurs étant par souci de lisibilité différentes pour les estimateurs affines (b) et (d).

La première remarque que l'on peut déduire de ces courbes est que la rotation a peu d'effet sur les performances des estimateurs minimisant une homographie. En revanche, les estimateurs affines voient leur performance diminuer très sensiblement pour des variations de ϕ , même légères. Dans un tel cas, la transformation n'étant plus modélisable par une affinité il s'agit d'un comportement tout à fait normal. On peut tout de même prendre

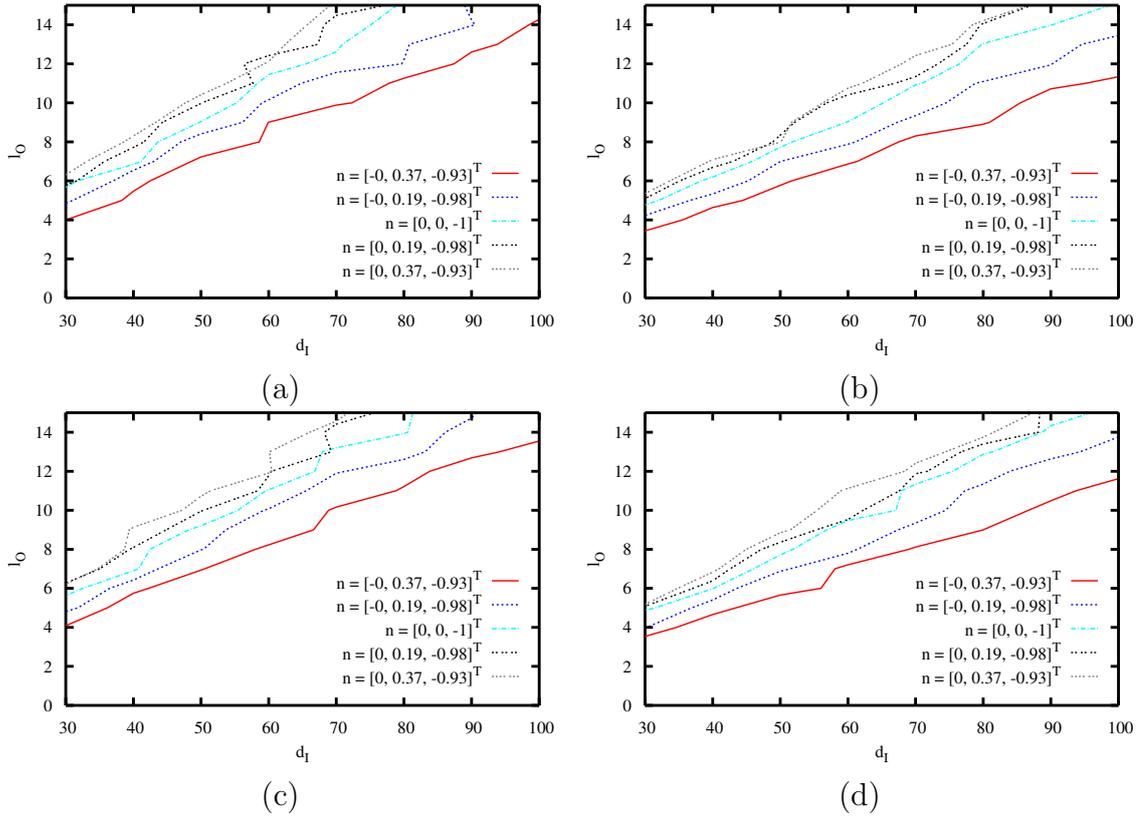


FIG. 3.7: Lignes de niveau de la surface Ω_O , associées à différents vecteurs normal au plan \mathbf{n}_I , en fonction de d_I et l_O pour $\Omega_O = 80\%$. Ces surfaces sont calculées avec un taux de recouvrement entre images $\rho = 80\%$, et $(\theta, \phi, \psi) = (0^\circ, 0^\circ, 0^\circ)$. Les figures (a) et (b) illustrent les performances des estimateurs sans normalisation des coordonnées alors que (c) et (d) présentent celles avec cette normalisation. Les estimateurs minimisant une homographie sont représentés par les figures (a) et (c), et ceux minimisant une affinité sont sur les figures (b) et (d).

en compte la rapidité de la baisse de performance engendrée par un angle de seulement 2.5° .

En outre, on remarquera encore une fois l'insensibilité de la normalisation des coordonnées sur les performances des estimateurs.

Performance des estimateurs

L'étude des figures précédentes a permis de mettre en évidence certaines particularités que l'on retrouve dans la figure 3.9. Les lignes de niveaux $\Omega(d_I, l_O) = 80\%$ avec $\rho = 80\%$ des estimateurs influant sur \mathcal{E}_O y sont représentées pour des variations de normales de $\mathbf{n}_I = (0, 0, -1)^\top$ à $\mathbf{n}_I = (0.348, 0.348, -0.87)^\top$ sur les figures (a) et (b) et pour des variations de la rotation de 0° à 5° sur (c) et (d).

Nous retrouvons donc avec les figures (c) et (d) la sensibilité des estimateurs affines

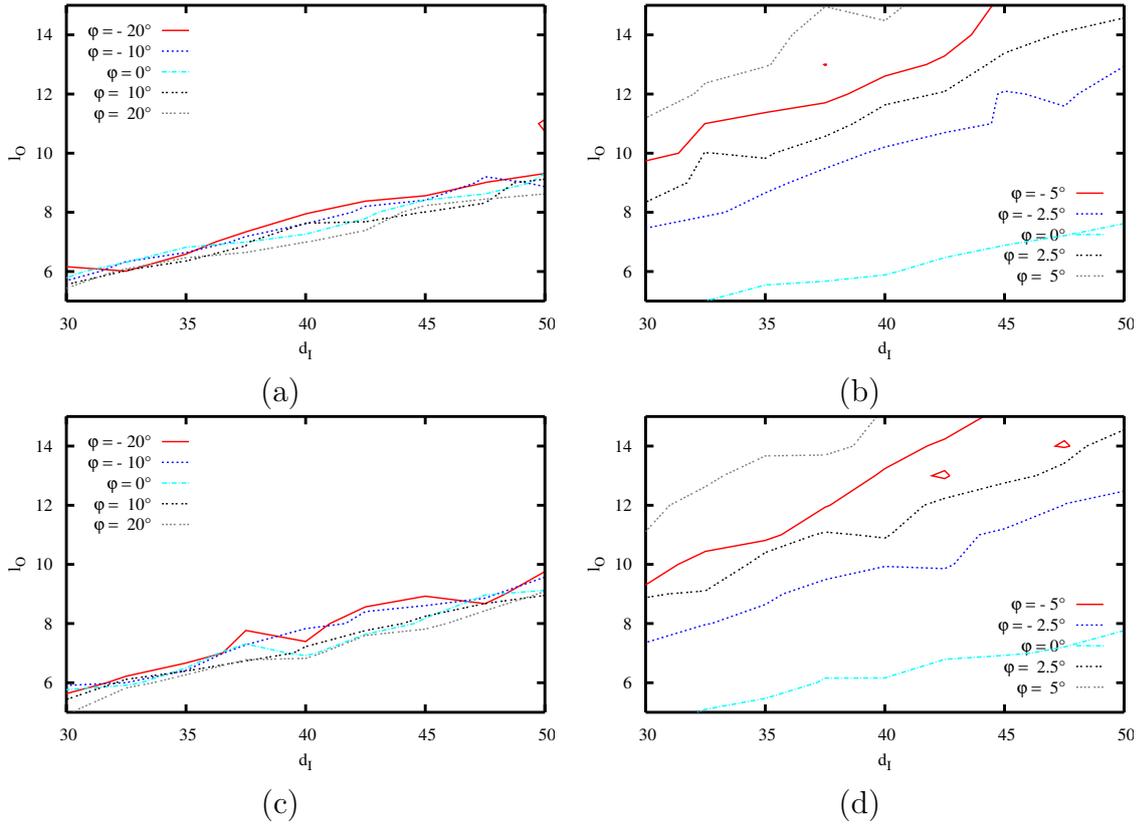


FIG. 3.8: Lignes de niveau de la surface Ω_O , associées à différents angles de rotation ϕ entre les caméras, en fonction de d_I et l_O pour $\Omega_O = 80\%$. Ces surfaces sont calculées avec un taux de recouvrement entre images $\rho = 80\%$, $\mathbf{n}_I = (0, 0, -1)^T$ et $(\theta, \psi) = (0^\circ, 0^\circ)$. Les figures (a) et (b) illustrent les performances des estimateurs sans normalisation des coordonnées alors que (c) et (d) présentent les performances avec normalisation. Les estimateurs minimisant une homographie sont représentés par les figures (a) et (c), et ceux minimisant une affinité sont sur les figures (b) et (d).

aux rotations et la bonne stabilité des performances des estimateurs d'homographies. Nous pouvons aussi constater qu'il n'en est pas de même pour la normale \mathbf{n}_I qui a une influence relativement forte sur leur performances.

Les figures (a) et (c) démontrent parfaitement les meilleures performances des estimations d'affinité dès lors que ce modèle est correct. On peut tout de même remarquer que le modèle d'homographie décrit aussi ce mouvement tout en incluant 2 degrés de libertés supplémentaires inexploités et donc pourrait produire des résultats de qualité similaires. Mais l'obtention de l'homographie optimale implique alors la minimisation d'une matrice \mathbf{A} qui n'est pas de rang plein. Nos expériences nous ont clairement montré qu'en présence de bruit sur les points $\{\mathbf{m}, \mathbf{m}'\}$ la minimisation du système $\mathbf{A}\mathbf{h}$ peut amener à des solutions impliquant ces deux degrés de libertés supplémentaires. Il s'avère même parfois possible de trouver une solution entièrement dégénérée où ces degrés de libertés sont prépondérants

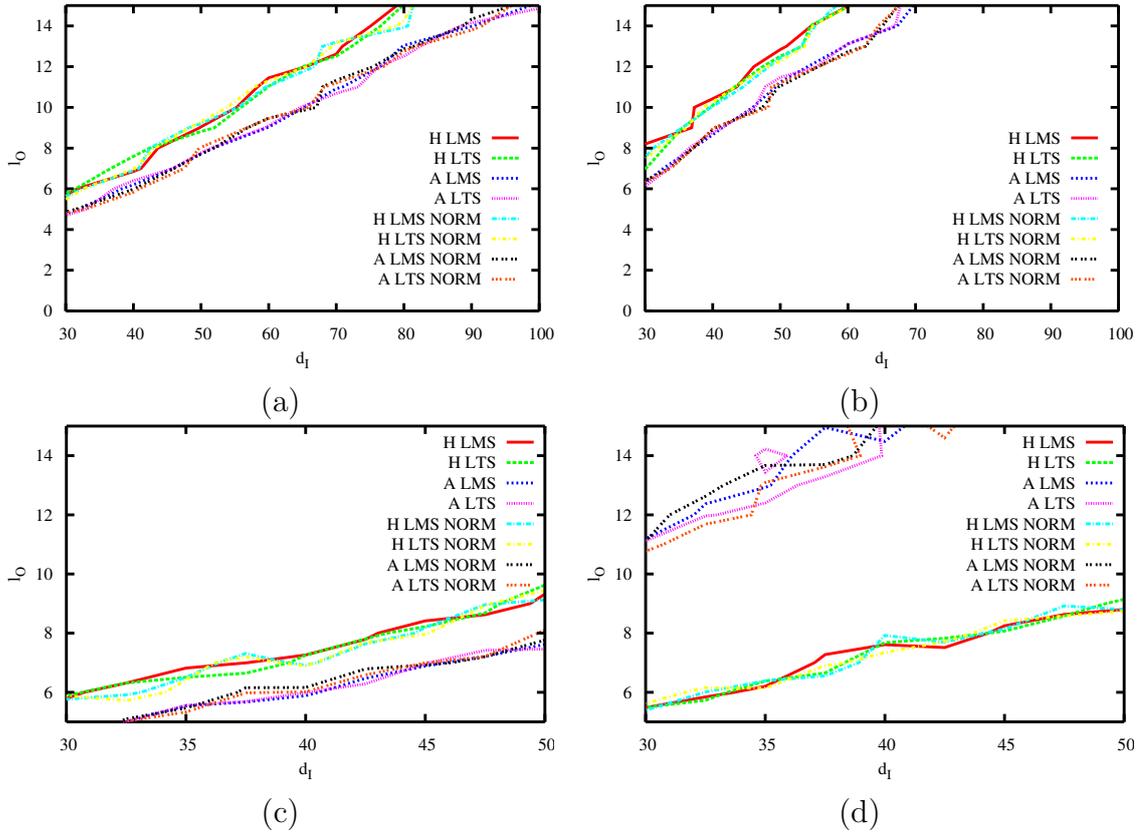


FIG. 3.9: Lignes de niveau des surfaces Ω_O de chaque estimateur robuste, en fonction de d_I et l_O pour $\Omega_O = 80\%$. Ces surfaces sont toutes calculées avec un taux de recouvrement entre images $\rho = 80\%$. Les figures (a) et (b) sont associées aux angles $(\theta, \phi, \psi) = (0^\circ, 0^\circ, 0^\circ)$ et ont pour normale respective $\mathbf{n}_I = (0, 0, -1)^\top$ et $\mathbf{n}_I = (0.348, 0.348, -0.87)^\top$. Les figures (c) et (d) sont associées à la normale $\mathbf{n}_I = (0, 0, -1)^\top$ et aux angles respectifs $(\theta, \phi, \psi) = (0^\circ, 0^\circ, 0^\circ)$ et $(\theta, \phi, \psi) = (0^\circ, 5^\circ, 0^\circ)$.

sur ceux de la translation et rotation.

Enfin cette même figure 3.9 ainsi que la figure 3.10 nous permettent de statuer sur l'influence de la normalisation des coordonnées. Comme on peut le voir celle-ci n'entraîne pas d'améliorations significatives des résultats dans les cas de figure que nous avons présentés. Cette dernière figure 3.10 qui compare les performances des estimateurs selon le recouvrement ρ a pour but d'exhiber des cas critiques où le nombre de points peut se retrouver très réduit et très localisé du fait d'un plus faible recouvrement. Toutefois ce cas de figure ne révèle pas non plus d'améliorations significatives des performances du à la normalisation.

La normalisation des coordonnées étant une méthode fréquemment adoptée pour l'estimation d'homographie, il s'agit là d'une conclusion relativement surprenante que nous devons pondérer. Nous la faisons d'abord dans le cas d'estimateurs robustes utilisant des méthodes de vote et cette conclusion ne nous semble donc pas pertinente dans le cadre d'une minimisation itérative du système. Le fait que les méthodes de votes réalisent un

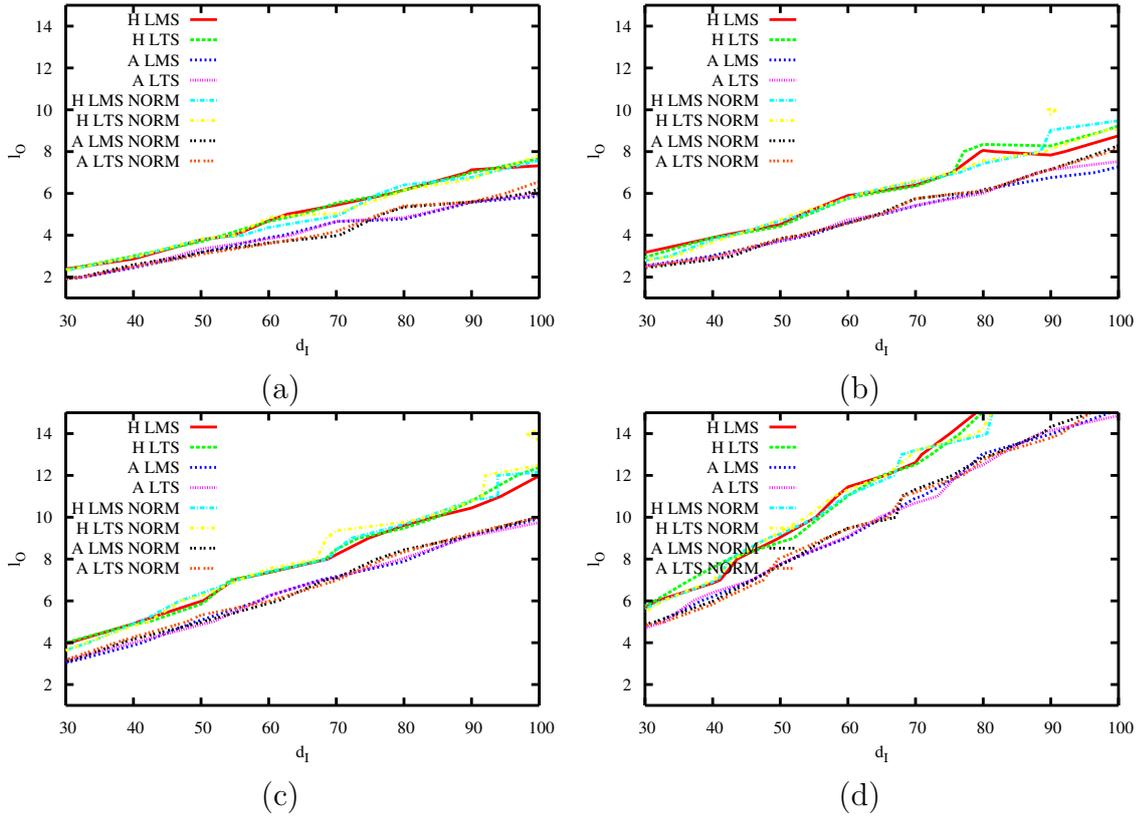


FIG. 3.10: Lignes de niveau des surfaces Ω_O de chaque estimateur robuste, en fonction de d_I et l_O pour $\Omega_O = 80\%$. Ces surfaces sont toutes calculées avec les angles $(\theta, \phi, \psi) = (0^\circ, 0^\circ, 0^\circ)$ et la normale $\mathbf{n}_I = (0, 0, -1)^\top$. Les figures (a), (b), (c) et (d) sont associées respectivement au taux de recouvrement entre images $\rho = 50\%$, $\rho = 60\%$, $\rho = 70\%$ et $\rho = 80\%$.

grand nombre de tirages pourrait s'avérer suffisant pour obtenir des estimées de H_I d'égale qualité. L'adoption dans nos algorithmes du *bucket-mapping* que nous détaillons au § 3.2.4 peut s'avérer être une autre piste d'explication. Enfin, nous avons réalisé ces tests sur des données de type aériennes donc bien que l'amélioration induite par la normalisation ne soit pas majeure lors de rotations élevées ou de normales éloignées de l'axe optique, cas où il est possible que la normalisation soit plus intéressante.

Enfin, nous avons pu constater que les estimateurs utilisant la fonction de coût LTS produisent les meilleures estimées de H_I dans tous les cas.

Influence du conditionnement du système linéaire

Comme nous l'avons abordé au § 2.3.1 il est nécessaire d'ajouter une contrainte supplémentaire au système que l'on minimise pour prendre en compte le facteur d'échelle. Cette contrainte peut s'ajouter en fixant h_{33} à 1 ou en supposant que le déterminant \mathbf{h} soit 1. Comme expliqué précédemment la contrainte $h_{33} = 1$ induit la recherche d'une

estimée de $\hat{\mathbf{h}}$ de degré 8, là où la seconde conduit à rechercher une estimée $\hat{\mathbf{h}}$ de degré 9. Nous n'avons pas comparé de façon exhaustive ces deux méthodes, car ceci nous aurait amené à étudier non plus 20 mais 40 estimateurs distincts. Tous nos estimateurs utilisent donc la première contrainte et minimisent des systèmes du type $\mathbf{A}\mathbf{h} = \mathbf{b}$. Nous avons toutefois introduit quelques estimateurs utilisant la seconde contrainte et n'avons constaté aucune différence notable pour les performances sur Ω_O . En revanche, les résidus obtenus avec les estimateurs utilisant la seconde contrainte ont tendance à être légèrement plus élevés comme illustré par la figure 3.11. Ce résultat nous semble dû au degré de liberté supplémentaire sur \mathbf{h} qui peut sous-contraindre le système et produire de moins bonnes estimées de \mathbf{H}_I .

Toutefois comme expliqué au § 2.3.1, l'utilisation de la contrainte $\|\mathbf{h}\| = 1$ peut induire de meilleures performances lorsque l'on estime une homographie \mathbf{H}_I associée à un plan Π_I dont la ligne de fuite tend vers l'horizon. Le paramètre h_{33} tendant alors vers 0, le fixer à 1 peut causer des solutions dégénérées.

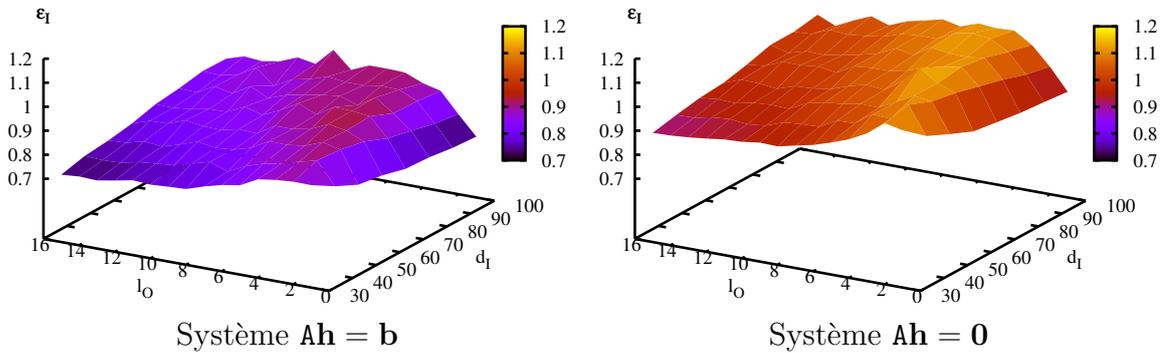


FIG. 3.11: Comparaison des résidus sur les inliers ϵ_I avec ou sans normalisation des coordonnées. Le résidu est représenté en fonction de d_I et l_O avec $\mathbf{t} = (0, 8, 0)^\top$, $(\theta, \psi, \phi) = (0^\circ, 0^\circ, 0^\circ)$ et $\mathbf{n} = (0.1, 0.1, 0.99)^\top$.

3.2.3 Analyse des estimateurs robustes avec des données réelles

Nous avons pu établir une première évaluation des estimateurs robustes par méthodes de vote en utilisant des données de synthèse. Afin de confirmer ces premiers résultats nous comparons les résidus ϵ_I associés à des estimations d'homographie sur plusieurs centaines de paires d'images acquises avec des robots aériens et terrestres. Les deux types d'environnement ainsi perçus ont pour principale distinction l'angle sous lequel est perçu le plan du sol. En effet, là où dans pour les images aériennes la normale à ce plan est proche de $[0, 0, -1]^\top$, pour les images terrestres elle tend à s'approcher de $[0, -1, 0]^\top$.

Nous illustrons ces résidus par la figure 3.12 pour les paires d'images aériennes et par la figure 3.13 pour les paires d'images terrestres. Afin d'améliorer la lisibilité de ces courbes, nous séparons sur ces deux figures les estimateurs d'homographie (a), (c) des estimateurs

d'affinité (b), (d). Les estimateurs n'utilisant pas de normalisation des coordonnées (a), (b) sont aussi représentés séparément de ceux l'utilisant (c), (d).

La valeur moyenne des résidus obtenus par chacun des estimateurs a aussi été reporté dans les tableaux 3.1 et 3.2.

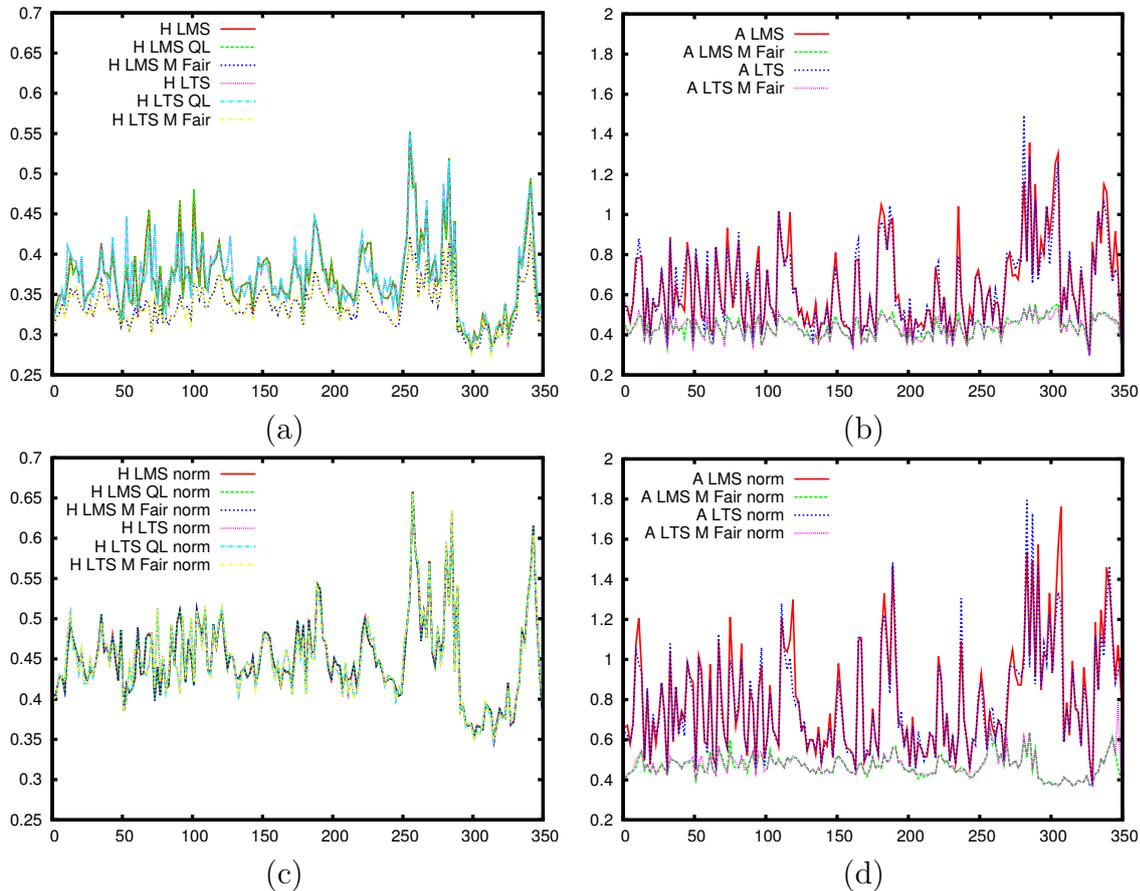


FIG. 3.12: Résidus ϵ_I des estimateurs robustes sur des paires d'images aériennes. (a) et (c) illustrent les estimateurs d'homographies, (b) et (d) les estimateurs d'affinité. Les estimateurs se basant sur des coordonnées normalisées sont illustrés par les figures (c) et (d) et ceux n'utilisant pas de normalisation par (a) et (b).

L'analyse de ces résultats nous permet de confirmer nos conclusions issues de l'étude des données de synthèse. Ainsi, les estimateurs n'utilisant pas de normalisation des coordonnées donnent toujours déjà de meilleurs résultats que ceux l'utilisant. Ceci confirme donc nos résultats précédents et permet de les valider dans un contexte d'images terrestres.

On peut aussi constater que l'optimisation des estimées d'homographies par le M-estimateur de Fair ou l'optimisation quasi-linéaire n'apporte pas de réelle amélioration. Pour bien comprendre le sens physique de ce résultat, il est intéressant de développer le processus d'optimisation itératif commun à ces méthodes : la méthode des moindres carrés repondérés (§ 2.3.2). On procède pour cette optimisation à une pondération des

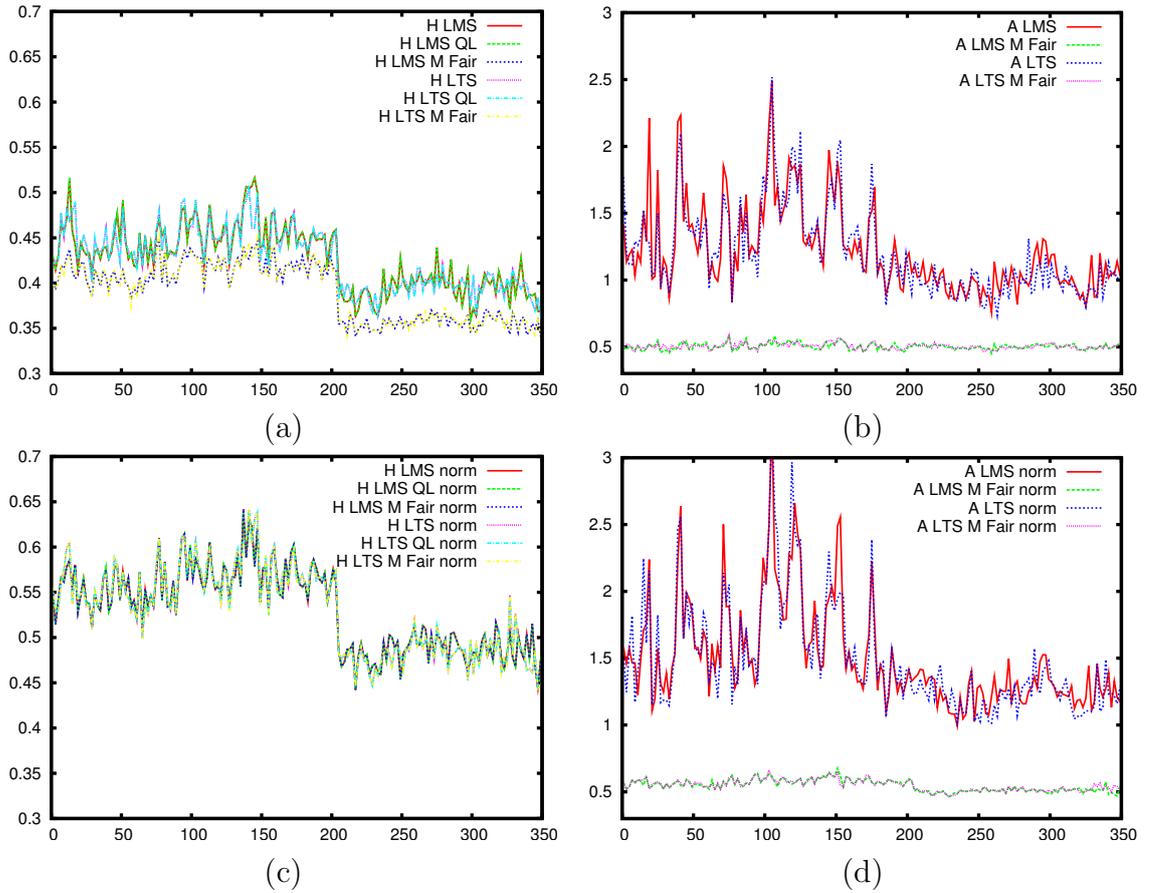


FIG. 3.13: Résidus ϵ_I des estimateurs robustes sur des paires d'images terrestres. (a) et (c) illustrent les estimateurs d'homographies, (b) et (d) les estimateurs d'affinité. Les estimateurs se basant sur des coordonnées normalisées sont illustrés par les figures (c) et (d) et ceux n'utilisant pas de normalisation par (a) et (b).

entrées du système linéaire jusqu'à convergence vers une solution. Cette convergence ne peut être établie avec un seuil sur le résidu, celui-ci étant inconnu. Nous utilisons donc deux coefficients relatifs à la convergence :

- k_1 qui est la somme des différences entre les des paramètres des estimées de \mathbf{h} successives :

$$k_1 = \sum_i |h_i^k - h_i^{k-1}|$$

- k_2 qui est la somme des différences entre les coefficients de pondération successifs du système :

$$k_2 = \sum_i |w_i^k - w_i^{k-1}|$$

Nous considérons que la convergence est atteinte si les coefficients k_1 et k_2 sont inférieurs à 0.0001. Ces optimisations ne permettent que de descendre vers un minimum à partir de l'estimation initiale de \mathbf{H}_I issue des estimateurs par méthode de votes. Ainsi,

ne pas obtenir d'amélioration sur le résidu après optimisation implique que l'estimée de H_I initiale tends déjà vers un minimum local.

