



HAL
open science

Reconstruction 3D infrarouge par perception active

Benoit Ducarouge

► **To cite this version:**

Benoit Ducarouge. Reconstruction 3D infrarouge par perception active. Informatique et langage [cs.CL]. INSA de Toulouse, 2011. Français. NNT : 2011ISAT0042 . tel-00681977

HAL Id: tel-00681977

<https://theses.hal.science/tel-00681977>

Submitted on 23 Mar 2012

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

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



THESE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *INSA de Toulouse*

Discipline ou spécialité : *Systèmes Informatiques*

Présentée et soutenue par *Benoît DUCAROUGE*

Le 26 Septembre 2011

Titre : *Reconstruction 3D infrarouge par perception active.*

JURY

Frédéric DEVERNAY, Chargé de Recherche INRIA, LIG, Grenoble
Michel DEVY, Directeur de Recherche CNRS, LAAS-CNRS, Toulouse
Jean-Yves FOURNIOLS, Professeur, LAAS-CNRS, INSA de Toulouse
Fabrice MERIAUDEAU, Professeur, Université de Bourgogne, Dijon
Thierry SENTENAC, Enseignant-Chercheur, ICA, EMAC, Albi
Peter STURM, Directeur de Recherche INRIA, Inrialpes, Grenoble

Ecole doctorale : *EDSYS*

Unité de recherche : *LAAS-CNRS et ICA-Albi*

Directeur(s) de Thèse : *Michel Devy, Thierry Sentenac*

Rapporteurs : *Fabrice Mériaudeau, Peter Sturm*

Remerciements

Tout d'abord je tiens à remercier l'homme sans qui tout cela n'aurait pu avoir lieu, celui qui m'a toujours poussé, du début à la fin, celui qui a toujours été là quel que soit l'endroit où j'étais, et qui, malgré toutes les péripéties que je lui ai fait subir ces quatre dernières années, a toujours gardé le sourire et sa bonne humeur, il est entre autre mon directeur de thèse : Mr Thierry Sentenac.

Ensuite je tiens à remercier mon autre directeur de thèse, Mr Michel Devy, qui malgré un agenda plus que chargé, a toujours été de bon conseil. Ses aventures, souvent aussi incroyables que les miennes, nous aurons bien fait rigolé, de plus sa bonne humeur et son calme légendaire ont largement contribué aux excellentes conditions de travail dans lesquelles j'étais plongé.

Je tiens à remercier Fabrice Mériaudeau et Peter Sturm qui ont accepté d'être rapporteurs et dont les remarques positives et constructives m'ont encouragé à terminer ce travail plus sereinement. Je remercie également Frédéric Devernay et Jean-Yves Fourniols pour l'intérêt qu'ils ont porté à ce travail en acceptant d'en être les examinateurs.

Mes remerciements vont également vers l'un des hommes de l'ombre de cette thèse, Mr Florian Bugarin, dont l'aide m'a été très précieuse. C'est aussi une personne qui a joué un rôle important dans l'ambiance de travail, il a su apporter une touche humoristique permettant de passer de bonnes journées et malgré ses humeurs ronchonnes, il nous aura bien fait rigolé.

Une autre personne que je remercie fortement, car sans elle je n'aurais jamais postulé pour ce sujet de thèse. Au départ un collègue d'école, puis un colocataire et finalement un collègue de bureau, mais un ami avant tout, Mr Julien Bonnal.

Je remercie aussi le reste de la dream team du bureau, Adrien et Benjamin pour les bon moments passés, les pauses dans le jardin japonais ou à la cafétéria et l'ensemble des conneries que l'on a pu faire dans le bureau.

Un grand merci à la fine équipe de l'école des mines d'Albi, Remy Gilblas pour sa bonne humeur et son humour décalé, Didier et Jean-Michel les techniciens de choc, Thomas et Vanessa pour m'avoir supporté lors des nombreux covoiturages et Jacques Harvent pour son style inimitable et ses aventures tout simplement magiques.

Je remercie évidemment toute ma famille pour son soutien et son aide. Toute la bande de toulousains, des amis qui m'ont accompagné durant ces dernières années et qui m'ont permis de décompresser en dehors du boulot.

Et finalement, the last but not the least, je tiens à remercier celle que je ne remercierai jamais assez, celle sans qui cette thèse aurait pu être plus simple mais beaucoup moins rigolote, celle qui m'a apporté et m'apporte tout ce dont j'ai besoin, celle qui m'a permis de finir ce manuscrit et qui aura vécu cette fin de thèse presque aussi intensément que moi, celle que j'aime plus que tout, Coraline Pascal.

Table des matières

1	Introduction	1
1.1	Contexte applicatif : métrologie thermique sur des objets 3D	2
1.2	Notre problématique	4
1.2.1	Reconstruction 3D en infrarouge	4
1.2.2	Reconstruction 3D active	5
1.3	Notre contribution	6
1.3.1	Auto-calibrage d'un système caméra infrarouge-robot	6
1.3.2	Reconstruction 3D multi-vues en infrarouge	7
1.4	Organisation du document	8
2	Rectification	11
2.1	Rappels et définitions	12
2.1.1	La géométrie épipolaire	12
2.1.2	Définition de la rectification	14
2.1.3	Les homographies compatibles	15
2.2	État de l'art	15
2.2.1	Méthode de type "Hartley"	16
2.2.2	Méthode de type "Loop et Zhang"	17
2.2.3	Méthode de type "Devernay"	18
2.2.4	Méthodes de type "Mallon"	18
2.2.5	Discussions	20
2.3	Contribution à la rectification	21
2.3.1	Méthode directe de rectification	21
2.3.2	Définition de la fonction coût	22
2.3.3	Définition des contraintes	23
2.3.4	Contraintes analytiques	25
2.3.5	Contraintes géométriques	26
2.3.6	Implémentation	29
2.3.7	Conclusion	29
2.4	Résultats et comparaisons	30
2.4.1	Critères d'évaluation	30
2.4.2	Évaluation sur des images stéréoscopiques visibles	30
2.4.3	Évaluation sur des images stéréo infrarouges	34
2.4.4	Les limites de l'algorithme	35
2.5	Conclusions	36
3	Auto-calibrage d'un système robot-caméra infrarouge	37
3.1	Définition et modélisation du système robot-caméra	40
3.1.1	Géométrie du système	40
3.1.2	Transformation pince-caméra	42

3.1.3	Discussions	43
3.2	Etat de l'art sur le calibrage Hand-Eye	43
3.2.1	Les méthodes classiques avec mires	43
3.2.2	Les méthodes sans mire : auto-calibrage	45
3.2.3	Conclusion	47
3.3	Proposition d'une méthode d'auto-calibrage basée sur la minimisation de distances dans l'espace $3D$	49
3.3.1	Principe	49
3.3.2	Expression du point P_c dans le repère de la caméra C	50
3.3.3	Problème à minimiser basé sur la distance entre les points $3D$	52
3.3.4	Conclusion	53
3.4	Proposition d'une méthode d'auto-calibrage basée sur la minimisation de distances dans l'espace $2D$	53
3.4.1	Principe	53
3.4.2	Système monoculaire	54
3.4.3	Système de stéréovision	55
3.4.4	Système de stéréovision rectifié	57
3.4.5	Conclusion	58
3.5	Evaluation des méthodes d'auto-calibrage sur des données de synthèse	58
3.5.1	Méthodologie	59
3.5.2	Contexte expérimental	59
3.5.3	Méthode de minimisation et initialisation	61
3.5.4	Critères d'évaluation	62
3.5.5	Évaluation sur des données de synthèse non bruitées	63
3.5.6	Influence du nombre de points	66
3.5.7	Influence du bruit sur les points des images	69
3.5.8	Influence de la répartition des points	72
3.5.9	Conclusion	73
3.6	Performances des méthodes de calibrage sélectionnées sur des données réelles	73
3.6.1	Contexte expérimental	74
3.6.2	Données issues d'images visibles	74
3.6.3	Données issues d'images infrarouges	75
3.7	Conclusion	76
4	Stéréovision infrarouge	77
4.1	Extraction et appariement	80
4.1.1	Détecteur "Phase Congruency"	80
4.1.2	Détecteur de "Harris"	81
4.1.3	Détecteur "SIFT"	82
4.1.4	Comparaison des détecteurs sur des images infrarouges	84
4.2	Filtrage des faux appariements	86
4.2.1	La méthode des "M-Estimeurs"	88
4.2.2	La méthode de "RANSAC"	88

4.2.3	La méthode "LTS" ou "LMS"	89
4.2.4	Comparaison	89
4.3	Corrélation	91
4.3.1	Principe de la corrélation	91
4.3.2	Méthodes locales	91
4.3.3	Méthodes globales	92
4.3.4	Conclusion	93
4.4	Reconstruction 3D avec des images infrarouges	93
4.5	Conclusion	94
5	Application	97
5.1	Système de perception actif robotisé	98
5.1.1	Robot cartésien six axes	99
5.1.2	Caméras	99
5.1.3	Objets chauffants	101
5.2	Vérification du modèle géométrique des caméras infrarouges	102
5.2.1	Problème d'extraction des points de la mire	102
5.2.2	Validation du modèle géométrique de la camera infrarouge	103
5.3	Reconstruction 3D infrarouge active	105
5.3.1	Étape de rectification	106
5.3.2	Étape d'autocalibrage	107
5.3.3	Corrélation, Triangulation et Recalage	110
5.4	Comparaison avec le modèle de référence	112
5.5	Conclusion	115
6	Conclusion	119
6.1	Bilan des travaux réalisés	119
6.2	Quelques perspectives	121
	Publications	123
A	Annexe 1 : Modélisation d'une caméra et d'un capteur stéréovision	125
A.1	Définition et modélisation du système mono caméra	125
A.2	Définition et modélisation d'un système stéréovision	126
A.2.1	Modèle	126
A.2.2	Matrice essentielle	127
A.2.3	Cas d'un système de stéréovision rectifié	127
	Bibliographie	129

Introduction

Nos travaux de recherche concernent la reconstruction 3D appliquée au contrôle sur la qualité de pièces ou d'objets fabriqués lors de processus industriels. Les exigences de qualité, de suivi et de reproductibilité, dans tous les secteurs industriels, en production ou en maintenance, imposent que le contrôle non destructif (CND) ne soit plus ponctuel et manuel mais qu'il soit accompagné de cartographies bidimensionnelles ou tridimensionnelles de l'objet observé réalisées de manière automatique. A titre d'exemple, les constructeurs aéronautiques sont amenés, afin de réduire leurs coûts, à automatiser leurs outils de contrôle et à s'orienter vers des solutions robotique mobiles et multi-procédés. Nous nous plaçons ici, dans cette perspective du CND automatisé, fondé sur la vision infrarouge ; deux motivations ont conduit à ce sujet d'étude.

D'une part, les caméras sont de plus en plus présentes dans les moyens de contrôle non destructif, et notamment les caméras infrarouges dès lors que des phénomènes thermiques doivent être caractérisés. L'évolution récente de leur technologie en terme de réductions de coûts et de gain en résolution spatiale, grâce à l'introduction de matrices bolométriques non refroidies, donne accès à des images de bonne résolution à un prix abordable. Il s'agit d'un paramètre clé pour de nombreuses applications en contrôle non destructif [Maldague 2001] qui nécessitent des mesures de température, mais cette plus grande accessibilité des caméras infrarouges a aussi donné naissance à de nombreuses applications : détection d'obstacles et de piétons de nuit [Bertozzi 2007], détection et reconnaissance de visages [Eveland 2003], vision nocturne en automobile [Owens 1999] . . .

D'autre part, les robots peuvent répondre au besoin général d'automatisation du contrôle appliqué sur des pièces ou d'objets de grandes dimensions, de forme et de structure de plus en plus complexes. Dans ces cas, une seule image ne suffit pas pour réaliser le contrôle ; il est nécessaire soit de déployer plusieurs capteurs autour de la pièce à contrôler, soit de déplacer un seul capteur selon une trajectoire, qui peut être adaptée en ligne en fonction des observations. Nous parlons alors de contrôle actif. Cela implique

- soit des procédures automatiques classiques en robotique, avec une boucle *observation de l'objet, choix du prochain point de contrôle, planification et exécution du mouvement du capteur*,
- soit des interactions avec un opérateur chargé de piloter un robot en fonction de sa propre expertise, notamment pour détecter des défauts : en ce cas le système peut être programmé pour automatiser un certain nombre de tâches,

comme suivi de surface à distance constante, positionnement du capteur pour que l'axe de vue soit orthogonal à un point de la surface. . .

L'association des deux technologies en un système de CND robotisé infrarouge autorise l'inspection de champs denses de températures sur des pièces de forme complexe pour obtenir une cartographie tridimensionnelle thermique. Ce thème est le sujet central de ma thèse qui est intitulée *Reconstruction tridimensionnelle infrarouge active*.

Nous présentons d'abord le contexte applicatif, lié au projet ANR R3T auquel nous avons participé, avant de détailler notre problématique, puis de présenter le plan du manuscrit.

1.1 Contexte applicatif : métrologie thermique sur des objets 3D

Cette thématique du Contrôle Automatisé sur des pièces 3D de forme complexe, a été traitée dans le cadre d'un projet applicatif appelé R3T, pour *Real Time and True Temperature field measurement based on optoelectronics devices*. Ce projet vise le développement d'un système optronique de mesure en ligne de champs de températures vraies sur un objet de forme tridimensionnelle.

Une caméra infrarouge permet seulement de mesurer des températures apparentes. Afin de transformer ces températures apparentes en températures vraies, il faut considérer le modèle radiométrique de la caméra. Dans ce modèle apparaissent deux facteurs qui sont intrinsèques à la zone de mesure sur l'objet : d'une part l'émissivité, d'autre part la forme 3D. Sur un objet de forme complexe et d'émissivité non uniforme, le calcul de la température vraie en un point donné de la surface de l'objet, requiert donc la fusion des mesures sur la température apparente, sur l'émissivité et sur la forme 3D de l'objet en ce point.

Afin de diminuer la complexité du système global, les informations spatiales, ou la forme de l'objet observé, sont mesurées avec le même système que celui pour la mesure de température apparente, donc avec un capteur stéréovision constitué de deux caméras infrarouges. Les informations d'émissivité sont obtenues par mesure de la réflectométrie de l'objet. Ce sujet est traité dans un autre lot de travail du projet R3T. Nous ne l'avons pas considéré dans nos travaux : néanmoins, cette mesure d'émissivité donne des contraintes

- sur la configuration complète du capteur, puisqu'il faut adjoindre aux caméras deux illuminateurs laser éclairant la surface dans deux longueurs d'onde différentes.
- et sur le positionnement du capteur par rapport à la surface de l'objet, puisque la mesure d'émissivité en un point requiert de connaître les angles d'incidence des rayons optiques et des faisceaux laser, donc cela nécessite la reconstruction locale de la forme 3D de cette surface.

La reconstruction 3D infrarouge a donc pour objectif, d'une part, la connais-

sance de la géométrie de l'objet observé, et d'autre part, la position relative caméra-objet (angle entre la normale à la surface de l'objet et le plan focal de la caméra). A terme, ces informations interviendront également dans la modélisation radiométrique de la scène observée pour déterminer les possibles phénomènes radiatifs entre les faces en regard de l'objet. Notons que pour mesurer la température vraie, le besoin en précision est faible ; nous les estimons à 1%, c'est-à-dire que nous devons pouvoir mesurer à 1cm près la taille d'un objet placé à 1m. du capteur.

Afin de mesurer un champ de températures vraies sur tout un objet, le capteur complet (deux caméras infrarouges, plus les illuminateurs laser) devra être monté sur un système de positionnement pour assurer une reconstruction la plus complète possible. Au stade actuel d'avancement du projet, il n'a pas été possible d'intégrer l'ensemble du capteur :

- nous avons donc uniquement considéré dans nos travaux, la mesure d'un champ de températures apparentes sur un objet de forme quelconque, à partir de caméras infrarouges en bande $8 - 12\mu m$.
- le système de positionnement est un robot. Les aspects robotique (planification et exécution des mouvements, asservissements ...) exploitent seulement les méthodes disponibles sur le robot.

Ce projet vise donc plutôt des applications de CND en vision infrarouge, sur des pièces avant ou en cours d'assemblage, par exemple pour des contrôles effectués dans des procédés sidérurgiques. Les résultats de ces travaux seront également exploités pour des opérations de maintenance : par exemple le projet Pi-cobot (pour *Pre-flight Inspection Collaborative Robot*) en cours de préparation, concerne l'assistance à l'inspection du fuselage d'un aéronef, soit une inspection réglementaire avant décollage sur le tarmac (assistance à un opérateur qui vérifie l'absence d'impacts importants. . .), soit une inspection complète dans un hangar (examen exhaustif du fuselage : suivi de l'évolution de défauts existants, détection de nouveaux défauts...).

Ces projets recouvrent donc cette problématique du diagnostic ou de l'inspection de pièces ou de structures de manière automatique, afin de satisfaire les exigences croissantes de qualité et de traçabilité. La perception ou vision 3D est un des enjeux majeur du diagnostic, car le contrôle de formes et la détection des défauts impliquent une reconstruction 3D automatisée qui soit la plus dense possible tout en minimisant l'intervention humaine dans le processus de contrôle. Les robots, associés aux moyens de contrôle non destructifs, deviennent un élément incontournable pour répondre à ce besoin d'automatisation.