Nous pouvons donc constater que les estimateurs d'homographies utilisant le LMS et le LTS produisent des estimées de H_I très proches et quasi minimales. En revanche, dans le cas de l'estimation d'affinité on peut constater une amélioration significative des résidus après optimisation par le M-estimateur Fair. Cependant, il faut pondérer cette observation notamment dans le cas d'images terrestres. Une homographie associée à un faible résidu est en effet trouvée mais il s'avère qu'elle est supportée par un faible nombre de points et est donc dégénérée.

Enfin, nous remarquons que l'estimation d'affinité est donc comme attendu inutile pour des images terrestres mais s'avère plus efficace pour des images aériennes. En pratique, nous obtenons de meilleurs résultats avec les estimateurs affine lors de mouvements particuliers de caméra tel que des mouvements planaires ou des translations lorsque le plan focal est parallèle à celui du sol. Ces cas sont assez peu fréquents mais méritent d'être pris en compte.

	H LMS	H LMS QL	H LMS M Fair	H LTS	H LTS QL
Sans normalisation	0.370	0.370	0.336	0.371	0.371
Avec normalisation	0.451	0.451	0.451	0.450	0.450

H LTS M Fair	A LMS	A LMS M Fair	A LTS	A LTS M Fair
0.337	0.639	0.436	0.631	0.434
0.450	0.796	0.468	0.782	0.470

TAB. 3.1: Moyennes des résidus ϵ_I des estimateurs robustes sur des paires d'images aériennes.

	H LMS	H LMS QL	H LMS M Fair	H LTS	H LTS QL
Sans normalisation	0.524	0.524	0.524	0.523	0.523
Avec normalisation	0.422	0.422	0.385	0.421	0.422

H LTS M Fair	A LMS	A LMS M Fair	A LTS	A LTS M Fair
0.523	1.478	0.543	1.468	0.544
0.385	1.199	0.505	1.185	0.506

TAB. 3.2: Moyennes des résidus ϵ_I des estimateurs robustes sur des paires d'images terrestres.

- Ces observations sur des données réelles nous amènent donc aux conclusions suivantes :
- L'utilisation du M-estimateur de Fair améliore un peu les estimées mais pas de manière significative.
 - L'estimation d'affinité peut s'avérer utile pour des images aériennes et occasionnellement permet d'obtenir de meilleurs résultats que l'estimation d'homographie.

- L'estimateur LTS a des performances très proches du LMS. Il est en pratique un peu plus performant que celui-ci en présence de beaucoup d'*outliers*.

3.2.4 Algorithme d'estimation mis en oeuvre

Nos précédentes conclusions sur les performances des estimateurs nous ont donc amené à utiliser l'estimateur d'homographie suivant. Il s'agit d'un estimateur par méthode de vote minimisant la fonction de coût LTS. Sa mise en oeuvre recourt aux optimisations suivantes dont la première est préconisée par [Zhang 95b].

Sélection des points : Comme expliqué au § 2.3.3, nous réalisons des tirages de sous-ensembles de points parmi l'ensemble des points mis en correspondance. Si les points d'un sous-ensemble sont spatialement trop proches, ceci fausse le processus d'estimation de l'homographie. Nous créons donc une grille d'occupation de la répartition des points dans les images et forçons le tirage de façon à tirer des sous-ensembles de points provenant de cellules différentes. Cette optimisation s'avère extrêmement bénéfique en améliorant très sensiblement la qualité et la robustesse des résultats.

Élimination des *outliers* à postériori : Pour chaque paire d'images I_{k-1} I_k où une homographie a été estimée, un ensemble de points $(m_{O_{k-1}}, m_{O_k})$ a été segmenté comme étant des *outliers*. Pour la paire d'images suivante, nous éliminons alors de l'ensemble des points (m_k, m_{k+1}) mis en correspondance les *outliers* m_{O_k} précédemment détectés. Cette technique permet d'accroître significativement la robustesse de nos estimées en enlevant des *outliers* en amont du processus d'estimation. Il est ainsi possible d'estimer convenablement des homographies dans des paires d'images contenant plus de 50 % d'*outliers*.

Sélection de modèle : Une double estimation d'homographie et d'affinité est réalisée pour chaque paire d'image. Nous effectuons ensuite une sélection assez frustrante car uniquement basée sur le résidu afin de sélectionner l'estimée optimale. En pratique cette méthode s'avère donner de bons résultats, les résidus associés aux affinités n'étant inférieurs à ceux des homographies que lorsque celles-ci sont dégénérées. Ce comportement est donc conforme à nos exigences.

3.3 Etude quantitative des possibilités de détection

Comme nous l'avons abordé au § 3.1 nous allons détecter les zones planes qui satisfont l'homographie H_I estimée en effectuant une corrélation entre l'image de référence et une image recalée par rapport à celle-ci. La transformation appliquée lors de ce recalage est H_I et doit donc recalculer au mieux la zone plane supportée par H_I tout en mettant en évidence la parallaxe par rapport à Π_I des zones non planes.

Cette approche de segmentation suppose donc d'obtenir un résidu de reprojection entre les images recalées dû à la parallaxe suffisamment élevé des zones non planes pour être détectable par corrélation. Nous proposons donc dans ce qui suit une quantification de ce résidu en fonction des paramètres de structure de l'environnement d_I et l_O afin de déterminer un ordre de grandeur des possibilités de détection de notre algorithme. Nous supposons pour cela un mouvement en translation du type $\mathbf{t} = (tx, ty, 0)^\top$ et sans rotation avec $\mathbf{R} = \mathbf{I}$. C'est un cas simplifié mais peu éloigné des conditions de vols habituelles d'un drone (figure 3.14).

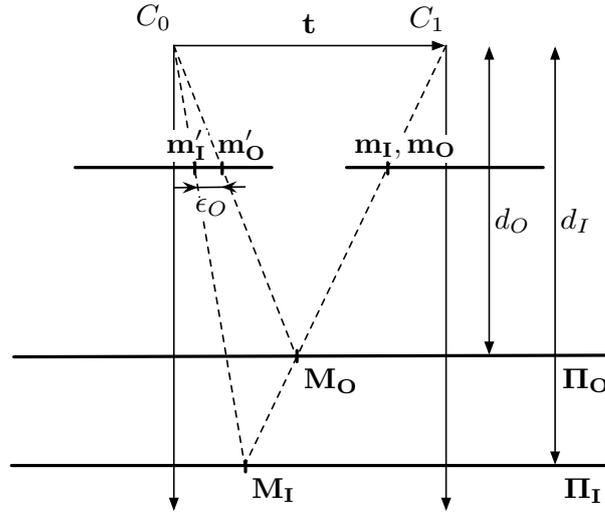


FIG. 3.14: Géométrie d'une paire d'image dans un cas simplifié où la rotation \hat{R} est une identité, la translation \mathbf{t} est $(tx, ty, 0)^\top$ et la normale au plan Π_I est $\mathbf{n}_I = [0, 0, -1]^\top$. Le point M_I est un inlier et \mathbf{m}_I est son projeté dans la seconde image. Le point M_O est un outlier tel que son projeté \mathbf{m}_O soit confondu avec \mathbf{m}_I dans la seconde image. Le plan Π_O est le plan coplanaire à Π_I et passant par M_O . L'écart entre \mathbf{m}'_I et \mathbf{m}'_O est donc l'erreur de reprojection ϵ_O et aussi la parallaxe de \mathbf{m}_O par rapport à l'homographie H_I induite par Π_I .

Afin de présenter le principe de notre méthode de calcul, nous proposons donc une première formulation de ce résidu dans l'hypothèse où il est calculé pour un outlier M_O situé sur un plan Π_O . Par une astuce mathématique, nous le généraliserons ensuite à n'importe quel type d'outliers.

Hypothèse d'outliers sur un plan : On notera $(\mathbf{m}_I, \mathbf{m}'_I, H_I)$ un triplet formé des inliers satisfaisant l'homographie H_I et $(\mathbf{m}_O, \mathbf{m}'_O, H_0)$ un triplet formé par les outliers. On peut écrire :

$$H_I = \mathbf{C} \left(\mathbf{R} - \frac{\mathbf{t}\mathbf{n}_I^\top}{d_I} \right) \mathbf{C}^{-1} \text{ et } H_0 = \mathbf{C} \left(\mathbf{R} - \frac{\mathbf{t}\mathbf{n}_O^\top}{d_O} \right) \mathbf{C}^{-1}$$

Maintenant considérons le résidu en pixels $\epsilon_O = d_A(\mathbf{m}', \mathbf{H}_I \mathbf{m})^2$ qui sera calculé en chaque point. Comme illustré par la figure 3.14, \mathbf{m} peut être soit la projection d'un point \mathbf{M}_I situé sur la surface plane recherchée, soit celle d'un point \mathbf{M}_O sur l'autre plan. On a donc $\mathbf{m} = \mathbf{m}_I = \mathbf{m}_O$. Si \mathbf{m} est la projection de \mathbf{M}_O alors on a :

$$\epsilon_O = d_A(\mathbf{m}'_O, \mathbf{H}_I \mathbf{m})^2 = d_A(\mathbf{H}_O \mathbf{m}, \mathbf{H}_I \mathbf{m})^2$$

Ce qui nous permet d'écrire en considérant que $\mathbf{R} = \mathbf{I}$ et $t_z = 0$:

$$\begin{aligned} \epsilon_O = \|\mathbf{H}_O \mathbf{m}, \mathbf{H}_I \mathbf{m}\| &= \|(\mathbf{H}_O - \mathbf{H}_I) \mathbf{m}\| \\ &= \left\| \left[\mathbf{C} \left(\mathbf{R} - \frac{\mathbf{t} \mathbf{n}_I^T}{d_I} \right) \mathbf{C}^{-1} - \mathbf{C} \left(\mathbf{R} - \frac{\mathbf{t} \mathbf{n}_O^T}{d_O} \right) \mathbf{C}^{-1} \right] \mathbf{m} \right\| \\ &= \left\| \left[\mathbf{C} \left(\frac{\mathbf{t} \mathbf{n}_O^T}{d_O} - \frac{\mathbf{t} \mathbf{n}_I^T}{d_I} \right) \mathbf{C}^{-1} \right] \mathbf{m} \right\| \\ &= \left\| \left[\mathbf{C} \left(\frac{\mathbf{t}(d_I \mathbf{n}_O^T - d_O \mathbf{n}_I^T)}{d_I d_O} \right) \mathbf{C}^{-1} \right] \mathbf{m} \right\| \end{aligned} \quad (3.2)$$

Ceci nous permet de déduire les conclusions suivantes :

1. Dans le cas où les plans Π_I et Π_O sont fronto-parallèles au plan focal ($\mathbf{n}_I = \mathbf{n}_O = [0, 0, 1]^T$), ϵ_O tend vers $\| [\mathbf{0}_{3 \times 2} \quad \mathbf{p}_{3 \times 1}] \mathbf{m} \|$ avec $\mathbf{p}_{3 \times 1} = [\alpha_u t_x, \alpha_v t_y, 0]^T$. ϵ_O est donc constant quelque soit la position de \mathbf{m} dans l'image (attendu que $\mathbf{m}[3] = 1$). Il ne dépend plus que de la translation \mathbf{t} et des paramètres de calibrage de la caméra.
2. Si les plans Π_I et Π_O sont parallèles, ϵ_O évolue linéairement en fonction de $\frac{d_I - d_O}{d_I d_O}$.

Cas général : On peut généraliser le raisonnement précédent. En effet, pour un point \mathbf{M}_O il existe une infinité de plans Π_O passant par ce point. On peut alors se permettre d'introduire un plan Π_O tel que $\mathbf{n}_O = \mathbf{n}_I$ en notant toujours d_O la distance normale de ce plan à la caméra. Ceci nous permet alors d'écrire :

$$\mathbf{H}_I = \mathbf{C} \left(\mathbf{R} - \frac{\mathbf{t} \mathbf{n}_I^T}{d_I} \right) \mathbf{C}^{-1} \text{ et } \mathbf{H}_O = \mathbf{C} \left(\mathbf{R} - \frac{\mathbf{t} \mathbf{n}_I^T}{d_O} \right) \mathbf{C}^{-1}$$

Et en utilisant la même démarche que précédemment en prenant toujours en compte que $\mathbf{R} = \mathbf{I}$ et $t_z = 0$ on obtient alors :

$$\begin{aligned} \epsilon_O = \|\mathbf{H}_O \mathbf{m}, \mathbf{H}_I \mathbf{m}\| &= \|(\mathbf{H}_O - \mathbf{H}_I) \mathbf{m}\| \\ &= \left| \frac{d_I - d_O}{d_I d_O} \right| \| (\mathbf{C} \mathbf{t} \mathbf{n}_I^T \mathbf{C}^{-1}) \mathbf{m} \| \end{aligned} \quad (3.3)$$

Ce résultat démontre que le résidu évolue linéairement avec le rapport noté K_d des distances normales à la caméra d_I et d_O . Il est donc possible de quantifier pour un plan Π_I et une translation \mathbf{t} donnée le résidu en fonction de d_O .

Nous pouvons choisir d'introduire $k = (d_I - d_O)/d_I = l_O/d_I$ qui est le rapport entre la distance de \mathbf{M}_O à Π_I et celle de \mathbf{C} à Π_I afin de tenter d'exprimer l'effet de la distance à Π_I sur le résidu. On peut alors noter :

$$K_d = \left| \frac{d_I - d_O}{d_I d_O} \right| = \left| \frac{k}{d_O} \right| = \frac{1}{d_I} \left| \frac{k}{1 - k} \right|$$

On peut déjà remarquer comme résultat notable que K_d et donc ϵ_O décroissent selon l'inverse de d_I . Considérons ensuite l'évolution de K_d pour $1/d_I = 1$ comme illustré par la figure 3.15. On peut ainsi voir que le comportement de K_d est asymptotique en $k = 1$. C'est un comportement très normal compte tenu que cela équivaut à $d_O = 0$, soit \mathbf{M}_O confondu avec le centre optique de la caméra \mathbf{C} .

Comme attendu, la valeur de K_d (et donc le résidu ϵ_O) est nulle en $k = 0$ soit pour $d_O = d_I$. Le point M_O est alors sur le plan Π_I .

Enfin, un résultat moins évident est le comportement de K_d si $k < 0$ c'est à dire si $d_O > d_I$ ou encore $l_O < 0$. On peut alors voir que K_d tends vers 1. L'influence de la parallaxe induite par \mathbf{M}_O tends donc vers 0 et ϵ_O vers $\|(\mathbf{Ctn}_I^T \mathbf{C}^{-1}) \mathbf{m}\|$.

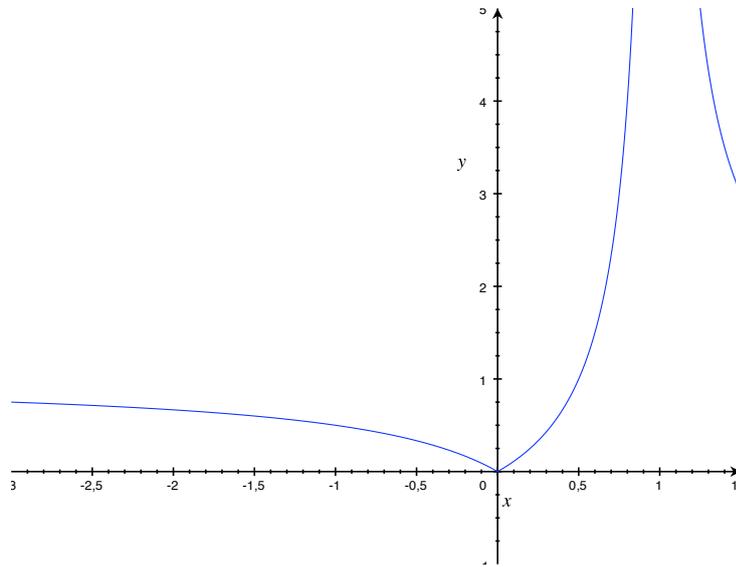


FIG. 3.15: Illustration de l'influence de d_O et d_I sur le résidu mise en évidence par l'évolution du rapport K_d en fonction de k pour $1/d_I = 1$.

Conclusions : Sans perte de généralité mais afin de simplifier les équations, nous supposons la translation comme étant simplement $\mathbf{t} = (tx, 0, 0)^T$. Les autres paramètres $\mathbf{R} = \mathbf{I}$ et $\mathbf{n} = (0, 0, 1)^T$ restant inchangés on peut alors obtenir :

$$\epsilon_O = \left| \frac{d_I - d_O}{d_I d_O} \right| \left\| \begin{bmatrix} 0 & 0 & \alpha_u t_x \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{m} \right\| = \left| \frac{l_O}{d_I(d_I - l_O)} \right| \alpha_u t_x$$

Le paramètre α_u dépendant linéairement de la focale f , nous pouvons constater que le résidu ϵ_O est proportionnel à la focale. En d'autres termes, pour des mêmes paramètres de mouvement et le même environnement, une focale courte qui implique un champ de vue large minimisera le résidu ϵ_O par rapport à une focale longue et un champ de vue réduit.

Le recouvrement entre deux images dans un tel cas s'obtient très simplement. En effet, la translation t_u subie par les pixels de l'image étant alors constante, on peut calculer l'homographie induite par la translation \mathbf{t} et l'appliquer en $\mathbf{m} = (0, 0, 1)^\top$, la translation obtenue par $\mathbf{m}' - \mathbf{m}$ est alors égale à t_u . En notant, L le nombre de pixels de l'image selon $(0u)$ on obtient :

$$\rho = 1 - \frac{t_u}{L} = 1 - \frac{\alpha_u t_x}{d_I L}$$

On peut ainsi exprimer ϵ_O en fonction de d_I , l_O et ρ :

$$\epsilon_O = \left| \frac{l_O}{(d_I - l_O)} \right| (1 - \rho)L \quad (3.4)$$

En utilisant la formule précédente et en considérant des images de 512 colonnes, il nous donc possible d'établir les valeurs théoriques du résidu ϵ_O en fonction de d_I , l_O et du taux de recouvrement τ de la paire d'images. Le tableau 3.3 présente ces résultats.

	$d_I = 25m$	$d_I = 50m$	$d_I = 75m$	$d_I = 100m$	$d_I = 200m$
$l_O = 0.5m, \tau = 90\%$	1.04	0.52	0.34	0.25	0.12
$l_O = 0.5m, \tau = 70\%$	3.18	1.55	1.03	0.77	0.38
$l_O = 0.5m, \tau = 50\%$	5.22	2.58	1.71	1.28	0.64
$l_O = 1m, \tau = 90\%$	2.13	1.04	0.69	0.51	0.25
$l_O = 1m, \tau = 70\%$	6.4	3.13	2.8	1.54	0.77
$l_O = 1m, \tau = 50\%$	10.66	5.22	3.45	2.08	1.28
$l_O = 5m, \tau = 90\%$	12.8	5.68	3.65	2.69	1.31
$l_O = 5m, \tau = 70\%$	38.4	17.06	10.97	8.08	3.93
$l_O = 5m, \tau = 50\%$	64	28.4	18.3	13.5	6.56

TAB. 3.3: Correspondance entre des résidus ϵ_O en pixels et des outliers à une distance l_O du plan de référence $\Pi_{\mathbf{I}}$. Ce plan est à une distance d_I de la caméra. Les taux de recouvrement ρ sont calculés en prenant en compte des images de 512 pixels de large. Le mouvement considéré entre les images est $(\mathbf{R} = \mathbf{I}, \mathbf{t} = (t_x, 0, 0)^\top)$.

Ces résultats sont intéressants mais il faut se garder d'y attribuer une valeur autre que qualitative. En effet, on peut voir que la parallaxe induite par l'application de l'homographie $\mathbf{H}_{\mathbf{I}}$ sur la projection \mathbf{m}_O d'un point \mathbf{M}_O à 50cm d'un plan $\Pi_{\mathbf{I}}$ à $d_I = 75m$ de la caméra, est d'un pixel si le recouvrement entre images est de 70%. Un tel résultat s'avèrerait extrêmement précis mais on se doit d'y ajouter les réserves suivantes :

- L'homographie $\mathbf{H}_{\mathbf{I}}$ doit être estimée parfaitement. Or on sait [Hartley 04] que sous l'hypothèse d'un bruit blanc gaussien de variance σ unitaire sur les points $(\mathbf{m}, \mathbf{m}')$ mis en correspondance, le résidu ϵ_I sur les N_I inliers est borné par :

$$\epsilon_I > \epsilon_{res} = \sigma \left(\frac{N_I - 4}{2N_I} \right)^{1/2}$$

où ϵ_{res} tends vers la limite asymptotique $\sigma/\sqrt{2}$. A cette erreur sur l'estimation de H_I s'ajoute celle induite par des outliers proches de Π_I .

- dans notre approche la parallaxe induite est détectée par corrélation sur les pixels. Dans le cas de faible différence de texture entre la surface plane et les outliers, la parallaxe doit alors être de plusieurs pixels pour être apparente lors de la corrélation.

On peut ainsi fixer une borne basse, se voulant conservative, sur les possibilités de détection de zones non planes, en imposant que le triplet (d_I, l_O, τ) induise une parallaxe d'au moins 4 pixels (1px pour l'estimation de H_I et 3px pour la corrélation).

3.4 Segmentation des zones planes

Le calcul d'une homographie et sa validation nous permettent de segmenter un ensemble \mathcal{E}_I de N_I correspondances $(\mathbf{m}, \mathbf{m}')$ parmi N , qui appartiennent à un même plan Π_I de l'environnement. On peut décrire cette information comme une connaissance éparse de l'environnement. La seule information ainsi obtenue sur sa structure est l'appartenance des *inliers* au plan associé à H_I .

Afin d'obtenir une connaissance plus dense de l'environnement nous proposons de densifier l'application de H_I non plus aux seuls *inliers* mais à chaque point de l'image. Pour cela, nous procédons au recalage des vues en utilisant l'homographie H_I estimée. Seuls les pixels du plan Π_I sont correctement recalés, les autres subissent la parallaxe relative au plan Π_I comme décrit précédemment par (3.1). Pour mettre en évidence ce que nous appellerons, par abus de langage, les zones planes (les zones de l'image dont les pixels sont issus de la projection du plan Π_I sur le plan image), nous utilisons le score de corrélation ZNCC tel que décrit au § 3.4.1. Après avoir discuté dans le § 3.4.2 des limitations de ce score nous proposerons au § 3.4.3 une méthode permettant de s'en affranchir et de segmenter de manière automatique les images de corrélation.

3.4.1 Score de corrélation ZNCC (Zero Normalized Cross Correlation)

Comme illustré par la figure 3.3 les images de référence et les images recalées sont ensuite corrélées. Le score de corrélation utilisé est le *Zero-mean Normalized Cross Correlation* (ZNCC). Il est calculé sur un voisinage local, appelé fenêtre de corrélation, et noté W_w . Il s'écrit :

$$ZNCC(\mathbf{m}) = \frac{\sigma_{1,2}(\mathbf{m})}{\sqrt{\sigma_1^2(\mathbf{m})\sigma_2^2(\mathbf{m})}} \quad (3.5)$$

où :

$$\sigma_{1,2}(\mathbf{m}) = \frac{1}{(2w+1)^2} \sum_{i=u-w}^{u+w} \sum_{j=v-w}^{v+w} (I_1(i, j) - \mu_1(u, v)) (I_2(i, j) - \mu_2(u, v))$$

$$\sigma_k^2(\mathbf{m}) = \frac{1}{(2w+1)^2} \sum_{i=u-w}^{u+w} \sum_{j=v-w}^{v+w} (I_k(i,j) - \mu_k(u,v))^2$$

$$\mu_k(\mathbf{m}) = \frac{1}{(2w+1)^2} \sum_{i=u-w}^{u+w} \sum_{j=v-w}^{v+w} I_k(i,j)$$

$\sigma_{1,2}(\mathbf{m})$ représente la covariance entre les pixels du voisinage W_w dans les images 1 et 2, $\sigma_k^2(\mathbf{m})$ la variance et $\mu_k(\mathbf{m})$ la moyenne des pixels sur le même voisinage pour l'image k .

L'avantage principal du score ZNCC est son indépendance à des changements affines de la luminosité entre images. D'un point de vue expérimental, avec des caméras réglées automatiquement en balance des blancs, gains et temps de pose, cette propriété est indispensable. L'échelle des scores va de -1 pour une corrélation opposée entre les voisinages considérés, jusqu'à 1 en cas d'égalité stricte entre tous les pixels des fenêtres considérées.

Nous nous attendons donc à des scores de corrélation très proches de 1 dans les zones planes et à des scores inférieurs pour les zones présentant une parallaxe.

3.4.2 Limitations du score ZNCC

L'étude expérimentale des résultats de corrélation met en évidence deux types de problèmes à l'issue de la corrélation. Ceux-ci sont illustrés sur la figure 3.16. Le premier type d'artefact est induit par des motifs texturés dont l'intervalle de répétition est de l'ordre de la précision du recalage des pixels. Les scores de corrélation présentent alors des extrémas erronés (cf (c) et (d) de la figure 3.16)

Le deuxième type de problème est plus difficile à mettre en évidence. Une analyse fine des scores de corrélation permet de constater que pour des pixels d'une même zone plane, les scores de corrélation varient en fonction de la variance locale. Les images (e) et (f) de la figure 3.16 illustrent ce comportement. On peut y constater une corrélation entre les variances et scores de corrélation faibles. Ainsi, malgré la normalisation de ce score par les variances des fenêtres de corrélation W_w , il subsiste une dépendance du score à ce paramètre.

Cette observation déjà réalisée par [Oriot 03] peut être visualisée plus quantitativement en représentant les scores de corrélation ZNCC en fonction de la variance. Ainsi la figure 3.17 illustre les densités de probabilité $f_l(s)$ et fonctions de répartition $F_l(s)$ des scores de corrélation pour différentes classes de variance \mathcal{C}_l (l identifie la classe de variance).

Les scores de corrélation qui y sont représentés sont calculés uniquement sur des *inliers* et doivent donc théoriquement tendre vers 1 . Ces scores ont ici été segmentés en 6 classes distinctes attachées à des intervalles de variance croissants. Comme on peut le voir, la distribution de ces scores diffère sensiblement selon la classe de variance \mathcal{C}_l . [Oriot 03] a montré ainsi une amélioration sensible de la corrélation en régularisant les scores de telle sorte que les fonctions de répartition ne dépendent plus de la classe de variance.

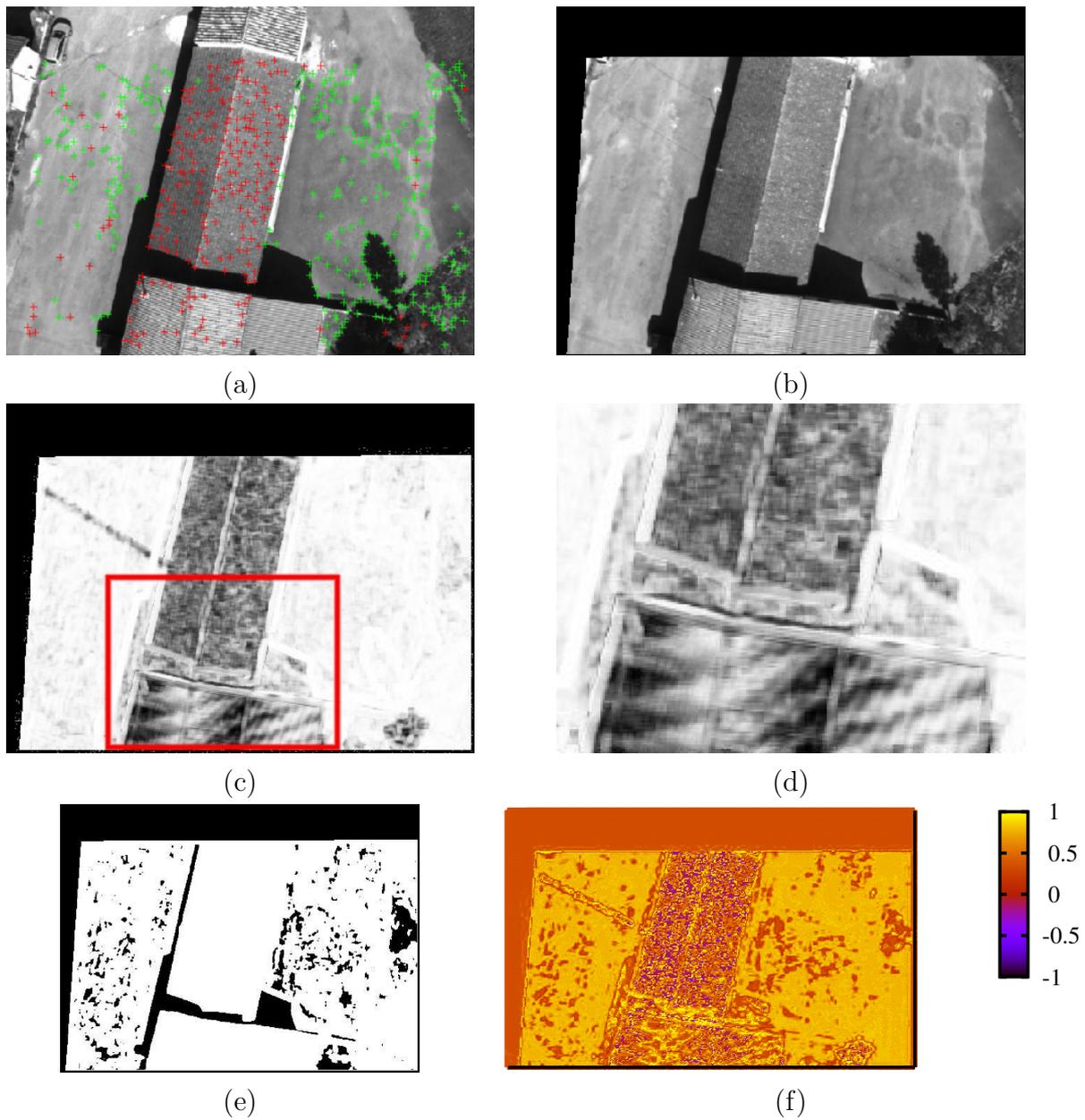


FIG. 3.16: Illustration des types d'artefacts de corrélation. (a) est l'image de référence. (b) est la seconde image recalée selon H_T^{-1} , (c) est l'image des scores de corrélation ZNCC entre (a) et (b), (d) magnifie la zone présentant des problèmes de corrélation, (e) est l'image des variances locales de (a) seuillée (les zones noires indiquent une faible variance, les blanches des variances élevées), (f) est l'image de corrélation (c) avec une palette de couleur qui permet de mieux différencier les scores ZNCC.