1.2 Notre problématique

1.2.1 Reconstruction 3D en infrarouge

Le premier enjeu pour résoudre un problème de CND impliquant des aspects thermiques, est donc de choisir le système de perception afin de disposer à la fois d'informations spatiale et thermique. Trois techniques sont particulièrement étudiées dans la littérature pour la reconstruction 3D infrarouge. Tout d'abord, la solution la plus immédiate est de procéder à la reconstruction 3D avec une tête de stéréovision fonctionnant dans le visible, puis d'appliquer sur les surfaces reconstruites, une texture en provenance d'une caméra infrarouge [Prakash 2006]. Cette méthode demande au moins trois caméras de technologies différentes. De plus, elle implique de traiter le délicat problème de la fusion des différentes images à l'aide d'un calibrage croisé entre images visible et infrarouge.

Une autre technique employée afin de texturer la surface est de recourir à un éclairage actif de type nappe laser associé aux caméras infrarouges. L'article [Eren 2009] présente des résultats de reconstruction 3D sur un exemple réputé délicat d'objets transparents. Cette technique requiert tout de même un laser fonctionnant en infrarouge lointain, ce qui en pratique, pose de nombreux problèmes technologiques. Dans notre projet R3T, nous utilisons aussi des illuminateurs laser, mais uniquement pour réaliser les mesures d'émissivité; pour l'heure nous exploitons pour ce faire, des illuminateurs ponctuels, dans la bande 1 infrarouge (entre 1 et $1.7\mu m$), adaptés uniquement avec des caméras infrarouges InGaAs également en bande 1.

Enfin, une approche analogue au domaine spectral visible [Orteu 2008] est de recourir à deux caméras infrarouges passives montées en stéréovision [Hajebi 2008]. Cette stratégie a retenu notre attention car elle minimise la complexité du système. Cependant l'article [Hajebi 2008] démontre la difficulté de mettre en correspondance des primitives communes aux deux images pour établir une carte de disparité dense c'est à dire une mise en correspondance pixel à pixel.

Nous proposons *un capteur basé sur un banc de stéréovision infrarouge passif* qui est composé de deux caméras bolométriques miniatures. Les caméras choisies embarquables sur robot fonctionnent en infrarouge lointain, $8 - 12\mu m$ qui permettent d'observer des luminances de température comprise entre 20 et $50^\circ C$. Leurs résolutions sont de 164×128 pixels et leurs distances focales de $11mm$. Ces caractéristiques engendrent évidemment des problèmes de précision pour les prises de vue lointaines, d'où l'intérêt supplémentaire d'un robot pour opérer une reconstruction 3D multivues.

Notons qu'apparaissent des capteurs bolométriques mégapixels; citons une nouvelle caméra réalisée par FLIR, conçue à partir d'un détecteur 1024×768 de ULIS, qui sera évaluée dans notre équipe au LAAS pour un projet de détection d'obstacles

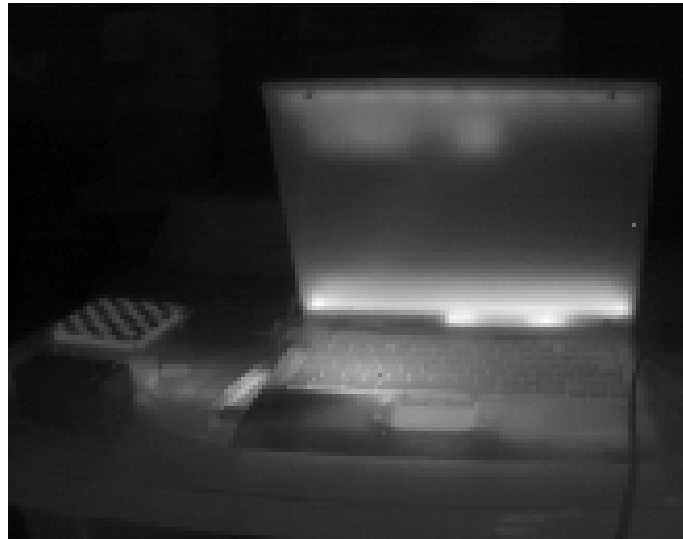


FIG. 1.1 – Exemple d’acquisition à partir d’une caméra bolométrique.

sur un aéronef dans les phases de roulage au sol, dans des conditions de visibilité difficiles.

1.2.2 Reconstruction 3D active

Le second enjeu est le choix du système robotique pour déplacer le capteur autour de l’objet. Nous avons uniquement considéré dans R3T, le contexte de CND sur des objets ou des pièces de petite taille (volume inférieur à $1m^3$). Il existe deux types de système pour réaliser cette tâche : les robots cartésiens et les robots polyarticulés. Les premiers sont dédiés à de l’observation de pièces simples ou peu évolutives et généralement utilisés de façon industrielle pour le contrôle de pièces de formes identiques. Pour l’observation de pièces de formes complexes et évolutives, les robots polyarticulés sont généralement privilégiés.

La mise en place et l’utilisation de tels robots est certes plus polyvalente mais elle est aussi plus complexe et amène de nouvelles contraintes. Pour cette première application, nous avons donc choisi un robot cartésien qui présente l’avantage d’accéder à la position de l’organe terminal de manière simple et rapide à travers la connaissance des positions articulaires fournies et de la cinématique connue du robot. Afin de pouvoir positionner notre capteur, quel que soit la position et l’orientation, nous proposons un robot cartésien à six axes de liberté, trois translations et trois rotations.

Donc finalement, nous proposons un système de mesure actif constitué d’un robot cartésien sur lequel est monté un capteur basé sur un banc de stéréovision infrarouge passif. La figure 1.2 présente ce système.



FIG. 1.2 – Présentation du système de mesure actif.

1.3 Notre contribution

1.3.1 Auto-calibrage d'un système caméra infrarouge-robot

La reconstruction tridimensionnelle d'objets à partir d'images prises par des caméras s'appuie sur un calibrage des caméras. Cette opération consiste à établir la relation entre les coordonnées d'un point dans le repère de l'image et ses coordonnées dans le repère du monde. Généralement, cette étape est conduite hors ligne à l'aide d'une mire. Dans des conditions sévères d'utilisation sur des procédés thermiques, il est difficile de demander à un opérateur de manipuler des mires de calibrage devant le système de vision. De plus, les applications dynamiques requièrent en général fréquemment une mise à jour des paramètres de calibrage. Le succès de l'introduction de tels systèmes de mesure dans le monde industriel implique de simplifier cette phase de calibrage hors ligne et même de pouvoir s'en affranchir.

Nous nous sommes orientés sur l'hypothèse que les caméras sont non calibrées au préalable et donc la reconstruction sera une reconstruction euclidienne non calibrée basée sur des méthodes d'autocalibrage. L'autocalibrage signifie d'une part que le calibrage du système de vision repose uniquement sur des images d'objets inconnus et non sur des mires et d'autre part que cette opération est réalisée en ligne. Les images acquises sont donc utilisées simultanément pour la reconstruction et pour le calibrage du système de vision.

L'autocalibrage de systèmes multivues mono caméra [Sturm 1997], [Hartley 1994], [Maybank 1992], ou d'un système de stéréovision passif [Devernay 1997], [Zhang 1993], [Horaud 1998] est bien connue. Cependant, aucune de ces méthodes n'a été appliquée avec des caméras infrarouges. En effet l'extraction d'amers ou de primitives visuelles est imprécise et beaucoup plus

difficile sur des images infrarouges. Les algorithmes doivent donc être repensés pour fonctionner de manière optimale avec des images infrarouges.

De même, le calibrage du système actif caméra-robot est généralement proposée sur la base de caméras préalablement calibrées fonctionnant dans le domaine spectral visible. Ce calibrage robot-caméra consiste à identifier les paramètres de la caméra et ceux de la transformation entre la caméra ("eye") et l'organe terminal du robot ("hand"). Celle-ci représente une inconnue supplémentaire qui est modélisée par une matrice composée d'une rotation et d'une translation. De nombreux articles [Tsai 1988], [R.Horaud 1995], [Daniilidis 1996]... traitent de ce problème de calibrage pince-caméra ou Hand-Eye. Ces approches supposent qu'une ou plusieurs caméras du système sont déjà préalablement calibrées et que le déplacement de la pince est connu. Le principe consiste alors à résoudre un système de type $AX = XB$ ayant comme données d'entrées : A , déplacement de la caméra autour de la mire (donné par le calibrage de la caméra) et B , déplacement de la pince (supposé connu) ; et comme inconnue, X , transformation rigide entre la pince et la caméra.

Avec des caméras visibles non calibrées, les méthodes pour calibrer simultanément les paramètres du système de vision et les paramètres de la transformation pince-caméra sans l'utilisation d'une mire sont plus rares. Les articles de Schmidt [Schmidt 2005] et Andreff [Andreff 1999] proposent des méthodes basées sur des algorithmes de *Structure from motion* permettant, dans un premier temps, de calibrer le système de vision et d'obtenir la position de la caméra, puis, dans un second temps, les paramètres du système pince-caméra sont calculés de façon classique. Jordt [Jordt 2009] et Wei [qing Wei 1998], quant à eux, proposent une approche basée sur un algorithme de *Bundle adjustment*, permettant de retrouver l'ensemble des paramètres en minimisant les reprojections entre les points $3D$ observés.

Avec des caméras infrarouges non calibrées qui fournissent des images peu texturées de faible résolution spatiale, il n'existe, à notre connaissance, aucune méthode adaptée qui consiste à calibrer directement le système complet à partir d'images avec un faible nombre de points et d'un suivi difficile de ces points sur plusieurs images.

1.3.2 Reconstruction $3D$ multi-vues en infrarouge

Le problème posé dans cette thèse est donc la reconstruction $3D$ d'objets, entièrement automatisée, à partir de plusieurs vues issues d'un banc de stéréovision infrarouge non calibré. La première problématique majeure rencontrée, est celle du calibrage du système robot-caméra complet, paramètres intrinsèques des caméras et paramètres de la transformation pince-caméra, avec des caméras infrarouges, non préalablement calibrées, de manière "automatique" sans avoir recours à des mires.

Deux méthodes différentes seront proposées pour résoudre ce problème. La première méthode est basée uniquement sur l'objet observé et plus précisément sur sa géométrie $3D$. La seconde est basée sur des propriétés internes du banc de

stéréovision, soit, sur une décomposition de la matrice fondamentale. Chacune de ses méthodes traite le problème dans son ensemble et optimise en une seule opération les paramètres intrinsèques des caméras et les paramètres de la transformation pince-caméra. Le calibrage complet du système robot-caméra comportant une vingtaine de paramètres à estimer, nous présentons tout l'intérêt de supposer au préalable que les caméras soient rectifiées pour diminuer ce nombre de paramètres.

Apparaît alors une deuxième problématique, la rectification d'images peu texturées et de faible résolution spatiale. En effet, l'utilisation de caméras infrarouges, implique que l'étape de rectification fonctionne avec un faible nombre de points extraits entachés d'erreur. Cette contrainte nous a conduit à proposer une nouvelle approche de calcul des matrices d'homographies de la rectification. En effet, afin de minimiser la propagation d'erreur, nous avons intérêt d'une part à ne pas procéder à un calcul préalable de la matrice fondamentale pour le calcul des homographies. Et d'autre part de minimiser les étapes de calcul, c'est à dire de regrouper l'étape de décomposition en transformations projectives et l'étape de minimisation des déformations projectives dans une même opération d'optimisation globale. Nous avons proposée une méthode de rectification permettant d'obtenir une paire d'homographies qui réduit les déformations des images, en une seule étape, en minimisant une fonction coût sous contraintes.

L'objectif final de cette thèse étant la reconstruction $3D$ d'objet, nous avons dû aussi analyser l'ensemble des étapes classiques de la reconstruction tridimensionnelle, telles que l'appariement de points d'intérêts, le filtrage de faux appariements ou encore la corrélation. Ces autres étapes ont donc été étudiées en prenant en compte les spécificités et la moindre qualité des images infrarouges pour déterminer les algorithmes les plus adaptés à ce type de caméra. Le recalage de modèles $3D$ partiels est traité avec un logiciel commercial.

1.4 Organisation du document

La problématique abordée dans cette thèse et nos contributions ayant été présentées, voici comment s'organise ce mémoire.

Le chapitre 2 propose une méthode de rectification non calibrée originale. Après un rappel des notions de géométrie projective nécessaires à la compréhension du chapitre, un état de l'art des différentes techniques de rectification non calibrées est présenté. Notre méthode est ensuite détaillée et comparée à celles de l'état de l'art.

Dans le chapitre 3, l'autocalibrage et le calcul des paramètres hand-eye sont abordés. Tout d'abord, nous présentons les méthodes existantes puis nous exposons l'une de nos contributions qui consiste à autocalibrer le banc stéréoscopique et les paramètres "hand-eye" du système caméra-robot en se basant sur le déplacement connu d'un robot cartésien.

Les différents concepts de la stéréovision infrarouge sont étudiés et détaillés dans le chapitre 4. Les spécificités de l'extraction de points d'intérêt et de la corrélation ainsi que leurs applications aux images infrarouges, sont présentées.

Le chapitre 5 clos ce mémoire avec une présentation des moyens expérimentaux et il expose les résultats généraux obtenus lors de plusieurs expérimentations dans différentes configurations.

Finalement, nous conclurons ce mémoire en rappelant les principales contributions obtenues et en présentant quelques perspectives à ces travaux.

Rectification

Sommaire

2.1	Rappels et définitions	12
2.1.1	La géométrie épipolaire	12
2.1.2	Définition de la rectification	14
2.1.3	Les homographies compatibles	15
2.2	État de l'art	15
2.2.1	Méthode de type "Hartley"	16
2.2.2	Méthode de type "Loop et Zhang"	17
2.2.3	Méthode de type "Devernay"	18
2.2.4	Méthodes de type "Mallon"	18
2.2.5	Discussions	20
2.3	Contribution à la rectification	21
2.3.1	Méthode directe de rectification	21
2.3.2	Définition de la fonction coût	22
2.3.3	Définition des contraintes	23
2.3.4	Contraintes analytiques	25
2.3.5	Contraintes géométriques	26
2.3.6	Implémentation	29
2.3.7	Conclusion	29
2.4	Résultats et comparaisons	30
2.4.1	Critères d'évaluation	30
2.4.2	Évaluation sur des images stéréoscopiques visibles	30
2.4.3	Évaluation sur des images stéréo infrarouges	34
2.4.4	Les limites de l'algorithme	35
2.5	Conclusions	36

Dans un contexte de reconstruction tridimensionnelle ($3D$) non calibrée réalisée avec un robot six axes et des caméras infrarouges, nous montrerons que la rectification préalable des images est une étape essentielle afin de "simplifier" la géométrie projective des méthodes de reconstruction $3D$ et d'autocalibrage présentées dans les chapitres suivants. En effet, la rectification simplifie l'appariement de la reconstruction $3D$ à partir d'images stéréo, car la recherche d'un stéréo-correspondant est alors sur la même ligne. Pour la méthode d'autocalibrage d'un banc stéréo infrarouge, uniquement à partir de la connaissance du mouvement du robot, la rectification diminue le nombre de paramètres du système de calibrage.

La plupart des méthodes proposées pour la rectification d'images stéréo non calibrées sont basées sur l'estimation préliminaire de la matrice Fondamentale et sur au moins deux étapes pour la détermination de l'ensemble des paramètres. Cependant, travailler avec des images infrarouges entraîne certaines contraintes comme une très faible résolution d'image, peu de texture, ou encore l'extraction d'un faible nombre de points. L'estimation de la matrice fondamentale avec des images infrarouges est alors entachée d'erreurs qui se propagent ensuite dans les différents calculs. De plus, comme la résolution de l'image est faible, de minimes déformations projectives entraînent d'importantes distorsions dans l'image qui compliquent l'étape de corrélation. Ces méthodes, avec une succession d'étapes distinctes appliquées aux images infrarouges, induisent donc des erreurs sur les paramètres de rectification ainsi que de fortes déformations projectives.

Ainsi, notre apport consiste à proposer une méthode qui ne nécessite qu'un faible nombre de points, qui minimise les déformations projectives et qui n'est pas basée sur une estimation préalable de la matrice Fondamentale. Cette méthode, dite directe, permet donc de minimiser la propagation d'erreurs en supprimant les étapes de calculs intermédiaires.

Ce chapitre s'articule comme suit : après un rappel sur la géométrie projective, nous présentons les bases du problème de la rectification et son état de l'art. Ensuite, nous détaillons notre approche d'estimation des homographies de rectification projective. Tout d'abord, nous exposons la fonction coût qui s'affranchit de l'estimation de la matrice Fondamentale tout en conservant les informations sur la géométrie épipolaire. L'espace des contraintes à imposer à l'espace des inconnues est ensuite explicité afin de contrôler les distorsions projectives de l'image. De même, la méthode d'optimisation sous contraintes retenue est présentée. Enfin, nous comparons notre approche avec les méthodes existantes afin de mettre en évidence les limites et les points forts de notre algorithme.

2.1 Rappels et définitions

2.1.1 La géométrie épipolaire

Notre approche étant basée sur des propriétés de l'espace projectif, il est nécessaire de rappeler les définitions de la géométrie épipolaire et les différentes propriétés de matrice Fondamentale.