Afin de bien saisir le principe et les limitations de cette méthode de régularisation, il est indispensable de s'attarder sur la notion d'échantillonnage des scores. En effet, afin d'utiliser correctement cette technique il convient de prendre en compte et d'éviter les

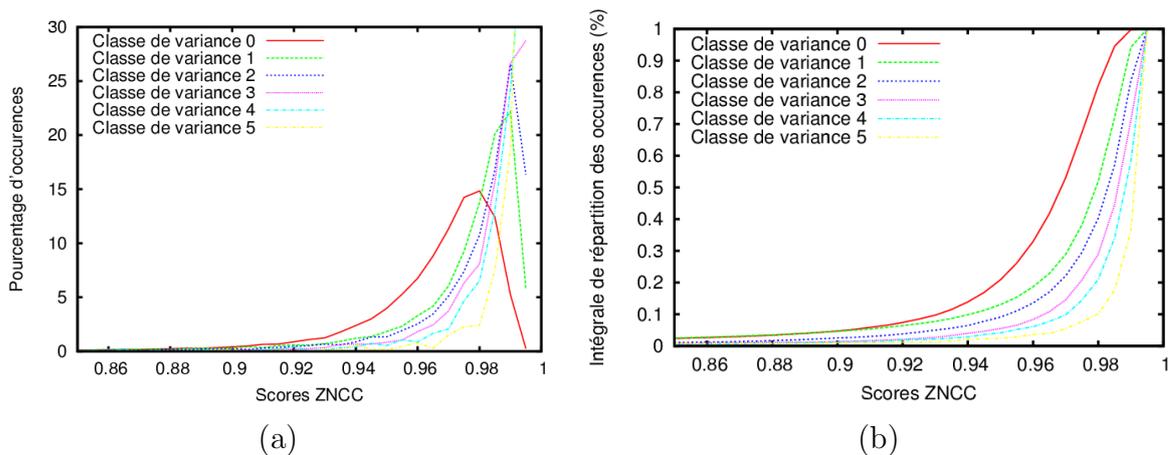


FIG. 3.17: Répartitions des scores de corrélation ZNCC en fonction de différentes classes de variance \mathcal{C}_l , établis sur environ 70000 points d'intérêts mis en correspondance. (a) Densité de probabilité discrète $f_l(s)$ (histogramme), (b) Fonction de répartition discrète $F_l(s)$ (intégrale de l'histogramme). Le pas d'échantillonnage sur les scores est de 0.01.

deux écueils suivants :

- D'une part le sous-échantillonnage des scores de corrélation. Le pas d'échantillonnage des scores ZNCC lors de la construction des histogrammes qui nous permettent de calculer les fonctions de répartition doit être suffisamment petit. Sans quoi les transitions des fonctions seront trop abruptes et les différentes classes de variance indifférenciables.
- D'autre part le problème inverse, c'est à dire le sur-échantillonnage des scores. En effet, ceci occasionne une sous-quantification des fonctions de répartition pour chaque pas d'échantillonnage. Il est alors impossible de mener une régularisation consistant à pondérer les scores afin d'obtenir des fonctions de répartition résultantes identiques quelque soit la classe de variance.

Cette méthode nécessite donc un nombre important de scores ZNCC pour chacune des classes \mathcal{C}_l afin de permettre un échantillonnage correct des fonctions de répartition. Elle est ainsi en pratique [Oriot 03] utilisée avec l'ensemble des scores d'images de corrélation dense.

3.4.3 Correction du score de corrélation ZNCC en ligne

Création des classes de variances \mathcal{C}_l

Une régularisation des scores de corrélation apparaît ainsi nécessaire afin de garantir une bonne segmentation des images de corrélation. Nous souhaitons seuiller les scores de corrélation pour mettre en évidence les zones planes satisfaisant l'homographie H_I . Si nous ne régularisons pas les scores, il apparaît donc que la mise en place d'un seuil τ_S unique

sur des scores non régularisés ne peut être une solution satisfaisante.

En outre, l'étude des scores pour différents types d'images nous a permis de constater que la répartition des scores au sein d'une même classe de variance diffère très nettement selon les images. La figure 3.18 présente quelques résultats. Ce constat implique de calculer une régularisation différente pour chaque paire d'images. Il n'est pas envisageable de calculer cette régularisation une fois pour toute.

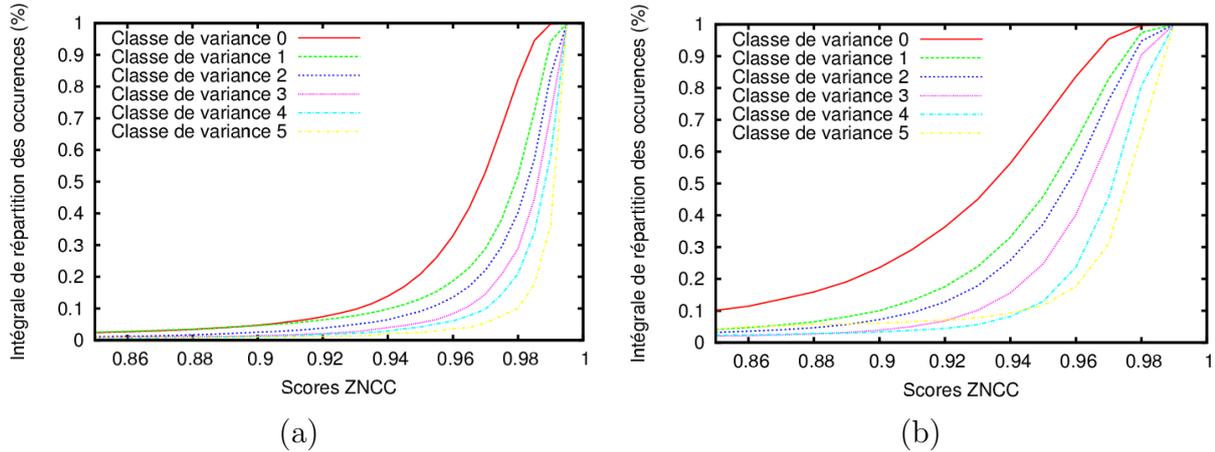


FIG. 3.18: Fonction de répartition discrète $F_l(s)$ des scores de corrélation ZNCC en fonction de différentes classes de variance \mathcal{C}_l pour deux séries d'images distinctes. Ces deux séries ont été acquises avec des caméras identiques et en des lieux différents.

La recherche d'un seuil τ_S permettant de segmenter les images conduit naturellement à s'intéresser aux scores des *inliers*. Étant par définition la projection de points du plan $\Pi_{\mathbf{I}}$, leur score régularisé doit être au delà du seuil τ_S . Dès lors, deux démarches sont envisageables :

1. Régulariser les scores en utilisant l'ensemble des scores de corrélation puis déterminer un τ_S unique comme un minorant des scores d'*inliers*.
2. Effectuer une régularisation en ne s'appuyant que sur les scores d'*inliers* avant de déterminer τ_S .

La première méthode implique d'abord une somme de calcul importante si l'on utilise tous les scores. Mais plus important, les densités de probabilités alors utilisées incluent les scores de pixels de zones non planaires, rendant la régularisation hasardeuse sur les densités de probabilité des *inliers*. Nous avons donc utilisé la seconde approche en n'utilisant que les scores d'*inliers*. Cela nous a amené à proposer une nouvelle méthode s'affranchissant de régularisation pour pallier à ce nombre réduit de scores et trouver automatiquement des seuils τ_{S_l} associés à chaque classe de variance \mathcal{C}_l .

Du fait du faible nombre de scores utilisés un phénomène supplémentaire est à prendre en compte : le sur-échantillonnage non plus des scores mais des variances. En effet à chaque score S est associé une variance (appartenant à une classe \mathcal{C}_l) et en créant des

classes de variance comprenant des intervalles trop étendus, certaines classes se retrouvent trop dépeuplées pour en extraire une quelconque information statistique. On a alors un sur-échantillonnage des scores. Ce phénomène est notablement accentué par le fait que les *inliers* sont des points de Harris et donc des pixels de variance locale relativement forte.

Nous proposons donc d'ordonner les variances des *inliers* et de créer les classes de variance de telle sorte qu'elles correspondent à un nombre égal de scores fixé par avance empiriquement comme étant suffisant. Le nombre de classes dépend donc du nombre d'*inliers* mais chaque classe contient le même nombre de scores.

La figure 3.19 illustre un exemple de densité de probabilité et de fonction de répartition utilisant cette méthode. Les densités de probabilité sont typiquement sur-échantillonnées, mais les fonctions de répartition tendent vers leur comportement attendu du fait de l'intégration des échantillons.

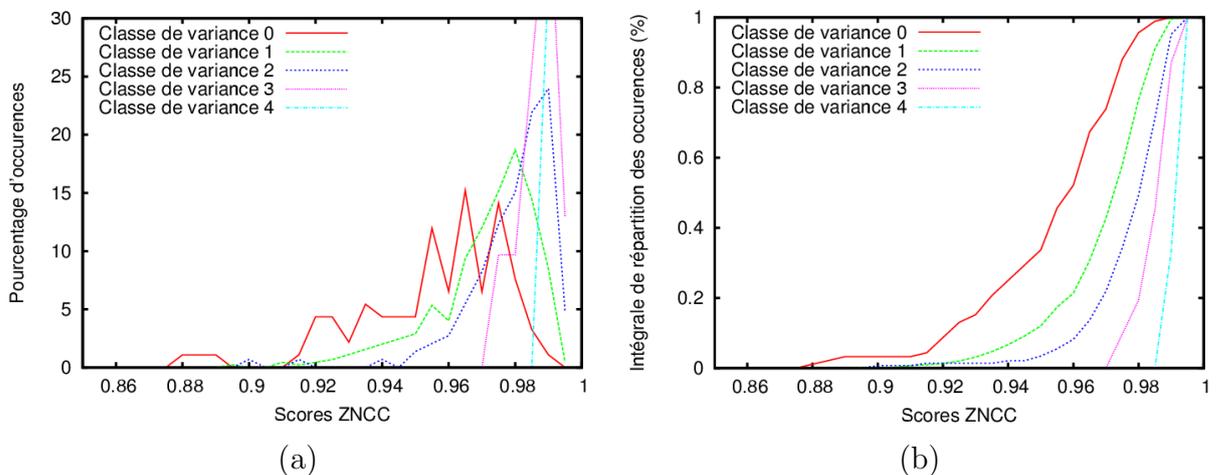


FIG. 3.19: Scores de corrélation ZNCC en fonction de différentes classes de variance pour une seule paire d'image établis sur environ 400 correspondances d'*inliers*. (a) Densité de probabilité discrète (histogramme), (b) Fonction de répartition discrète (intégrale de l'histogramme). Le pas d'échantillonnage sur les scores est de 0.005.

Calcul des seuils τ_{S_i}

Plutôt que de régulariser les densités de probabilités $f_l(s)$ nous avons choisi de calculer un seuil τ_{S_i} par classe de variance \mathcal{C}_l . Chaque seuil est choisi tel que $F_l(\tau_{S_i}) = 1\%$.

Du fait du sur-échantillonnage, l'intégrale de répartition que nous obtenons pour les *inliers* est trop imprécise pour permettre de trouver ces seuils. Nous proposons une modélisation de ces intégrales $F_l(s)$ s'appuyant sur la distribution de Gumbel [Gumbel 35] (aussi appelée distribution de Fisher-Tippett [Fisher 28]), qui est un cas particulier de la distribution de Weibull lorsque $k = e$ (tableau 3.4).

	Distribution de Weibull	Distribution de Gumbel
Densité de probabilité	$(k/\lambda)(x/\lambda)^{(k-1)}e^{-(x/\lambda)^k}$	ze^{-z}/β avec $z = e^{-(x-\mu)/\beta}$
Fonction de répartition	$1 - e^{-(x/\lambda)^k}$	e^{-z} avec $z = e^{-(x-\mu)/\beta}$

TAB. 3.4: Définition des distributions de Weibull et Gumbel (aussi appelée distribution de Fisher-Tippett)

La modélisation que nous avons retenue possède les propriétés suivantes :

$$\text{Densité de probabilité : } \hat{f}_l(s) = e^{-e^{(s-\mu_l)/\beta_l}} * e^{(s-\mu_l)/\beta_l} / \beta_l \quad (3.6)$$

$$\text{Fonction de répartition : } \hat{F}_l(s) = 1 - e^{-e^{(s-\mu_l)/\beta_l}} \quad (3.7)$$

Les paramètres de ce modèle s'obtiennent de manière naturelle, en posant tout d'abord $\beta_l = (1 - \mu_l)/2$. En effet, en fixant ainsi β_l nous modifions la distribution de telle sorte que $F_l(1) = 0.999382$ et assurons une convergence extrêmement forte vers la valeur 1 pour $s > 1$.

L'expérimentation nous a montré que, du fait de l'intégration réalisée, la largeur à mi-hauteur S_{M_l} est un paramètre relativement constant des intégrales de répartition $F_l(s)$.

Nous calculons donc μ_l en posant $\hat{F}_l(S_{M_l}) = 50\%$ ce qui nous donne :

$$\mu_l = \frac{\ln(\ln(2)) - 2S_{M_l}}{\ln(\ln(2)) - 2}$$

Cette modélisation, qui ne s'appuie que sur le seul paramètre S_{M_l} , s'avère très satisfaisante, comme on peut le voir en comparant $F_l(s)$ et $f_l(s)$ avec leur modélisation sur les figures 3.20 et 3.21.

Les seuils τ_{S_l} s'obtiennent très simplement en posant $\hat{F}_l(\tau_{S_l}) = 1\%$, ce qui nous permet d'obtenir la formulation suivante :

$$\tau_{S_l} = \frac{\ln(-\ln(1 - 0.01)) * (S_{M_l} - 1) + \ln(\ln(2)) - 2S_{M_l}}{\ln(\ln(2)) - 2} \quad (3.8)$$

Comme on peut le voir sur la figure 3.20, ce seuil peut s'avérer un peu trop haut, mais nous nous en satisfaisons dans le sens où il n'entraîne qu'une segmentation un peu trop conservatrice des scores.

L'algorithme (2) résume l'ensemble de la démarche nous permettant de trouver les seuils τ_{S_l} . Le calcul des scores sur les images est réalisé en utilisant les optimisations décrites en annexe § A.

3.4.4 Correction du score de corrélation ZNCC hors-ligne

Comme nous l'avons vu les scores de corrélation S dépendent de la variance $\sigma_{W_w}^2$ de la fenêtre de corrélation W_w pour laquelle ils sont calculés. L'utilisation de classes de variance

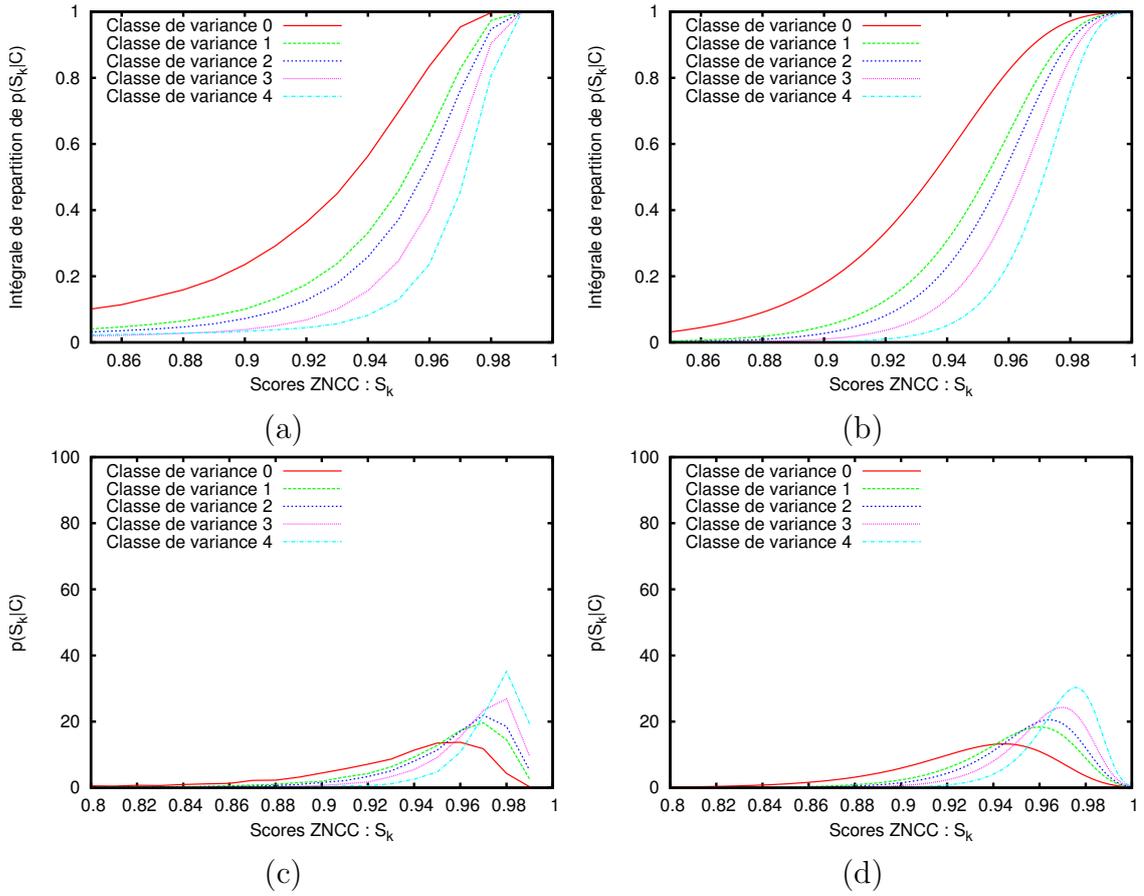


FIG. 3.20: Comparaison entre intégrales de répartition et densités de probabilités expérimentales (a) (c) et modélisées (b) (d). Les données expérimentales reposent sur 60000 points d'intérêts mis en correspondance.

Algorithme 2 : Calcul automatique des seuils τ_{S_i} des scores de corrélation ZNCC.

1. Calculer la variance locale σ_i^2 autour de chaque inlier \mathbf{m}_i
 2. Trier les N_I variances et déterminer L intervalles de variance de façon qu'ils recouvrent le même nombre de σ_i^2
 3. Regrouper les scores S en classes \mathcal{C}_l en fonction de l'intervalle recouvrant leur variance σ_i^2
 4. Calculer la valeur médiane S_{M_l} de chaque classe de scores \mathcal{C}_l soit $F(S_{M_l}) = 50\%$
 5. Calculer le seuil τ_{S_i} pour chaque classe de scores \mathcal{C}_l en utilisant l'équation (3.8)
-

\mathcal{C}_l est un moyen de prendre en compte et de s'affranchir de cette dépendance. Toutefois, en utilisant la même fonction F_l pour tout un intervalle de variance nous effectuons en quelque sorte un sous-échantillonnage de cette dépendance. Il est évident que l'utilisation

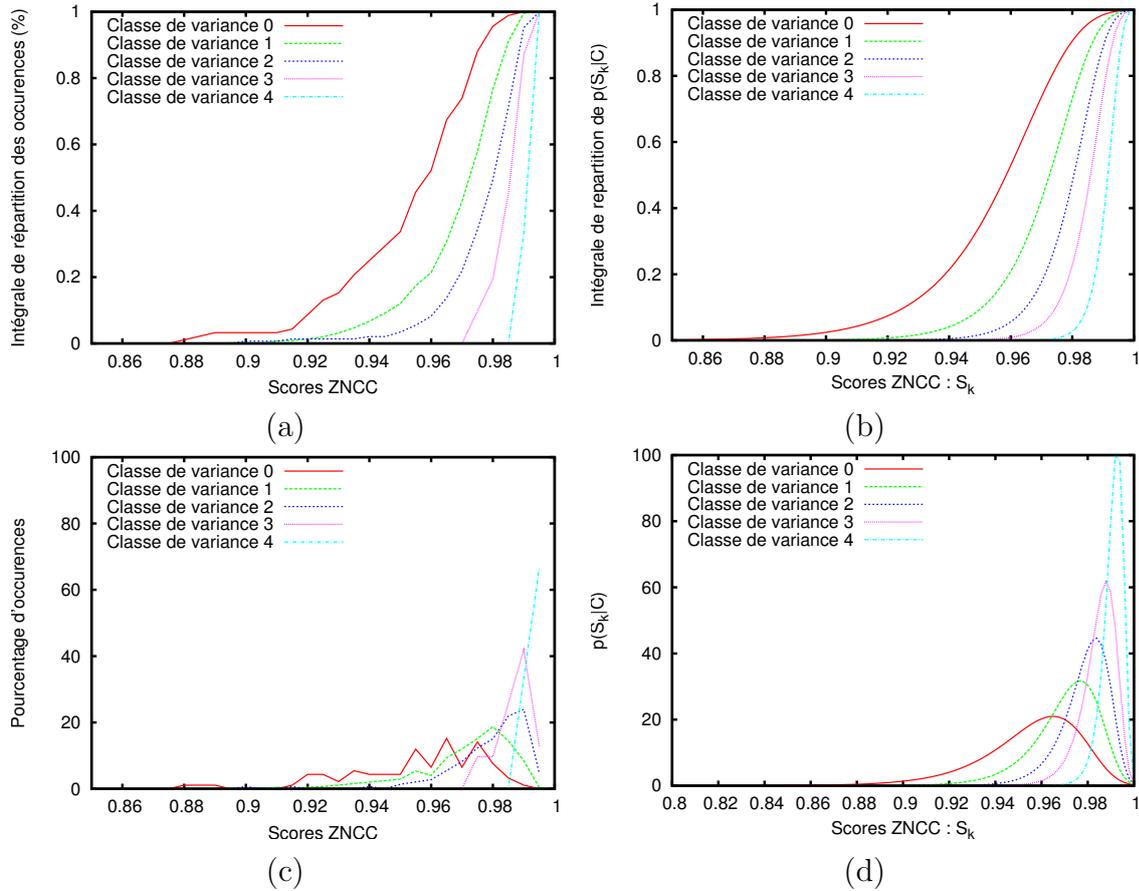


FIG. 3.21: Comparaison entre intégrales de répartition et densités de probabilités expérimentales (a) (c) et modélisées (b) (d). Les données expérimentales reposent sur environ 300 inliers mis en correspondance dans une seule paire d'images.

de fonctions F_{σ^2} dépendant directement de la variance et non d'un intervalle de variance serait une modélisation plus fine de cette dépendance.

Nous avons donc introduit une modélisation de la relation entre S_M , la médiane d'une fonction F_{σ^2} recherchée et la variance σ^2 . Ce modèle est de la forme :

$$S_M(\sigma^2) = -A(\sigma^2)^{-\frac{3}{2}} + B$$

Bien qu'imparfait, ce modèle se rapproche de la véritable relation physique existant entre ces deux grandeurs comme le montre la figure 3.22 illustrant deux exemples d'utilisation de cette formule. Comme attendu, le jeu de paramètres A, B diffère entre ces deux cas, les paramètres de la relation entre variance et distribution des scores variant selon le type d'images.

Il est à noter que l'estimation de ces résultats nécessite, en l'état, l'expertise d'un opérateur humain. Dans la pratique, nous calculons les S_{M_i} correspondants à différentes classes de variance C_i pour différentes paires d'images. Certains scores S_{M_i} pouvant être

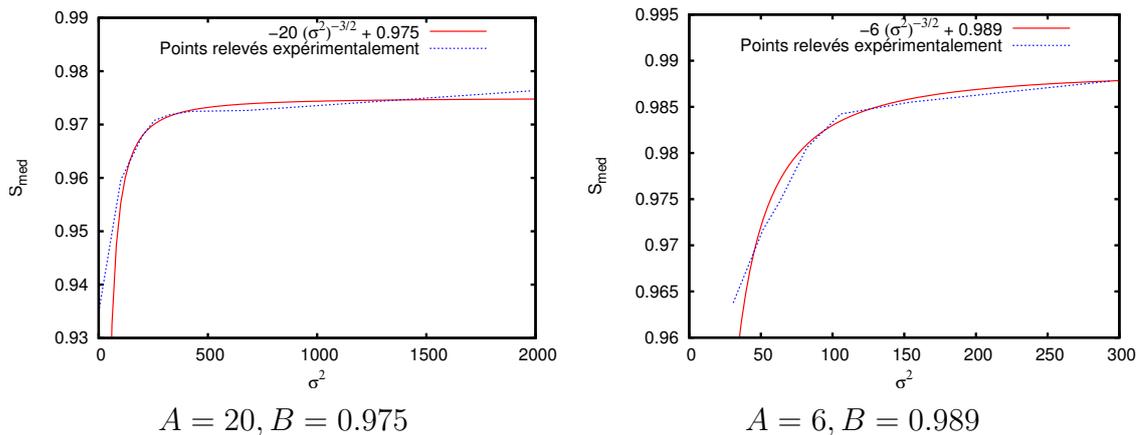


FIG. 3.22: Mise en relation de la médiane S_M et de la variance σ^2 par la fonction $S_M(\sigma^2) = -A(\sigma^2)^{-\frac{3}{2}} + B$.

mal évalués, ils sont alors rejetés manuellement. Lorsqu'est obtenu un ensemble de $\{S_{M_i}, \sigma_i^2\}$ considéré comme cohérent sur un intervalle de variance assez large, nous estimons alors les paramètres $\{A, B\}$ de la fonction $S_M(\sigma^2)$.

Une fois ces paramètres estimés, il est alors trivial de calculer S_M à partir de la variance $\sigma_{W_w}^2$ et d'en déduire un seuil τ_{σ^2} à appliquer au score S considéré en utilisant l'équation (3.8).

Nous présentons dans ce qui suit une comparaison entre cette méthode de correction et la méthode en-ligne. Toutefois, nous pouvons déjà noter que cette méthode pourrait être améliorée avec une modélisation plus fine de $S_M(\sigma^2)$.

3.5 Résultats

Les résultats suivants ont pour but d'illustrer les performances et le comportement de notre algorithme de segmentation de zones planes intégrant une correction des scores de corrélation S . Nous avons isolé des résultats mettant en évidence les possibilités de cet algorithme en terme de robustesse et d'influence du facteur d'échelle et du recouvrement entre images.

Ces analyses s'appuient uniquement sur l'utilisation de la correction de score en ligne dans la mesure où seule cette méthode est entièrement automatique et ne requiert pas d'intervention manuelle. Une comparaison entre correction en ligne et hors-ligne est réalisée séparément.

3.5.1 Robustesse

Ce premier cas illustré par la figure 3.23 présente une vue avec un nombre élevé d'*outliers* de l'ordre de 40% (ici non représentés). Ces *outliers* ont pour particularité

d'avoir une parallaxe relative à Π_I élevée du fait de la faible élévation de la caméra par rapport à la hauteur des outliers. Dans un tel cas, le fort nombre d'outliers ne nuit pas à l'estimation car la différence entre les résidus d'*inliers* et d'*outliers* est très importante.

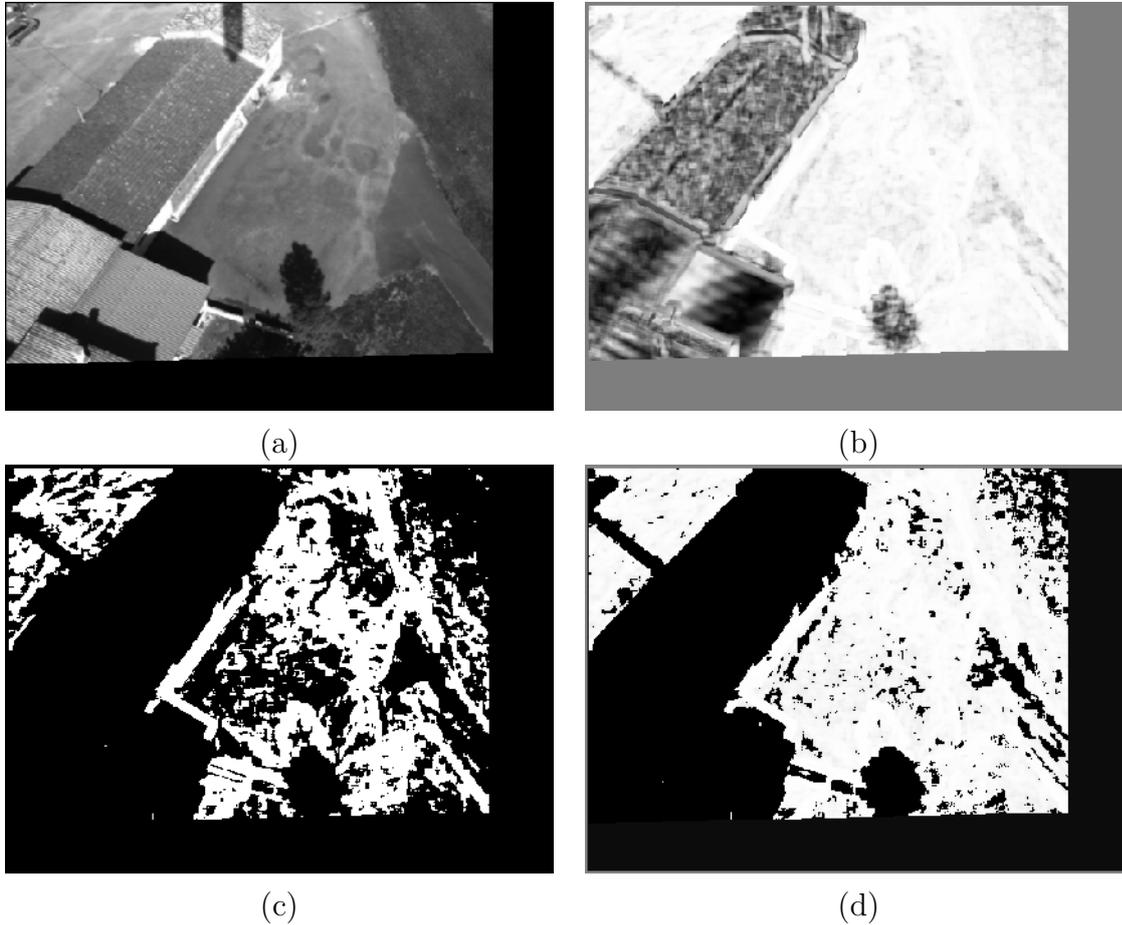


FIG. 3.23: *Illustration de la segmentation de zones planes. (a) image recalée, (b) image des scores de corrélations non corrigés, (c) image des scores seuillée manuellement par un seuil unique afin d'éliminer les zones non planes, (d) image seuillée automatiquement par notre algorithme.*

Les scores de corrélation des zones non planes, du fait de leur grande parallaxe relative et de leur texture très différente de celle du sol se détachent de manière significative des scores de la zone plane. Un tel cas de figure est en fait “quasi-idéal” : parallaxe élevée, texture des zones non planes très différente de celle des zones planes.