Considérons un banc stéréoscopique composé de deux caméras de centres optiques C et C' . Par convention, on appelle I (resp. I') l'image issue de la caméra gauche (resp. droite). Comme le montre la figure 2.1, M un point 3D de la scène observée simultanément par les deux caméras, se projette en deux pixels m et m' dans les images I et I' . Considérons le pixel m , le lieu géométrique où se situe son stéréo-correspondant est une droite de I' , appelée ligne épipolaire et notée l'_{ep} . Cette droite est définie par l'intersection du plan défini par (C, C', M) et du plan support de I' . Lorsque m décrit l'image I , les lignes épipolaires de l'image I' forment un faisceau dont le point d'intersection, noté e' , est appelé épipôle droit. Ce point peut aussi être défini comme l'intersection de la droite définie par les points C, C' et du plan support de l'image I' . L'épipôle gauche e est défini de la même manière.

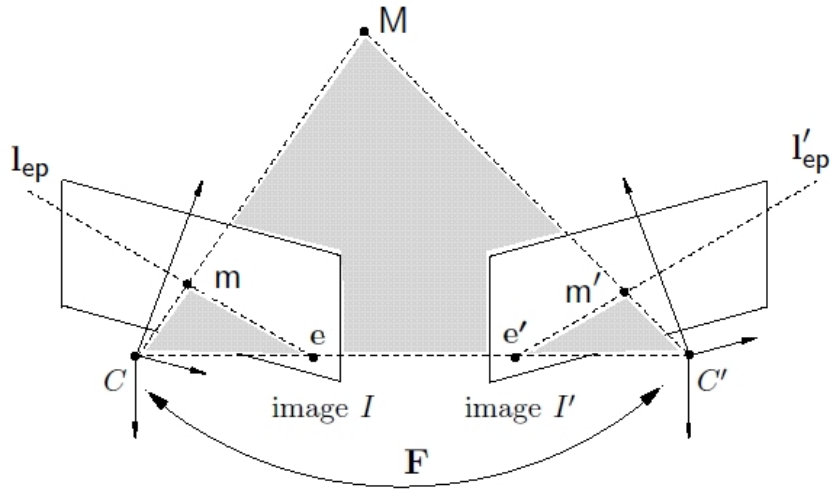


FIG. 2.1 – Géométrie épipolaire.

On définit ensuite la matrice Fondamentale \mathbf{F} comme l'application de l'espace projectif, qui à tout point m de l'image I associe la ligne épipolaire l'_{ep} . Dans ce cas, il est aussi possible de montrer que \mathbf{F}^\top est l'application qui à tout point m' de l'image I' associe la ligne épipolaire gauche l_{ep} . Ainsi l'_{ep} et l_{ep} ont pour équation :

$$l'_{ep} = \mathbf{F}m' \quad (2.1)$$

$$l_{ep} = \mathbf{F}^\top m \quad (2.2)$$

Par définition, le point $m = (u, v, 1)^\top$ appartient à la ligne épipolaire droite l'_{ep} dans l'image I' . De même, le point $m' = (u', v', 1)^\top$ appartient à la ligne épipolaire gauche l_{ep} dans l'image I . Ainsi m et m' vérifient les équations suivantes :

$$m'^\top (\mathbf{F}m) = 0 \text{ dans } I' \quad (2.3)$$

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

2.1.2 Définition de la rectification

L'étape de rectification consiste à réorienter les lignes épipolaires pour qu'elles soient parallèles avec l'axe horizontal de l'image. Cette réorientation s'effectue grâce à une transformation qui projette les épipoles à l'infini. Par conséquent, toutes les paires de stéréo-correspondants sont nécessairement sur la même ordonnée (cf. figure 2.2).

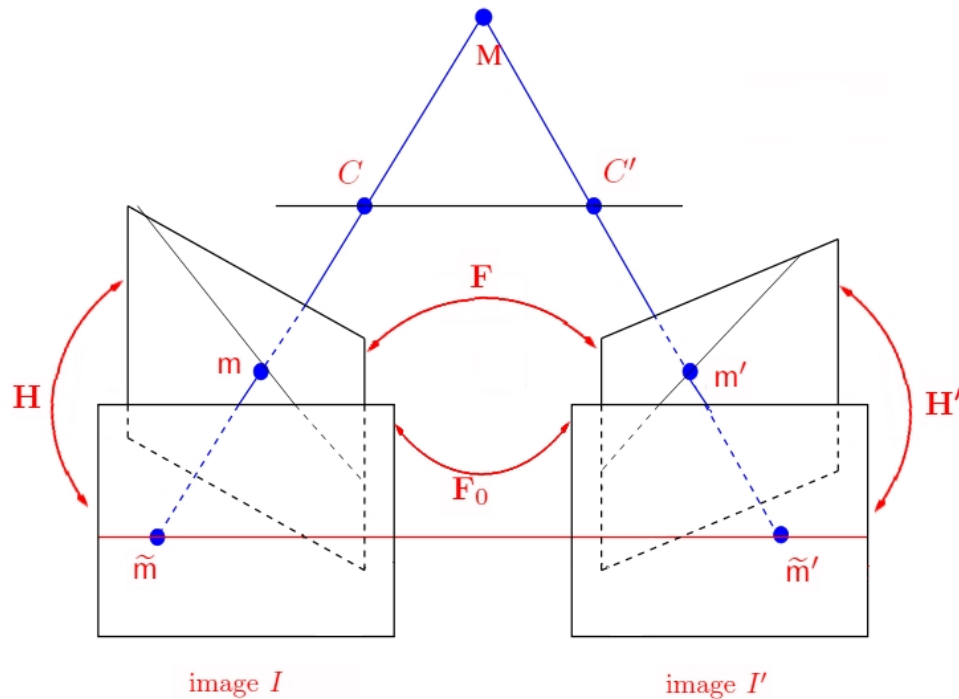


FIG. 2.2 – Opération de rectification.

Il existe deux classes de méthodes de rectification : la rectification d'images avec un système déjà calibré et la rectification d'images provenant d'un système non calibré. Dans ce chapitre, nous nous intéressons uniquement à la deuxième classe de méthodes. Ces méthodes sont fondées sur la connaissance de la géométrie épipolaire et donc sur la connaissance de la matrice Fondamentale.

Lorsque deux images sont rectifiées, la matrice Fondamentale, aussi appelée matrice Fondamentale rectifiée \mathbf{F}_0 , est de la forme suivante :

$$\mathbf{F}_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad (2.5)$$

D'un point de vue purement algébrique, à partir de la connaissance de la matrice Fondamentale \mathbf{F} et de la matrice Fondamentale rectifiée \mathbf{F}_0 , il est possible de déterminer deux homographies \mathbf{H} et \mathbf{H}' telles que :

$$\mathbf{F} = \mathbf{H}'^\top \mathbf{F}_0 \mathbf{H} \quad (2.6)$$

De plus, si un ensemble de couples de points $(\mathbf{m}_k, \mathbf{m}'_k)$, extrait dans les images d'origines, est transformé en un ensemble de points $(\tilde{\mathbf{m}}'_k, \tilde{\mathbf{m}}_k)$ après rectification, la contrainte épipolaire est alors conservée et s'écrit :

$$\tilde{\mathbf{m}}_k'^\top \mathbf{F}_0 \tilde{\mathbf{m}}_k = \mathbf{m}_k'^\top \mathbf{H}'^\top \mathbf{F}_0 \mathbf{H} \mathbf{m}'_k = \mathbf{m}_k'^\top \mathbf{F} \mathbf{m}_k = 0 \quad \forall k \quad (2.7)$$

A partir de la connaissance de la géométrie épipolaire et donc de la matrice fondamentale, l'objectif de la rectification projective, est de déterminer les homographies \mathbf{H} et \mathbf{H}' satisfaisant l'équation (2.7) qui réorientent les lignes épipolaires parallèlement aux lignes des images. Les paramètres de ses homographies sont estimés de manière à minimiser les déformations entre images d'origine et images rectifiées.

2.1.3 Les homographies compatibles

La résolution du système d'équations (2.7) permet d'obtenir une paire d'homographies compatibles \mathbf{H} et \mathbf{H}' sans nécessairement devoir préalablement calculer la matrice Fondamentale \mathbf{F} [Isgro 1999]. Cependant, comme l'extraction de points appariés n'est jamais parfaite, le système d'équations (2.7) ne possède en général jamais de solutions. Ainsi en pratique, on considère un ensemble de N stéréo-correspondants $(\mathbf{m}_k, \mathbf{m}'_k)_{k=1, \dots, N}$ et on détermine une paire d'homographies compatible en minimisant l'ensemble des équations $\mathbf{m}_k'^\top \mathbf{H}'^\top \mathbf{F}_0 \mathbf{H} \mathbf{m}_k = 0$ au sens des moindres carrés :

$$\min_P \sum_k \left(\mathbf{m}_k'^\top \mathbf{H}'^\top \mathbf{F}_0 \mathbf{H} \mathbf{m}_k \right)^2 \quad (2.8)$$

où : $P \in \mathbb{R}^{18}$ est le vecteur d'inconnues composé des 9 paramètres de l'homographie \mathbf{H} et des 9 paramètres de l'homographie \mathbf{H}' .

En outre, les homographies satisfaisant le système d'équations (2.8) ne sont pas uniques. En effet, à cause de la structure de la matrice \mathbf{F}_0 , les premières lignes de chacune des homographies \mathbf{H} et \mathbf{H}' ne peuvent être déterminés. Ces différents paramètres ne jouent donc pas sur l'alignement des lignes épipolaires, en revanche ils peuvent entraîner des déformations projectives. Le problème est alors de trouver une paire d'homographies réduisant le plus possible les déformations des images rectifiées.

2.2 État de l'art

Dans un contexte non calibré, les approches proposées pour trouver un jeu d'homographies minimisant les distorsions projectives s'articulent autour de deux étapes :

- le calcul de la géométrie épipolaire (cf. équation (2.7))
- l'extraction d'homographie et la minimisation des déformations projectives qu'elles induisent.

Il s'agit donc de déterminer tout d'abord une première paire d'homographies compatibles. Ainsi, les épipoles sont projetés à l'infini et les droites épipolaires deviennent parallèles aux lignes des images ainsi transformées. Ensuite, pour minimiser les déformations projectives, les approches adoptées diffèrent en fonction des critères de minimisation géométriques appliqués et des décompositions de matrices affines associées à la matrice de transformation projective. De plus, ces opérations sont soit menées de manière symétrique sur les deux images, soit de manière non symétrique, avec la minimisation des déformations appliquée à une seule image.

2.2.1 Méthode de type "Hartley"

Hartley et al. [Hartley 1999] proposent une approche non symétrique dans laquelle le calcul de chacune des homographies est réalisé en deux étapes. La première étape consiste à trouver une projection qui soit la plus proche possible d'une transformation affine et qui projete l'épipoles à l'infini. Cette projection est de la forme :

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/f & 0 & 1 \end{pmatrix} \quad (2.9)$$

où l'épipoles a pour coordonnées $\mathbf{e} = (f, 0, 1)$.

Afin de rendre les lignes épipolaires horizontales et de centrer l'image rectifiée autour de $(u_0, v_0, 1)$, une rotation \mathbf{R} et une translation \mathbf{t} sont appliquées pour obtenir une première homographie :

$$\mathbf{H}' = \mathbf{GRt} \quad (2.10)$$

Ensuite, l'obtention de la seconde homographie se fait par la minimisation de la distance entre les lignes épipolaires correspondantes :

$$\sum d(\mathbf{Hm}_i, \mathbf{H}'\mathbf{m}'_i)^2 \quad (2.11)$$

où \mathbf{H} est de la forme :

$$\mathbf{H} = \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{H}'\mathbf{M} \quad \text{avec} \quad \mathbf{F} = [\mathbf{e}']_{\times} \mathbf{M}. \quad (2.12)$$

\mathbf{e}' représente l'épipoles associé à l'image I' et $[\mathbf{e}']_{\times}$ la matrice antisymétrique qui lui est associée.

En intégrant l'équation (2.12) dans l'équation (2.11), le critère suivant est obtenu :

$$\sum (a\hat{u}_i + b\hat{v}_i + c - \hat{u}'_i)^2 \quad \text{avec} \quad \hat{\mathbf{m}}'_i = \mathbf{H}'\mathbf{m}'_i \quad \text{et} \quad \hat{\mathbf{m}}_i = \mathbf{H}'\mathbf{M}\mathbf{m}_i = \mathbf{H}_0\mathbf{m}_i \quad (2.13)$$

Il est à noter que la méthode peut être améliorée en prenant en compte la minimisation de l'effet de ré-échantillonnage et de remise à l'échelle.

2.2.2 Méthode de type "Loop et Zhang"

Contrairement à Hartley, Loop et Zhang [Loop 1999] suggèrent une approche symétrique basée sur la décomposition de chaque homographie en transformations projectives \mathbf{H}_p et en transformations affines ("similarity transform" \mathbf{H}_r et "shearing" \mathbf{H}_s) :

$$\mathbf{H} = \mathbf{H}_s \mathbf{H}_r \mathbf{H}_p. \quad (2.14)$$

\mathbf{H} est alors de la forme :

$$\mathbf{H} = \begin{pmatrix} u_a & u_b & u_c \\ v_a & v_b & v_c \\ w_a & w_b & 1 \end{pmatrix} \quad (2.15)$$

Plus précisément, la matrice de projection \mathbf{H}_p (resp. \mathbf{H}'_p) déplace l'épipole \mathbf{e} (resp. \mathbf{e}') à l'infini. Elle est de la forme :

$$\mathbf{H}_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ w_a & w_b & 1 \end{pmatrix} \quad (2.16)$$

avec $\mathbf{w} = [\mathbf{e}]_{\times} \mathbf{z}$ $\mathbf{w}' = \mathbf{F} \mathbf{z}$ et $\mathbf{z} = (\lambda \ \mu \ 0)$ Le vecteur \mathbf{z} est défini de façon à minimiser les distorsions et à rendre la transformation la plus affine possible. Il est obtenu en minimisant un critère non linéaire.

Les matrices de "similarity transform" \mathbf{H}_r , qui alignent les lignes épipolaires correspondantes, sont ensuite définies de la manière suivante :

$$\mathbf{H}_r = \begin{pmatrix} F_{32} - w_b F_{33} & w_a F_{33} - F_{31} & 0 \\ F_{31} - w_a F_{33} & F_{32} - w_b F_{33} & F_{33} + v'_c \\ 0 & 0 & 1 \end{pmatrix} \quad (2.17)$$

$$\mathbf{H}'_r = \begin{pmatrix} F_{23} - w'_b F_{33} & w'_a F_{33} - F_{13} & 0 \\ F_{13} - w'_a F_{33} & F_{23} - w'_b F_{33} & v'_c \\ 0 & 0 & 1 \end{pmatrix}. \quad (2.18)$$

L'application des matrices \mathbf{H}_r et \mathbf{H}'_r peut générer des déformations sur l'axe des abscisses. Ces déformations sont alors minimisées par les matrices dites de "shearing" \mathbf{H}_s , de la forme :

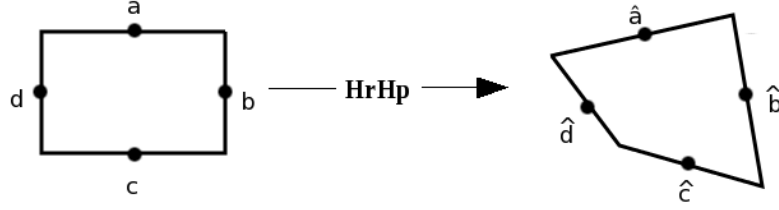
$$\mathbf{H}_s = \begin{pmatrix} a & b & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.19)$$

avec a et b calculés en fonction de la taille de la fenêtre rectifiée (voir fig.2.2.2) :

$$a = \frac{h^2 x_v^2 + w^2 y_v^2}{hw(x_v y_u - x_u y_v)} \quad ; \quad b = \frac{h^2 x_u x_v + w^2 y_u y_v}{hw(x_v y_u - x_u y_v)} \quad (2.20)$$

avec $x = \hat{b} - \hat{d}$ et $y = \hat{c} - \hat{a}$

Tout comme Hartley [Hartley 1999], la méthode peut être améliorée en prenant en compte la minimisation de l'effet de ré-échantillonnage et de remise à l'échelle.



2.2.3 Méthode de type "Devernay"

Dans sa thèse [Devernay 1997], Devernay propose également de traiter les images de manière symétrique à partir d'une décomposition en valeurs singulières de la matrice Fondamentale. Cette matrice étant de rang 2 et définie à un facteur d'échelle près, elle peut s'écrire :

$$\mathbf{F} = \begin{pmatrix} \mathbf{e}' & \mathbf{U}_1 & \mathbf{U}_2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{pmatrix} \begin{pmatrix} \mathbf{e}^\top \\ \mathbf{V}_1^\top \\ \mathbf{V}_2^\top \end{pmatrix} \quad (2.21)$$

où σ est un nombre réel positif et où les vecteurs \mathbf{e} et \mathbf{e}' représentent les épipoles associés à chacune des images. Une simple manipulation sur les trois matrices permet d'obtenir \mathbf{H}_0 et \mathbf{H}'_0 :

$$\mathbf{F} = \begin{pmatrix} \mathbf{e}' & \mathbf{U}_1 & \sqrt{\sigma}\mathbf{U}_2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{e}^\top \\ \sqrt{\sigma}\mathbf{V}_2^\top \\ -\mathbf{V}_1^\top \end{pmatrix} = \mathbf{H}'_0^\top \mathbf{F}_0 \mathbf{H}_0. \quad (2.22)$$

Ensuite, Devernay démontre que, pour minimiser la déformation des images, il est possible de multiplier \mathbf{H}_0 (resp. \mathbf{H}'_0) par une matrice \mathbf{T} (resp. \mathbf{T}') sans modifier l'alignement des lignes épipolaires :

$$\mathbf{m}'^\top \mathbf{F} \mathbf{m} = \mathbf{m}'^\top \mathbf{H}'_0^\top \mathbf{T}'^\top \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \mathbf{T} \mathbf{H}_0 \mathbf{m} = 0 \quad (2.23)$$

avec :

$$\mathbf{T} = \begin{pmatrix} a_1 & b_1 & c_1 \\ 0 & e & f \\ 0 & h & i \end{pmatrix} ; \quad \mathbf{T}' = \begin{pmatrix} a_2 & b_2 & c_2 \\ 0 & e & f \\ 0 & h & i \end{pmatrix} \quad (2.24)$$

où $a_1, a_2, b_1, b_2, c_1, c_2, e, f, g, h, i$ sont calculés en fonction de différentes fonctions objectifs permettant de minimiser les déformations de l'image.