Toutefois l'image de corrélation (b) n'en est pas pour autant parfaite et aisément exploitable. Ainsi nous pouvons constater qu'une partie du toit a un score de corrélation élevé (en bas à gauche de (b)). Ceci est dû à une répétition de la texture du toit qui met en échec la corrélation. De plus, le muret se détache très peu de par sa faible variance (en bas au centre de (a) et (b)). Enfin le fil électrique est relativement atténué et alterne les

scores de corrélations élevés et faibles (au centre gauche de (b)).

La figure (c) est issue de l'application d'un seuil unique sur l'image de corrélation. Ce seuil est ici déterminé manuellement afin d'obtenir une segmentation des zones non planes équivalente à celle obtenue avec notre algorithme. On peut ainsi constater qu'un tel seuil unique offrant des performances similaires pour la segmentation des zones non planes s'avère classifier de façon trop importante les zones planes comme étant non planes.

Comme l'illustre la figure (d) notre algorithme réussit à passer outre ces défauts, tout en segmentant correctement les zones non planes et planes. Il tend vers une sursegmentation des zones non planes mais ce comportement conservatif est préférable dans le sens où une non détection d'obstacle peut s'avérer critique dans nos expérimentations.

3.5.2 Facteur d'échelle

Nous avons évalué la dépendance de notre méthode aux changements d'échelle. La figure 3.24 illustre deux prises de vues correspondants à des survols à deux altitudes distinctes du même environnement. La première prise correspondant à la première rangée d'images a été acquise à une altitude de 58m et la seconde à 36m.

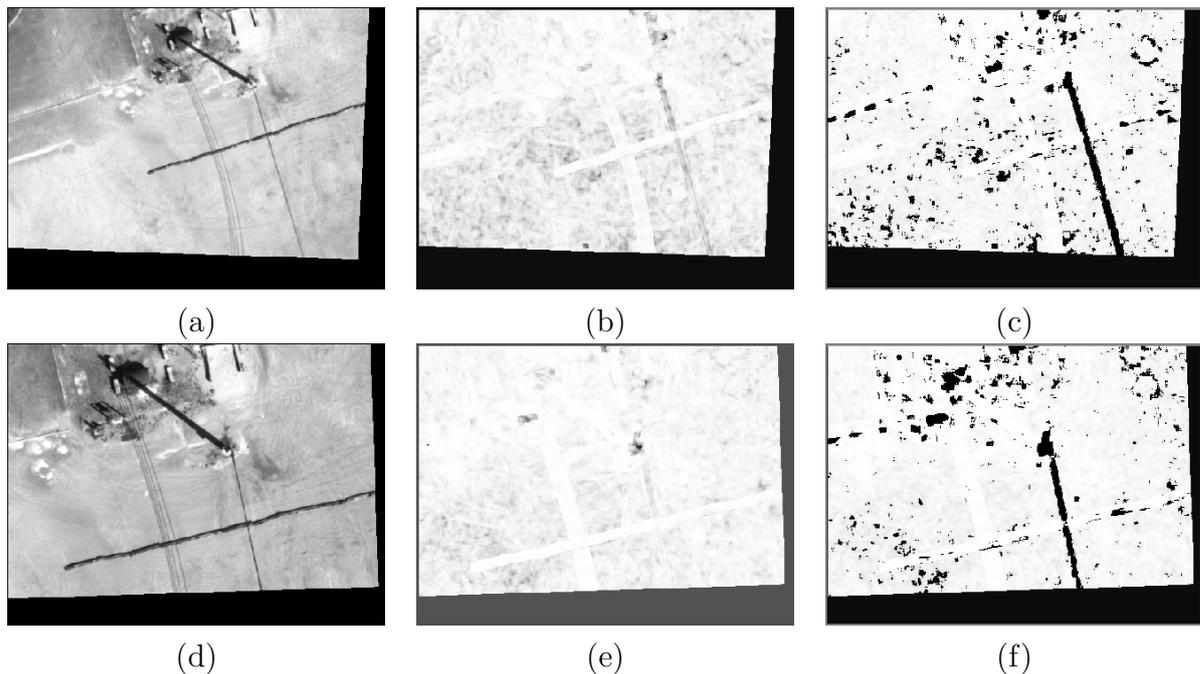


FIG. 3.24: Influence du facteur d'échelle sur la segmentation de zones planes. (a) et (d) images de référence, (b) et (e) images des scores de corrélations non corrigés, (c) et (f) images des scores seuillées automatiquement par notre algorithme.

Ces deux séries ont un faible nombre d'outliers réellement démarqué du plan Π_I mais le sol n'est pas parfaitement plan. L'homographie H_I estimée "moyenne" donc ces inliers

et l'image de corrélation résultante ne présente pas autant de scores élevés qu'attendu. Ainsi, les fils électriques traversant les deux images de bas en haut sur la partie droite des images (a) et (d) sont sur les images de corrélation peu (b) voir très peu (e) distinguables. On remarquera que les lignes qui lui sont parallèles et plus à gauche ne sont que l'ombre de ces fils électriques.

Toutefois notre algorithme, toujours en tendant vers une sursegmentation des zones non planes, détecte parfaitement ces fil électriques tout en ayant une bonne répétabilité de ces résultats entre les deux échelles.

3.5.3 Influence du recouvrement

Il est enfin intéressant de comparer les performances réelles de notre algorithme à celles, théoriques, donc nous avons donné une appréciation qualitative au § 3.3.

Ainsi la figure 3.25 présente les résultats de notre algorithme sur deux paires d'images utilisant la même image de référence et deux autres images à des baselines différentes. Le recouvrement δ_1 de la première paire est de 91%, et celui de la seconde est $\delta_2 = 78\%$.

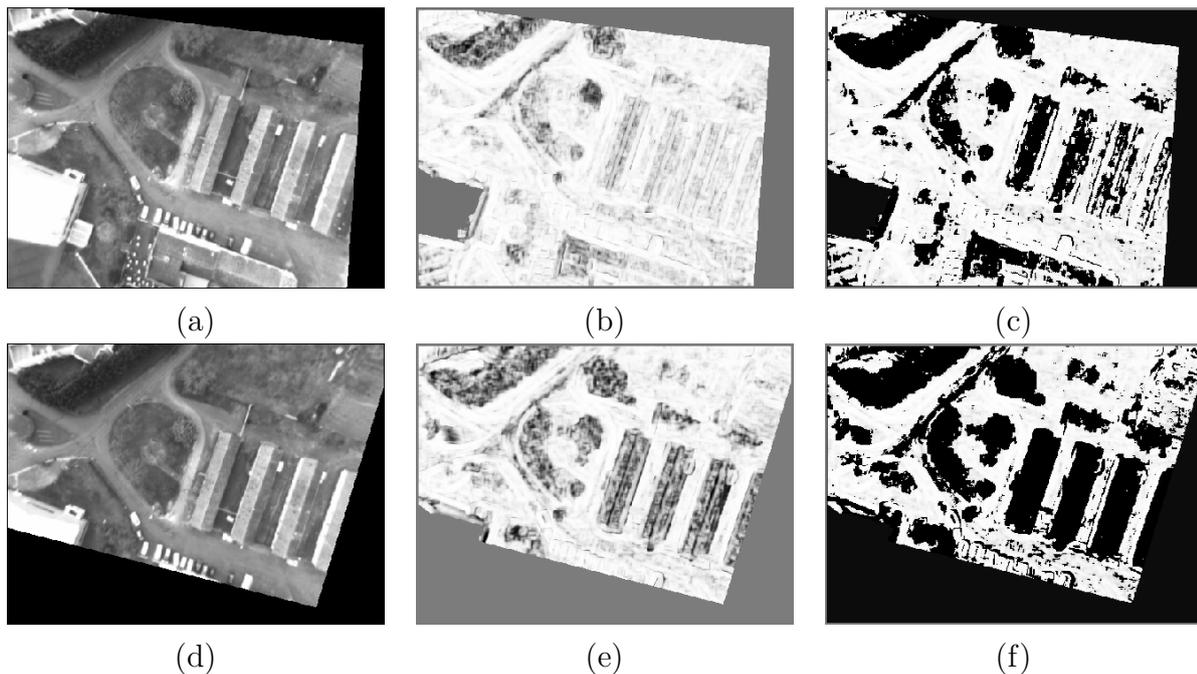


FIG. 3.25: *Influence de la baseline sur la segmentation de zones planes. (a) et (d) images de référence, (b) et (e) images des scores de corrélations non corrigés, (c) et (f) images des scores seuillées automatiquement par notre algorithme. La première rangée d'image illustre un recouvrement de 91%, la seconde un recouvrement de 78%*

La formule (3.4) peut nous permettre d'estimer ϵ_O et ainsi de quantifier la distance minimale $l_{O_{min}}$ des *outliers* relativement à Π_I telle que ces outliers soit segmentés. Même si

dans ce cas précis cette formule n'est pas réellement adaptée (la présence d'une rotation non nulle entre les images invalidant les hypothèses de ce résultat) il est possible d'en utiliser les résultats de façon qualitative.

En effet, notre expérience des résultats expérimentaux nous a montré que dans un tel cas où la rotation est dans le même plan de la translation, l'estimée de ϵ_O est du même ordre de grandeur que dans la réalité. Ces calculs ont été menés mais ne sont pas détaillés en raison de leur lourdeur. La rotation influence la valeur du résidu en fonction de la position des points dans l'image, mais ceux ci se répartissent finalement autour d'une valeur relativement proche de la valeur théorique de ϵ_O .

Pour des outliers à $5m$ de Π_I , c'est à dire pour la série des quatre hangars alignés, nous obtenons donc des valeurs théoriques de 2.6 pixels pour la première paire d'image et de 6.3 pixels pour la seconde. A $8m$ c'est à dire pour la haie en haut à gauche, on obtient alors respectivement 4.3 et 10.6 pixels. Ces valeurs bien qu'à prendre en compte de manière qualitative reflètent néanmoins assez fidèlement les performances de notre algorithme et confirment notre hypothèse de détectabilité : une parallaxe d'au moins 4 pixels est nécessaire pour assurer la détection des zones non planes.

3.5.4 Comparaison entre les deux méthodes de correction

Comme les résultats de la figure 3.26 l'attestent, la méthode de correction hors-ligne tend à produire des résultats de qualité très supérieurs à ceux de la méthode en-ligne. On peut tout de même remarquer une légère sous-segmentation du toit directement imputable à la modélisation de $S_M(\sigma^2)$ et à la difficulté de selection des $\{S_{M_l}, \sigma_l\}$.

Cette méthode reste tout de même fortement conseillée lorsqu'il est possible de recourir à l'expertise d'un opérateur humain, la sélection des $\{S_{M_l}, \sigma_l\}$ s'avérant délicate.

3.5.5 Problèmes en suspens

Les résultats obtenus par notre algorithme sont satisfaisants dans les limites de détectabilité que nous avons précisées. Néanmoins, il existe un phénomène récurrent de sursegmentation des zones non planes. Le positionnement de ces mauvaises segmentations étant aléatoire, une solution naturelle à ce problème serait de prendre en compte plusieurs paires d'images et de fusionner les résultats ainsi obtenus dans le but d'améliorer la segmentation mais aussi de la rendre simplement plus fiable.

Cette solution est développée et mise en oeuvre par une approche probabiliste dans le chapitre suivant.

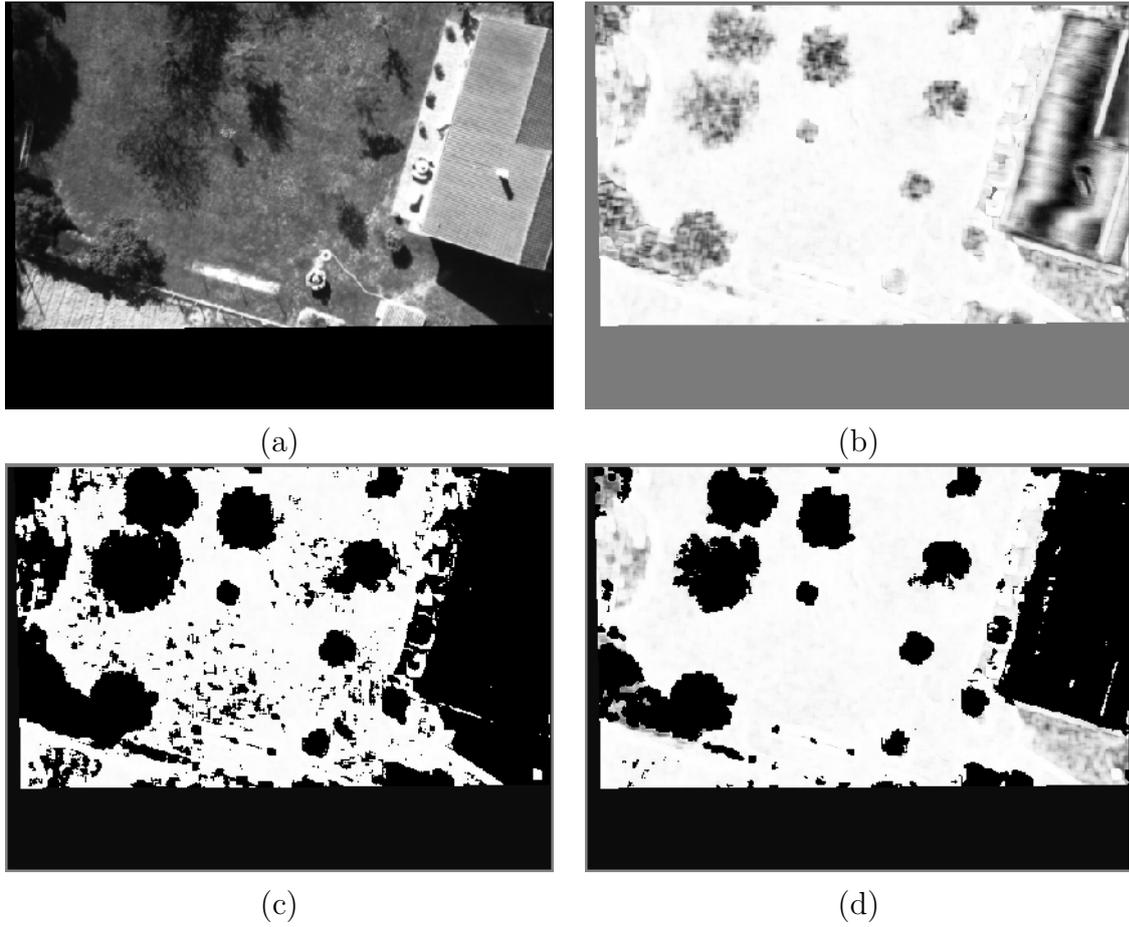


FIG. 3.26: Comparaison entre correction des scores de corrélation en ligne et hors ligne. (a) est l'image de référence, (b) l'image des scores de corrélations non corrigés, (c) l'image des scores seuillées par la correction en-ligne et (d) l'image des scores seuillées par la correction hors-ligne.

Chapitre 4

Mise à jour d'un modèle probabiliste de traversabilité

4.1 Introduction

Comme nous venons de le présenter il est possible de segmenter des images de façon relativement fiable afin de mettre en évidence les zones planes de l'environnement ainsi que ses obstacles. Toutefois cette méthode est limitée car elle n'opère que sur une paire d'images et ne permet donc pas de statuer sur la persistance d'un plan perçu dans deux paires d'images successives ou encore de prendre en compte le maximum d'observations d'une même zone. Afin de pallier ces défauts nous avons introduit un modèle probabiliste de traversabilité mis à jour à chaque itération de la segmentation de zones planes dans une nouvelle paire d'images. Ce modèle nous permet d'assurer un suivi des zones planes sur une séquence d'images et d'enrichir une grille de traversabilité locale et globale exploitable par différentes autres fonctionnalités (sélection de site d'atterrissage ou de zones traversables pour un robot terrestre par exemple). La figure 4.6 illustre ce principe. On peut y voir deux types de grilles :

- des grilles locales : G_L . Ces grilles quantifient des probabilités de planéité de l'ordre du pixel, chacune correspondant à un seul plan détecté. Elles sont construites dans le repère de la première image d'une séquence où ce plan est détecté. La mise à jour de cette grille est décrite au § 4.2.
- une grille globale : G_G . Cette grille est la projection des grilles locales sur le plan du sol dans des cellules de résolution donnée (de l'ordre du mètre par exemple). Sa mise à jour est présentée au § 4.3.

4.2 Fusion des informations

4.2.1 Construction de la grille locale G_L

L'algorithme de segmentation de zones planes fournit une observation après seuillage des scores de corrélation $S(u, v)$, que l'on peut considérer comme étant une observation binaire $O(u, v)$ sur la probabilité de planéité de chaque pixel $\mathbf{m} = (u, v, 1)^T$. Nous construisons donc une grille G_L dont la valeur de chaque cellule $\mathbf{c}(u, v)$ correspond à la probabilité que le pixel \mathbf{m} soit la projection d'un point du plan Π_I .

La segmentation étant réalisée sur une série d'images, il est alors possible de mettre à jour ces probabilités lors de nouvelles observations. La grille G_L est positionnée dans le repère de la première image $I_{k=0}$ et les observations de la k -ième paire d'images sont projetées dans le repère de G_L (équivalent à celui de $I_{k=0}$). Pour cela, nous combinons les homographies H_I estimées entre l'instant 0 et k . Ces observations reprojétées $O_k(u, v)$ sont alors utilisées pour mettre à jour la probabilité contenue dans $\mathbf{c}(u, v)$.

4.2.2 Cas d'observations binaires

Dans ce qui suit nous allons considérer que nous nous intéressons à une même cellule $\mathbf{c}(u, v)$ notée \mathbf{c} afin d'alléger les notations.

Nous notons c le fait que cette cellule soit plane et \bar{c} qu'elle ne le soit pas. S_k sont les scores de corrélations obtenus à l'instant k pour cette cellule. Les observations binaires $O_k \subset \{S_k \geq \tau_l, S_k < \tau_l\}$ y sont associées. Nous pouvons définir la probabilité à posteriori qu'une cellule \mathbf{c} de la grille G_L soit plane compte tenu d'une observation courante $O_k = S_k \geq \tau_l$ à l'instant k comme $p(c|O_k)$. La formule de Bayes nous permet d'écrire :

$$p(c|O_k) = \frac{p(O_k|c)p(c)}{p(O_k)} = \frac{p(O_k|c)p(c)}{p(O_k|c)p(c) + p(O_k|\bar{c})(1-p(c))}$$

En introduisant la probabilité d'erreur de première espèce dite de *fausse alarme* ou de *faux positif* (la probabilité d'observer une cellule comme étant plane alors qu'elle ne l'est pas) $P_f = p(S_k \geq \tau_l|\bar{c})$ ainsi que la probabilité de bonne détection $P_d = p(S_k \geq \tau_l|c)$ nous pouvons alors écrire :

$$p(c|S_k \geq \tau_l) = \frac{P_d p(c)}{P_d p(c) + P_f (1 - p(c))} \quad (4.1)$$

Les observations O_k étant indépendantes, on peut appliquer la formule précédente itérativement en considérant que $p(c) = p(c|O_{k-1})$.

Nous pouvons de même calculer la probabilité que la cellule soit plane \mathbf{c} sachant que l'observation $O_k = S_k < \tau_l$ indique que non comme :

$$p(c|S_k < \tau_l) = \frac{(1 - P_d) p(c)}{(1 - P_d) p(c) + (1 - P_f)(1 - p(c))} \quad (4.2)$$

Nous mettons donc à jour incrémentalement les probabilités $p(c|O_k)$ selon la dernière observation O_k par la simple application des formules (4.1) et (4.2), selon que la nouvelle observation confirme la planéité de \mathbf{c} ou non. Les probabilités $p(c)$ pour $k = 0$ sont initialisées à $\frac{1}{2}$ traduisant ainsi l'absence d'information à priori.

Les valeurs de P_d et P_f utilisées dans nos expérimentations ont été obtenues empiriquement. N'étant pas corrélées avec des grandeurs telles que le résidu ou un quelconque descripteur de parallaxe et étant très similaires quelque soit la série d'images, nous avons pu les fixer à $P_d = 0.8$ et $P_f = 0.12$. Pour cela, nous avons évalué le nombre de fausses alarmes et de bonnes détections sur diverses sorties de segmentation telles les images du § 3.5.

Résultats

La fusion des informations issues de deux segmentations successives est illustrée par la figure 4.1. Nous y présentons les images utilisées, les résultats de notre algorithme de segmentation appliquée aux trois paires d'images successives, et enfin les mises à jour de la grille locale G_L . La représentation de ces grilles doit être lue comme suit :

- Les cellules vertes illustrent des probabilités $p(c) > \frac{1}{2}$ et donc considérées comme associées au plan $\Pi_{\mathbf{I}}$
- Les cellules rouges sont associées à des probabilités $p(c)$ inférieure à $\frac{1}{2}$ et donc n'appartiennent pas au plan $\Pi_{\mathbf{I}}$ considéré
- Plus une cellule est sombre, plus sa probabilité tend vers $\frac{1}{2}$ et donc plus la question de son appartenance au plan $\Pi_{\mathbf{I}}$ est incertaine. A l'inverse plus une cellule est claire, plus sa probabilité tend vers 1 ou 0 selon sa couleur.

L'image (h) est donc la première mise à jour de la grille probabiliste G_L , soit la fusion entre l'absence d'information initiale pour chaque cellule et l'observation de planarité de la segmentation. Les probabilités résultantes découlent donc directement de la segmentation. Comme l'indique la couleur sombre des cellules, une seule observation ayant été réalisé par cellule les probabilités sont encore relativement proche de $\frac{1}{2}$.

La seconde mise à jour de G_L est obtenue après recalage par l'homographie inter-images $H_{3,1} = H_{3,2} H_{2,1}$ de la segmentation (e) dans le repère de G_L . Nous constatons comme attendu que les cellules associées à deux segmentations successives concordantes voient leur probabilité se renforcer en tendant vers 1 ou 0. A l'inverse, les fausses détections dont la position peut être considérée comme aléatoire, ne sont pas confirmées et les probabilités des cellules associées de la grille se rapprochent ainsi de $\frac{1}{2}$.

Enfin, la dernière mise à jour de G_L (j) accentue ces comportements pour les cellules ayant été mis à jour trois fois. Ainsi, les figures (k),(l) et (m) représentant un agrandissement de certaines de ces cellules permettent d'apprécier l'évolution positive de la fusion des observations.

La figure 4.2 permet d'apprécier l'évolution de la grille après une quinzaine de fusions successives.

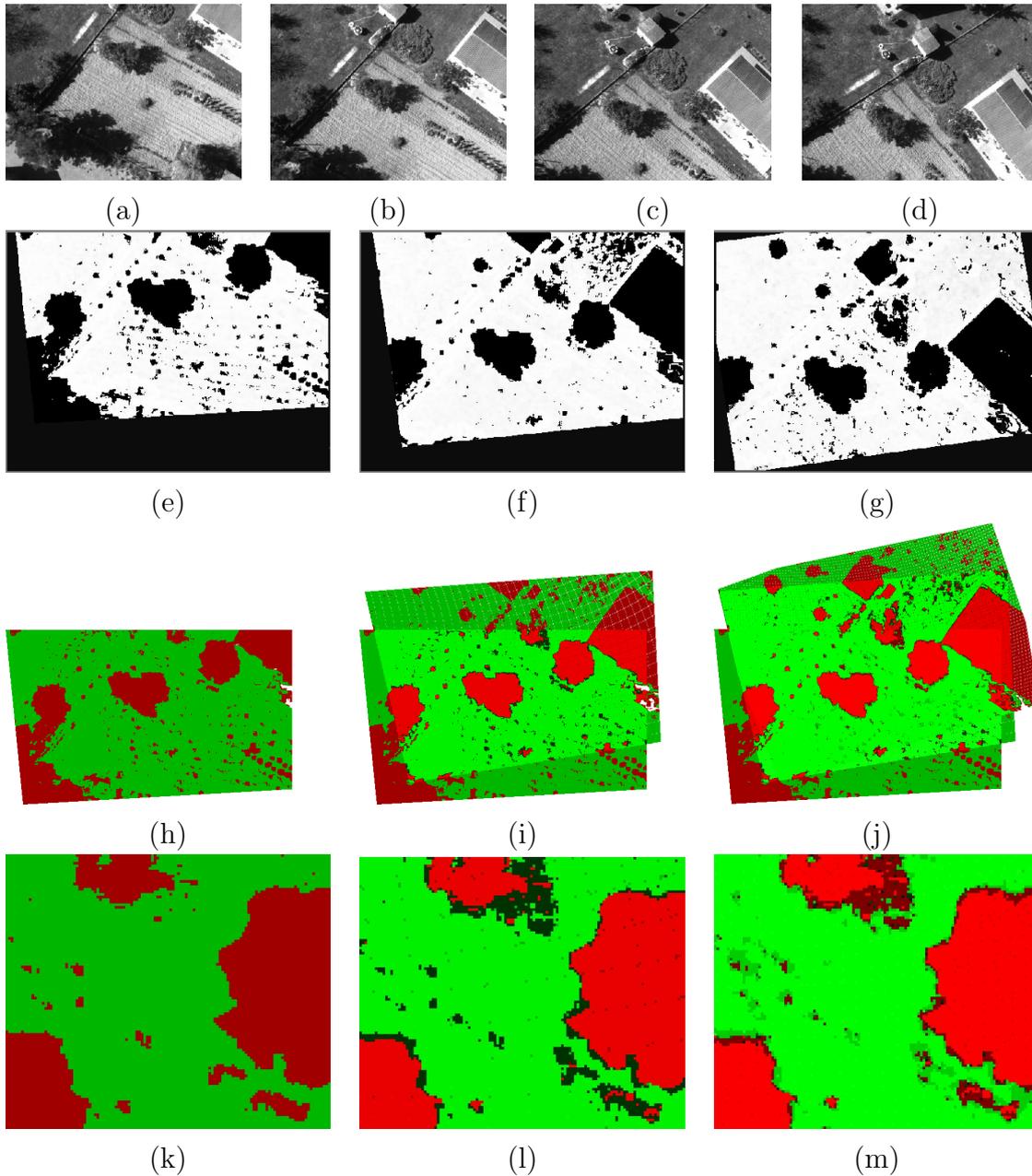


FIG. 4.1: *Illustration de fusions successives d'observations binaires dans une grille locale G_L . (a), (b), (c) et (d) sont les images utilisées, (e), (f) et (g) les segmentations de zones planes réalisées à partir des paires correspondantes, (h), (i) et (j) les représentations de G_L après la fusion des obserbations binaires issues de ces segmentations. Les probabilités $p(c|O_k)$ associée à chacune des cellules y sont représentées. La couleur rouge clair indique une probabilité de 0, le vert clair de 1 et le rouge ou vert sombre une probabilité qui tends vers $\frac{1}{2}$. Les images (k), (l) et (m) sont des agrandissements de cellules de G_L qui ont été mises à jour trois fois.*

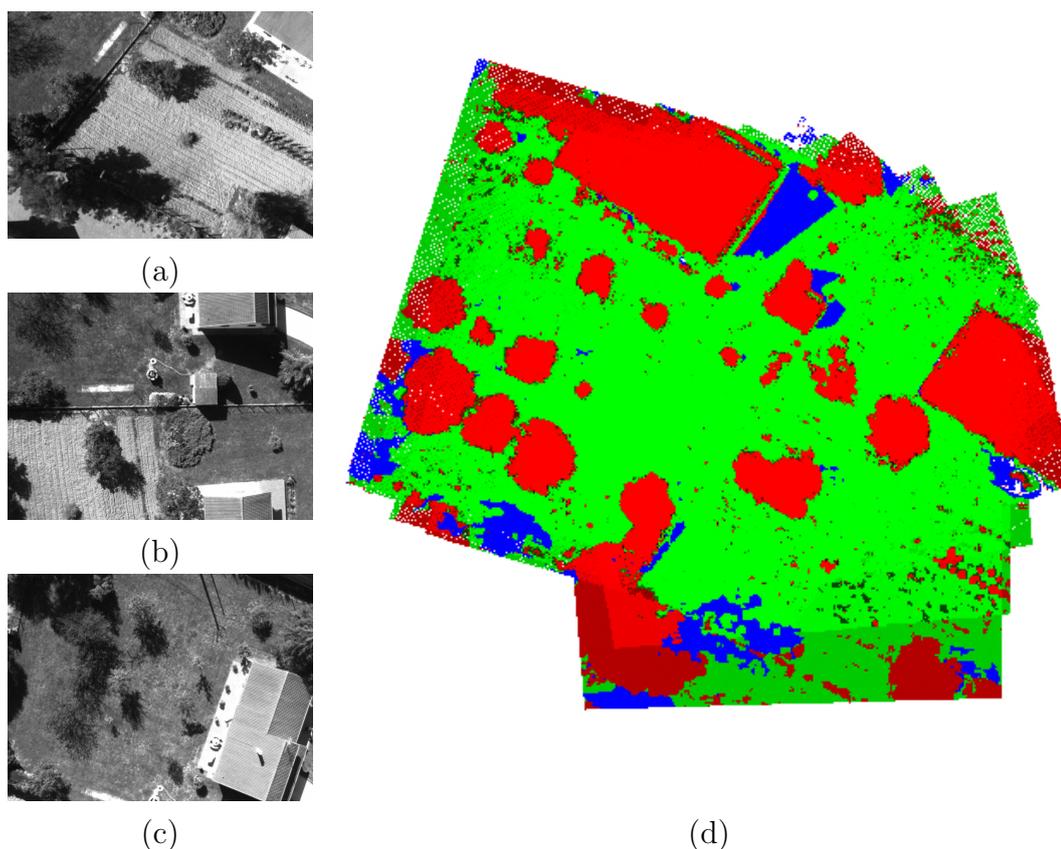


FIG. 4.2: *Illustration d'une grille locale construite à partir d'observations binaires. (a)(b)(c) sont les images respectives du début, milieu et fin de la séquence utilisée. (d) représente linéairement les probabilités finales $p(c|O_k)$ où le rouge clair indique une probabilité de 0, le vert clair de 1 et le rouge ou vert sombre une probabilité qui tends vers $\frac{1}{2}$. Le bleu indique des régions où la variance des images originales est trop faible pour en tirer des conclusions fiables.*

4.2.3 Cas d'observations continues

Au lieu d'associer la probabilité à une observation binaire O_k dépendante du seuillage, il est aussi possible de l'associer directement au score de corrélation S_k en prenant en compte la classe de variance \mathcal{C}_l à laquelle elle appartient. En effet, nous pouvons écrire :

$$p(c|S_k) = \frac{p(S_k|c)p(c)}{p(S_k)} = \frac{p(S_k|c)p(c)}{p(S_k|c)p(c) + p(S_k|\bar{c})(1-p(c))}$$

La probabilité $p(S_k|c)$ est ici une loi continue que l'on peut exprimer en fonction des classes de variance \mathcal{C}_l . En effet, comme nous l'avons vu au § 3.4.3, il est possible de modéliser les densités de probabilités d'un score S_l associé à une classe \mathcal{C}_l par l'équation (3.6). On obtient alors :

$$p(S_k|c) = e^{-e^{(S_k-\mu_l)/\beta_l}} * e^{(S_k-\mu_l)/\beta_l} / \beta_l$$

Nous avons quantifié expérimentalement la probabilité $p(S_k|\bar{c})$ en isolant les scores de corrélation associés à des zones non planes par la segmentation avec les seuils τ_l . Ceci nous a permis d'obtenir les résultats présentés figure 4.3. Comme on peut le voir cette probabilité présente un pic pour des valeurs de corrélation très forte ce qui est tout à fait anormal. Nous avons émis les hypothèses suivantes concernant ce pic :

- D'une part, notre segmentation s'affranchit de certains problèmes de répétition de texture qui occasionnent des scores de corrélation élevés pour des zones non planes. Ces scores participent donc au pic constaté.
- D'autre part, les classes de variance élevées ont des densités de probabilité étroites qui rendent leur estimation, conduisant au seuil τ_l , plus sensible. Les scores segmentés à tort comme associés à des pixels non plans contribuent donc eux aussi à ce pic.