2.2.4 Méthodes de type "Mallon"

Sur la base des travaux de Hartley, Mallon [Mallon 2005] propose d'améliorer l'étape de minimisation des déformations projectives. L'intérêt de cette approche est

que la minimisation est réalisée en considérant tous les points de l'image. En effet, afin de minimiser les distorsions, la méthode utilisée consiste à multiplier chacune des homographies par une matrice de la forme :

$$\mathbf{A} = \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.25)$$

où le paramètre c ne joue pas sur la distorsion, il peut donc être choisi de manière à recentrer l'image rectifiée. Pour déterminer les valeurs de a et de b , il est analysé comment les transformations \mathbf{H} et \mathbf{H}' déforment l'image, c'est-à-dire, étirent ou compressent localement des pixels. L'outil utilisé est le jacobien de la transformation qui associe un pixel de l'image originale à un pixel de l'image rectifiée :

$$\begin{aligned} f_{\mathbf{H}} : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ m &\rightarrow \begin{bmatrix} (Hm)_x & (Hm)_y \\ (Hm)_z & (Hm)_z \end{bmatrix} \end{aligned}$$

Il permet de caractériser en chaque point la façon dont ces deux transformations projectives "créent" ou "détruisent" localement des pixels. Gluckman et al. [Gluckman 2001] ont proposé de quantifier de façon scalaire cette déformation locale par le calcul du déterminant du jacobien, et de chercher la transformation dont le déterminant s'approche le plus de 1. Mais on peut montrer en exhibant des contres-exemples que ceci ne conduit pas toujours au meilleur résultat. En effet, le jacobien de la transformation $f_{\mathbf{H}}$ s'écrit :

$$\mathbf{J}_{\mathbf{H}} = \begin{bmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \end{bmatrix} = \begin{bmatrix} \frac{\delta \hat{u}}{\delta u} & \frac{\delta \hat{u}}{\delta v} \\ \frac{\delta \hat{v}}{\delta u} & \frac{\delta \hat{v}}{\delta v} \end{bmatrix} \quad (2.26)$$

Le déterminant s'écrit alors $\det(\mathbf{J}_{\mathbf{H}}) = j_{11}j_{22} - j_{12}j_{21}$. Nous constatons qu'il est tout à fait possible d'avoir un déterminant égal à 1, alors même que la transformation projective peut créer ou détruire localement des pixels.

Mallon propose donc d'utiliser à la place du déterminant les valeurs singulières du jacobien $\mathbf{J}_{\mathbf{H}}$, qui caractérisent de façon plus fine la transformation projective. En effet, ces valeurs singulières sont égales à 1 pour une transformation rigide. Pour chacune des deux matrices \mathbf{H} et \mathbf{H}' , la méthode consiste donc à trouver par un processus itératif les valeurs de a et b qui minimisent l'expression suivante :

$$f(a, b) = \sum [(\sigma_1 - 1)^2 + (\sigma_2 - 1)^2] \quad (2.27)$$

avec σ_1 et σ_2 étant issus de la décomposition en valeurs singulières de la matrice $J(H)$:

$$\mathbf{J}_{\mathbf{H}} = \mathbf{U} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \mathbf{V}^T \quad (2.28)$$

L'intérêt de cette approche est que la minimisation se fait sur un ensemble de points (n points à gauche, et n points à droite) librement choisis. Dans le cas général, on peut prendre une grille de points sur l'image, mais il est aussi possible

de minimiser la distorsion sur une zone de l'image. Ce dernier point est intéressant si l'on y a identifié un élément auquel on s'intéresse plus particulièrement. De récentes recherches montrent comment améliorer cette méthode grâce à des méthodes d'optimisation globale [Bugarin 2010].

Finalement, Isgro [Isgro 1999] propose une méthode pour obtenir les matrices d'homographies sans calcul préalable de la matrice Fondamentale. Cette méthode est basée sur la décomposition des matrices d'homographies de Hartley [Hartley 1999] mais résout le problème de manière différente. La première étape consiste à minimiser la fonction coût suivante :

$$\min_{P \in \mathbb{R}^{18}} \sum_{k=1}^N [m'_k{}^\top \mathbf{H}'^\top \mathbf{F}_0 \mathbf{H} m_k]^2 \quad (2.29)$$

avec $\mathbf{H}' = \mathbf{GRt}$ (voir equation (2.10)).

Cependant, comme la première ligne de la matrice \mathbf{F}_0 est le vecteur nul, cette équation permet seulement d'obtenir les deux dernières lignes des matrices d'homographies. Afin d'obtenir la première ligne de la matrice \mathbf{H} tout en minimisant les déformations projectives l'auteur choisit de minimiser la disparité entre les images en résolvant le problème d'optimisation suivant :

$$\min_{P \in \mathbb{R}^6} \sum_{k=1}^N [(\mathbf{H} m_k)_1 - (\mathbf{H}' m'_k)_1]^2 \quad (2.30)$$

où $(\cdot)_1$ représente la première coordonnée du point rectifié $\mathbf{H} m_k$ (resp. $\mathbf{H}' m'_k$).

2.2.5 Discussions

Les approches proposées consistent à dissocier les deux étapes classiques de la rectification (à savoir (a) une transformation qui projette les épipoles à l'infini, (b) une minimisation des déformations projectives). Nous avons donc deux étapes d'optimisation qui se succèdent. La première conséquence est une propagation d'erreur entre chaque étape. En effet, la première phase débute par la détermination de la matrice Fondamentale qui est souvent entachée d'erreur. La seconde conséquence est une correction de ces homographies à partir de propriétés mathématiques. La méthode consiste à multiplier chacune des homographies par une matrice dite de "shearing" qui étire ou compresse localement des pixels et donc "créé" ou "détruit" des pixels. Les paramètres de cette matrice sont calculés à partir des propriétés du jacobien de l'image (déterminant proche de 1 [Gluckman 2001], ou les valeurs singulières du jacobien proche de 1 [Mallon 2005]). De plus, cette approche est d'autant plus efficace que le nombre de points est grand.

L'application de ces méthodes avec des images infrarouges donne des résultats insuffisants (voir paragraphe 2.4.3). En effet, malgré un calcul de la matrice fondamentale soigné, le faible nombre de points extraits et leurs précisions entraînent tout de même des erreurs sur son estimation qui se répercutent sur le calcul des matrices de projections (permettant d'envoyer les épipoles à l'infini) et génèrent des erreurs

de rectification finales encore plus grandes. De plus, la faible résolution de nos caméras ajoutée à un petit angle d'ouverture, nécessitant une plus grande convergence du banc stéréo, créent, dans la plupart des cas testés, de fortes distorsions projectives sur les images rectifiées.

Notre priorité est donc de minimiser le nombre de processus itératifs où ces points sont utilisés comme mesures. Nous avons donc intérêt :

- à ne pas procéder à un calcul explicite préalable de la matrice Fondamentale pour le calcul des homographies compatibles [Isgro 1999]
- à regrouper l'étape de décomposition en transformations projectives et celle de minimisation des déformations en une même opération.

Nous proposons donc une méthode de rectification permettant d'obtenir une paire d'homographies qui réduit les déformations des images, en une seule étape, en minimisant une fonction coût sous contraintes. La section suivante présente cette méthode originale de rectification basée sur un processus d'optimisation qui résout directement l'équation classique sans introduire explicitement la matrice Fondamentale. Cette optimisation sous contraintes est réalisée sans décomposition particulière des homographies ce qui permet de minimiser simultanément les distorsions sur chacune d'elles. Nous présentons tout d'abord le choix de la fonction coût permettant de conserver au mieux la géométrie épipolaire. Puis nous exposons la définition de l'espace des contraintes qui permet de contrôler les déformations projectives. Enfin, une section est consacrée à la comparaison des résultats de notre méthode avec les approches existantes.

2.3 Contribution à la rectification

Cette section décrit une approche directe et symétrique pour estimer les homographies nécessaires à la rectification projective d'images non calibrées infrarouges. La méthode proposée est réalisée en une seule étape, sans estimation initiale de la matrice Fondamentale. Cette étape est basée sur une méthode d'optimisation sous contraintes qui permet à la fois une bonne estimation de la géométrie épipolaire, tout en limitant les distorsions induites sur les images rectifiées. La fonction coût et l'espace des contraintes sont choisis afin d'obtenir une erreur sur la rectification qui soit similaire à celle obtenue sur l'estimation de la matrice Fondamentale. Les résultats numériques obtenus montrent que les distorsions résiduelles projectives sur les images rectifiées, évaluées à partir de mesures indépendantes, sont très faibles.

2.3.1 Méthode directe de rectification

La méthode de rectification présentée est directe et permet d'obtenir, en une seule étape, les deux homographies grâce à la minimisation sous contraintes d'une fonction coût. La méthode consiste à résoudre le système d'équations (2.7) au sens des moindres carrés. Cette minimisation se fait via une fonction coût qui pénalise la moyenne des distances aux droites épipolaires sous des contraintes géométriques qui contrôlent les distorsions projectives induites par le processus de rectification.

Nous décrivons tout d'abord la fonction coût comme étant une distance physique de l'image. Ensuite une méthode originale est proposée afin de décrire l'effet des distorsions projectives sur les images pour définir un espace de contraintes. Puis nous montrons que cet espace de contraintes peut être décrit de manière analytique et de manière géométrique. Enfin l'implémentation et l'évaluation de ces méthodes sont présentées.

2.3.2 Définition de la fonction coût

L'objectif est de résoudre l'équation (2.7) pour trouver les deux homographies sur la base d'un ensemble de couples de points appariés $(\mathbf{m}_k, \mathbf{m}'_k)$. Cette équation est alors résolue, comme proposé par [Isgro 1999], au sens des moindres carrés en minimisant la fonction coût suivante :

$$\min_P \sum_k \left(\mathbf{m}'_k{}^\top \mathbf{H}'^\top (P_2) \mathbf{F}_0 \mathbf{H} (P_1) \mathbf{m}_k \right)^2 \quad (2.31)$$

où $P \in \mathbb{R}^{18}$ est le vecteur d'inconnues composé des entrées des matrices \mathbf{H} et \mathbf{H}' .

Cependant, le problème (2.31) est très sensible au bruit. Une quantité mesurable dans le plan image est la distance Euclidienne d'un point \mathbf{m}' à la ligne épipolaire associée \mathbf{Fm} . Cette distance est calculée de manière croisée dans les deux images. Cela conduit à résoudre le problème de minimisation non linéaire suivant :

$$\min_P \sum_k \left[d(\mathbf{m}'_k, \mathbf{Fm}_k)^2 + d(\mathbf{m}_k, \mathbf{F}'^\top \mathbf{m}'_k)^2 \right] \quad (2.32)$$

En outre, la distance d'un point \mathbf{m}' à la droite épipolaire \mathbf{Fm} s'exprime dans \mathbb{R}^2 comme suit :

$$d(\mathbf{m}', \mathbf{Fm})^2 = \frac{|\mathbf{m}'^\top \mathbf{Fm}|^2}{\|\pi(\mathbf{Fm})\|_2^2} \quad (2.33)$$

où $\pi : (x_1, x_2, x_3) \rightarrow (x_1, x_2)$ la projection orthogonale canonique. On définit de manière analogue la distance d'un point \mathbf{m} à la ligne épipolaire correspondante $\mathbf{F}'^\top \mathbf{m}'$ par :

$$d(\mathbf{m}, \mathbf{F}'^\top \mathbf{m}')^2 = \frac{|\mathbf{m}^\top \mathbf{F}'^\top \mathbf{m}'|^2}{\|\pi(\mathbf{F}'^\top \mathbf{m}')\|_2^2} \quad (2.34)$$

Le problème (2.32) se reformule donc en :

$$\min_P \sum_k \frac{|\mathbf{m}'_k{}^\top \mathbf{Fm}_k|^2}{\|\pi(\mathbf{Fm}_k)\|_2^2} + \frac{|\mathbf{m}'_k{}^\top \mathbf{F}'^\top \mathbf{m}'_k|^2}{\|\pi(\mathbf{F}'^\top \mathbf{m}'_k)\|_2^2} \quad (2.35)$$

Ce critère est utilisé à l'origine pour estimer la matrice Fondamentale \mathbf{F} , cependant Luong [Luong 1992] a montré qu'il était plus pertinent de remplacer ce critère, basé sur la distance euclidienne, par un autre critère normalisé basé sur la distance de Sampson [Hartley 2003] :

$$\min_P \sum_k \frac{|\mathbf{m}'_k{}^\top \mathbf{Fm}_k|^2}{\|\pi(\mathbf{Fm}_k)\|_2^2 + \|\pi(\mathbf{F}'^\top \mathbf{m}'_k)\|_2^2} \quad (2.36)$$

Afin d'estimer les homographies \mathbf{H} et \mathbf{H}' , le critère est réécrit en remplaçant la matrice Fondamentale \mathbf{F} par l'équation (2.6). De plus, ce critère est minimisé sous l'espace de contraintes C pour trouver une paire d'homographies permettant de réduire les distorsions perspectives. Finalement ce critère est donné par la relation suivante :

$$\min_{P \in C} \sum_k \frac{|m'_k{}^\top \mathbf{H}'{}^\top \mathbf{F}_0 \mathbf{H} m_k|^2}{\|\pi(\mathbf{H}'{}^\top \mathbf{F}_0 \mathbf{H} m_k)\|_2^2 + \|\pi(\mathbf{H}{}^\top \mathbf{F}_0' \mathbf{H}' m'_k)\|_2^2} \quad (2.37)$$

avec C définit par : $C = \bigcap_{i=1}^n [g_i^{-1}(I_i) \cap g'_i{}^{-1}(I_i)]$

et

$$\begin{aligned} g_i^{-1}(I_i) &:= \{P \in \mathbb{R}^{18} | g_i(P) \in I_i\} \\ g'_i{}^{-1}(I_i) &:= \{P \in \mathbb{R}^{18} | g'_i(P) \in I_i\} \end{aligned}$$

où $g_i : \mathbb{R}^9 \times \mathbb{R}^9 \mapsto \mathbb{R}$ et $g'_i : \mathbb{R}^9 \times \mathbb{R}^9 \mapsto \mathbb{R}$ sont des fonctions deux fois dérivables, n est le nombre de contraintes et I_i ($i = 1 \dots n$) sont les intervalles de ces contraintes dans \mathbb{R}^+ . Il est à noter que les fonctions g_i (resp. g'_i) représentent les contraintes imposées sur l'image de gauche (resp. pour l'image de droite).

La résolution du problème (2.37) est réalisée avec une méthode d'optimisation locale sous contraintes. Cependant, de manière générale, le choix d'un vecteur de paramètres initiaux P_0 satisfaisant les contraintes est parfois aussi difficile que résoudre le problème lui-même. Ainsi notre apport est double puisqu'il consiste d'une part à définir un ensemble de contraintes contrôlant les distorsions projectives et d'autres à les formuler de telles sortes que l'on puisse choisir dans tous les cas une estimée initiale convenable. Le détail du choix des valeurs de C et P_0 est exposé dans le paragraphe suivant.

2.3.3 Définition des contraintes

Comme indiqué précédemment, l'objectif d'une rectification idéale est que chaque pixel des images originales correspond à un seul pixel dans les images rectifiées. Cependant, la rectification entraîne des distorsions perspectives, ce qui signifie que certaines parties de l'image peuvent se déformer provoquant alors une création ou une perte de pixels. Dans [Hartley 1999], cette altération de l'image est modélisée par la modification d'une petite partie de l'image, autour des points appariés, avant et après l'étape de rectification. Cette modification peut être déterminée grâce à des propriétés numériques tel que le jacobien de la matrice \mathbf{H} . Ces propriétés numériques peuvent être contrôlées par des opérateurs comme le déterminant de la matrice [Gluckman 2001] ou la décomposition en valeurs singulières de cette même matrice [Mallon 2005]. Ainsi, en minimisant ces opérateurs, il est possible de réduire les distorsions dans l'image entière.

Toutefois, l'application de ces opérateurs comme contraintes est numériquement instable et demande beaucoup de ressources. Ainsi, notre approche est de définir un

ensemble de contraintes C numériquement stable qui renforce les propriétés physiques de l'image. Les distorsions de l'image sont alors définies comme des modifications de la structure de l'image : diagonale, la taille (largeur et hauteur) et l'angle.