Ces deux observations ne remettent pas en cause la qualité de la segmentation, puisque dans un cas on a effectivement une bonne segmentation et dans l'autre une sur-segmentation qui est donc conservatrice. En revanche, il n'est donc pas possible de modéliser $p(S_k|\bar{c})$ par ce moyen.

Nous avons donc finalement opté pour l'hypothèse de l'équiprobabilité de S entre -1 et 1 . Ceci nous conduit à poser que la probabilité $p(S_k|\bar{c})$ est égale à $\frac{1}{2}$ pour tout score S .

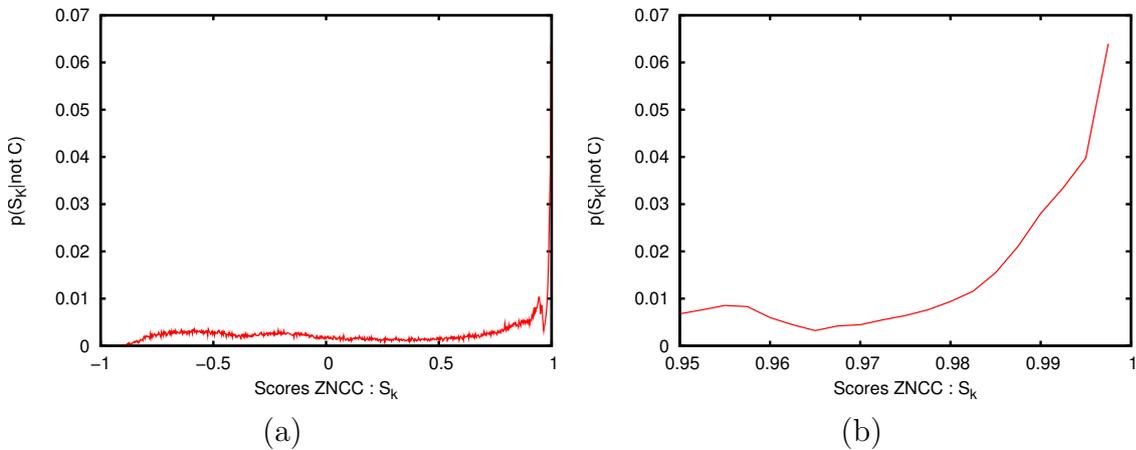


FIG. 4.3: Représentation de la densité de probabilité $p(S_k|\bar{c})$ déterminée expérimentalement à partir des résultats de la segmentation. La figure (a) illustre cette probabilité sur l'ensemble de l'intervalle de définition de S , et la figure (b) se concentre sur les scores les plus élevés.

Résultats

A l'exception de l'utilisation directe des scores de corrélation plutôt que des sorties de segmentation, cette méthode de fusion donne des résultats très similaires à celle illustrée par la figure 4.1. On observe les mêmes comportements de correction des probabilités. La

figure 4.5 en comparaison de la figure 4.2 permet en revanche d'apprécier que l'évolution des probabilités est beaucoup plus rapide. Ce phénomène est aussi très clair sur la figure 4.4. On peut ainsi y voir que la segmentation des zones non planes est très forte, voir excessive comme on peut le constater sur le fil électrique. Ce résultat est dû à la faiblesse de $p(S_k|\bar{c})$ en regard de $p(S_k|c)$ conduisant à une évolution très rapide des probabilités. On peut donc s'interroger sur la pertinence de la modélisation proposée pour $p(S_k|\bar{c})$. Quoiqu'il en soit, la fusion par observations binaires produit des résultats très satisfaisants que nous avons retenu dans la suite de nos travaux.

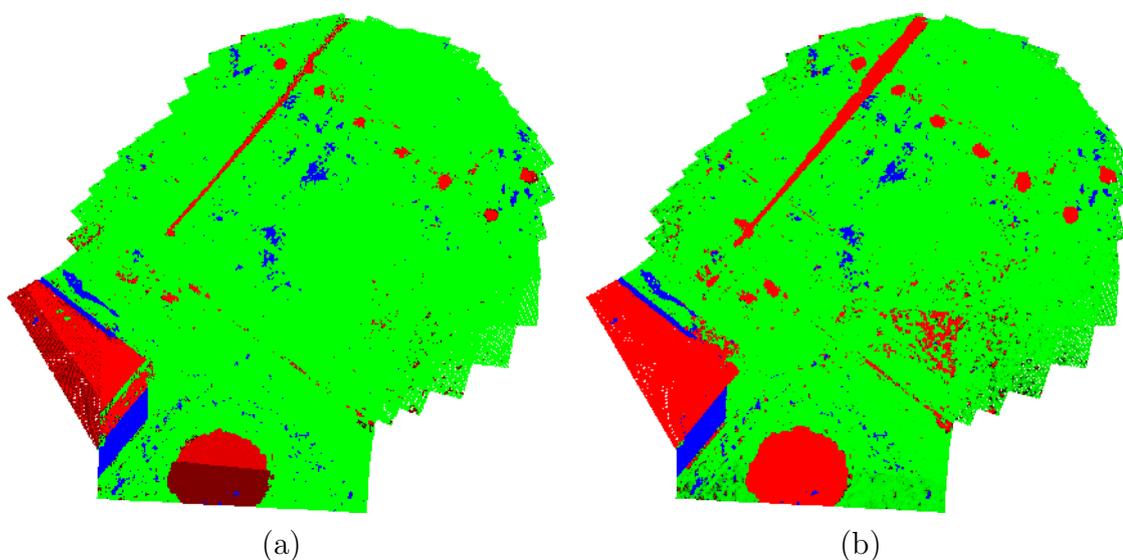


FIG. 4.4: *Comparaison des méthodes de fusion binaire et continue. La figure (a) illustre la fusion d'observations binaires et la figure (b) la fusion d'observations continues. Le trait rouge au centre haut des images s'avère être un fil électrique.*

4.3 Géoréférencement

Des grilles locales G_L contenant les probabilités de planéité des environnements survolés sont donc estimables et apportent une réelle amélioration par rapport aux estimées de base obtenues initialement par la segmentation. Toutefois pour en permettre l'exploitation par un robot terrestre il est nécessaire de projeter ces observations du référentiel capteur auxquelles elles sont liées à un référentiel commun. Pour cela, il est possible de reprojeter chacune de ces grilles locales sur une grille globale G_G , horizontale, géoréférencée et positionnée à une altitude fixée équivalente à celle du sol.

Pour assurer cette reprojektion il est nécessaire d'utiliser les données de nos capteurs comme détaillé dans le § 4.3.1. Toutefois, les reprojektions de ces différentes grilles locales G_L souffrent des incertitudes imputables aux capteurs ce qui, dans le cas de recouvrement,

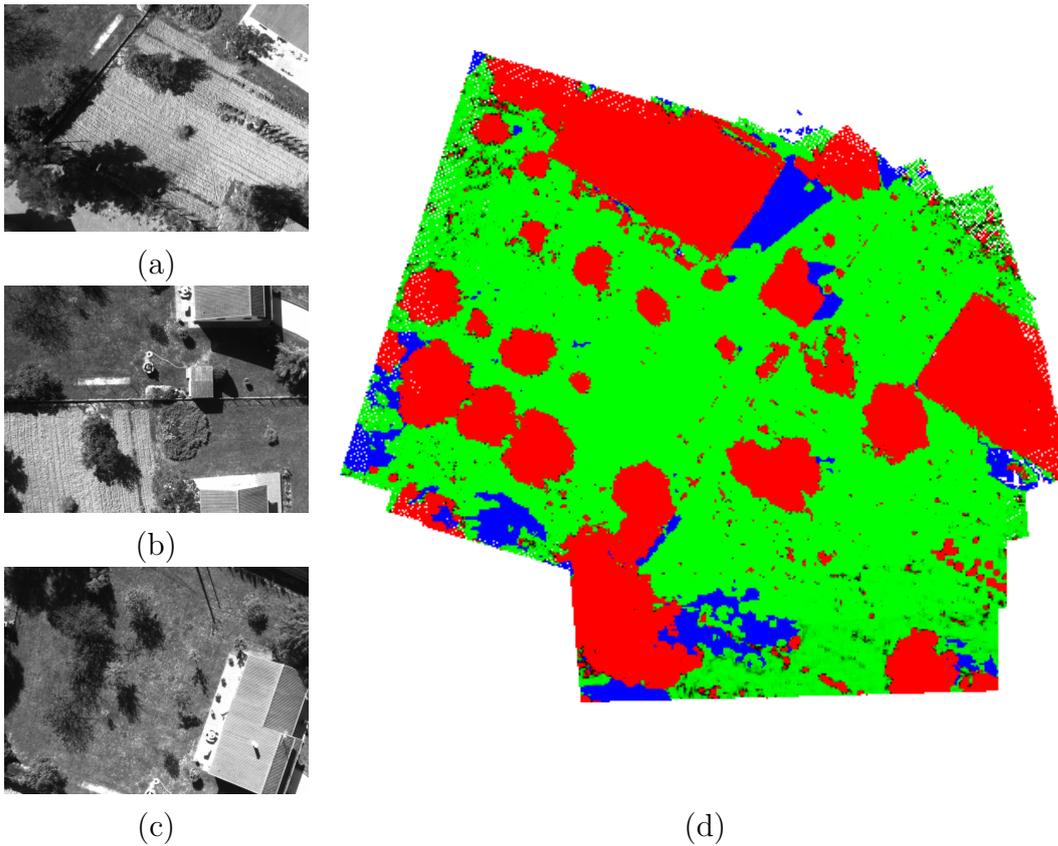


FIG. 4.5: *Illustration d'une grille locale construite à partir d'observations continues. (a)(b)(c) sont les images respectives du début, milieu et fin de la séquence utilisée. (d) représente les probabilités finales $p(c|O_k)$.*

s'avère très dommageable. Nous proposons donc au § 4.3.2 une autre méthode s'appuyant une fois encore sur l'estimation d'homographie.

4.3.1 Fusion utilisant les capteurs

La géométrie d'une scène telle que la projection d'une grille locale G_L sur une grille globale G_G est illustrée sur la figure 4.6. Cette projection peut s'exprimer comme suit.

Si on note $\mathbf{M}_{G_L|R_{C_O}}$ un point de la grille locale G_L exprimé dans le repère R_{C_O} on peut alors écrire :

$$\mathbf{M}_{G_L|R_{C_O}} = K \mathbf{m}_{G_L} : \quad \text{c'est à dire :} \quad \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{C}^{-1} \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$$

Afin de projeter ce point sur le plan du sol, équivalent au plan (Oxy) du repère R_{G_G} , il est nécessaire d'obtenir les coordonnées de \mathbf{M}_{G_L} dans le repère R_{C_O,G_G} centré en C_O

mais orienté pareillement à R_{G_G} . Cette transformation s'écrit :

$$\mathbf{M}_{G_L|R_{C_0,G_G}} = \begin{bmatrix} \mathbf{R} & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \mathbf{K} \mathbf{m}_{G_L}$$

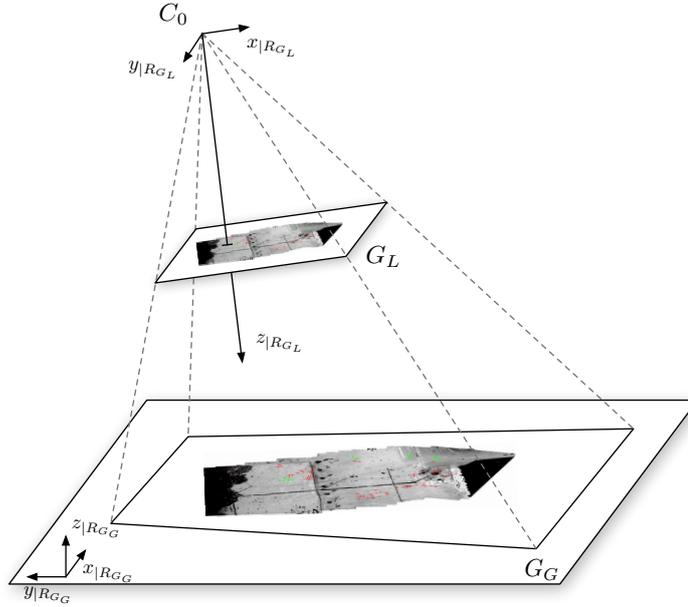


FIG. 4.6: Géométrie de la projection de la grille locale G_L sur la grille globale G_G .

Pour projeter ce point sur le plan (Oxy) , il suffit alors d'effectuer la projection centrale de $\mathbf{M}_{G_L|R_{C_0,G_G}}$ selon l'axe z de R_{C_0,G_G} sur le plan défini par $z_{|R_{C_0,G_G}} = -z_{C_0|R_{G_G}}$ (soit $z_{|R_{G_G}} = 0$). Cette projection peut être modélisée de la manière suivante :

$$s\mathbf{M}_{G_G|R_{C_0,G_G}} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3 \times 1} \\ 0 & 0 & \frac{1}{-z_{C_0|R_{G_G}}} & 0 \end{bmatrix} \mathbf{m}_{G_G|R_{C_0,G_G}}$$

Il suffit alors d'appliquer la translation $t = C_0|R_{G_G}$ à $\mathbf{M}_{G_G|R_{C_0,G_G}}$ pour obtenir les coordonnées de \mathbf{M}_{G_G} , point de la grille globale G_G exprimé dans le repère R_{G_G} . On peut donc écrire la relation entre \mathbf{m}_{G_L} et $\mathbf{M}_{G_G|R_{G_G}}$ sous la forme :

$$s\mathbf{M}_{G_G} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{t} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3 \times 1} \\ 0 & 0 & \frac{1}{-z_{C_0|R_{G_G}}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{C}^{-1} \\ 0 & 0 & 1 \end{bmatrix} \mathbf{m}_{G_L}$$

La figure 4.7 présente ainsi des résultats de projection d'une grille G_L vers une grille globale G_G ayant une résolution au sol de 25 cm et étant orienté dans un repère ENU (East North Up). La grille locale G_L a été déterminée à partir des 4 images représentées. La projection utilise donc les informations de position relatives à l'instant d'acquisition de la première vue (a).

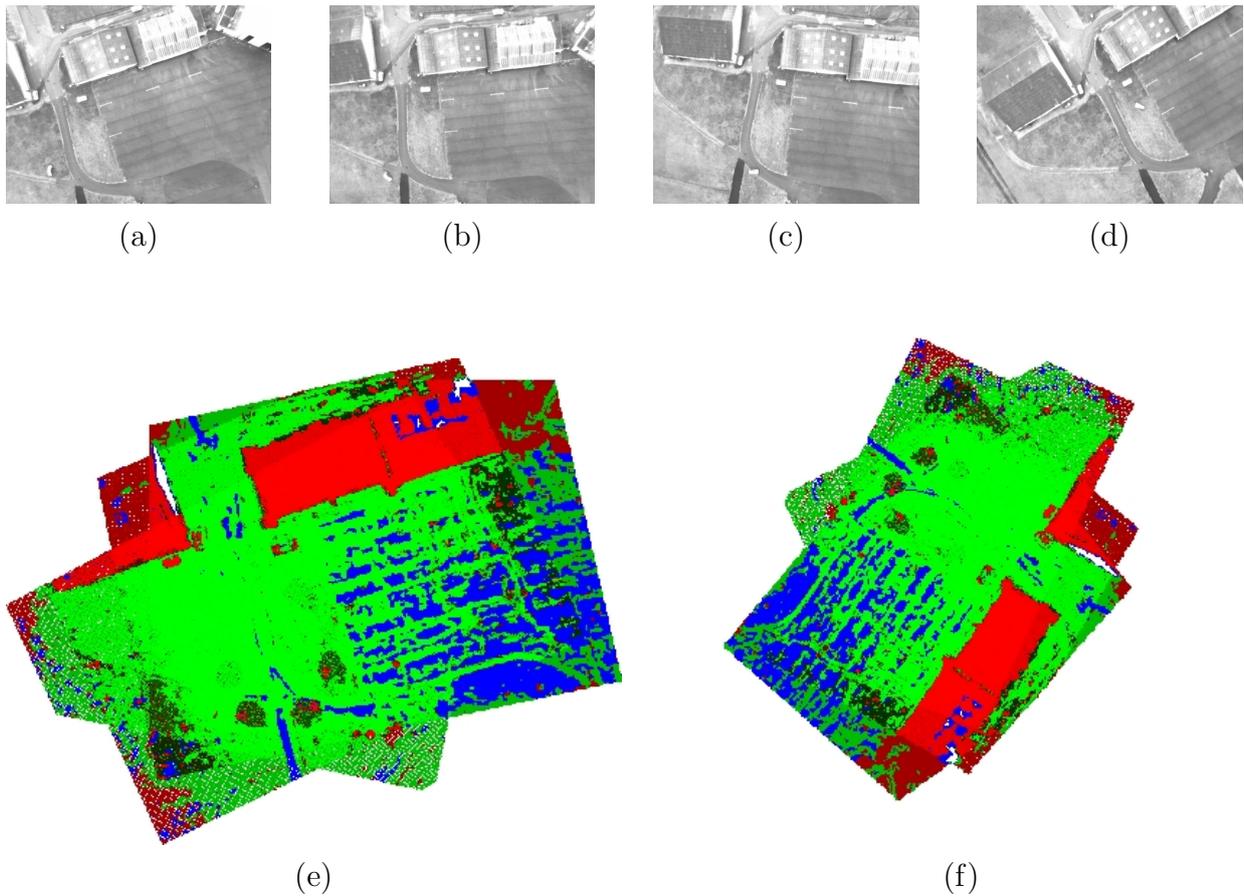


FIG. 4.7: *Illustration de la projection d'une grille locale G_L vers une grille globale G_G . Les images utilisées pour la construction de G_L sont représentées par les figures (a), (b), (c) et (d). La grille locale est la figure (e) et la grille globale la figure (f).*

La figure 4.8 présente un exemple de problème de mise à jour de cette grille globale. Afin de mettre ceci en évidence, nous avons procédé à la construction de deux grilles locales G_{L_1} et G_{L_2} déterminées respectivement à partir des vues (a) et (b) puis (c) et (d). Ces deux grilles locales G_{L_1} et G_{L_2} sont illustrées par les figures (e) et (f) et leur projections respectives dans la grille globale G_G sont (i) et (j). Comme on peut le voir, ces projections sont loin de se recouvrir. C'est un comportement qui est dû à la mauvaise précision de notre capteur inertielle qui estime les angles d'attitude lors d'une phase de vol où notre dirigeable est fortement ballotté par le vent. Si dans ce cas précis notre centrale inertielle est en cause, il s'avère aussi que notre GPS est souvent trop imprécis pour nous fournir une estimée fiable du paramètre $z_{C_0|R_{G_G}}$ de la projection. Lors de survol d'une même zone, nous avons déjà ainsi pu constater une différence de plus de 40% de cette estimée d'altitude.

L'imprécision trop grande nos capteurs nous conduit donc à proposer une autre approche s'appuyant sur l'estimation d'homographies.

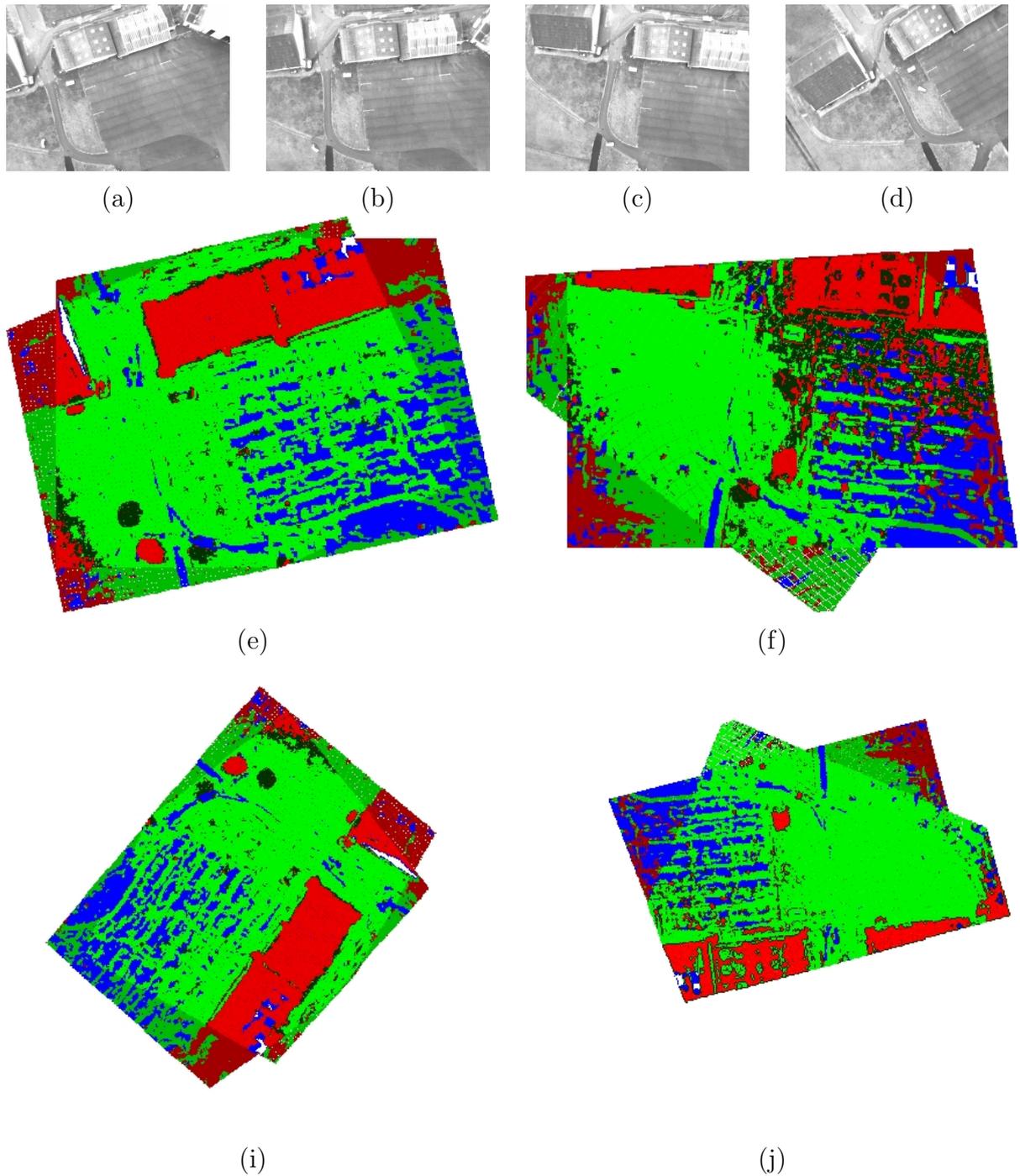


FIG. 4.8: Illustration des problèmes de mises à jour de la grille globale G_G . Les grilles globales G_{L_1} (e) et G_{L_2} (f) sont respectivement construites à partir des images (a) et (b) puis (c) et (d). Leurs projections respectives basées sur les estimations d'attitude de notre central inertiel sont illustrées sur les figures (i) et (j). On voit qu'une rotation en cap serait ici nécessaire pour assurer la cohérence de la grille globale.

4.3.2 Fusion par homographie

En l'absence de capteurs suffisamment précis, une solution naturelle au problème de la fusion de différentes grilles locales est l'utilisation des mosaïques correspondantes aux grilles locales. La construction de telles mosaïques consiste à employer les homographies inter-images calculées et les appliquer sur les images pour obtenir une image augmentée dans le repère de la grille locale correspondante. Ces mosaïques contenant donc l'information spatiale des images, nous pouvons les mettre en correspondance afin de déterminer leur positions respectives et fusionner les grilles locales.

Construction de mosaïque

Les mosaïques ainsi construites correspondent à une projection centrale sur un capteur de même position et résolution que pour la première image mais de dimension supérieure.

Les caméras que nous utilisons étant réglées en balance des blancs automatique, il est nécessaire de corriger la luminosité des pixels avant de les intégrer au sein de la mosaïque. En notant I_k et I_{k+1} la luminosité des images originales aux instants k et $k + 1$ on peut calculer M_k la luminosité corrigée de l'image comme étant :

$$M_k = I_k + \tilde{I}_{k-1} - \tilde{I}_k$$

où \tilde{I} est la moyenne de la luminosité des pixels segmentés comme étant plans. Cette connaissance de la structure de l'environnement nous permet de fiabiliser sensiblement cette correction d'intensité en ne prenant en compte que les pixels observés dans les deux images.

En ne construisant les mosaïques qu'avec les zones segmentées, comme appartenant au plan du sol, nous obtenons des mosaïques éparées correspondant à la zone plane de l'orthoimage qui serait construite pour l'environnement survolé.

L'avantage majeur de cette technique est que les mosaïques ainsi construites ont un réel sens physique. En effet, comme illustré par la figure 4.9 en appliquant le recalage homographique sur les zones non planes, celles-ci apparaissent "floutées". Nos mosaïques offrent donc un résultat plus éparé mais aussi plus exact. Il est à noter que les zones de faible variance, dont nous ne pouvons juger la planéité, sont pour la construction de la mosaïque considérée comme étant plane. En effet, du fait de leur faible variance, cela n'a pas d'effet sur la qualité radiométrique de la mosaïque.

Fusion

Une fois les mosaïques construites, il suffit donc d'estimer l'homographie entre celles-ci pour les mettre à jour. Ces mosaïques, étant par construction, constituées de zones planes, l'estimation en est très aisée. Le problème principal est amont, lors de la mise en correspondance des points d'intérêts : un facteur d'échelle ou un changement de point de vue trop importants peuvent compliquer sensiblement cette étape.

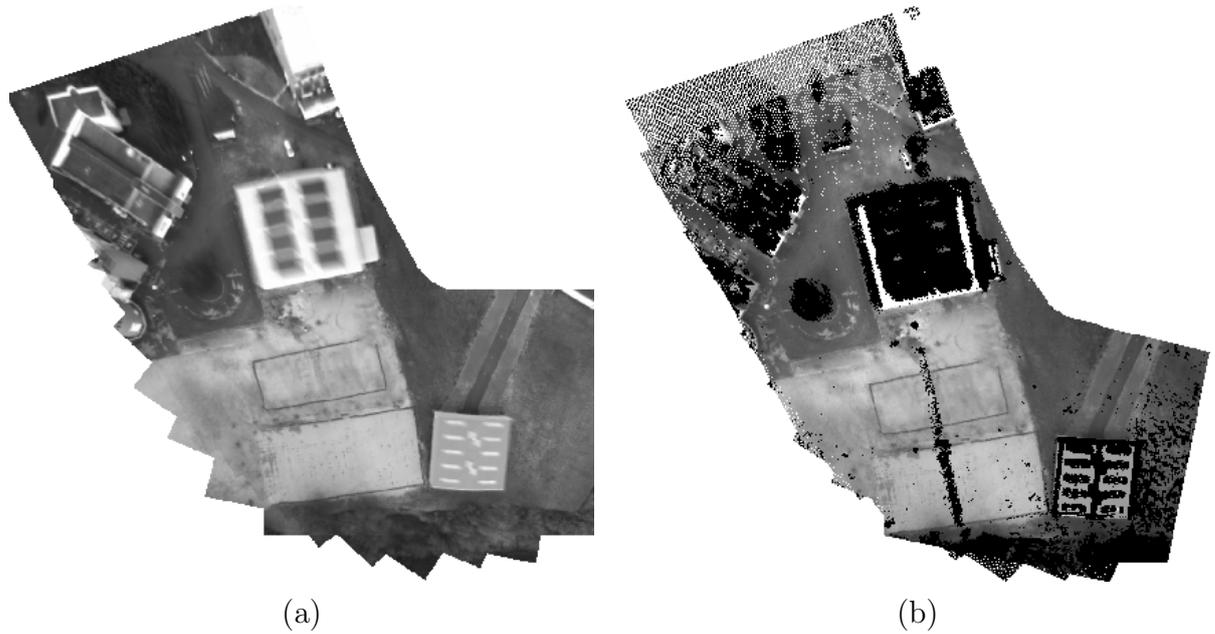


FIG. 4.9: *Comparaison des mosaïques avec (a) et sans (b) zones planes. Le toit du bâtiment blanc est distinctement “floutté” dans l’image (a).*

Les figures 4.10 et 4.11 illustrent une fusion de deux G_L . Comme on peut le constater, il n’y a quasiment aucun outlier. La fusion présentée sur la figure (d) étant une simple moyenne de pixels, il est aisé de juger de la qualité du recalage avant fusion. On peut ainsi constater que ce recalage est bien meilleur que celui effectué en utilisant les capteurs de positionnement

On notera, de plus que si des points ont pu être mis en correspondance entre les deux G_L (plus de 300 dans cet exemple) et une homographie a pu être estimée, ceci signifie que les mosaïques correspondant à chaque G_L sont de bonne précision.