2.3.3.1 Définition du rapport d'aspect

Les contraintes sur la structure de l'image sont exprimées par l'invariance du rapport des diagonales des images. Seuls quatre coins de l'image sont nécessaires pour calculer cette contrainte. Considérant \mathbf{p}_j (resp. \mathbf{p}'_j), $j = 1 \dots 4$ les coins des images \mathcal{I} (resp. \mathcal{I}'). Ainsi, le rapport d'aspect E_a , par exemple pour l'image \mathcal{I} , est alors défini par :

$$E_a = \frac{\|\mathbf{p}_1 - \mathbf{p}_3\|_2}{\|\mathbf{p}_2 - \mathbf{p}_4\|_2} \quad (2.38)$$

2.3.3.2 Définition du rapport de taille de l'image

L'invariance de la largeur et de la hauteur entre les images originales et les images rectifiées est définie à partir du milieu des côtés des images. On note $\mathbf{q}_1, \mathbf{q}_3$ (resp. $\mathbf{q}'_1, \mathbf{q}'_3$) les milieux de la partie horizontale de l'image \mathcal{I} (resp. \mathcal{I}') et $\mathbf{q}_2, \mathbf{q}_4$ (resp. $\mathbf{q}'_2, \mathbf{q}'_4$) les milieux de la partie vertical de l'image \mathcal{I} (resp. \mathcal{I}'). Les points correspondants peuvent être définis dans les images rectifiées : $\tilde{\mathbf{q}}_j$ (resp. $\tilde{\mathbf{q}}'_j$), $j = 1 \dots 4$. Les rapports de la largeur et de la hauteur entre les images originales et corrigées sont alors définis comme suit :

$$E_h = \frac{\|\mathbf{q}_1 - \mathbf{q}_3\|_2}{\|\tilde{\mathbf{q}}_1 - \tilde{\mathbf{q}}_3\|_2} \quad E_w = \frac{\|\mathbf{q}_2 - \mathbf{q}_4\|_2}{\|\tilde{\mathbf{q}}_2 - \tilde{\mathbf{q}}_4\|_2} \quad (2.39)$$

2.3.3.3 Définition de l'orthogonalité de l'image

L'invariance des angles entre les images originales et rectifiées est calculée à partir de l'orthogonalité des bi-médianes de images. Les articles [Loop 1999] et [Wu 2007] estiment que l'application de contraintes sur les milieux des quatre côtés permet de préserver l'orthogonalité de l'image. En conservant les notations précédentes, l'orthogonalité E_o , dans l'image \mathcal{I} , est alors exprimée comme l'angle entre les vecteurs $\mathbf{q}_1 - \mathbf{q}_3$ et $\mathbf{q}_2 - \mathbf{q}_4$:

$$E_o = \cos^{-1} \left(\frac{\langle \mathbf{q}_1 - \mathbf{q}_3 | \mathbf{q}_2 - \mathbf{q}_4 \rangle}{\|\mathbf{q}_1 - \mathbf{q}_3\| \|\mathbf{q}_2 - \mathbf{q}_4\|} \right) \quad (2.40)$$

Les distorsions de l'image sont réduites au cours du processus de rectification en utilisant des contraintes s'appliquant sur la taille, le rapport d'aspect et de l'orthogonalité de l'image (voir figure 2.3).

La section suivante traite de la mise en oeuvre des contraintes avec deux méthodes différentes : l'une analytique et l'autre géométrique.

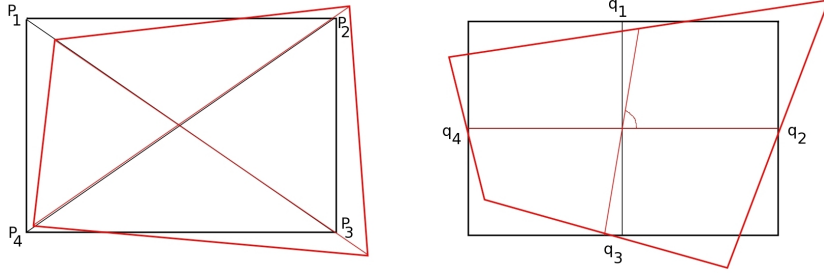


FIG. 2.3 – Distorsions avant et après rectification : à gauche, les déformations dues à une modification du rapport d'aspect, et à droite, les déformations dues à une modification d'orthogonalité.

2.3.4 Contraintes analytiques

La mise en oeuvre des contraintes analytiques correspond à l'évaluation des formules précédentes sur un intervalle de tolérance.

2.3.4.1 Contraintes sur le rapport d'aspect

En appliquant l'équation (2.38) aux images rectifiées droite et gauche, les fonctions g_1 et g'_1 de l'équation (2.37) sont données par :

$$g_1(P) = \frac{\|\tilde{p}_1 - \tilde{p}_3\|_2}{\|\tilde{p}_2 - \tilde{p}_4\|_2} \quad g'_1(P) = \frac{\|\tilde{p}'_1 - \tilde{p}'_3\|_2}{\|\tilde{p}'_2 - \tilde{p}'_4\|_2} \quad (2.41)$$

Idéalement le rapport d'aspect des images rectifiées est égal à l'unité.

2.3.4.2 Contraintes sur la taille de l'image

En appliquant l'équation (2.39) aux images rectifiées droite et gauche, les fonctions g_2 , g'_2 , g_3 et g'_3 de l'équation (2.37) sont données par :

$$g_2(P) = \frac{\|\mathbf{q}_1 - \mathbf{q}_3\|_2}{\|\tilde{\mathbf{q}}_1 - \tilde{\mathbf{q}}_3\|_2} \quad g'_2(P) = \frac{\|\mathbf{q}'_1 - \mathbf{q}'_3\|_2}{\|\tilde{\mathbf{q}}'_1 - \tilde{\mathbf{q}}'_3\|_2} \quad (2.42)$$

$$g_3(P) = \frac{\|\mathbf{q}_2 - \mathbf{q}_4\|_2}{\|\tilde{\mathbf{q}}_2 - \tilde{\mathbf{q}}_4\|_2} \quad g'_3(P) = \frac{\|\mathbf{q}'_2 - \mathbf{q}'_4\|_2}{\|\tilde{\mathbf{q}}'_2 - \tilde{\mathbf{q}}'_4\|_2} \quad (2.43)$$

Les rapports des largeurs et hauteur sont idéalement égaux à l'unité.

2.3.4.3 Contraintes sur l'orthogonalité

En appliquant l'équation (2.40) aux images rectifiées droite et gauche, les fonctions g_4 et g'_4 de l'équation (2.37) sont données par :

$$g_4(P) = \cos^{-1} \left(\frac{\langle \tilde{\mathbf{q}}_1 - \tilde{\mathbf{q}}_3 | \tilde{\mathbf{q}}_2 - \tilde{\mathbf{q}}_4 \rangle}{\|\tilde{\mathbf{q}}_1 - \tilde{\mathbf{q}}_3\| \|\tilde{\mathbf{q}}_2 - \tilde{\mathbf{q}}_4\|} \right) \quad (2.44)$$

$$g'_4(P) = \cos^{-1} \left(\frac{\langle \tilde{\mathbf{q}}'_1 - \tilde{\mathbf{q}}'_3 | \tilde{\mathbf{q}}'_2 - \tilde{\mathbf{q}}'_4 \rangle}{\|\tilde{\mathbf{q}}'_1 - \tilde{\mathbf{q}}'_3\| \|\tilde{\mathbf{q}}'_2 - \tilde{\mathbf{q}}'_4\|} \right) \quad (2.45)$$

L'orthogonalité est idéalement de 90° .

2.3.4.4 Conclusions

L'espace des contraintes C peut être défini comme :

$$C_A = \bigcap_{i=1}^4 g_i^{-1}(I_i) \cap g_i'^{-1}(I_i) \quad (2.46)$$

où $I_1 = [1 - r, 1 + r]$, $I_2 = I_3 = [1 - s, 1 + s]$, $I_4 = [90 - \theta, 90 + \theta]$ avec r , s et θ la tolérance à la modification du rapport d'aspect, de la taille de l'image et de l'orthogonalité. Les contraintes sont explicitement exprimées par huit fonctions servant à limiter les distorsions définies par le calcul du rapport d'aspect, du changement de taille et de l'orthogonalité.

2.3.5 Contraintes géométriques

Une autre façon de définir l'espace des contraintes est d'utiliser trois anneaux concentriques dans lesquels les coins de l'image (un anneau) et les milieux des côtés (deux anneaux, un pour les côtés horizontaux et l'autre pour les côtés verticaux de l'image) doivent rester après rectification. La mise en oeuvre est basée sur les propriétés des transformations rigides comme les rotations et les translations dans le plan projectif.

2.3.5.1 Explication de l'approche

Dans un premier temps, nous nous plaçons dans le contexte d'une transformation sans déformation. Soit une homographie \mathbf{H} appliquée aux quatre coins \mathbf{p}_j et quatre milieux des côtés \mathbf{q}_j ($j = 1 \dots 4$) d'une image \mathcal{I} :

1. $\mathbf{H}\mathbf{p}_j$, $j = 1 \dots 4$ sont sur un cercle \mathcal{C}_1 de centre \mathbf{c}_1 et de rayon r_1 .
2. $\mathbf{H}\mathbf{q}_2$, $\mathbf{H}\mathbf{q}_4$ sont sur un cercle \mathcal{C}_2 (\mathbf{c}_1, r_2), avec $r_2 < r_1$.
3. $\mathbf{H}\mathbf{q}_1$, $\mathbf{H}\mathbf{q}_3$ sont sur un cercle \mathcal{C}_3 (\mathbf{c}_1, r_3), avec $r_3 < r_2 < r_1$.

La figure 2.4 montre les trois cercles concentriques dans lesquels les distances entre les points sélectionnés restent invariantes. Les deux points $\mathbf{H}\mathbf{p}_1$ et $\mathbf{H}\mathbf{p}_2$ sont sur le même cercle \mathcal{C}_1 . De plus, le segment $[\mathbf{H}\mathbf{p}_1, \mathbf{H}\mathbf{p}_2]$ est tangent à \mathcal{C}_3 en son milieu $\mathbf{H}\mathbf{q}_1$. Par conséquent, comme \mathcal{C}_1 et \mathcal{C}_3 sont concentriques, le segment $[\mathbf{H}\mathbf{p}_1, \mathbf{H}\mathbf{p}_2]$ a nécessairement une longueur constante, indépendamment de la position de $\mathbf{H}\mathbf{p}_1$ et $\mathbf{H}\mathbf{p}_2$ sur \mathcal{C}_1 . En appliquant le même raisonnement au segment $[\mathbf{H}\mathbf{p}_2, \mathbf{H}\mathbf{p}_3]$ avec son milieu $\mathbf{H}\mathbf{q}_2$ et en utilisant les cercles concentriques \mathcal{C}_1 et \mathcal{C}_2 , nous pouvons conclure que sa longueur est constante, indépendamment de la position de $\mathbf{H}\mathbf{p}_2$ et $\mathbf{H}\mathbf{p}_3$ sur \mathcal{C}_1 . Par symétrie, des résultats similaires sont obtenus sur $[\mathbf{H}\mathbf{p}_3, \mathbf{H}\mathbf{p}_4]$ et $[\mathbf{H}\mathbf{p}_4, \mathbf{H}\mathbf{p}_1]$. Enfin, les côtés de l'image rectifiée étant deux à deux identiques et constants, on en déduit que la transformation \mathbf{H} ne peut être qu'une rotation (cf. figure 2.4).

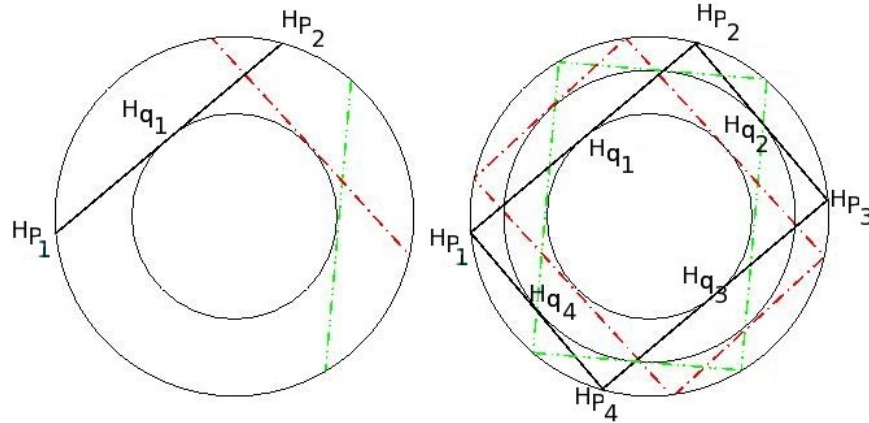


FIG. 2.4 – Invariance de la taille d'un segment se déplaçant sur deux cercles (image gauche) et invariance d'une image se déplaçant sur trois cercles (image droite).

Toutefois, les centres des cercles n'ont aucune raison d'être identiques, nous introduisons donc un degré de liberté permettant aux homographies d'aligner les lignes épipolaires des images rectifiées. Il est donc possible d'ajouter un paramètre de translation sans changer les propriétés géométriques. Cette translation est représentée par le déplacement du centre des cercles.

Enfin, les contraintes définies ci-dessus limitent l'homographie \mathbf{H} à des transformations rigides et par conséquent de conserver une taille d'image constante, un rapport d'aspect unitaire et une orthogonalité de 90° . Cependant, ces contraintes sont trop fortes pour assurer une solution satisfaisante au problème de minimisation. Afin d'élargir l'espace des solutions, les cercles sont remplacés par des anneaux de largeur variable qui autorisent alors quelques déformations.

2.3.5.2 Définition des fonctions de contrainte

Les quatre coins des images rectifiées gauche et droite sont contraints de se déplacer dans un anneau \mathcal{R}_1 ayant pour centre celui de l'image \mathbf{c} , avec un diamètre intérieur égal à $R_1 - r_1$ et un diamètre extérieur égal à $R_1 + r_1$. La figure 2.5 montre la contrainte d'anneau sur ces points.

Les milieux des côtés horizontaux des images rectifiées gauche et droite sont contraints de se déplacer sur un anneau \mathcal{R}_2 , de centre \mathbf{c} , de rayon R_2 et de demi-largeur r_2 . Par analogie, les milieux des côtés verticaux des images rectifiées gauche et à droite sont contraints de se déplacer sur un anneau \mathcal{R}_3 , de centre \mathbf{c} , de rayon R_3 et de demi-largeur r_3 . La figure 2.6 montre l'ensemble des contraintes d'anneaux appliquées aux huit points spécifiques de l'image. En conservant les notations pré-

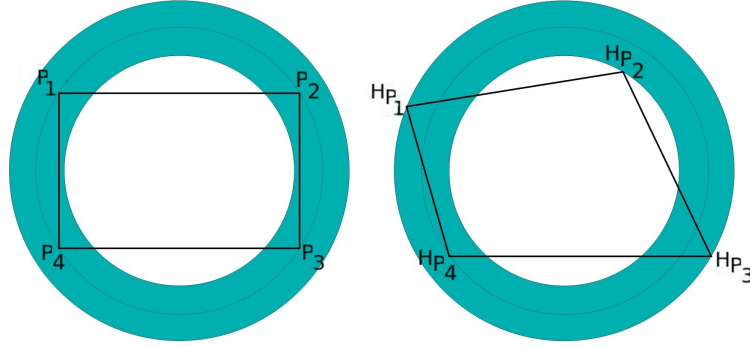


FIG. 2.5 – Une contrainte d’anneau où seuls les coins doivent rester après rectification (image originale à gauche et image rectifiée à droite).

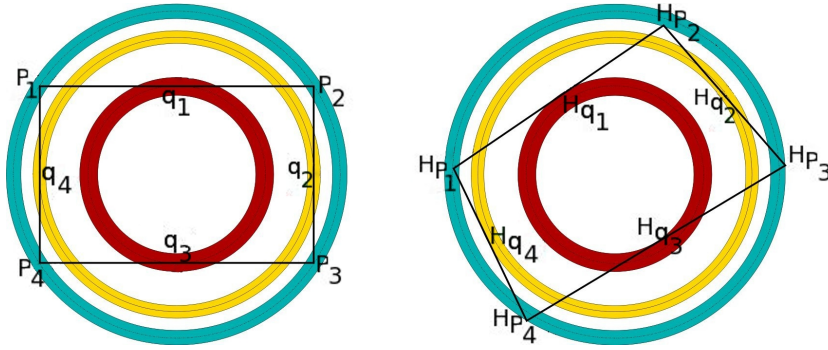


FIG. 2.6 – Deux anneaux additionnels où les milieux des cotés doivent rester après rectification (image originale à gauche et image rectifiée à droite).

cédentes, les fonctions g_j et g'_j peuvent être définies comme :

$$\begin{aligned} g_j(P) &= \|\tilde{p}_j - c\|_2 & g_{j+4}(P) &= \|\tilde{q}_j - c\|_2 \\ g'_j(P) &= \|\tilde{p}'_j - c\|_2 & g'_{j+4}(P) &= \|\tilde{q}'_j - c\|_2 \end{aligned}$$

avec : $j = 1 \dots 4$

Il est à noter que les fonctions g_j^{-1} et g'_{j+4}^{-1} , $j = 1 \dots 4$ correspondent aux contraintes appliquées aux quatre coins de l’image. Et que les fonctions g_{j+4}^{-1} et g'_j^{-1} , $j = 5 \dots 8$ correspondent aux contraintes appliquées aux quatre milieux des cotés de l’image. L’espace des contraintes est donné par :

$$C_B = \bigcap_{j=1}^8 \left[g_j^{-1}(I_j) \cap g'_{j+4}^{-1}(I_j) \right] \quad (2.47)$$

avec les intervalles définis comme :

$$\begin{aligned} I_j &= [R_1 - r_1, R_1 + r_1] & j = 1 \dots 4 \\ I_j &= [R_2 - r_2, R_2 + r_2] & j = 5, 6 \\ I_j &= [R_3 - r_3, R_3 + r_3] & j = 7, 8 \end{aligned}$$

Les contraintes sont explicitement exprimées par seize fonctions servant à limiter les distorsions projectives. Ces fonctions garantissent que huit points spécialement choisis restent dans des anneaux dont la largeur est définie en terme de compromis entre la précision de rectification et la distorsion projective (i.e. plus les intervalles sont petits, plus les erreurs résiduelles de rectification sont élevées et inversement, si les anneaux sont plus grands, les erreurs résiduelles diminuent au détriment des distorsions projectives).

2.3.6 Implémentation

La minimisation de l'équation (2.37) est un problème d'optimisation non linéaire sous contraintes définies par les équations (2.46) ou (2.47). Le processus de minimisation sous contraintes est effectué par l'algorithme "Feasible Sequential Quadratic Programming" mis en oeuvre par Andre Tits [Lawrence 2001] [C.T. Lawrence 1997]. De plus, les algorithmes d'optimisation locale requièrent une bonne estimée initiale qui doit appartenir à l'espace des contraintes pour garantir une solution satisfaisante. Comme l'image d'origine n'a pas subi de déformation, elle satisfait toutes les contraintes et le vecteur de paramètres initiaux le plus simple est une paire d'homographies égale à l'identité. En outre, l'identité a la propriété d'être toujours située dans le bassin d'attraction d'un minimum local correspondant à une réalité physique.

Il est à noter que certains résultats préliminaires ont été obtenus avec des algorithmes d'optimisation globale appliqués à la rectification d'image [Bugarin 2010]. Plusieurs développements sont en cours pour l'utilisation de l'optimisation globale polynomiale avec la théorie des moments [Lasserre 2001] avec une fonction de coût rationnelle de degré 2 à 18 variables. Cependant, la complexité de cet algorithme est en $O(d^n)$, où n est le nombre de variables et d le degré maximal de la fonction de coût, ce qui le rend inapplicable dans notre cas.

2.3.7 Conclusion

Les homographies de rectification sont trouvées par la minimisation d'une fonction coût non linéaire dans un espace de contraintes. La fonction de coût pénalise l'erreur sur la géométrie épipolaire (distance d'un point à la ligne épipolaire associée) de l'image rectifiée. Dans le même temps, l'espace des contraintes limite les distorsions projectives de l'image inhérentes au processus de rectification. Ces contraintes appliquées sur des points choisis de l'image permettent de conserver la taille, le rapport d'aspect et l'orthogonalité de l'image.

2.4 Résultats et comparaisons

La méthode proposée a été évaluée sur plusieurs types d'images. Pour commencer, la méthode a été appliquée sur des images stéréoscopiques visibles pour établir une base de référence. Les résultats sont ensuite comparés aux résultats obtenus à l'aide de différentes méthodes issues de [Mallon 2005]. Puis, la méthode a été appliquée à des images stéréoscopiques infrarouges issues de notre application. Enfin, la méthode a été testée sur des images visibles quelconques de la communauté pour définir les limites de son domaine de validité.

2.4.1 Critères d'évaluation

La qualité de la rectification d'images est tout d'abord analysée par la précision de rectification E_r , représentée par la moyenne (*moy*) et par l'écart-type (σ) de la distance de chaque point d'une image à la ligne épipolaire correspondante dans l'autre image. Ce critère est comparé à l'erreur sur matrice Fondamentale E_f . Pour les méthodes de rectification basées sur un calcul préalable de la matrice Fondamentale, cette erreur représente l'erreur de rectification minimum atteignable. D'autres critères sont calculés comme le rapport d'aspect E_a défini par l'équation (2.38) et l'orthogonalité E_o (voir l'équation (2.40)). Idéalement, le rapport d'aspect E_a doit être de 1 et de l'orthogonalité E_o doit être de 90 °.

2.4.2 Évaluation sur des images stéréoscopiques visibles

L'évaluation des méthodes proposées est d'abord basée sur une comparaison avec l'algorithme de Mallon présenté dans [Mallon 2005]. Cette analyse comparative a été effectuée sur des images visibles de résolution 640×480 , acquises dans des contextes différents : intérieur, extérieur... Les images ont été acquises par un banc de stéréovision classique. Les images sont disponibles sur le site web <http://www.eeng.dcu.ie/~vsl/vsgcode.html>. Le tableau 2.1 donne les résultats de la comparaison entre cinq méthodes différentes (Mallon [Mallon 2005], Loop and Zhang [Loop 1999], Hartley [Hartley 1999]) et les méthodes proposées ci-dessus. La méthode appelée "Méthode 1" (resp. "Méthode 2") est basée sur les contraintes mises en oeuvre en utilisant la formule analytique (resp. géométrique). Pour tous ces tests, la tolérance de la déformation de la taille, du rapport d'aspect et de l'orthogonalité des images, a été maintenue constante et égale à 4%. De même la demi largeur des anneaux reste également inchangée pour tous les tests. Elle est égale à un cinquantième du rayon de l'anneau.

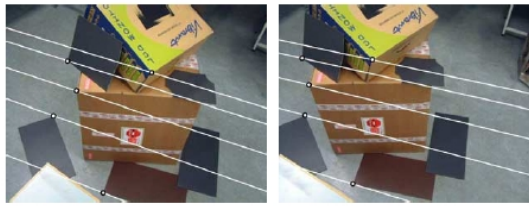
2.4.2.1 Erreur de rectification

La moyenne (resp. l'écart-type) sur l'erreur de rectification des méthodes proposées est comprise entre 0.11 (resp. 0.06) et 0.83 (resp. 0.62). En comparaison avec les autres méthodes où, la moyenne (resp. l'écart-type) de l'erreur de rectification

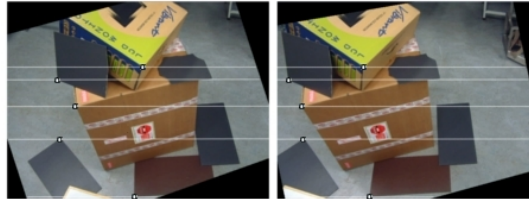
Échantillon	Erreur Fond. E_f		Méthode	Orthogonalité E_o		Rapport d'aspect E_a		Erreur Rect. E_r	
	<i>moy</i>	σ		H'	H	H'	H	<i>moy</i>	σ
Boxes	0.5068	0.3630	Méthode 1	89.44	90.66	0.9897	1.0105	0.11	0.06
			Méthode 2	88.55	89.41	0.9785	0.9936	0.19	0.10
			Mallon	88.78	89.33	0.9785	0.9889	0.44	0.33
			Loop	97.77	95.69	1.1279	1.0900	4.35	9.20
			Hartley	86.56	94.99	0.9412	1.0846	33.36	8.65
Slate	0.5987	0.4593	Méthode 1	90.16	90.30	1.0069	1.0094	0.41	0.30
			Méthode 2	88.97	88.75	0.9861	0.9827	0.26	0.17
			Mallon	89.12	89.13	0.9852	0.9855	0.59	0.56
			Loop	37.29	37.15	0.2698	0.2805	1.14	3.84
			Hartley	89.96	88.54	1.0000	0.9769	2.27	5.18
Roof	1.6422	1.7085	Méthode 1	90.94	89.36	1.0158	0.9894	0.83	0.62
			Méthode 2	88.41	89.58	0.9737	0.9936	0.72	0.57
			Mallon	88.35	88.23	1.1077	0.9700	1.96	2.95
			Loop	69.28	87.70	0.6665	1.0497	0.84	11.01
			Hartley	122.77	80.89	1.5256	0.8552	11.89	18.15
Arch	0.3244	0.3123	Méthode 1	89.67	89.88	0.9955	0.9983	0.59	0.38
			Méthode 2	90.27	88.79	1.0045	0.9800	0.17	0.16
			Mallon	91.22	90.26	1.0175	1.0045	0.22	0.33
			Loop	95.40	98.94	1.0991	1.1662	131.3	20.63
			Hartley	100.74	93.05	1.2077	1.0546	39.21	13.85
Yard	0.6365	0.4776	Méthode 1	89.92	89.01	1.0046	0.9975	0.30	0.19
			Méthode 2	89.30	89.02	0.9881	0.9847	0.30	0.19
			Mallon	89.91	90.26	0.9987	1.0045	0.53	0.54
			Loop	133.62	134.27	2.1477	2.4045	8.91	13.19
			Hartley	101.95	91.91	1.2303	1.0335	48.19	11.49
Drive	0.5684	0.7568	Méthode 1	89.00	89.00	0.9847	0.9842	0.49	0.49
			Méthode 2	90.61	89.32	1.0128	0.9876	0.55	0.44
			Mallon	90.44	90.12	1.0060	1.0021	0.18	0.91
			Loop	98.73	101.42	1.1541	1.2052	10.41	3.24
			Hartley	107.66	90.87	1.3491	1.015	3.57	3.43

TAB. 2.1 – Comparaison des méthodes proposées sur des images visibles.

varie de 0.18 (resp. 0.33) à 48.1 (resp. 20.63). La moyenne et l'écart-type de l'erreur de rectification E_r des méthodes proposées sont inférieurs ou comparables à la moyenne et l'écart-type des autres méthodes. En outre, si l'on compare la moyenne et l'écart-type de l'erreur de la matrice Fondamentale E_f , l'erreur de rectification E_r issue de nos méthodes est toujours inférieure. La matrice fondamentale n'est pas explicitement définie dans la fonction coût de l'équation (2.37) car elle est remplacée par l'équation (2.6). L'expression de cette dernière permet de forcer le rang de la matrice à 2. Cette contrainte n'est généralement pas respectée lors du calcul de la matrice Fondamentale. L'erreur de rectification est alors plus faible que l'erreur de la matrice Fondamentale.



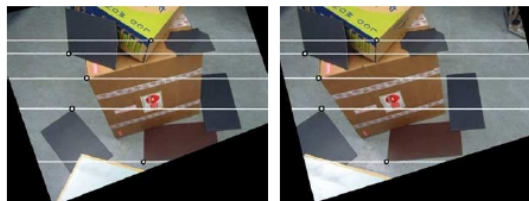
a) images originales.



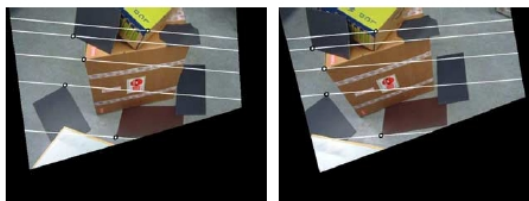
b) images rectifiées avec la méthode 1.



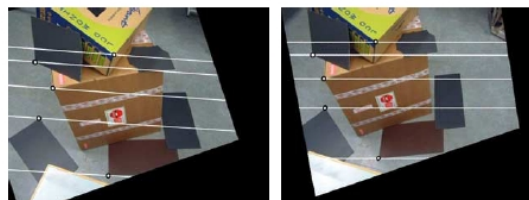
c) images rectifiées avec la méthode 2.



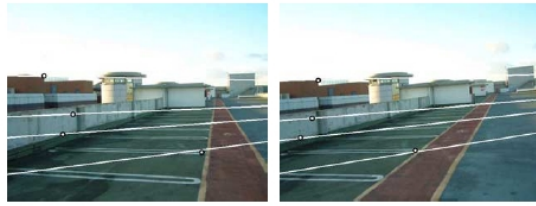
d) images rectifiées avec Mallon.



e) images rectifiées avec Loop et Zhang.



f) images rectifiées avec Hartley.



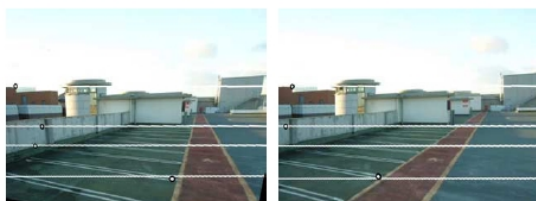
a) images originales



b) images rectifiées avec la méthode 1.



c) images rectifiées avec la méthode 2.



d) images rectifiées avec Mallon



e) images rectifiées avec Loop et Zhang



f) images rectifiées avec Hartley

FIG. 2.7 – Images de l'exemple "Boxes".

FIG. 2.8 – Images de l'exemple "Roof".



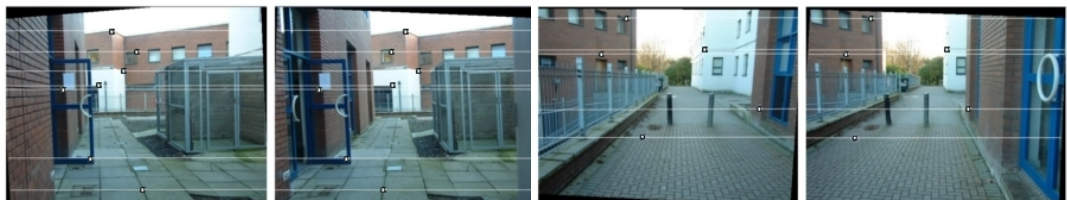
a) images originales

a) images originales



b) images rectifiées avec la méthode 1.

b) images rectifiées avec la méthode 1.



c) images rectifiées avec la méthode 2.

c) images rectifiées avec la méthode 2.



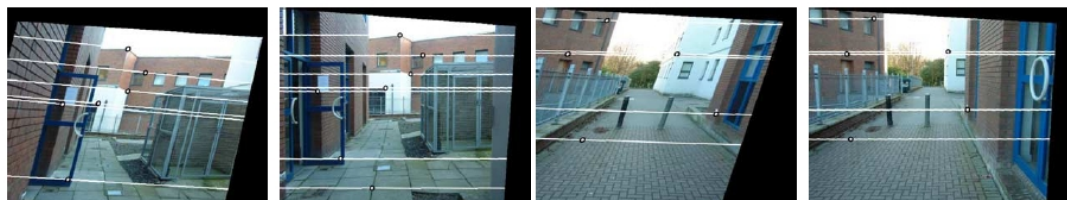
d) images rectifiées avec Mallon

d) images rectifiées avec Mallon



e) images rectifiées avec Loop et Zhang

e) images rectifiées avec Loop et Zhang



f) images rectifiées avec Hartley

f) images rectifiées avec Hartley

FIG. 2.9 – Images de l'exemple "Yard".

FIG. 2.10 – Images de l'exemple "Drive".

2.4.2.2 Réduction des distortions

L'erreur d'orthogonalité maximale des méthodes proposées est de 1.59° pour les deux homographies. Notez que cette erreur est, pour la meilleure des autres méthodes, de 1.77° . En comparaison, les méthodes d'Hartley ou de Loop et Zhang introduisent une erreur d'angle plus importante. De même, l'erreur maximale de rapport d'aspect pour les méthodes proposées est de 0.0263 alors qu'elle est de 0.1077 pour la meilleure méthode alternative. Là encore, les méthodes d'Hartley et de Loop et Zhang introduisent des erreurs plus élevées.

2.4.3 Évaluation sur des images stéréo infrarouges

Les résultats des tests présentés ci-dessous ont été obtenus à partir d'images infrarouges (fig. 2.11). L'acquisition provient d'un banc de stéréovision classique composé de deux caméras infrarouges (bande spectrale $8\text{-}12\mu\text{m}$) avec une faible résolution spatiale (164×128 pixels). Les deux caméras ont des paramètres intrinsèques similaires et les distortions dues à la lentille sont négligeables. Les performances des méthodes proposées ont été comparées avec la méthode la plus efficace du tableau 2.1 sur un petit nombre de points appariés. Les résultats sont exposés dans le tableau 2.2.

Échantillon	Erreur Fond. E_f		Méthode	Orthogonalité E_o		Rapport d'aspect E_a		Erreur Rect. E_r	
	mo	σ		H'	H	H'	H	mo	σ
Objet chaud	1.34	1.19	Méthode 1	90.99	90.76	1.0999	0.900	0.92	0.75
			Méthode 2	90.14	90.29	0.9985	0.9992	1.13	0.82
			Mallon	110.6	108.3	1.4048	1.2836	1.34	1.19
Objet transparent	3.19	3.69	Méthode 1	90.99	89.00	1.0998	1.0197	1.20	0.89
			Méthode 2	90.57	89.26	1.0068	0.9848	1.35	0.86
			Mallon	154.7	156.0	3.4962	2.1662	3.19	3.69

TAB. 2.2 – Comparaison des méthodes proposées sur des images stéréo-infrarouges (Objet chaud et Transparent).