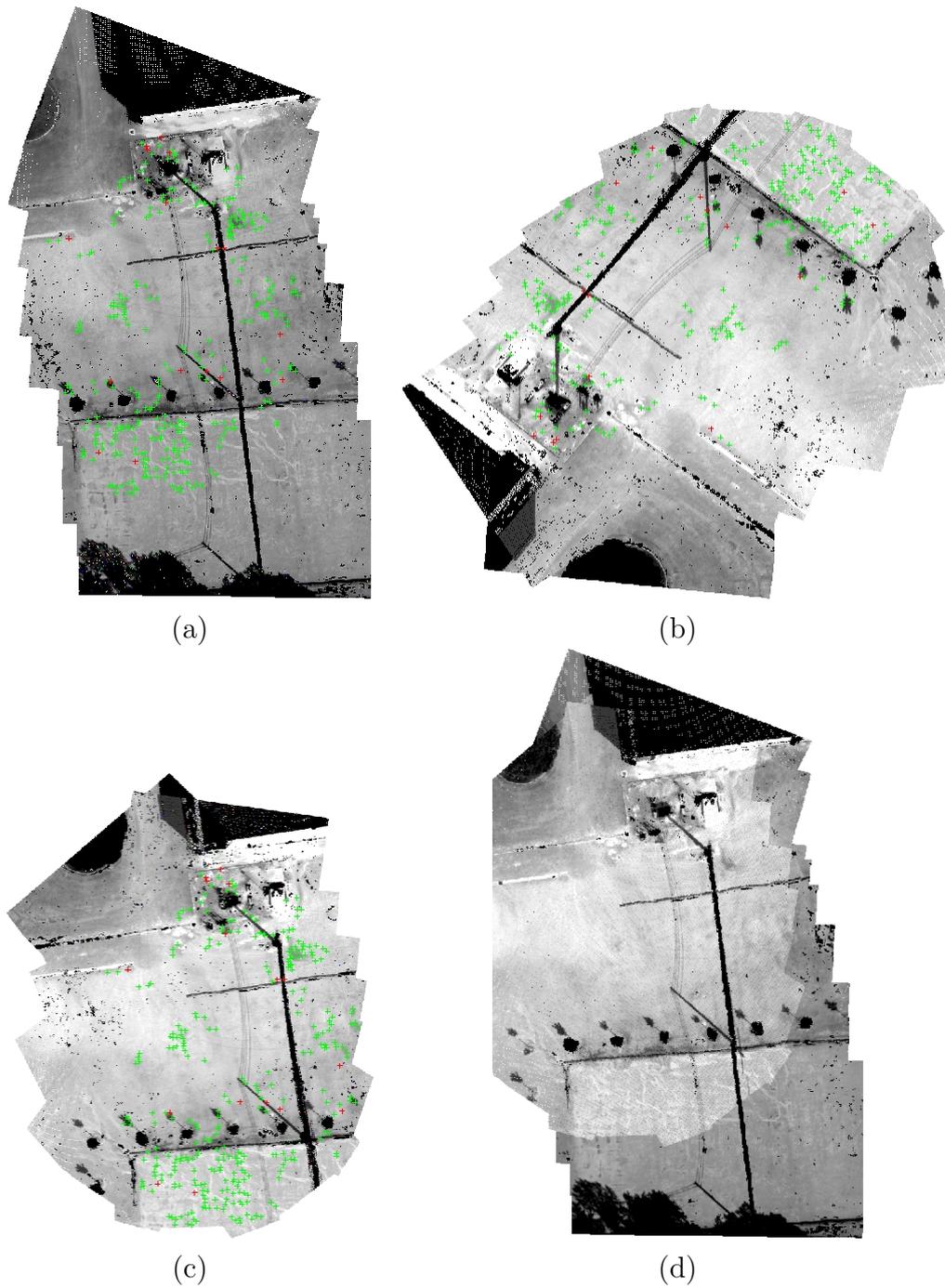


FIG. 4.10: Illustration de la fusion par homographie de grilles locales G_L . Les images (a) et (b) sont les mosaïques associées à deux grilles locales distinctes. L'image (c) est la mosaïque (b) recalée dans le repère de (a). L'image (d) est la mosaïque associée à la grille locale fusion des deux grilles de départ.

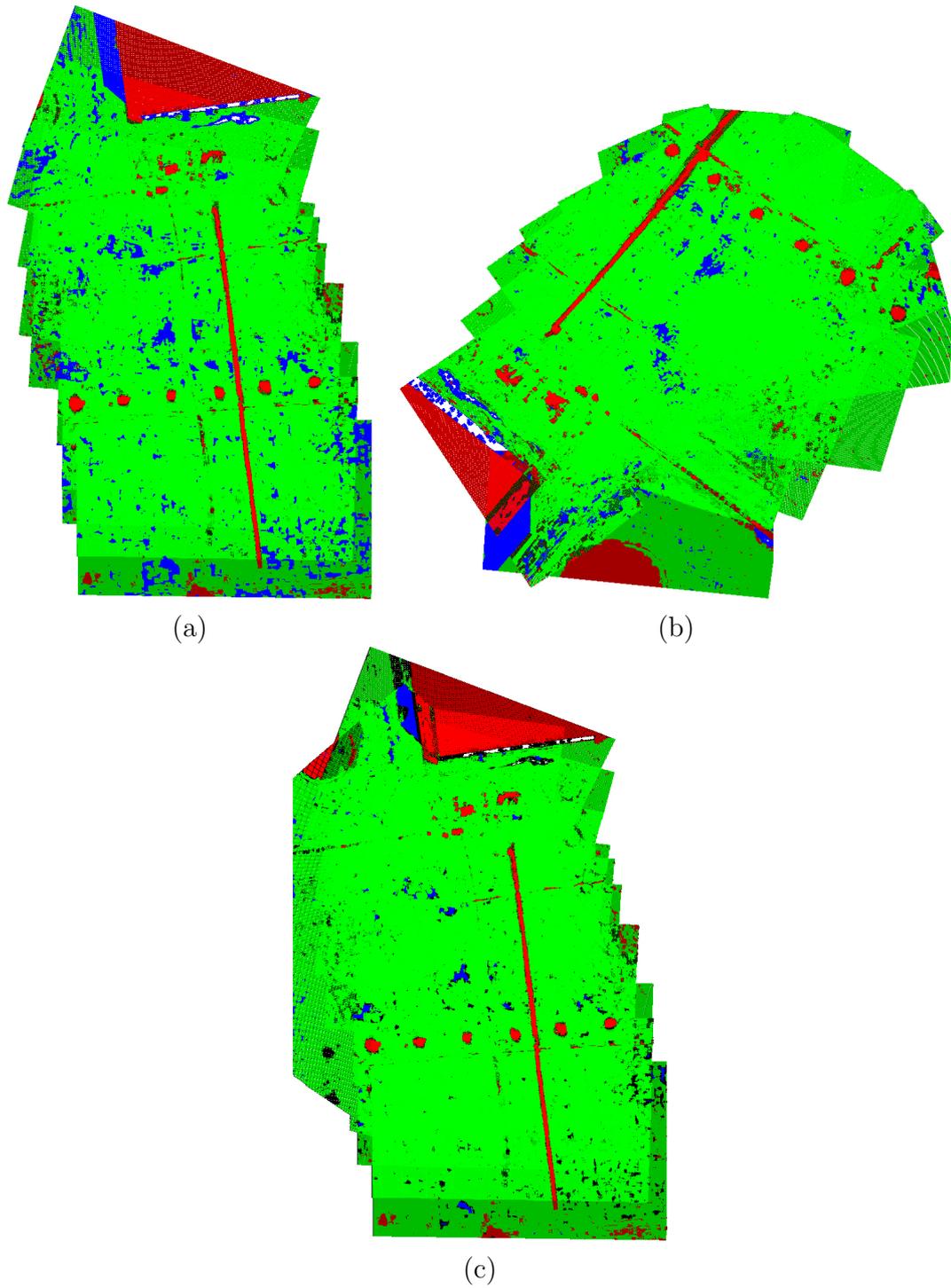


FIG. 4.11: Illustration de la fusion par homographie des probabilités de grilles locales G_L . Les images (a) et (b) sont les probabilités associées à deux grilles locales distinctes. L'image (c) représente les probabilités associées à la grille locale fusion des deux grilles de départ.

Chapitre 5

Vers une modélisation intégrant données terrestres et aériennes

5.1 Introduction

La fusion de données aériennes et terrestres en sein d'un même modèle nécessite tout d'abord la connaissance de leurs positions respectives : cette connaissance peut être obtenue grâce au géo-référencement des modèles construits. Mais, comme nous l'avons vu précédemment, si il est possible de déterminer un modèle aérien de traversabilité G_L avec une bonne cohérence spatiale, ce modèle est positionné dans un repère caméra, et donc non géoréférencé. Les imprécisions des capteurs de position (GPS) et d'orientation (centrale inertielle) de notre dirigeable nous empêchent d'obtenir un modèle global G_G correctement géoréférencé. Il en va de même pour les modèles de l'environnement construits à partir de données acquises par un robot terrestre : de nombreuses techniques de localisation (et notamment de SLAM) permettent de pallier l'imprécision des capteurs et de construire des modèles spatialement cohérents. Cependant, mais le géo-référencement de ces modèles nécessite l'exploitation d'un GPS précis.

En l'absence de tel moyens de géo-référencement, il est donc nécessaire de *mettre en correspondance* des éléments extraits des données afin de déterminer la transformation entre ceux-ci. Entre des données terrestres et des données aériennes, les différences de points de vue et de résolution rendent inopérantes les méthodes de mise en correspondance d'images : c'est sur la base d'informations contenues dans les modèles construits séparément que nous pouvons espérer établir des correspondances. Différents types de modèles sont envisageables :

- Les modèles numériques de terrain (MNT), qui représentent la géométrie de l'environnement par une grille cartésienne en chaque point de laquelle une élévation est estimée.
- Des modèles structurés de l'environnement, qui contiennent des primitives géométriques telles que des plans ou des segments, dont la position 3D est estimée. Les

plans et segments observés au sol sont plus faciles à ré-observer en altitude, et semblent pouvoir permettre de mettre en relation les modèles afin d’obtenir leur positionnement relatif.

- Enfin, des orthoimages découlant de la détection de zones planes peuvent être employés pour une mise en relation s’appuyant sur des techniques de mise en correspondance d’images multi-échelle.

Dans ce chapitre, nous allons évaluer les types de modèles qu’il est possible de construire aussi bien avec des robots terrestres qu’aériens à partir de vision, et les moyens de les mettre en correspondance. La section § 5.2 discute des modèles non structurés (MNT), et la section § 5.3 présente les moyens de construire des modèles basés sur des segments 3D. Enfin, la section § 5.4 présente la construction d’un modèle du plan du sol à partir de données terrestres, modèle analogue à celui que nous avons présenté dans les deux chapitres précédents.

5.2 Modèles numériques de terrain

Les MNT sont très couramment exploités pour représenter la géométrie de l’environnement, tant dans le domaine des Systèmes d’Information Géographique, où il sont construits sur la base de données aériennes, qu’en robotique, où ils sont construits sur la base de données de profondeur fournies par télémétrie laser ou stéréovision [Huber 00, Kweon 92] (figure 5.1). Dans [Vandapel 06], une approche qui met en correspondance des MNT construits à partir de données terrestres et aériennes est présentée : cette approche exploite deux modèles construits à partir de télémétrie laser.

Cependant, notre contexte de vision monoculaire aérienne rend difficile la construction de MNT de résolution compatible avec les modèles construits au sol (de l’ordre de 0.1 mètre). De nombreux travaux permettent la construction de MNT sur la base d’imagerie aérienne monoculaire, mais ils exploitent des capteurs de positionnement précis, et font appel à des calculs lourds, réalisés hors-ligne (voir par exemple [Zhu 01, Sanfourche 05]).

Dans notre cas, la détermination d’une reconstruction dense de la structure de l’environnement nécessite d’utiliser la géométrie projective multi-vues. L’étape essentielle avant reconstruction est donc l’estimation de la matrice fondamentale F . Deux difficultés se posent alors :

- D’une part, l’existence de solutions dégénérées lors de l’estimation de F . En effet, dans le cas d’un environnement principalement plan, le système dont la minimisation nous fournit une estimation de F est sous-contraint. Les estimées de F en résultant vérifient donc la géométrie épipolaire mais n’ont pas de sens physique dans le sens où nous l’entendons, c’est à dire comme une matrice induite par les paramètres des caméras et de mouvement [Torr 98, Kanatani 98, Luong 96].
- D’autre part, les mouvements de la caméra de notre ballon dirigeable sont très proches des mouvements critiques tels que définis par [Sturm 96]. Ces mouvements

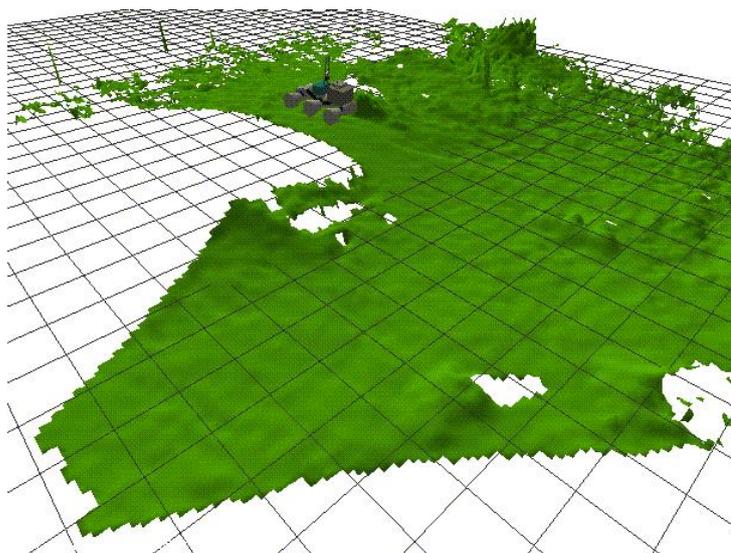


FIG. 5.1: *Modèle numérique de terrain construit par intégration de données stéréoscopiques acquises par un robot terrestre (la localisation précise du robot lors des déplacements est nécessaire pour construire un tel modèle – elle peut être fournie par une approche de type SLAM par exemple).*

imposent eux aussi des contraintes au système linéaire que l'on veut minimiser, et impliquent l'obtention d'estimées erronées de F si elles ne sont pas prises en compte.

Ces deux contraintes rendent ce problème d'estimation, qui est en fait apparenté à un problème d'autocalibration, très difficile à résoudre dans le contexte expérimental qui sont les nôtres. Ceci nous a donc orienté vers la construction de modèles plus structurés de l'environnement, constitués de primitives géométriques telles que des segments 3D.

5.3 Modèles de l'environnement basés sur les segments

Les segments de droites présents dans l'environnement sont des primitives géométriques qui semblent intéressantes pour la mise en correspondance entre données terrestres et aériennes : en effet, ils sont détectables sur de grandes variations d'échelle et de point de vue.

La construction d'un modèle basé sur des segments 3D passe par une première étape d'extraction puis de mise en correspondance de segments dans les images. Il faut ensuite réaliser une triangulation des segments appariés afin de déterminer leur localisation dans l'environnement :

- L'extraction de segments dans une séquence d'image est un processus classique : nous utilisons pour cela un algorithme associant un détecteur de contour de Canny-

Deriche et un processus de squelettisation.

- La mise en correspondance dépend des conditions d'acquisition des images (monoculaires ou stéréoscopiques, positionnement précis connu ou non).
- Enfin, la triangulation de ces segments est un processus plus délicat, qui dépend fortement des conditions d'acquisition des images.

5.3.1 Modélisation par segments 3D pour un robot terrestre

Cas de la stéréovision. Si le robot terrestre est muni d'un banc stéréoscopique calibré, chaque acquisition de paire d'images permet d'estimer la position 3D des segments appariés [Zhang 95a]. Des techniques de SLAM pourraient alors permettre de mettre à jour les coordonnées des segments au fur et à mesure des déplacements¹.

Cas monoculaire. De récents travaux tel que [Gee 06] ou au sein de notre laboratoire [Lemaire 07] ont introduit la notion de SLAM avec des segments détectés dans des séquences monoculaires. Ces approches permettent de minimiser la dérive spatiale du modèle et d'estimer la localisation des segments 3D en fusionnant leurs différentes observations.

La figure 5.2 présente les résultats obtenus par [Lemaire 07] sur une série d'images acquises avec un robot terrestre le long d'une trajectoire autour de deux boîtes en cartons. Sur ces données, nous pouvons vérifier certaines propriétés géométriques : par exemple les arêtes des boîtes sont bien perpendiculaires.

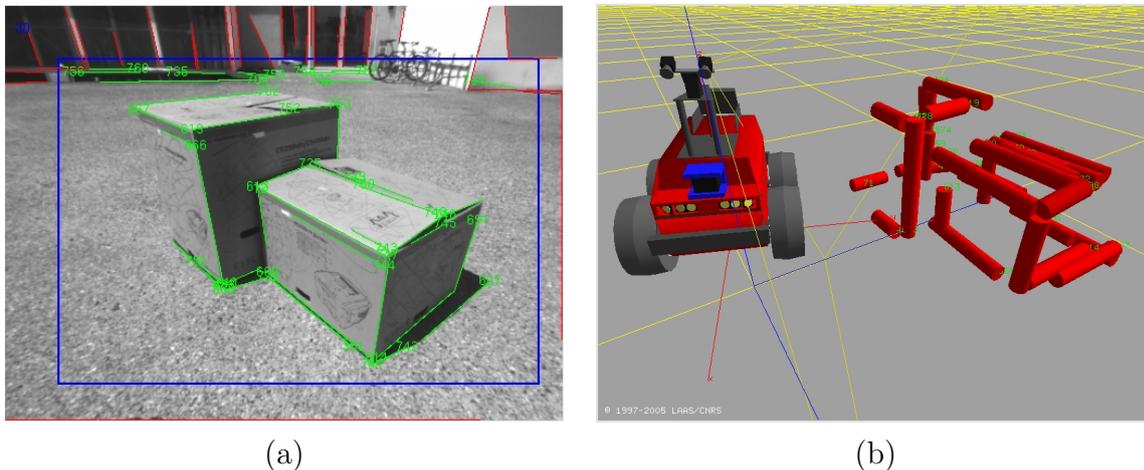


FIG. 5.2: *Résultat de SLAM monoculaire avec segments : (a) est une image de la séquence avec les segments extraits. (b) est le modèle 3D obtenu (d'après [Lemaire 07]).*

¹De manière surprenante, la littérature existante sur le SLAM ne propose pas de telles méthodes, sans doute parce que la stéréovision basée sur des appariements de segments est maintenant très rarement utilisée en robotique.

5.3.2 Modélisation par segments 3D pour un robot aérien

L'approche de SLAM monoculaire semble toute indiquée pour exploiter les segments extraits dans des séquences d'images aériennes. Cependant, les capteurs de positionnement et surtout d'attitude dont nous disposons délivrent des estimées de positionnement trop peu précises pour fournir une prédiction du mouvement exploitable par un algorithme de SLAM : les estimées 3D des positions des segments sont trop erronées et ne constituent pas un modèle exploitable.

Par contre, il est possible d'estimer les coordonnées des segments appartenant au sol. En effet, l'estimation d'homographie à partir de segments appariés apparaît comme une méthode naturelle pour estimer les segments qui sont portés par un même plan, de manière tout à fait analogue avec la méthode que nous avons utilisée.

Estimation d'homographie à partir de segments.

Comme nous l'avons vu, la relation entre points mis en correspondance vérifiant une homographie est $\mathbf{m}' = \mathbf{H}\mathbf{m}$. Si on considère un ensemble de droites mises en correspondance $\{\mathbf{l}, \mathbf{l}'\}$ tel que ces droites soient coplanaires au plan Π défini par l'homographie \mathbf{H}_1 on peut alors écrire la relation suivante :

$$\mathbf{l} = \mathbf{H}_1 \mathbf{l}' = \mathbf{H}^T \mathbf{l}' \quad (5.1)$$

Il est donc possible d'estimer \mathbf{H} à partir des droites porteuses de segments appariés, en réutilisant les algorithmes robustes décrits au § 2.3. Toutefois, cette estimation s'avère plus sensible que celle à partir de points d'intérêt. Nous en détaillons donc les spécificités ainsi que les points sensibles de la technique d'estimation dans ce qui suit.

Calcul des paramètres des droites porteuses. Nous considérons que nous avons un ensemble de segments préalablement mis en correspondance² $\{\mathbf{s}, \mathbf{s}'\}$ où chaque segment est simplement modélisé par ses extrémités. Soit pour un segment, $\mathbf{s} : \{\mathbf{s}_1 = (s_{1u}, s_{1v}, 1)^T, \mathbf{s}_2 = (s_{2u}, s_{2v}, 1)^T\}$. La droite porteuse $\mathbf{l} = (a, b, c)^T$ associée au segment s est simplement calculable comme la droite passant par \mathbf{s}_1 et \mathbf{s}_2 , soit $\mathbf{l} = \mathbf{s}_1 \times \mathbf{s}_2$. Afin d'améliorer la stabilité numérique de ce calcul et donc des suivants, *il est nécessaire* de commencer par effectuer un changement de repère sur les coordonnées des segments en déplaçant l'origine au centre de l'image (u_c, v_c) . Ceci permet en effet de centrer les paramètres (a, b, c) et d'obtenir une répartition plus uniforme de ces paramètres dans l'espace qu'ils décrivent. On obtient alors $\tilde{\mathbf{l}} = \tilde{\mathbf{s}}_1 \times \tilde{\mathbf{s}}_2$. La droite $\tilde{\mathbf{l}}$ étant définie à un facteur près, on choisira de la normaliser en fixant \tilde{c} à 1.

Un dernier point est à prendre en compte : cette modélisation est très imprécise pour toute droite porteuse passant près de l'origine. Nous paramétrisons donc en (ρ, θ) la droite passant par les segments $\tilde{\mathbf{s}}$ et éliminons les segments avec un ρ trop faible. En pratique

²La méthode de mise en correspondance des segments est présentée dans l'annexe § D.

nous avons fixé un seuil de 30 pixels qui nous permet de ne retenir pour nos calculs que les segments qui sont suffisamment éloignés du centre de l'image.

Estimation de H . L'ensemble des correspondances $\{\tilde{\mathbf{I}}, \tilde{\mathbf{I}'}\}$ étant défini, il est alors possible d'estimer H_1 par un algorithme robuste. Nous avons choisi le même algorithme que celui que nous avons utilisé pour notre estimation d'homographie à partir de points (§ 3.2.4) mais en optant pour la normalisation des coordonnées. En regard de la grande inhomogénéité entre les valeurs de \tilde{a} , \tilde{b} et \tilde{c} cette normalisation est indispensable afin d'obtenir un conditionnement correct de la matrice \mathbf{A} .

Une autre spécificité de cette estimation est l'obligation d'utiliser un modèle d'homographie complet, le modèle affine étant à proscrire. En effet, l'homographie recherchée étant $H = H_1^T$, l'utilisation du modèle affine fixerait les paramètres $h_{l_{31}}$ et $h_{l_{33}}$ de H_1 à 0 et donc les paramètres de translation $h_{l_{13}}$ et $h_{l_{23}}$ de H .

L'homographie estimée étant donc \tilde{H}_1^T , il reste donc à la dénormaliser afin d'obtenir H . En considérant que la normalisation initiale s'écrit :

$$\tilde{\mathbf{m}} = \mathbf{L}\mathbf{m} \quad \text{avec } \mathbf{L} = \begin{bmatrix} 1 & 0 & -u_c \\ 0 & 1 & -v_c \\ 0 & 0 & 1 \end{bmatrix}$$

On obtient alors la relation $H = L^{-1}\tilde{H}L$. Comme $\tilde{H} = \tilde{H}_1^T$, on a donc :

$$H = L^{-1}\tilde{H}_1^T L$$

Résultats. Comme l'illustre la figure 5.3, les images recalées par l'utilisation des homographies de segments sont proches de celles obtenues par les homographies de points. Toutefois les scores de corrélation entre les images recalées par l'homographie estimée montre que la précision de ce recalage est très inférieure. On remarquera qu'on effectue une corrélation directe, donc le moindre décalage d'un pixel suffit à obtenir une mauvaise corrélation.

La difficulté de cette approche est qu'il est difficile d'évaluer la qualité d'un résidu calculé sur des coefficients de droites porteuses. L'estimation d'homographie à partir de points a pour avantage de fournir un résidu possédant un sens physique évident comme étant l'erreur de reprojexion des points (assimilable à la parallaxe relative au plan). En restant dans un formalisme de droites, nous manquons donc de quantification sur l'erreur résiduelle de l'estimée d'homographie.

En outre, les segments appariés n'étant pas nécessairement de mêmes longueurs dans les deux images, il n'est pas possible d'utiliser les coordonnées de leurs extrémités pour estimer un résidu plus exploitable.

Ces conclusions nous amènent donc à l'approche suivante.

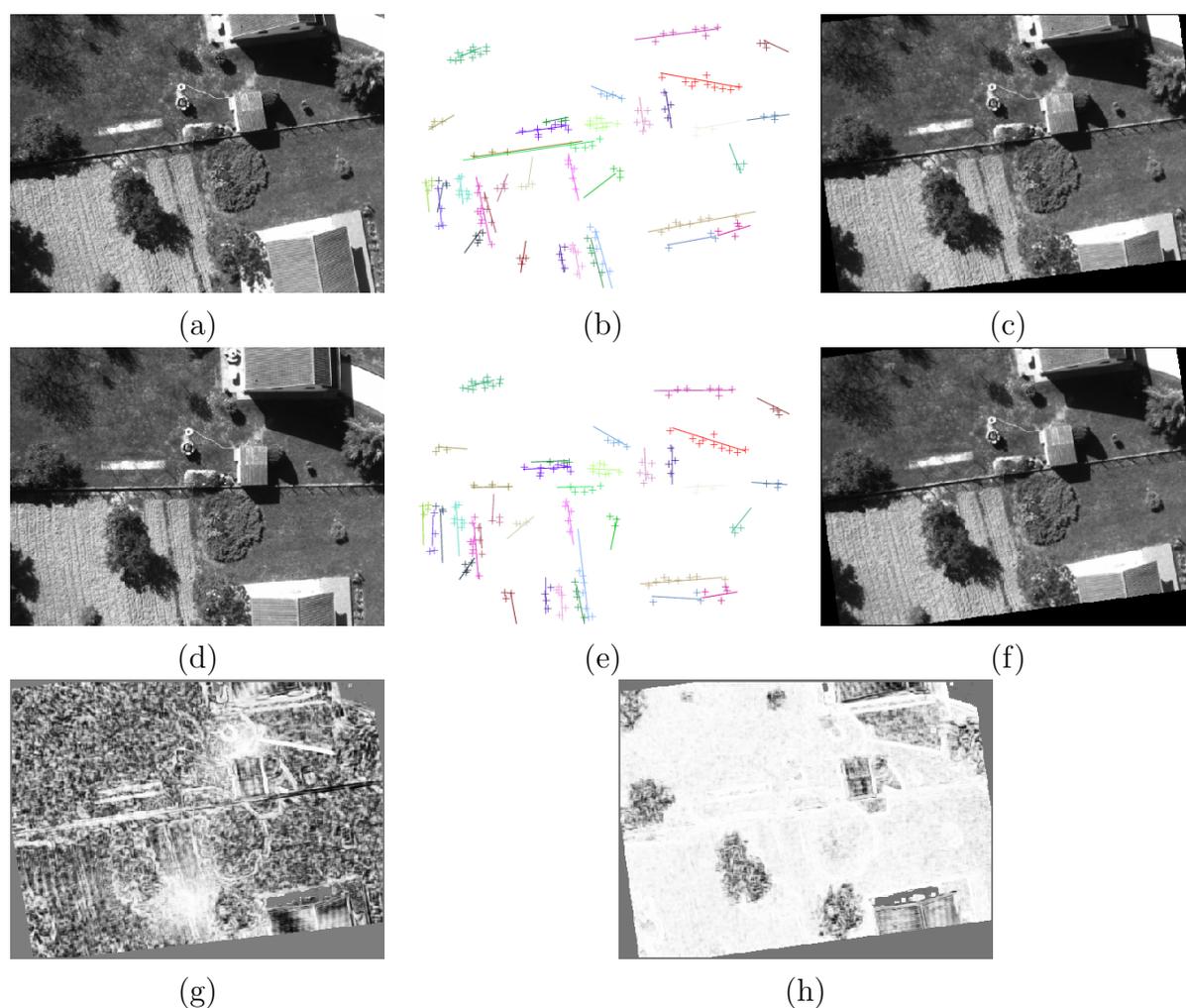


FIG. 5.3: Illustration de l'estimation d'homographie à partir de segments : (a) image de référence, (d) image suivante, (b) et (e), segments appariés – les segments correspondants ont la même couleur, notons la présence de faux appariements, (c) est l'image (d) recalée dans le repère de l'image de référence à l'aide de l'homographie estimée à partir des segments. L'image (f) est l'image (d) recalée dans le repère de l'image de référence à l'aide de l'homographie estimée à partir des points. Les images (g) et (h) illustrent les corrélations respectives entre les images (c) et (a) et entre les images (f) et (a).

Estimation des segments validant une homographie

La dualité de la relation (5.1) nous suggère d'estimer l'homographie H associée à un ensemble de points mis en correspondance puis de sélectionner l'ensemble des segments $(\mathbf{s}_1, \mathbf{s}_2)$ validant $H_1 = H$.

Mais comme nous venons de l'aborder nous manquons de métrique fiable. Pour cela, plusieurs scores prenant en compte l'orientation, la position et l'intégrale entre les segments \mathbf{s}_2 et la projection de leurs correspondants $\hat{\mathbf{s}}_2$ sont calculés et nous permettent de

déterminer les paires de segments (s_1, s_2) correspondants entre les deux images. Notons que ces scores ne peuvent être raisonnablement utilisés par l'estimation de H_1 à partir des segments, car il faudrait les mettre aussi en oeuvre lors des tirages aléatoires de l'estimation robuste ce qui s'avérerait trop coûteux.