Ces exemples montrent les limites de la méthode de Mallon et des autres méthodes basées sur une pré-estimation de la matrice Fondamentale. En effet, l'utilisation d'images infrarouges ne permet pas d'extraire un ensemble de points, nécessaires à la rectification, suffisamment précis. L'erreur lors du calcul de la matrice Fondamentale est donc plus importante. A titre de comparaison, sur l'exemple de l'objet transparent, la moyenne (resp. l'écart-type) sur l'erreur de rectification des méthodes proposées est de 1.20 (0.89), soit trois fois plus faible que la méthode de Mallon. De plus, dans les deux exemples sur les images infrarouges, les distortions sont beaucoup plus faibles avec les méthodes proposées. Nous avons moins d'un degré d'erreur d'orthogonalité et une erreur de rapport d'aspect de 0.1, comparé à la technique Mallon, qui dans le meilleur des cas obtient 8.3° d'erreur d'orthogonalité et une erreur de 0.28 pour le rapport d'aspect. La qualité des résultats pour les deux

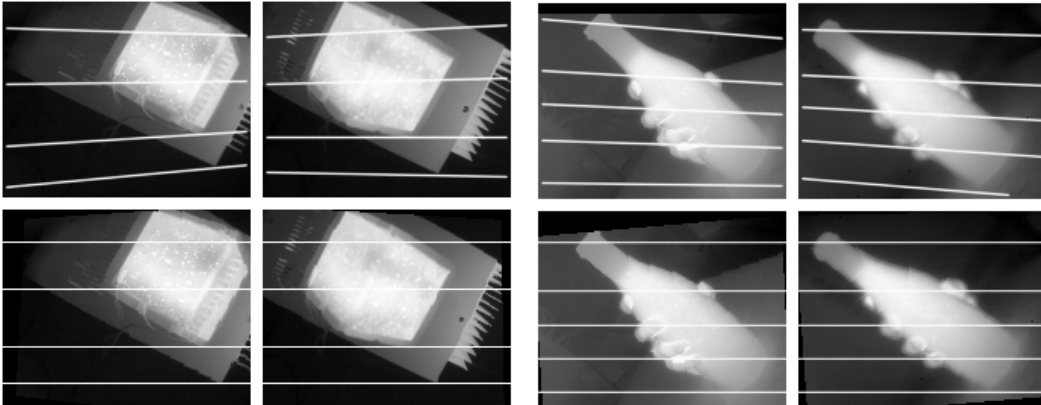


FIG. 2.11 – Rectification des images acquises sur un objet chaud et un objet transparent.

méthodes proposées est comparable. Cependant, le choix de l'intervalle de tolérance, pour les contraintes de la première méthode proposée, semble être plus instinctif. En effet, le choix d'une valeur d'orthogonalité ou d'une valeur de rapport d'aspect est plus intuitif que le choix de la largeur des bandes des anneaux.

2.4.4 Les limites de l'algorithme

Afin d'évaluer le comportement des deux approches sur des images quelconques, les méthodes ont été appliquées à des images visibles non stéréoscopiques avec un changement d'échelle, une translation avant et une rotation due à de forts déplacements de la caméra (images de www.robots.ox.ac.uk/~vgg/data/data-mview.html). Les résultats du tableau 2.2 proviennent de deux séries d'images : la première est "Library" et la seconde est "Merton".

Échantillon	Erreur Fond. E_f		Méthode	Orthogonalité E_o		Rapport d'aspect E_a		Erreur Rect. E_r	
	<i>moy</i>	σ		H'	H	H'	H	<i>moy</i>	σ
Library_23	1.09	0.69	Méthode 1	89.91	90.99	0.9992	1.0778	9.37	5.00
			Méthode 2	90.45	90.24	1.0089	0.9996	13.55	6.74
			Mallon	90.01	94.14	1.0002	1.0436	1.09	0.69
Merton3_23	0.29	0.25	Méthode 1	90.99	89.00	1.0268	0.9893	16.46	10.19
			Méthode 2	91.43	87.99	1.0262	0.9670	18.44	12.62
			Mallon	1.019	1.058	1.1782	0.9098	0.29	0.25

TAB. 2.3 – Comparaison des méthodes proposées sur des images non stéréoscopiques (Library_23 and Merton3_23).

L'erreur de rectification montre que les méthodes proposées ne fonctionnent pas

très bien en présence de déplacements importants. En comparaison avec la méthode de Mallon, nos méthodes ont moins de distorsions géométriques car les contraintes permettent de les bloquer, cependant l'erreur de rectification est plus beaucoup plus importante en raison d'une difficile convergence de l'algorithme. Il est à noter que ces résultats ont été obtenus avec la même tolérance sur les contraintes que dans les tests précédents. En conclusion, dans la majorité des cas, la géométrie qui résulte de telles situations, comme un changement d'échelle ou d'importantes translations, n'est généralement pas rencontrée dans une configuration stéréoscopique.

2.5 Conclusions

Ce chapitre décrit une méthode de rectification d'images stéréoscopiques infrarouges non calibrées n'utilisant qu'un petit nombre de points appariés avec une faible précision. Les homographies nécessaires à la rectification sont estimées en une seule étape à partir d'un algorithme d'optimisation sous contraintes. La fonction coût ne repose pas sur l'estimation au préalable de la matrice Fondamentale à la différence des autres méthodes. Par conséquent, cette approche limite le nombre de processus itératifs. Ceux-ci sont, en général, entachés d'erreurs lorsqu'ils utilisent un faible nombre de points appariés. La méthode proposée minimise directement la distance des points à la ligne épipolaire associée. De plus la déformation projective est contrôlée par un ensemble de contraintes, analytiques ou géométriques, sur l'image. Ces contraintes sont basées sur la limitation de la perte ou de la création locale de pixels. Elles sont exprimées comme des critères physiques de la structure de l'image : le rapport d'aspect, la taille et l'orthogonalité. Ceci permet d'avoir une quantification physique du niveau de distorsion. Ces contraintes sont imposées aux deux homographies et sont implémentées en utilisant soit une expression analytique soit une forme géométrique. La forme géométrique est basée sur des anneaux dans lesquels les points sélectionnés doivent rester après rectification. Finalement, il a été montré que lors de l'utilisation d'un faible nombre de points appariés, estimer les homographies de rectification de manière directe dans un processus d'optimisation appliqué aux deux images prenant en considération des contraintes physiques, était plus pertinent qu'une estimation préalable de la matrice Fondamentale. Nous avons montré que

- l'utilisation des contraintes ne réduit pas l'espace des solutions. En effet l'erreur de rectification finale est toujours plus petite que l'erreur sur l'estimation de la matrice Fondamentale : la solution est alors proche de l'optimum.
- l'espace des contraintes est assez restreint pour limiter les distorsions projectives.

De plus, l'initialisation de notre méthode est triviale car elle utilise l'identité. En effet, l'image originale est une image sans distorsion et est donc une initialisation viable comprise dans l'espace de contraintes. L'identité fournit une bonne estimation initiale et garantit toujours la convergence vers une solution physiquement acceptable.

Auto-calibrage d'un système robot-caméra infrarouge

Sommaire

3.1	Définition et modélisation du système robot-caméra	40
3.1.1	Géométrie du système	40
3.1.2	Transformation pince-caméra	42
3.1.3	Discussions	43
3.2	Etat de l'art sur le calibrage Hand-Eye	43
3.2.1	Les méthodes classiques avec mires	43
3.2.2	Les méthodes sans mire : auto-calibrage	45
3.2.3	Conclusion	47
3.3	Proposition d'une méthode d'auto-calibrage basée sur la minimisation de distances dans l'espace 3D	49
3.3.1	Principe	49
3.3.2	Expression du point P_c dans le repère de la caméra C	50
3.3.3	Problème à minimiser basé sur la distance entre les points 3D	52
3.3.4	Conclusion	53
3.4	Proposition d'une méthode d'auto-calibrage basée sur la minimisation de distances dans l'espace 2D	53
3.4.1	Principe	53
3.4.2	Système monoculaire	54
3.4.3	Système de stéréovision	55
3.4.4	Système de stéréovision rectifié	57
3.4.5	Conclusion	58
3.5	Evaluation des méthodes d'auto-calibrage sur des données de synthèse	58
3.5.1	Méthodologie	59
3.5.2	Contexte expérimental	59
3.5.3	Méthode de minimisation et initialisation	61
3.5.4	Critères d'évaluation	62
3.5.5	Évaluation sur des données de synthèse non bruitées	63
3.5.6	Influence du nombre de points	66
3.5.7	Influence du bruit sur les points des images	69
3.5.8	Influence de la répartition des points	72
3.5.9	Conclusion	73

3.6 Performances des méthodes de calibrage sélectionnées sur des données réelles	73
3.6.1 Contexte expérimental	74
3.6.2 Données issues d'images visibles	74
3.6.3 Données issues d'images infrarouges	75
3.7 Conclusion	76

Le système robotique considéré est composé d'une caméra ou d'un banc de deux caméras infrarouges montées en stéréo vision et d'un bras cartésien à six axes. Calibrer (ou étalonner) un système de vision consiste à établir la relation entre les coordonnées d'un point dans le repère de l'image et ses coordonnées dans le repère du monde. Dans le cas d'un système monoculaire, le calibrage consiste à identifier ses paramètres intrinsèques et sa position et son orientation par rapport au référentiel du monde. Pour un système de stéréovision, il s'agit de déterminer les paramètres intrinsèques de chaque caméra, la position et l'orientation relatives entre les deux caméras et la position et l'orientation d'une des deux caméras par rapport au référentiel du monde. Le modèle de paramètres choisis pour représenter la caméra est le modèle sténopé. Il représente la relation perspective d'un objectif par la projection centrale. Dans ce modèle, tous les rayons lumineux passent par le centre optique intersectant le plan image sans changer de direction. Le calibrage d'une caméra seule ou d'un banc de stéréo vision est bien connu à l'aide d'une mire, calibrage dit "fort", ou sans mire, dit auto-calibrage de la caméra. Dans le calibrage du système robot-caméra, il faut également considérer le déplacement, entre deux positions, de l'organe terminal du robot, appelé pince. Celui-ci est généralement connu et il est modélisé par une matrice de déplacement composée par des rotations et des translations. En revanche, dans ce système robot-caméra, la transformation entre la caméra et l'organe terminal du robot est généralement inconnue. Cette transformation est modélisée par une matrice composée d'une rotation et d'une translation. Le calibrage d'un système robot-caméra consiste donc à identifier les paramètres du modèle de la caméra et les paramètres de la transformation pince-caméra. Ce problème est traditionnellement résumé par la terminologie étalonnage pince-caméra. Nous trouvons également le terme calibrage Hand-Eye (la caméra, dénommée "eye" est fixée à l'organe terminal du robot, appelé "hand").

Dans la littérature, les approches classiques du calibrage pince-caméra consiste à supposer que une ou plusieurs caméras du système sont déjà préalablement calibrées et que le déplacement de la pince est connu. Le principe consiste alors à résoudre le système ayant comme données d'entrées : le déplacement de la caméra autour de la mire (donné par le calibrage de la caméra) et le déplacement de la pince (supposé connu); et comme inconnue la transformation rigide entre la pince et la caméra. Généralement, ces méthodes reposent sur une mire (objet étalon) qui est constituée d'un ensemble de points dont les coordonnées sont parfaitement connues dans le repère objet. Chaque point de la mire se projette dans l'image et ses coordonnées sont mesurées dans le repère image. Connaissant les points de la mire et leurs projetés dans l'image, une méthode d'optimisation non linéaire permet de résoudre le système. Dans la section 3.2, les différentes stratégies sont présentées pour estimer les paramètres des translations et des rotations de la transformation pince-caméra de manière séparée ou simultanément.

Dans certains cas, milieux restreints ou hostiles à l'homme (tore de fusion, ...), il est impossible d'introduire une mire de calibrage. Une première solution consiste à auto-calibrer préalablement les caméras. De nombreux articles abordent ce problème d'identifier les paramètres intrinsèques de la caméra uniquement à partir de

l'observation d'une scène quelconque. Ensuite, comme précédemment les paramètres de la transformation pince-caméra peuvent être identifiés. La deuxième solution serait de supposer des caméras non préalablement calibrées et de déterminer en même temps les paramètres intrinsèques de la caméra et ceux de la transformation pince-caméra. Comme le montre la section 3.2 peu d'articles présentent des solutions pour calibrer simultanément les paramètres de la caméra (paramètres intrinsèques) et les paramètres de la transformation pince-caméra sans l'utilisation d'une mire. De plus, l'utilisation de caméras infrarouges qui fournissent des images peu texturées et de faible résolution spatiale, pose le problème de calibrer le système complet avec un faible nombre de points.

En conclusion, le problème posé dans ce chapitre est de calibrer le système robot-caméra complet, paramètres intrinsèques des caméras et paramètres de la transformation pince-caméra, avec des caméras infrarouges, non préalablement calibrées, de manière "automatique" sans avoir recours à des mires. Le calibrage complet du système robot-caméra implique vingt paramètres à estimer. Nous présenterons tout l'intérêt de supposer au préalable que les caméras sont rectifiées pour diminuer ce nombre de paramètres.

Nous proposerons ainsi deux méthodes différentes afin de résoudre ce problème. La première méthode est basée uniquement sur l'objet observé et plus précisément sur sa géométrie 3D. La seconde est basée sur des propriétés internes du banc de stéréovision, soit, sur une décomposition de la matrice Fondamentale. Chacune de ses méthodes traite le problème dans son ensemble et optimise en une seule opération les paramètres intrinsèques des caméras et les paramètres de la transformation pince-caméra.

La section 3.1 décrit le modèle et les paramètres du système robotique (système de positionnement et système de vision). La section 3.2 présente une étude bibliographique des méthodes de calibrage d'un système de vision mono et stéréo caméra et d'un système robotique et vision qui sont beaucoup moins nombreuses. La section 3.4 propose une méthode originale d'auto-calibrage du système robotique global (système de positionnement et système de vision). Finalement, la section 3.5 établit une évaluation de la méthode proposée.

3.1 Définition et modélisation du système robot-caméra

Nous décrivons dans cette section les différentes transformations reliant les divers éléments du robot. Nous présentons comment la transformation pince-caméra relie les déplacements de l'ensemble du système et comment se pose le problème de l'étalonnage pince-caméra.

3.1.1 Géométrie du système

Nous allons tout d'abord rappeler la notation classique adoptée pour représenter l'ensemble des transformations nécessaires à la résolution du problème de l'étalonnage pince-caméra. La transformation notée \mathbf{M}_{AB} permet de calculer les coordon-

nées d'un point $3D$ dans le repère A par rapport aux coordonnées de ce même point connues dans le repère B . Si on considère tous les repères dans un système robotique, ici notre système robot-caméra, on peut construire un graphe des repères dans lequel deux repères A et B sont liés par une arête orientée de A vers B , annotée par la transformation \mathbf{M}_{AB} , de sorte que :

$$A \xrightarrow{\mathbf{M}_{AB}} B \text{ permet d'écrire } P_A = \mathbf{M}_{AB}P_B \quad (3.1)$$

Dans ce graphe, il est possible de calculer la transformation entre deux repères quelconques, en multipliant les transformations annotant les arêtes d'un chemin entre les noeuds correspondants ; si une arête est suivie à contresens, il faut inverser cette transformation :

$$A \xrightarrow{\mathbf{M}_{AB}} B \xrightarrow{\mathbf{M}_{BC}} C : \mathbf{M}_{AC} = \mathbf{M}_{AB}\mathbf{M}_{BC} \quad (3.2)$$

$$A \xrightarrow{\mathbf{M}_{AB}} B \xleftarrow{\mathbf{M}_{CB}} C : \mathbf{M}_{AC} = \mathbf{M}_{AB}\mathbf{M}_{CB}^{-1} \quad (3.3)$$

La transformation \mathbf{M}_{AB} est définie par une matrice de rotation 3×3 \mathbf{R}_{AB} et par un vecteur de translation \mathbf{t}_{AB} . La matrice ainsi formée est notée de la manière suivante :

$$\mathbf{M}_{AB} = \begin{pmatrix} \mathbf{R}_{AB} & \mathbf{t}_{AB} \\ 0_3 & 1 \end{pmatrix}. \quad (3.4)$$

Cette matrice \mathbf{M}_{AB} est donc représentée par six paramètres (trois paramètres de rotation et trois de translation). Une telle transformation donne aussi la position du repère B par rapport au repère A : les colonnes de la matrice de rotation \mathbf{R}_{AB} sont les vecteurs unitaires des axes Bx , By et Bz exprimés dans le repère A , tandis que \mathbf{t}_{AB} donne la position de l'origine de B dans le repère A .

La figure 3.1 représente le graphe des repères considéré pour un système avec un robot et le banc de stéréovision dans deux positions différentes.

Ces transformations sont définies de la manière suivante :

- $\mathbf{M}_{P_{ij}}$ représente le déplacement de l'organe terminal du robot de la position j à la position i . Cette transformation est composée de la rotation $\mathbf{R}_{P_{ij}}$ et de la translation $\mathbf{t}_{P_{ij}}$.
- $\mathbf{M}_{C_{ij}}$ (resp. $\mathbf{M}_{C'_{ij}}$) représente le déplacement de la caméra C (resp. C') de la position j à la position i . Cette transformation est composée de la rotation $\mathbf{R}_{C_{ij}}$ (resp. $\mathbf{R}_{C'_{ij}}$) et de la translation $\mathbf{t}_{C_{ij}}$ (resp. $\mathbf{t}_{C'_{ij}}$).
- \mathbf{M}_{PC} représente la position de la caméra C par rapport à l'organe terminal. Cette transformation est composée de la rotation \mathbf{R}_{PC} et de la translation \mathbf{t}_{PC} . C'est la transformation rigide *Hand-Eye* ou Pince-Caméra qu'il convient de calibrer.
- \mathbf{M}_{ext} représente la matrice de passage allant du repère de la caméra C' à la caméra C . Cette transformation est composée de la rotation \mathbf{R}_{ext} et de la translation \mathbf{t}_{ext} . Elle donne la position de C vis-à-vis de C' .