La figure 5.4 illustre les segments ainsi mis en évidence. En exécutant ce processus conjointement aux estimations de grilles locales G_L il est alors aisé de positionner les segments 3D dans celles-ci.

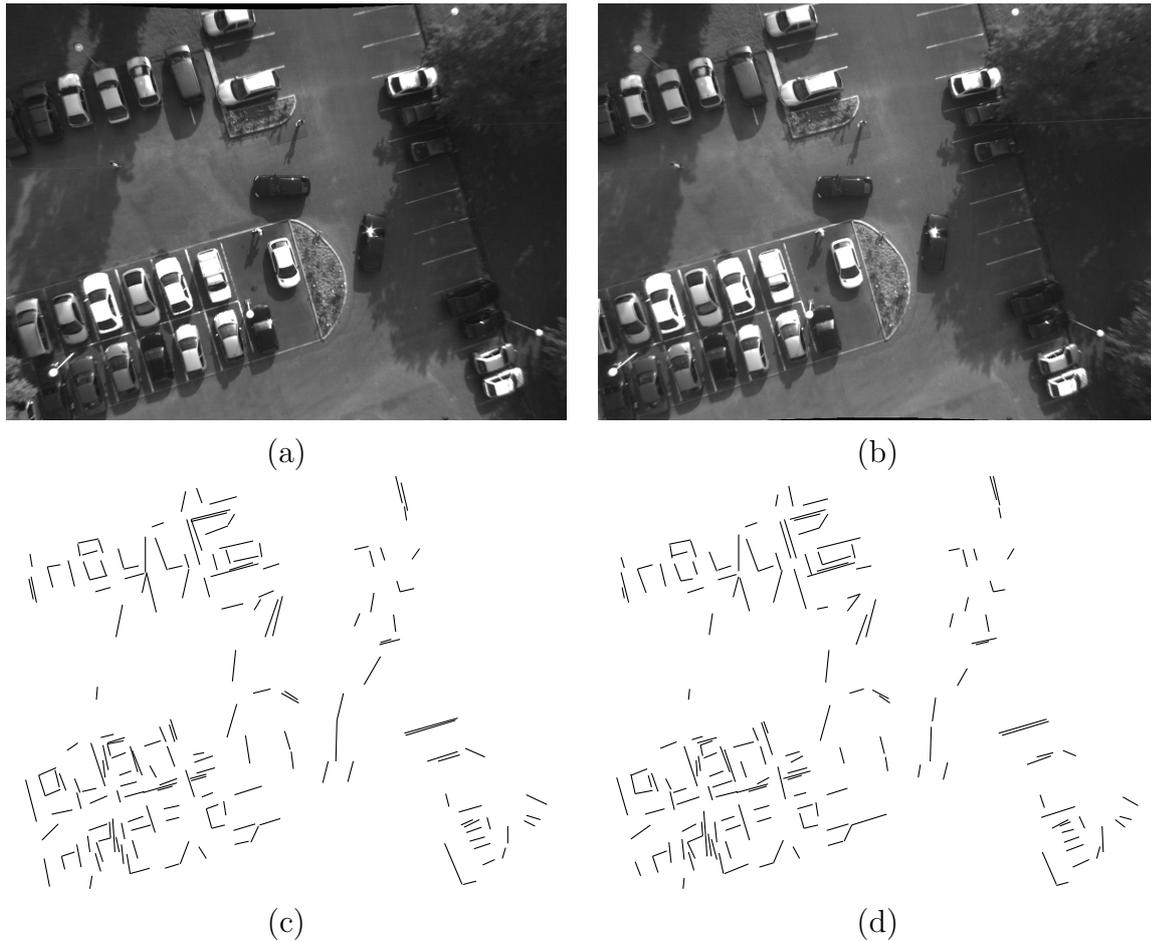


FIG. 5.4: *Illustration de la mise en correspondance de segments par homographie : (a) image de référence, (b) seconde image, (c) segments issus de l'image (a), (d) segments correspondants issus de l'image (b).*

5.4 Modèles de l'environnement basés sur les plans

Les premiers chapitres de cette thèse ont exposé notre approche de détection de zones traversables pour un robot aérien, qui correspondent au plan du sol : nous nous intéressons

ici à la détection de ce même plan à partir des données acquises par un robot terrestre, afin de pouvoir mettre en correspondance la mosaïque associée avec celle qui est construite par le robot aérien.

L'extraction de zones planes peut être réalisée par l'analyse des données de profondeur fournies par stéréovision ou télémétrie laser, ou même par l'analyse du modèle numérique de terrain construit sur la base de ces données (§ 5.2). Mais si de telles méthodes sont couramment exploitées pour déterminer l'espace navigable pour un robot terrestre, elles ne fournissent pas de mosaïque qui puisse être mis en correspondance avec le modèle construit par l'analyse de séquences monoculaires.

Par contre, l'application des travaux développés pour des images aériennes à des données acquises par un robot terrestre fournit de modèles comparables, et est relativement directe : l'unique différence s'avère être le type de mouvement de la caméra, qui nous oblige à orienter différemment le repère de notre grille locale G_L .

En effet, lors de la construction de la grille locale G_L à partir d'images aériennes cette grille était initialisée et orientée dans le même repère que celui de la première image. Ce choix était naturel car, la caméra se déplaçant quasiment dans un plan parallèle à celui du sol, on minimisait ainsi les déformations dues à la reprojection des images dans ce repère.

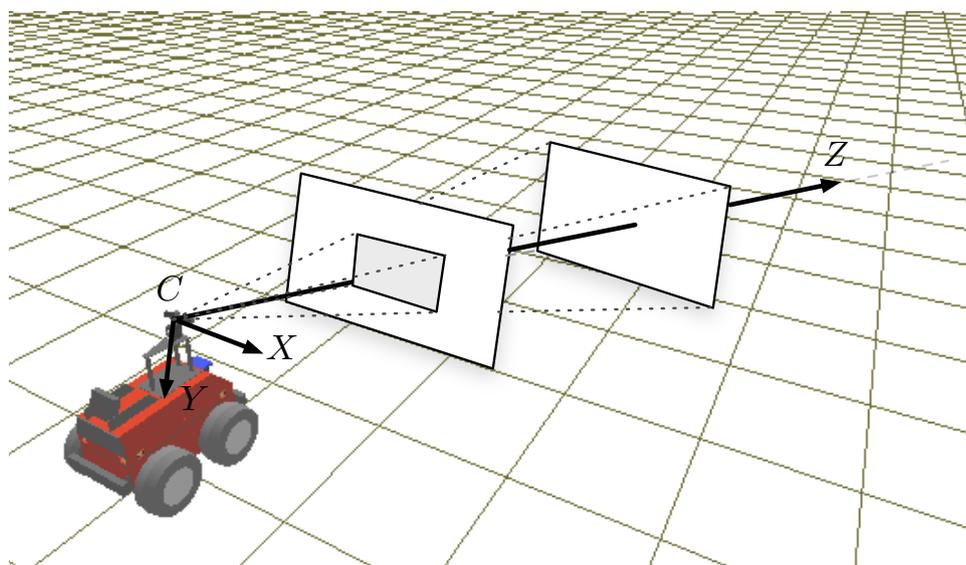


FIG. 5.5: Illustration du déplacement pour un robot terrestre. Le cadre grisé est la projection de la seconde image dans la 1ère.

Mais comme l'illustre la figure 5.5, dans le cas d'images terrestre, ce choix n'est pas adapté. En effet, si le robot effectue une translation vers l'avant, l'homographie inter-images traduit alors un facteur d'échelle. Ce type de déplacement induisant beaucoup d'incertitudes dans son estimation, ceci influe sur la qualité des résultats. En outre, en continuant à positionner la grille locale dans le repère de la première image, les nouvelles

observations utilisées pour sa mise à jour sont alors associées à un nombre de cellules de la grille décroissant à chaque nouvelle image utilisée.

Nous avons donc naturellement choisi de positionner la grille locale G_L sur le plan Π_{G_L} du sol de coordonnées $[0, -1, 0, d]^T$ dans le repère caméra R_{C_2} . Ce repère est centré en \mathbf{C} et orienté tel que le plan $0xz$ de ce repère soit parallèle à Π_{G_L} .

Afin de déterminer la projection entre les points de l'image et les cellules de la grille locale G_L nous utilisons la notation de Plücker (§ 3.2.2 [Hartley 04]) des droites. Cette notation nous permet de définir la projection d'un point \mathbf{m} sur un plan $\Pi = [n_x, n_y, n_z, d]^T$ comme étant :

$$\mathbf{M} = \mathbf{L} \Pi$$

où \mathbf{M} est la projection du point \mathbf{m} et \mathbf{L} la ligne (ou raie) passant par le centre optique $\mathbf{C} = [0, 0, 0, 1]^T$ et le point \mathbf{m} .

On peut tout d'abord détailler l'écriture de \mathbf{L} qui est égale à $\mathbf{C}\bar{\mathbf{m}}^T - (\mathbf{C}\bar{\mathbf{m}}^T)^T$ ($\bar{\mathbf{m}} = [\bar{m}_x, \bar{m}_y, \bar{m}_z = 1, 1]^T$ est la projection du point image \mathbf{m} dans le plan focal normalisé). En notant $\tilde{\mathbf{m}} = [\tilde{m}_x, \tilde{m}_y, \tilde{m}_z]^T$ et en considérant :

$$\mathbf{C}\bar{\mathbf{m}}^T = \begin{bmatrix} 0_{3 \times 4} \\ \tilde{\mathbf{m}}^T & 1 \end{bmatrix}$$

On peut alors écrire :

$$\mathbf{L} = \begin{bmatrix} 0_{3 \times 3} & -\tilde{\mathbf{m}} \\ \tilde{\mathbf{m}}^T & 1 \end{bmatrix}$$

Avec $\tilde{\mathbf{n}} = [\tilde{n}_x, \tilde{n}_y, \tilde{n}_z]^T$, on obtient alors l'écriture de la projection suivante :

$$\mathbf{M} = \mathbf{L} \Pi = \begin{bmatrix} 0_{3 \times 3} & -\tilde{\mathbf{m}} \\ \tilde{\mathbf{m}}^T & 1 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{n}} \\ -d \end{bmatrix} = \begin{bmatrix} d\tilde{\mathbf{m}} \\ \tilde{\mathbf{n}}\tilde{\mathbf{m}} \end{bmatrix}$$

Cette dernière reformulation peut s'écrire sous la forme suivante, ce qui permet de mettre en évidence une matrice de projection \mathbf{P} à appliquer aux points \mathbf{m} :

$$\mathbf{M} = \bar{\mathbf{P}} \mathbf{K}^{-1} \mathbf{m} = \mathbf{P} \mathbf{m} \text{ avec } \bar{\mathbf{P}} = \begin{bmatrix} d\mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 1} \\ \tilde{\mathbf{n}} & 0 \end{bmatrix} \text{ et } \mathbf{K}^{-1} = \begin{bmatrix} \mathbf{C}^{-1} \\ 0 & 0 & 1 \end{bmatrix}$$

En considérant que l'on souhaite projeter \mathbf{m} sur le plan Π_{G_L} on doit alors appliquer à \mathbf{m} avant projection la rotation \mathbf{R} qui relie le repère de la caméra R_C au repère R_{C_2} . Ceci nous permet alors d'écrire la relation suivante entre \mathbf{m} et son projeté $\mathbf{M}_{G_L|R_{C_2}}$ sur le plan Π_{G_L} :

$$\mathbf{M}_{G_L|R_{C_2}} = \begin{bmatrix} d_{|R_{C_2}} \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 1} \\ \tilde{\mathbf{n}}_{|R_{C_2}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{1 \times 3} & 0 \end{bmatrix} \mathbf{K}^{-1} \mathbf{m} \quad (5.2)$$

où $\tilde{\mathbf{n}}_{|R_{C_2}} = [0, -1, 0]^T$ et $d_{|R_{C_2}}$ la distance de la caméra au sol selon l'axe ($0y$).

La mise à jour de chaque cellule de G_L , à la suite de nouvelles segmentations de zone plane, s'effectue de façon identique au cas aérien. La position des cellules à mettre à jour

s’obtenant donc en combinant les homographies inter-images de façon à obtenir la position des nouvelles observations dans le repère de la première image, puis en effectuant à nouveau la transformation (5.2).

Afin d’obtenir la grille locale G_G positionnée dans le repère R_{ENU} (East (x), North (y), Up (z)) centré en un point du sol et orienté suivant la convention ENU il est nécessaire de reprojeter la grille G_L . En considérant la rotation \mathbf{R}' et la translation \mathbf{t}' entre le repère R_{C_2} et le repère R_{ENU} , on peut appliquer le même raisonnement qu’au § 4.3.1. Ceci nous permet ainsi d’écrire la transformation suivante :

$$s\mathbf{M}_{G_G} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{t}' \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} & \mathbf{I}_3 & \mathbf{0}_{3 \times 1} \\ 0 & 0 & \frac{1}{-z_{C_0|R_{G_G}}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}' & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \mathbf{m}_{G_L}$$

Résultats. La figure 5.6 présente des résultats obtenus par un robot terrestre en employant ces transformations. Nous pouvons constater que la nouvelle orientation de G_L par rapport au repère caméra permet d’obtenir le même comportement de notre algorithme que dans le cas aérien.

Toutefois l’homographie inter-images utilisée estime toujours le facteur d’échelle entre les images, ce qui se traduit par des imprécisions de la grille locale G_L .

Il serait donc préférable d’utiliser des grilles locales induites par seulement deux ou trois vues, puis de les positionner dans une grille globale à l’aide des estimées de position fournies par un algorithme de SLAM par exemple. Ce modèle structuré d’environnement gagnerait ainsi en consistance spatiale.

5.5 Bilan

Nous avons évoqué, dans ce chapitre, quelques moyens de mettre en correspondance des données terrestres et aériennes. L’exploitation de la géométrie 3D de l’environnement semble possible et prometteuse, mais requiert la capacité de construire des modèles suffisamment précis, qu’ils soient structurés en primitives géométriques du premier ordre (droites et plans) ou non (MNT). Grâce à la possibilité d’acquérir des données de profondeur, la construction de tels modèles par des robots terrestres est possible, comme le montrent de nombreuses contributions dans l’état de l’art. Mais ça n’est pas encore le cas à partir de séquences monoculaires aériennes pour lesquelles le positionnement est imprécis. Les méthodes actuellement disponibles nécessitent des algorithmes uniquement exploitables hors-ligne.

Par contre, nous avons montré la possibilité de construire aisément une mosaïque d’images correspondant au sol avec des images acquises par des robots terrestres, en adaptant la méthode exploitée pour des images aériennes. Les deux méthodes fournissent alors un modèle de l’environnement de nature semblable, qu’il paraît possible de recalibrer et donc de fusionner à l’aide de techniques de mise en correspondance d’images – et

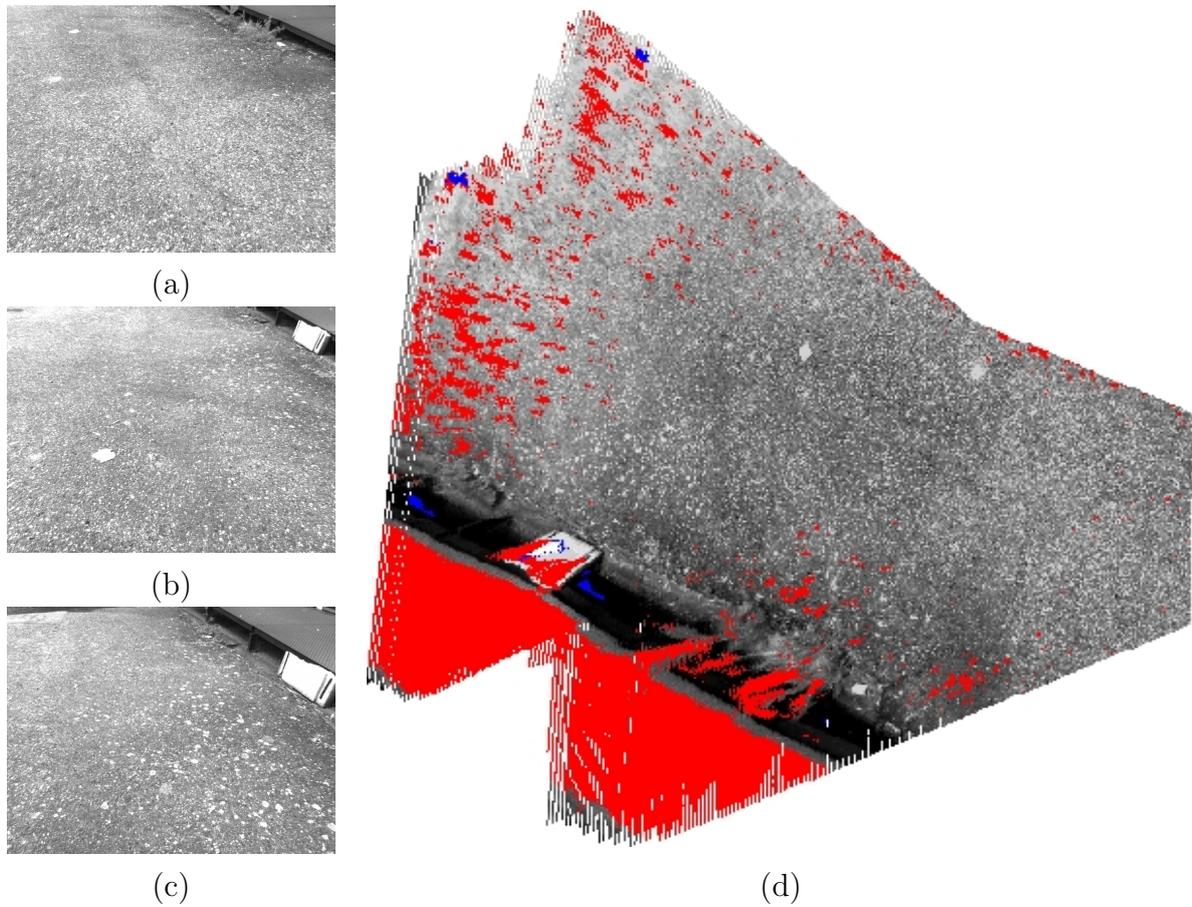


FIG. 5.6: *Illustration d'une grille locale construite à partir d'images terrestres. (a)(b)(c) sont les images respectives du début, milieu et fin de la séquence utilisée. (d) représente la mosaïque d'images correspondant au plan du sol sur laquelle sont représentées en rouge les zones telles que $p(c|O_k) < 0.5$ et en bleu les régions où la variance des images originales est trop faible pour en tirer des conclusions fiables.*

notamment celles qui sont robustes à des changements d'échelle importants [Lowe 04].

Chapitre 6

Conclusion et perspectives

Nous avons exposé nos travaux sur la modélisation d'environnements par vision monoculaire dans un contexte de robotique aéro-terrestre. Les principales contributions de ces travaux sont les suivantes :

- **Estimation robuste d'homographie sur la base de points mis en correspondance dans des séquences d'images.** Nous avons présenté une étude approfondie des propriétés des estimateurs robustes dans le contexte de l'estimation d'homographie. Un résultat notable de cette étude est le faible apport de la normalisation des coordonnées introduite par Hartley pour de tels estimateurs. Un algorithme rapide et performant a ensuite été proposé afin d'estimer les homographies et de segmenter les points appariés, selon qu'ils appartiennent, ou non, au plan majoritaire.
- **Densification des informations de traversabilité.** Sur la base des appariements de points attachés à des zones traversables, nous avons proposé une méthode permettant de déterminer l'ensemble des pixels qui correspondent à des zones traversables. La modélisation complète du comportement du score de corrélation utilisé est une contribution importante de cette thèse. Elle nous a permis de proposer une approche de segmentation des zones planes ne requérant aucun seuil ou variable d'ajustement.
- **Construction d'un modèle probabiliste de traversabilité.** Nous avons introduit deux approches permettant la fusion, en un modèle global, des zones planes détectées au fur et à mesure de l'acquisition des images. Ce modèle exprime la probabilité que les zones perçues soient traversables. Des résultats concluants ont ainsi été présentés ainsi qu'une méthode permettant d'obtenir l'équivalent d'une orthoimage par la création d'une mosaïque des zones traversables.
- **Vers la construction d'un modèle plus complet basé sur des primitives planes.** Différentes approches visant à construire des modèles utiles à la mise en correspondance entre données terrestres et aériennes ont été étudiées. Nous avons

proposé un moyen de construire un modèle de l'environnement à partir des données terrestres comparable à celui que nous avons proposé pour des données aériennes.

Cependant, le problème de la fusion de données terrestres et aériennes reste encore difficile. Dans la prolongation des travaux que nous avons proposés, nous pouvons envisager les extensions suivantes :

- **Détection de l'ensemble des plans.** Nous avons proposé un algorithme robuste de détection de plan, mais cet algorithme reste dépendant de l'existence d'une majorité de points associé au plan correspondant à l'homographie estimée. Il nous est ainsi actuellement impossible d'estimer une homographie décrite par seulement 40% des points d'intérêt mis en correspondance. Récemment, [Silveira 06] a proposé une nouvelle approche permettant de passer outre cette limitation en détectant l'ensemble des homographies existant dans une paire d'images. Il serait donc souhaitable de combiner ces travaux à notre estimateur afin d'améliorer ses performances.
- **Vers des grilles locales plus consistantes spatialement.** Notre processus de construction de grilles locales est limité par la dérive causée par le cumul des homographies inter-images. Bien que notre algorithme d'estimation soit très robuste, il est préférable de se limiter à quelques dizaines d'images consécutives si l'on souhaite conserver une bonne consistance spatiale. Toutefois, une nouvelle voie pourrait être envisagée en mettant en pratique les travaux de [Caballero 07] qui affine les incertitudes d'estimation des homographies inter-images grâce à un filtre de Kalman. Lors des fermetures de boucle, il est alors possible de propager les corrections sur l'ensemble des homographies et ainsi d'éliminer les dérives d'estimation. Il serait ainsi possible de créer des grilles locales de dimension très supérieures tout en conservant une bonne localisation.
- **Vers une meilleure odométrie en aérien.** Si nous parvenons à créer des grilles locales spatialement consistantes, la création de grilles globales reste délicate, notamment lorsque que les cartes locales ne se recouvrent pas. Il nous est alors impossible de mettre en correspondance nos mosaïques pour contourner nos problèmes d'incertitudes de géoréférencement. Comme nous l'avons vu, ce problème de localisation est délicat à résoudre par l'emploi de techniques de SLAM, les prédictions du mouvement étant trop erronées. Les travaux de [Triggs 98] semblent être une voie intéressante pour résoudre ce problème. En effet, ils introduisent une méthode d'extraction des paramètres du mouvement (\mathbf{R}, \mathbf{t}) à partir d'une homographie \mathbf{H} . Nous présentons en (annexe § C) quelques résultats préliminaires de cette approche. S'il est alors possible, de déterminer par les homographies inter-images, une odométrie dans le plan image consistante en employant les travaux de [Caballero 07], il serait possible de déterminer une estimation du géoréférencement tout aussi consistante.

Annexe A

Optimisation du calcul de score de corrélation ZNCC

Les optimisations que nous avons implémentées sont inspirées en grande partie de la procédure de calcul décrite par McDonnell [McDonnell 81] et Faugeras[Faugeras 93a]. Toutefois, dans notre cas particulier de corrélation directe (où la fenêtre de recherche est nulle), nous avons pu simplifier la méthode en ne gardant qu'une seule technique utilisée pour le calcul des moyennes μ_k , des variances σ_k^2 et aussi des covariances $\sigma_{1,2}^2$ avec k indice de l'image. Nous allons en expliquer le fonctionnement en prenant l'exemple du calcul des moyennes qui a été généralisé par la suite aux autres grandeurs.

Cette technique permet de calculer la moyenne des pixels d'une fenêtre, indépendamment de la taille de la fenêtre, en quatre opérations par valeur de sortie. Rappelons l'équation définissant la moyenne en $\mathbf{m} = (u, v, 1)^\top$ pour une fenêtre \mathbb{W} :

$$\mu_k(\mathbf{m}) = \frac{1}{(2w+1)^2} \sum_{i=u-w}^{u+w} \sum_{j=v-w}^{v+w} I_k(i, j) \quad (\text{A.1})$$

Les moyennes tout comme les variances sont calculées en tout pixel \mathbf{m} de l'image. Si on s'intéresse à celles d'une même ligne, la redondance des calculs est évidente entre deux pixels adjacents. L'optimisation proposée est illustré par la figure A.1. Si on a $\mu_{v,k}(u)$ moyenne sur la colonne u de la fenêtre courante telle que :

$$\mu_{v,k}(u) = \frac{1}{(2w+1)^2} \sum_{j=v-w}^{v+w} I_k(u, j)$$

On obtient alors :

$$\mu_k(u, v) = \mu_k(u-1, v) + \mu_{v,k}(u+w) - \mu_{v,k}((u-1)-w) \quad (\text{A.2})$$

Il est ainsi possible de s'abstraire de toute redondance sur le calcul des moyennes ce qui améliore très fortement les temps de calcul.

Ce procédé est donc utilisé pour le calcul de μ_k , σ_k^2 et $\sigma_{1,2}^2$ et sa mise en oeuvre est décrite dans le cas des calculs de moyenne dans l'algorithme (3).

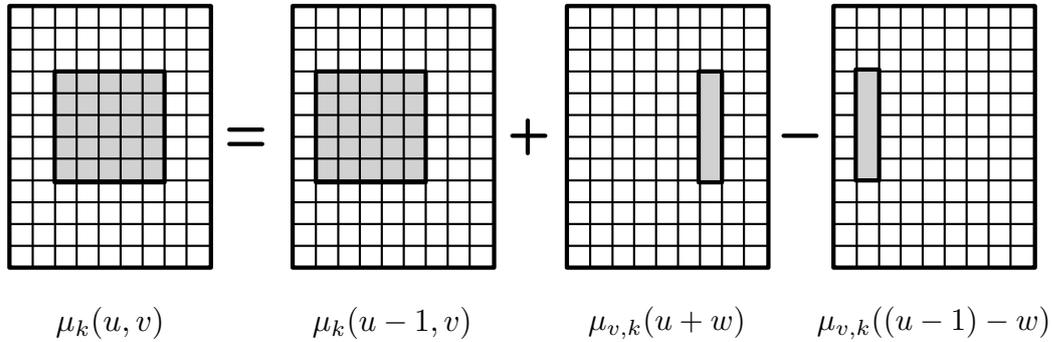


FIG. A.1: Schéma de principe de l'optimisation du calcul ZNCC

Algorithme 3 : Optimisation du calcul du score de corrélation ZNCC

Données : L'image I_k de M lignes et N colonnes

pour chaque $v \in [w \dots N - w]$ **faire**

calculer $\mu_k(w, v)$, moyenne de la 1ère fenêtre de la ligne v selon l'équation (A.1)

pour chaque $i \in [w \dots M - w]$ **faire**

calculer et stocker $\mu_{v,k}(i)$, moyenne sur la ligne i de la fenêtre courante

fin

pour chaque $i \in [w \dots ncol - w]$ **faire**

calculer $\mu_k(u, v)$ par l'équation (A.2) :

fin

fin

Annexe B

Relation entre parallaxe et score de corrélation ZNCC

Nous avons essayé d'évaluer la parallaxe due au recalage des pixels d'une zone ne supportant l'homographie H_I en fonction des scores de corrélation. Nous avons donc généré des images avec des parallaxes pré-définies et établi des fonctions de répartition pour évaluer l'évolution du score ZNCC. Ceci nous a permis de déterminer, en fonction d'un score, la probabilité qu'il soit associé à une certaine parallaxe. La parallaxe étant associée à la distance au plan, cela nous apporte aussi une information sur la structure 3D des zones non planes. La figure B.1 illustre ces résultats.

Comme nous pouvons l'observer cette information n'est rapidement plus discriminante. En effet dès que le score S est inférieur à 0.97 les différentes probabilités associées aux différentes parallaxes possibles deviennent quasiment équiprobables. En outre, seules les parallaxes de 0.1 et 0.2 pourraient être éventuellement identifiables mais de tels décalages sont situés dans la marge d'erreur de l'estimation de l'homographie H_I . Ces résultats n'ont donc pas été exploités.

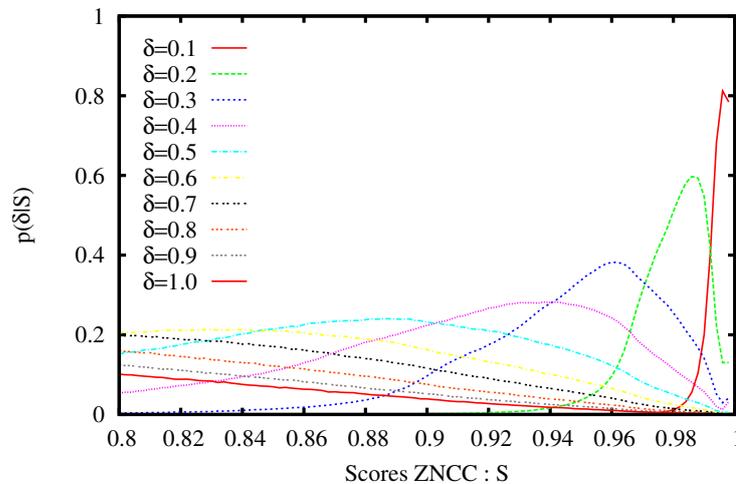


FIG. B.1: Probabilité que la parallaxe par rapport à Π_I soit de δ pixels sachant le score de corrélation ZNCC est S .

Annexe C

Extraction du mouvement à partir des homographies

Une homographie étant définie par $H = C(\mathbf{R} - \frac{\mathbf{t}\mathbf{n}^T}{d})C^{-1}$ elle contient tous les paramètres du mouvement (\mathbf{R}, \mathbf{t}) ainsi que ceux du plan induisant cette homographie (\mathbf{n}, d) .

[Triggs 98] propose une méthode permettant d'extraire deux jeux de paramètres $(\mathbf{R}, \mathbf{t}, \mathbf{n}, d)$ compatibles avec cette homographie. Nous ne décrivons pas cet algorithme qui est déjà largement commenté dans son article et accompagné d'une implémentation en code MATLAB.

Nous avons donc implémenté cette technique en y ajoutant un historique des solutions qui nous permet de sélectionner les jeux de paramètres $(\mathbf{R}, \mathbf{t}, \mathbf{n}, d)$ tels que les estimations successives de (\mathbf{n}, d) soient consistantes.

La figure C.1 illustre, pour des images aériennes, les estimées de translation (b) extraites des homographies entre l'image de départ et l'image de courante. Ces homographies sont issues du cumul des homographies inter-images d'une série de 130 images. Ces estimées sont proches de notre mesure de vérité terrain réalisée à l'aide d'un GPS métrique (a).

En revanche pour des images terrestres, les estimées de translation (b) sont nettement moins précises et sujettes à une dérive importante comme l'illustre la figure C.2. Ces résultats sont directement imputables à la qualité des estimation des homographies dans un tel contexte de prise de vues, comme nous l'avons décrit précédemment (§ 5.4).

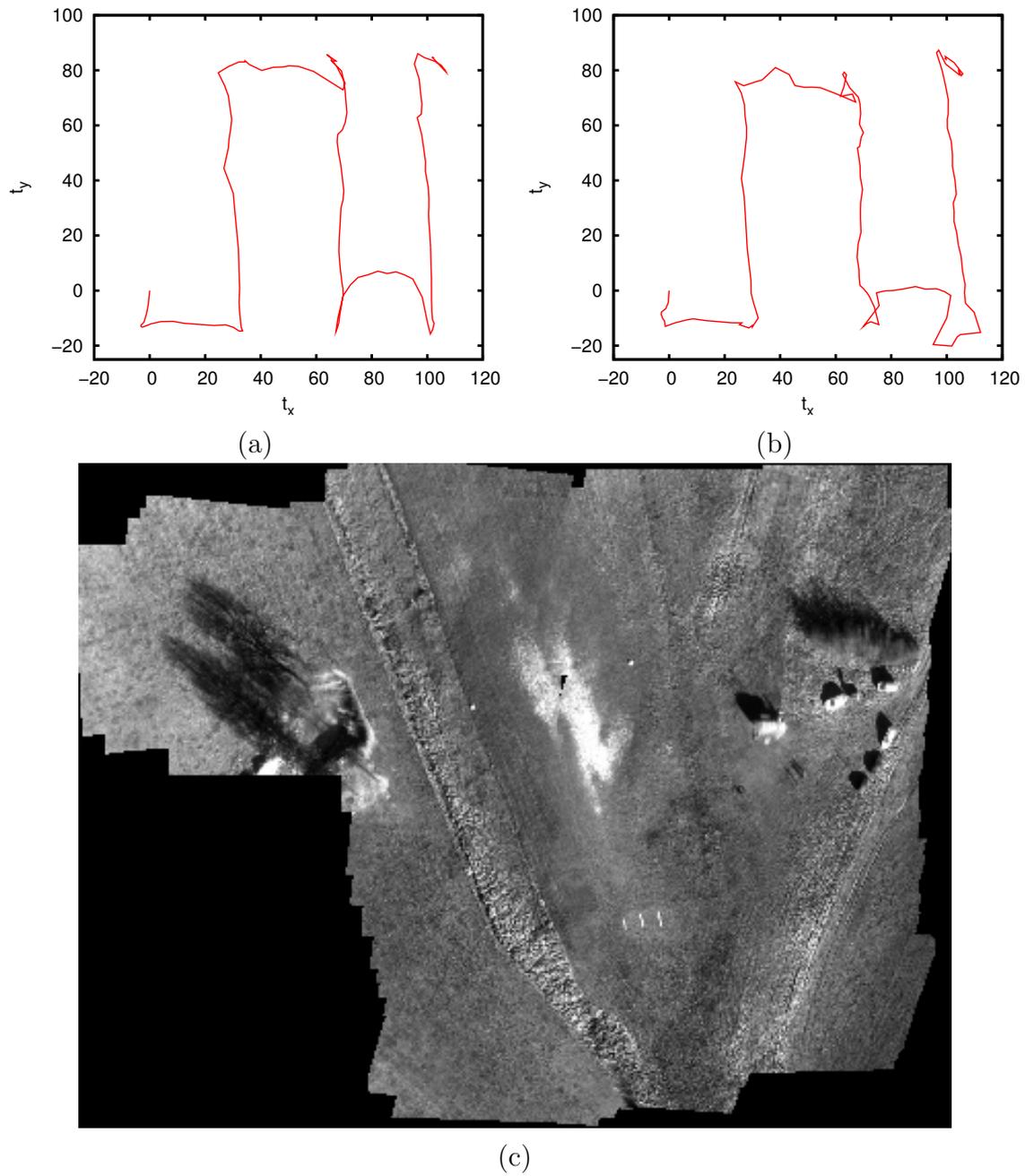


FIG. C.1: *Extraction du mouvement à partir des homographies pour des images aériennes. (a) est la mesure du GPS, (b) la translation estimée à partir des homographies. (c) est une mosaïque sans détection de zones planes déterminée à partir des homographies.*

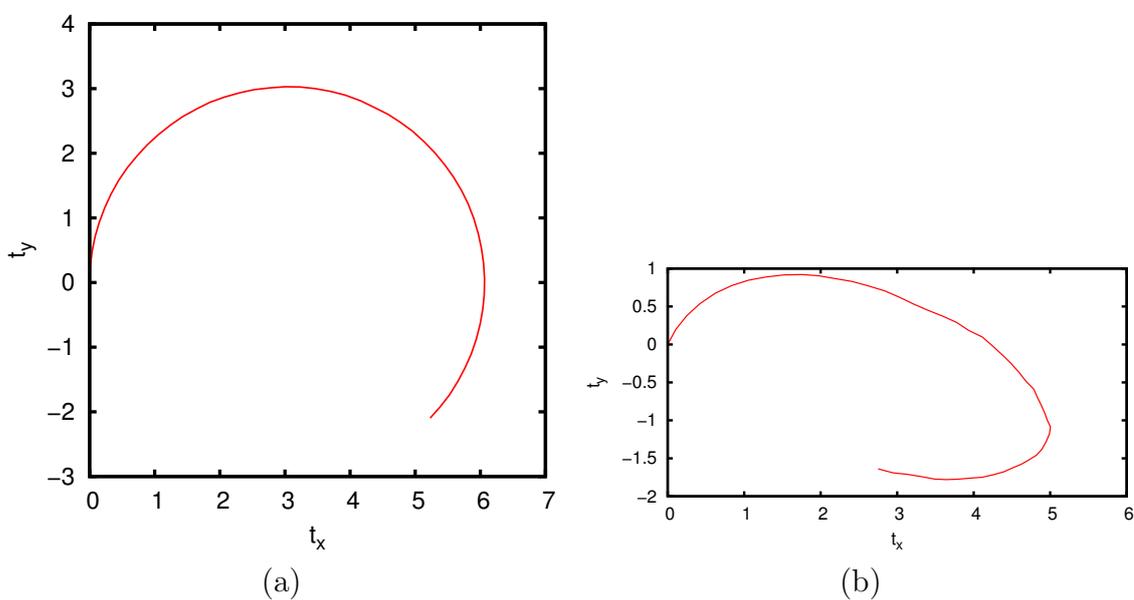


FIG. C.2: *Extraction du mouvement à partir des homographies pour des images terrestres. (a) est l'odométrie mesurée, (b) la translation estimée à partir des homographies.*

Annexe D

Algorithmes de mise en correspondance exploités

Mise en correspondance de points d'intérêt

Pour mettre en correspondance les points d'intérêt détectés dans deux images, nous utilisons un algorithme d'appariement qui a été précédemment développé au laboratoire [Jung 01]. L'algorithme se base sur des groupements locaux de points d'intérêt, combinant ainsi des informations relatives au signal et à la géométrie des images. Les étapes de l'algorithme sont les suivantes :

1. À partir d'un point sélectionné au hasard dans la première image, des hypothèses d'appariement sont générées sur la base d'une mesure de similarité entre les vecteurs propres de la matrice d'auto-corrélation du détecteur de Harris
2. Des groupements locaux de points sont déterminés autour du point considéré et des appariements potentiels, en considérant les n voisins les plus proches¹.
3. Des hypothèses d'appariement entre les voisins sont générées avec la même mesure de similarité qu'en 1, et une estimation de la transformation affine entre les groupes considérés est déterminée. La transformation qui produit la plus grande répétabilité est confirmée par une mesure de corrélation entre les points.
4. Une fois qu'une hypothèse d'appariement entre deux groupes est confirmée, les étapes 1 à 3 sont réitérées, en commençant par les points les plus proches du groupement considéré. La transformation affine estimée pour le premier appariement de groupes est exploitée pour focaliser la recherche de nouveaux appariements.

Le processus stoppe lorsque plus aucun appariement ne peut être déterminé. La figure D.1 présente un résultat de l'algorithme sur une paire d'images aériennes.

¹ $n = 6$ dans l'implémentation de l'algorithme.

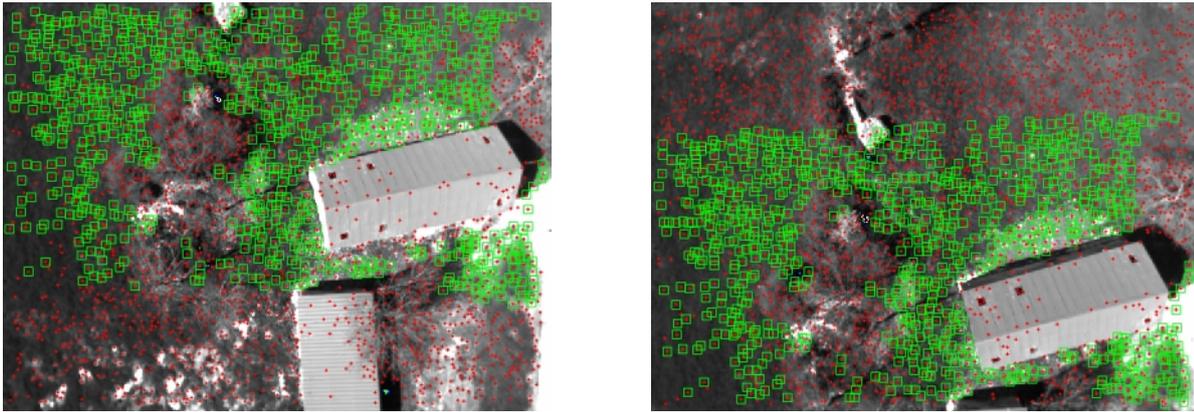


FIG. D.1: *Résultat de l'algorithme de mise en correspondance de points. Les croix rouges sont l'ensemble des points de Harris détectés, ceux qui sont encadrés de vert ont été appariés.*

Mise en correspondance de segments

La détection des segments est un processus très instable, la longueur et le nombre de segments détectés variant considérablement d'une image à l'autre, même pour de très faibles changements de points de vue : il est donc délicat de les mettre en correspondance. Nous avons donc développé un algorithme de mise en correspondance de segments qui se base sur l'algorithme de mise en correspondance de points, ces derniers étant plus stables. Le principe est d'associer aux segments détectés les points les plus proches, et d'exploiter les points mis en correspondance pour générer et vérifier des hypothèses de mise en correspondance de segments. La figure D.2 montre un résultat de mise en correspondance de segments.

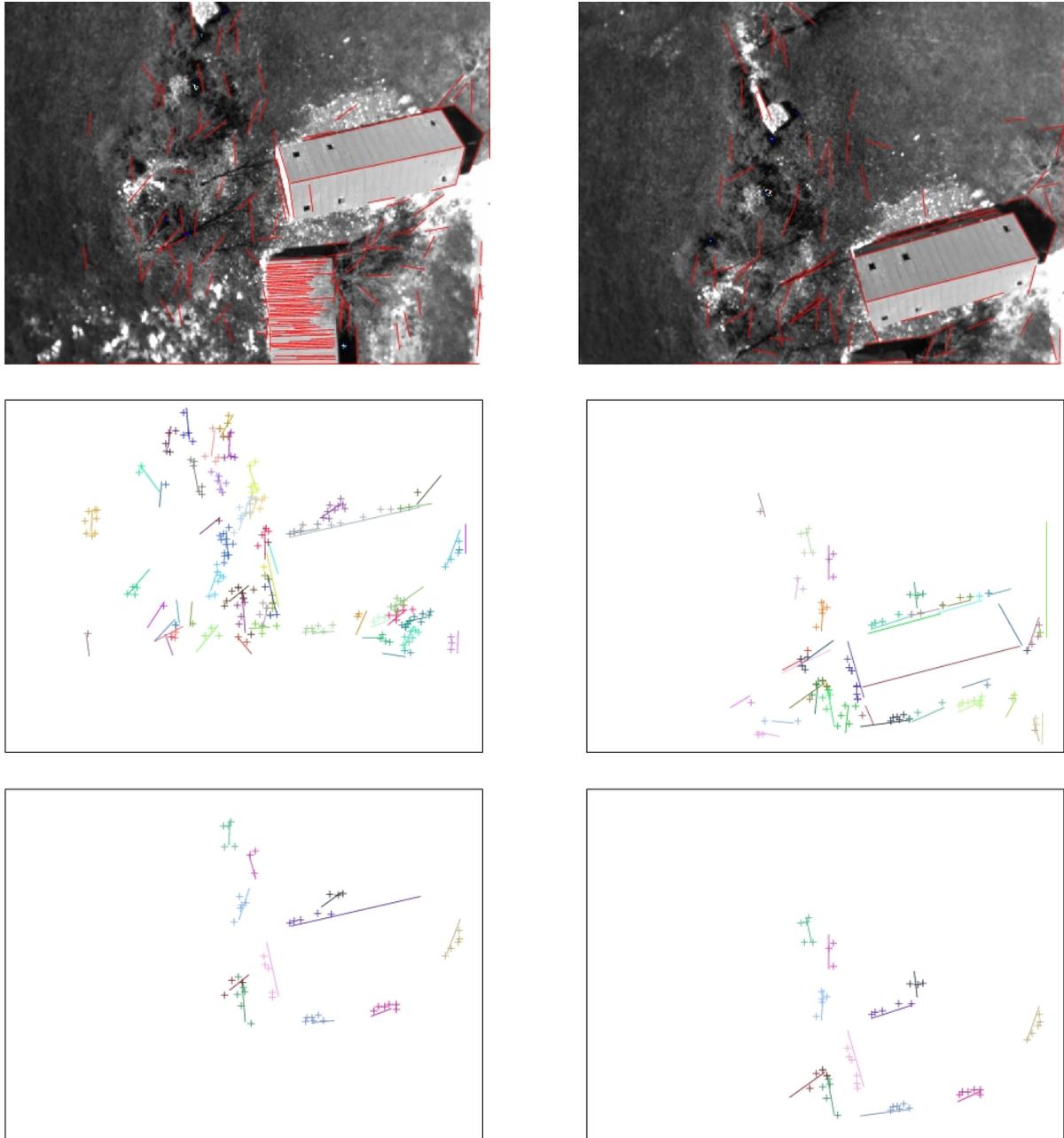


FIG. D.2: Résultat de l'algorithme de mise en correspondance de segments. De haut en bas : segments détectés dans les images, segments associés aux points qui ont été mis en correspondance (figure D.1), et segments mis en correspondance. Les couleurs indiquent les segments associés, et on peut voir une erreur d'appariement (segments noirs).

Bibliographie

- [Baillard 99] C Baillard & A. Zimmerman. *Automatic Reconstruction of Piecewise Planar Models from Multiple Views*. In Proceedings of the Conference on Computer Vision and Pattern Recognition, pages 559–565, June 1999. 26
- [Bartoli 01] Adrien Bartoli. *Piecewise Planar Segmentation for Automatic Scene Modeling*. In In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, volume II, pages 283–289, Dec 2001. Hawaii, USA. 15
- [Bartoli 03] Adrien Bartoli. *Reconstruction et alignement en vision 3D : points, droites, plans et caméras*. PhD thesis, INPG - Institut National Polytechnique de Grenoble, Grenoble, France, September 2003. 15, 18
- [Benhimane 04] S. Benhimane & E. Malis. *Real-time image-based tracking of planes using efficient second-order minimization*. IEEE/RSJ International Conference on Intelligent Robots Systems in Sendai, Japan, October 2004. 19
- [Besnerais 05] Guy Le Besnerais & Frédéric Champagnat. *Dense optical flow by iterative local window registration*. In ICIP (1), pages 137–140, 2005. 16
- [Caballero 07] F. Caballero, L. Merino, J. Ferruz & A. Ollero. *Homography Based Kalman Filter for Mosaic Building. Applications to UAV position estimation*. In Proceedings of the IEEE International Conference on Robotics and Automation. IEEE, April 2007. 90
- [Chaimowicz 05] L. Chaimowicz, A. Cowley, D. Gomez-Ibanez B. Grocholsky, M. Hsieh, H. Hsu, J. Keller, V. Kumar, R. Swaminathan & C. Taylor. *Deploying air-ground multi-robot teams in urban environments*. In International Workshop on Multi-Robot Systems, Washington, DC (USA), 2005. 2
- [Faugeras 93a] Faugeras, Vieville, Theron, Vuillemin, Hotz, Zhang, Moll, Bertin, Mathieu & Fua. *Real-time correlation-based stereo : algorithm, im-*

- plementations and applications*. Rapport de recherche INRIA RR-2013, vol. 2013, 1993. 93
- [Faugeras 93b] Olivier Faugeras. Three-dimensional computer vision, a geometric viewpoint. MIT Press, 1993. 12, 14
- [Fischler 87] Martin A. Fischler & Robert C. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography, pages 726–740. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987. 22
- [Fisher 28] R. A. Fisher & L. H. Tippett. *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. In Proceedings of the Cambridge Philosophical Society, numéro 24, pages 180–190, 1928. 51
- [Frueh 01] Christian Frueh & Avidesh Zakhor. *3D Model Generation for Cities Using Aerial Photographs and Ground Level Laser Scans*. In IEEE Conference on Computer Vision and Pattern Recognition, Kauai (USA), pages 31–38, 2001. 2
- [Garcia-Padro 01] P. Garcia-Padro, G. Sukhatme & J. Montgomery. *Towards vision-based safe landing for an autonomous helicopte*. Robotics and Autonomous Systems, vol. 38, no. 1, pages 19–29, 2001. 16
- [Gee 06] A.P. Gee & W. Mayol Cuevas. *Real-Time Model-Based SLAM Using Line Segments*. In International Symposium on Visual Computing, pages II : 354–363, 2006. 80
- [Gelgon 97] Marc Gelgon & Patrick Bouthemy. *A region-level graph labeling approach to motion-based segmentation*. In CVPR '97 : Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97), page 514, Washington, DC, USA, 1997. IEEE Computer Society. 16
- [Gumbel 35] E.J. Gumbel. *Les valeurs extrêmes des distributions statistiques*. In Annales de l'institut Henri Poincaré, volume 5, pages 115–158. 1935. 51
- [Gurdjos 03] P. Gurdjos & P.F. Sturm. *Methods and geometry for plane-based self-calibration*. In CVPR03, pages I : 491–496, 2003. 15
- [Harbour 05] J. Harbour, S. Bauer, D. Bruemmer, D. Carroll, E. Pacis, K.D. Mullens & H.R. Everett. *Enabling technologies for unmanned protection systems*. In SPIE Proc. 5804 : Unmanned Ground Vehicle Technology VII, Orlando, FL (USA), 2005. 2
- [Hartley 97] Richard Hartley. *In Defense of the Eight-Point Algorithm*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 6, pages 580–593, 1997. 18

- [Hartley 04] Richard Hartley & A. Zisserman. Multiple view geometry in computer vision. Cambridge University Press, ISBN : 0521540518, second edition, 2004. 5, 12, 14, 17, 18, 19, 45, 86
- [Hettmansperger 92] TP Hettmansperger & SJ Sheather. *A cautionary note on the method of least median squares*. American Statistician, 1992. 23
- [Horn 90] Berthold K.P. Horn. *Recovering Baseline and Orientation from Essential Matrix*. MIT Courses, January 1990. 14
- [Hough 59] P.V.C. Hough. *Machine analysis of bubble chamber pictures*. In 554-556, editeur, International Conference on High Energy Accelerators and Instrumentation. CERN, 1959. 22
- [Huber 81] P.J. Huber. Robust statistics. Wiley, New York, 1981. 20
- [Huber 00] D. Huber, O. Carmichael & M. Hebert. *3D Map Reconstruction from Range Data*. In IEEE International Conference on Robotics and Automation, San Francisco, Ca (USA), pages 891–897. IEEE, 2000. 78
- [JFR 06] *Special issues on the DARPA Grand Challenge 2005, Parts I and II*. Journal of Field Robotics, Vol. 23, Num. 8/9, 2006. 1
- [Jung 01] Il-Kyun Jung & Simon Lacroix. *A robust Interest Point Matching Algorithm*. International Conference on Computer Vision, July 2001. 19, 101
- [Kanatani 98] Kenichi Kanatani. *Geometric Information Criterion for Model Selection*. Int. J. Comput. Vision, vol. 26, no. 3, pages 171–189, 1998. 15, 78
- [Kweon 92] I.S. Kweon & T. Kanade. *High-Resolution Terrain Map from Multiple Sensor Data*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pages 278–292, Feb. 1992. 78
- [Lacroix 05] S. Lacroix. *Air / Ground Robotics Ensembles for Risky Applications*. Invited talk at the 3rd International Symposium on MEXT DDT Rescue Robot project, Kobe (Japan), Jan. 2005. 2
- [Lacroix 06] S. Lacroix, S. Joyeux, T. Lemaire, S. Bosch, P. Fabiani, C. Tessier, O. Bonnet, D. Dufourd & E. Moline. *Projet Acrobat : Algorithmes pour la coopération entre robots terrestres et aériens*. In Quatrièmes journées du programme Robea, Paris (France), Mars 2006. 2
- [Lemaire 07] Thomas Lemaire & Simon Lacroix. *Monocular-vision based SLAM using line segments*. In IEEE International Conference on Robotics and Automation, April 2007. 80
- [Lourakis 02] M.I.A. Lourakis, A.A. Argyros & S.C. Orphanoudakis. *Detecting Planes In An Uncalibrated Image Pair*. In BMVC02, page Poster Session, 2002. 15

- [Lowe 04] D. Lowe. *Distinctive Features from Scale-Invariant keypoints*. International Journal on Computer Vision, vol. 60, no. 2, pages 91–110, 2004. 88
- [Luong 96] Q.T. Luong & O.D. Faugeras. *The Fundamental Matrix : Theory, Algorithms, and Stability Analysis*. International Journal of Computer Vision, vol. 17, no. 1, pages 43–75, January 1996. 78
- [Madsen 04] K. Madsen, H. B. Nielsen & O. Tingleff. *Methods for Non-Linear Least Squares Problems (2nd ed.)*, 2004. 19
- [Malis 05] E. Malis & E. Marchand. *Méthodes robustes d'estimation pour la vision robotique*. In Journées nationales de la recherche en robotique, JNRR'05, Guidel, France, October 2005. 19, 20
- [McDonnell 81] M. J. McDonnell. *Box-filtering techniques*. Computer Graphics and Image Processing, vol. 17, pages 65–70, 1981. 93
- [Mullens 04] K. Mullens, B. Troyer, R. Wade, B. Skibba & M. Dunn. *Experiments in multirobot air-ground coordination*. In IEEE International Conference on Robotics and Automation, New Orleans, LA (USA), 2004. 2
- [Mullens 06] K. Mullens. *Collaborative engagement experiment*. In SPIE Proc. 6230 : Unmanned Systems Technology VIII, Defense Security Symposium, Orlando, FL, USA, 2006. 2
- [Odobez 97] J. Odobez & P. Bouthemy. Separation of moving regions from background in an image sequence acquired with a mobile camera. 1997. 16
- [Oriot 03] Hélène Oriot. *Extraction de bâtiments de formes complexes à partir de couples d'images aériennes stéréoscopiques*. Rapport technique RT 1/06761 DTIM, Onera, 2003. 47, 49
- [PEA 07] *PEA Action : Etude de coopération de multivéhicules hétérogènes*. <http://action.onera.fr/>, 2007. 2
- [Rey 83] W. J. Rey. Introduction to robust and quasi-robust statistical methods. Springer-Verlag, Berlin, 1983. 21
- [Rousseeuw 84] PJ Rousseeuw. *Least Median of Squares Regression*. American Statistics Associated, vol. 79, pages 871–880, 1984. 19
- [Rousseeuw 87] PJ Rousseeuw. Robust regression and outlier detection. John Wiley & Sons, Inc., New York, NY, USA, 1987. 19, 20, 23
- [Sanfourche 05] Martial Sanfourche. *Traitement de séquences d'images pour l'estimation jointe de la structure et du mouvement. Application à l'imagerie aérienne*. PhD thesis, Université de Cergy-Pontoise, 2005. 78

-
- [Silveira 06] Geraldo Silveira, Ezio Malis & Patrick Rives. *Real-time Robust Detection of Planar Regions in a Pair of Images*. In IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 49–54, China, 2006. 90
- [Sturm 96] P. Sturm. *Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction*. Rapport technique, 1996. 15, 78
- [Szeliski 98] Richard Szeliski & P. H. S. Torr. *Geometrically Constrained Structure from Motion : Points on Planes*. European Workshop on 3D Structure from Multiple Images of Large-Scale Environments (SMILE), pages 171–186, June 1998. 15
- [Torr 98] P. H. S. Torr, A. Zisserman & S. J. Maybank. *Robust Detection of Degenerate Configurations while Estimating the Fundamental Matrix*. Computer Vision and Image Understanding : CVIU, vol. 71, no. 3, pages 312–333, 1998. 15, 78
- [Triggs 98] Bill Triggs. *Autocalibration from Planar Scenes*. In Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany, 1998. 15, 16, 90, 97
- [Triggs 00] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley & Andrew W. Fitzgibbon. *Bundle Adjustment - A Modern Synthesis*. In ICCV '99 : Proceedings of the International Workshop on Vision Algorithms, pages 298–372, London, UK, 2000. Springer-Verlag. 19
- [Tsai 81] Robert Y. Tsai & Thomas S. Huang. *Estimating three-dimensional motion parameters of a rigid planar patch, II : singular value decomposition*. IEEE Trans. Acoustic, Speech and Signal Processing, vol. 29, no. 6, pages 1147–1152, 1981. 15
- [Tsai 82] Robert Y. Tsai, Thomas S. Huang & Wei-Lee Zhu. *Estimating three-dimensional motion parameters of a rigid planar patch, II : singular value decomposition*. IEEE Transactions on Acoustic, Speech and Signal Processing, vol. 30, no. 4, pages 525–524, 1982. 15
- [Tsai 84] Robert Y. Tsai & Thomas S. Huang. *Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, January 1984. 8
- [Vandapel 06] N. Vandapel, R.R. Donamukkala & M. Hebert. *Unmanned Ground Vehicle Navigation Using Aerial Ladar Data*. International Journal of Robotics Research, vol. 25, no. 1, pages 31–51, Jan. 2006. 78

- [Vincent 01] E. Vincent & R. Laganière. *Detecting Planar Homographies in an Image Pair*. In IEEE International Symposium on Image and Signal Processing and Analysis, pages 182–187, Croatia, 2001. 15
- [Zhang 95a] Zhengyou Zhang. *Estimating Motion and Structure from Correspondences of Line Segments between Two Perspective Images*. Pattern Analysis and Machine Intelligence, vol. 17, no. 12, pages 1129–1139, December 1995. 80
- [Zhang 95b] Zhengyou Zhang. *Parameter Estimation Techniques : A Tutorial with Application to Conic Fitting*. Rapport technique Rapport de recherche INRIA RR-2676, INRIA Sophia Antipolis, 1995. 20, 21, 41
- [Zhu 01] Z. Zhu, E.M. Riseman & A. R. Hanson. *Parallel-perspective stereo mosaics*. In Eighth IEEE International Conference on Computer Vision, Vancouver (Canada), pages 345–352, July 2001. 78

Abstract :

This thesis mainly deal with planar areas detection from monocular images and low-budget sensors. These works allow safe landing area detection for UAV and ground/air coopération with traversability maps fusion.

In using homographies' properties, we segment the regions of images that are the projections of a plane surface. An extensive study has been made on robust linear estimators in a context of homography estimation from matched points. Synthesis datas were used in order to highlight the influence of the parameters of the environment on the quality of estimated homographies. These results were confronted thereafter with real data.

These techniques allow the segmentation of matched points belonging to a plane zone. In order to produce a dense description (continues) of planar zones, we proposed an improvement of the Zero-mean Normal Cross Correlation score. By considering an adaptive model of the influence of the luminosity variances on the scores of correlation, we introduced an automatic segmentation method of planar zones which offer good results.

The introduction of probabilistic models enabled us to fuze iteratively the observations, and to build local grids which represent the probability of flatness on zones.

Finally with the aim of the co-operation between air and terrestrial robots, we extended this work to the context of terrestrial robotics and the fusion of aéro-terrestrial models.

Keywords :

homography, monocular vision, traversability map, ZNCC, aéro-terrestrial

Auteur : Sébastien BOSCH

Titre : Contribution à la modélisation d'environnements par vision monoculaire dans un contexte de robotique aéro-terrestre

Directeur de thèse : Simon LACROIX

Lieu et date de soutenance : Salle Europe - LAAS/CNRS, le 1 octobre 2007

Résumé :

Les travaux de cette thèse s'articulent principalement autour de la détection de surfaces planes à partir de séquences d'images faiblement calibrées. La détection de telles surfaces offre des possibilités pour l'atterrissage autonome de drones, la coopération air/sol par la fusion de modèles de traversabilité ou encore la cartographie aérienne.

Considérant d'abord le contexte d'images aériennes faiblement localisées (GPS métrique, centrale inertielle bas coût...), nous exploitons les propriétés des homographies, qui définissent le déplacement entre deux images de points appartenant à un même plan du monde. Afin de segmenter les régions des images contenant la projection d'un plan, nous avons évalué différents algorithmes d'estimation linéaire robuste dans le cadre de l'estimation d'homographie à partir de points appariés.

Ces techniques permettent de segmenter les points appariés selon qu'ils appartiennent ou non à une zone plane. Afin de produire une description dense (continue) des zones détectées, en considérant un modèle adaptatif de l'influence des variances des niveaux de gris sur les scores de corrélation normalisée croisée, nous avons introduit une méthode de segmentation automatique de zones planes offrant de bons résultats.

L'introduction de modèles probabilistes nous a permis de fusionner les observations au fur et à mesure du déplacement de la caméra, et de construire des grilles locales qui représentent la probabilité de planéité sur des zones.

Enfin dans l'optique de la coopération entre robots aériens et terrestres, nous avons étendu ces travaux au contexte de la robotique terrestre et de la fusion de modèles aéro-terrestres.

Mots-Clefs : homographie, vision monoculaire, traversabilité, ZNCC, robotique, aéro-terrestre

École Doctorale : Système

Spécialité : Informatique

Laboratoire d'Analyse et d'Architecture des Systèmes du CNRS

7, Avenue du Colonel Roche

31077 Toulouse Cedex 4