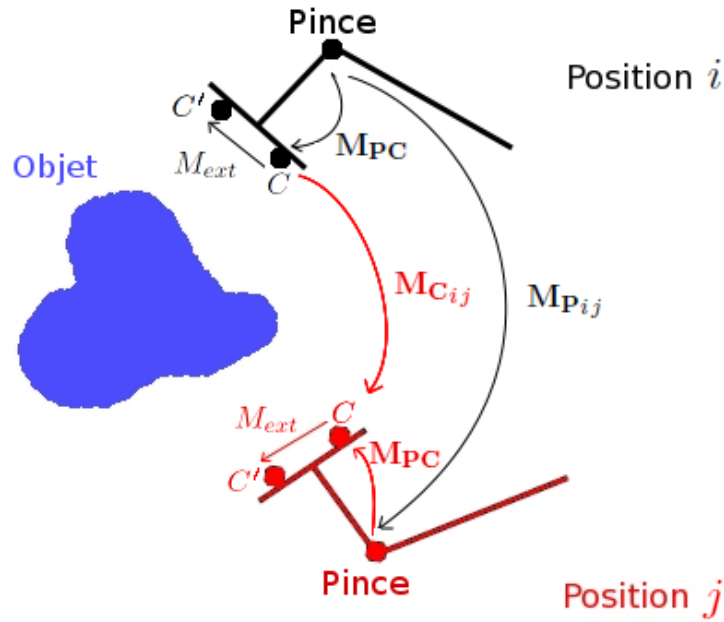


FIG. 3.1 – Géométrie du système et orientation des transformations.

Il est usuel de définir dans une application de vision, un repère du monde, repère unique dans lequel sont exprimées les coordonnées des points reconstruits par la vision. Le choix de ce repère dépend de l'application ; par exemple, il peut être lié à l'objet, à la base du robot, ou encore à une position initiale de l'organe terminal.

3.1.2 Transformation pince-caméra

Comme le montre la figure 3.1, lorsque la pince se déplace avec une transformation rigide \mathbf{M}_{Pij} entre les étapes j et i , la caméra, elle, se déplace selon la transformation rigide \mathbf{M}_{Cij} . Ces deux transformations sont liées par la transformation Pince-Caméra, \mathbf{M}_{PC} , selon la relation (3.5).

$$\mathbf{M}_{Pij}\mathbf{M}_{PC} = \mathbf{M}_{PC}\mathbf{M}_{Cij} \quad (3.5)$$

La matrice Pince-Caméra \mathbf{M}_{PC} à déterminer est complètement décrite par six paramètres, trois paramètres de rotation et trois de translation.

Dans le cas d'un système de stéréovision, il faut déterminer le plan image de référence pour cette transformation. Nous choisissons la transformation rigide comprise entre l'organe terminal du robot et la caméra C comme transformation Hand-Eye du système stéréovision.

3.1.3 Discussions

L'équation (3.5) est à résoudre connaissant le déplacement du robot \mathbf{M}_P grâce au modèle cinématique du robot et le déplacement de la caméra \mathbf{M}_C par calibrage de la caméra ou du système de stéréovision. L'approche pour la résolution de cette équation appelle plusieurs questions. Compte tenu des contraintes fixées par le projet, le calibrage sans mire est-il suffisamment précis ? Ensuite, il reste le problème du choix du système de vision, monoculaire ou stéréovision. Un système monoculaire reste le plus simple, cependant, un système de stéréovision présente le double avantage : d'une part de pouvoir trianguler directement les points des images et d'autre part d'augmenter le nombre de points d'extraits dans des images infrarouges. Dans ce dernier cas, le calibrage simultané modèle complet du système robot-caméra implique donc vingt paramètres à estimer. Une rectification préalable peut présenter un intérêt pour diminuer ce nombre de paramètres.

La section suivante établit un état de l'art des stratégies couramment mises en oeuvre et elle propose une solution compatible avec les exigences de l'application.

3.2 Etat de l'art sur le calibrage Hand-Eye

Cette section établit un bref état de l'art des méthodes les plus courantes de calibrage de la transformation pince-caméra avec une mire et de quelques rares méthodes d'auto-calibrage sans mire. Les premières se distinguent par la représentation des mouvements, des rotations et des translations qui les composent (angle, quaternion) et par les méthodes de résolution linéaire ou non linéaire. Si l'auto-calibrage est une problématique bien connue, avec une bibliographie fournie, en vision classique, elle est, à l'inverse beaucoup moins utilisée pour la détermination des paramètres de la matrice de Pince-Caméra. Pourtant, quelque soit l'application, il est intéressant que le système robotique soit le plus autonome possible. Pour que cette autonomie soit totale, le calibrage de la transformation Pince-Caméra doit être réalisé d'une part sans mire de calibrage mais d'autre part sans connaissance a priori de sa valeur. Nous présentons les méthodes de la littérature qui abordent ces deux problèmes.

3.2.1 Les méthodes classiques avec mires

3.2.1.1 Représentation axe/angle - Résolution linéaire

Les méthodes présentées par Tsai et Lenz [Tsai 1988], ou Shiu et Ahmad [Shiu 1989], décomposent l'équation (3.5) en deux équations : l'une qui contient uniquement les rotations et l'autre les translations, selon la relation (3.6) :

$$\mathbf{R}_{P_{ij}}\mathbf{R}_{PC} = \mathbf{R}_{PC}\mathbf{R}_{C_{ij}} \quad (3.6)$$

$$\mathbf{R}_{P_{ij}}\mathbf{t}_{PC} - \mathbf{t}_{PC} = \mathbf{R}_{PC}\mathbf{t}_{C_{ij}} - \mathbf{t}_{P_{ij}} \quad (3.7)$$

La première équation est résolue par un algorithme classique de minimisation par

moindres carrés linéaires avec une représentation des rotations en vecteur axe/angle :

$$[\overrightarrow{n_{C_{ij}}} + \overrightarrow{n_{P_{ij}}}]_{\times} \overrightarrow{n_{PC}} = \overrightarrow{n_{C_{ij}}} - \overrightarrow{n_{P_{ij}}} \quad (3.8)$$

$$\theta_{PC} = 2 \arctan(\|\overrightarrow{n_{PC}}\|) \quad (3.9)$$

où $\overrightarrow{n_{\phi}}$ représente le vecteur unitaire de l'axe et θ l'angle de la rotation \mathbf{R}_{ϕ} . La fonction $[\]_{\times}$ renvoie la matrice antisymétrique associée à un vecteur de dimension 3 tel que $[t]_{\times} x = t \times x$ avec \times représentant le produit vectoriel. Une fois la rotation \mathbf{R}_{PC} connue, sa valeur est remplacée dans la seconde équation qui est un système linéaire.

Cette approche présente l'avantage d'être simple mais elle demande d'amples déplacements de rotation pour pouvoir en extraire les axes. De plus, une estimation par les moindres carrés est sensible aux mesures aberrantes.

3.2.1.2 Représentation quaternion - Résolution linéaire et non linéaire

La méthode présentée dans l'article [Daniilidis 1996] est basée sur la représentation des transformations rigides à l'aide des quaternions duaux unitaires. Ces derniers sont formés par un couple $\tilde{q}_{\chi} = (q_{\chi}, q'_{\chi})$ de quaternions, avec q_{χ} le quaternion associé à la rotation \mathbf{R}_{χ} et q'_{χ} , son dual, lui aussi un quaternion, mais il représente la translation \mathbf{t}_{χ} de la transformation rigide \mathbf{M}_{χ} . Pour représenter cette transformation le quaternion dual doit être unitaire, ce qui implique :

$$q_{\chi}^T q_{\chi} = 1 \quad \text{and} \quad q_{\chi}^T \overline{q'_{\chi}} = 0 \quad (3.10)$$

En intégrant ces nouvelles notations dans l'équation (3.5), on obtient le système d'équation (3.11).

$$\begin{pmatrix} q_{C_{ij}} - q_{P_{ij}} & [q_{C_{ij}} + q_{P_{ij}}]_{\times} & 0_{3 \times 1} & 0_{3 \times 3} \\ q'_{C_{ij}} - q'_{P_{ij}} & [q'_{C_{ij}} + q'_{P_{ij}}]_{\times} & q_{C_{ij}} - q_{P_{ij}} & [q_{C_{ij}} + q_{P_{ij}}]_{\times} \end{pmatrix} \begin{pmatrix} q_{PC} \\ q'_{PC} \end{pmatrix} = 0 \quad (3.11)$$

Si nous supposons n déplacements du système robotique, la matrice obtenue est de dimension $6n \times 8$ et de rang 6 dont la solution peut être obtenue par décomposition en valeurs singulières (SVD) (cf. articles [Zhuang 1992] et [Chou 1991]).

Dans l'article [R.Horand 1995] une méthode minimisant une fonction coût non linéaire est proposée :

$$\begin{aligned} f(q_{PC}, \mathbf{t}_{PC}) &= \lambda_1 \sum_{i=1}^n |n_{C_{ij}} - q_{PC} n_{P_{ij}} \overline{q_{PC}}|^2 \\ &+ \lambda_2 \sum_{i=1}^n |q_{PC} \mathbf{t}_{P_{ij}} \overline{q_{PC}} - (\mathbf{R}_{C_{ij}} - I) \mathbf{t}_{PC} - \mathbf{t}_{C_{ij}}|^2 \\ &+ \lambda (1 - q_{PC}^T q_{PC})^2 \end{aligned} \quad (3.12)$$

avec q_{PC} le quaternion associé à \mathbf{R}_{PC} et $\overline{q_{PC}}$ son conjugué ; λ est un multiplicateur de Lagrange ; λ_1 et λ_2 sont des coefficients choisis par l'utilisateur et n le nombre déplacements du système robotique. Le premier terme de cette fonction représente l'erreur sur l'orientation Pince-Caméra dans l'espace des quaternions, le second terme représente la translation et le dernier terme est une contrainte due

à l'utilisation des quaternions et permet au quaternion q_{PC} de rester un quaternion unitaire. La représentation par des quaternions permet une résolution facilitée avec les multiplicateurs de Lagrange. Une résolution est également proposée avec l'algorithme de Levenberg-Marquardt. Les auteurs concluent que cette approche de détermination simultanée de la rotation et de la translation améliore l'erreur de reprojection.

3.2.1.3 Discussions

La première représentation de la rotation par "axe/angle" est simple mais sa résolution en deux étapes, d'abord la rotation puis la translation, en présence de bruit, peut engendrer un système mal conditionné et des solutions non valides. De plus, l'erreur obtenue lors du calcul de la rotation est propagée sur l'estimation de la translation. Dans un contexte infrarouge, avec des images bruitées, une représentation des matrices de rotations par des quaternions serait à privilégier pour résoudre le problème en une seule étape. De plus, une minimisation non-linéaire permet une grande précision de résolution à condition que l'initialisation soit proche de la solution. Il faut donc une estimée initiale correcte de la matrice Pince-Caméra. Enfin, nous retiendrons que l'ensemble des méthodes avec des représentations réduites des rotations (axe/angle ou quaternions) nécessitent des déplacements avec des rotations de grande amplitude. Ce dernier point a fait l'objet de nombreuses études [Andreff 1999], il en ressort tous types de déplacements contribuent à la détermination de l'orientation relative de la pince par rapport à la caméra alors que seules les rotations participent à la détermination de sa position relative. Nous serons donc vigilants aux types de déplacements du robot.

3.2.2 Les méthodes sans mire : auto-calibrage

Comme cela a été écrit en introduction de cette section, l'autonomie totale du système robotique implique que le calibrage de la transformation pince-caméra soit réalisé d'une part sans mire de calibrage et d'autre part sans connaissance à priori de sa valeur. En l'absence de mire, le déplacement de la caméra ne peut être calculé que par une reconstruction tridimensionnelle qui est dans ce cas à un facteur d'échelle près inconnu. En effet, avec une mire dont on connaît parfaitement la forme et la taille, le calcul de pose nous donne l'orientation/position de la mire par rapport au repère de la caméra, et le déplacement est obtenu par la composition de changement de repère entre deux poses successives. Sans mire, nous observons l'Objet inconnu sous plusieurs points de vues, la reconstruction tridimensionnelle établit la forme géométrique de l'Objet et son orientation/position par rapport au repère de la caméra. Le déplacement est calculé comme précédemment mais comme l'Objet est inconnu il reste une ambiguïté qui est la distance entre l'Objet et la caméra, appelé la profondeur. Un facteur d'échelle apparaît donc dans les positions et les

déplacements de la caméra. Dans ces conditions les équations (3.6) deviennent :

$$\begin{aligned}\mathbf{R}_{P_{ij}}\mathbf{R}_{PC} &= \mathbf{R}_{PC}\mathbf{R}_{C_{ij}} \\ \mathbf{R}_{P_{ij}}\mathbf{t}_{PC} - \mathbf{t}_{PC} &= s_{PC}\mathbf{R}_{PC}\mathbf{t}_{C_{ij}} - \mathbf{t}_{P_{ij}}\end{aligned}\quad (3.13)$$

avec s_{PC} représentant le facteur d'échelle. Nous constatons que le facteur d'échelle n'a pas d'influence sur la rotation. Cette équation peut donc être résolue comme dans le paragraphe précédent. En revanche, le facteur d'échelle influence l'équation de translation. Il faudra donc en tenir compte pour éviter un biais dans la solution.

Cette sous-section présente les différentes stratégies pour calibrer la transformation pince-caméra sans mire et pour considérer le facteur d'échelle.

3.2.2.1 Structure from motion

Les articles Schmidt [Schmidt 2005] ou Andreff [Andreff 1999], montrent comment retrouver les paramètres Pince-Caméra sans l'aide de mire. Dans un premier temps, afin de retrouver les paramètres des caméras et l'estimation de la pose de celles-ci, ils utilisent un algorithme classique de "structure from motion" [Heigl 2004][Christy 1996]. Ensuite, les équations (3.13) sont résolues en deux étapes : tout d'abord la rotation, puis la translation et le facteur d'échelle, de différentes manières. L'article [Andreff 1999] propose une résolution avec au moins deux rotations et une translation indépendantes et à l'aide d'une méthode de résolution basée sur les moindres carrés linéaires. L'article [Schmidt 2005] compare la solution précédente avec une ré-écriture de l'équation (3.13) à l'aide des quaternions duaux unitaires et avec une méthode de résolution non linéaire. Cette approche donne une meilleure précision sur le facteur d'échelle et la translation de matrice de Hand-Eye.

Ces méthodes présentent l'avantage d'être utilisées quelques soit l'objet observé, mais elles sont cependant dépendantes de la qualité de l'algorithme de "structure from motion" utilisé. Elles impliquent également des déplacements de type translation pure du robot. Comme dans le cas précédent, la résolution en deux étapes, d'abord la rotation puis la translation et le facteur d'échelle, peut engendrer une propagation de l'erreur obtenue lors du calcul de la rotation sur l'estimation de la translation et du facteur d'échelle.

3.2.2.2 Suivi d'un point unique

L'article de Jordt [Jordt 2009] propose une méthode basée sur le suivi d'un unique point. La résolution se déroule en trois étapes : trois déplacements suivant des translations pures pour obtenir l'orientation de la caméra, puis trois rotations pures pour estimer la distance du point de calibrage et finalement plusieurs déplacements quelconques afin d'optimiser l'ensemble des paramètres. Le critère optimisé ici est basé comme [qing Wei 1998] sur les méthodes de "bundle adjustment" car il minimise les erreurs de reprojections du point cible dans les différentes position de la caméra.

$$F(P) = \sqrt{\sum_{i \in I} e_i(P)^2} \quad (3.14)$$

avec

$$e_i(P) = \sqrt{(u_i - u_i^*)^2 + (v_i - v_i^*)^2} \quad (3.15)$$

où P représente le vecteur de 11 paramètres composé des six paramètres de la relation Pince-Caméra, des paramètres α_u et α_v de la caméra et de la position du point cible en $3D$. (u_i, v_i) représente la reprojection de la cible dans l'image à la position i et (u_i^*, v_i^*) représente l'estimation de la reprojection de la cible en fonction du déplacement de la caméra et du vecteur de paramètres P . De plus, l'auteur précise qu'il est possible de changer le modèle de la caméra en prenant en compte les distorsions et en ajoutant deux paramètres à optimiser sans modifier la convergence de son algorithme.

Cette méthode a l'avantage d'optimiser l'ensemble des paramètres en une seule étape et donc de minimiser la propagation d'erreur. Cependant la cible choisie devra être visible et extraite de façon très précise lors de chaque déplacement de la caméra. De plus, une initialisation précise est importante afin de converger et la procédure mise en oeuvre pour initialiser le système nécessite des rotations pures difficiles à obtenir.

3.2.2.3 Bundle adjustment

L'article Wei and al. [qing Wei 1998] propose une méthode basée sur du "bundle adjustment" en minimisant les distances entre les reprojections des points $3D$ observés. Son critère de minimisation, non linéaire, prend entre compte les paramètres de distorsions des caméras. Cependant, il utilise une unique vue de référence, afin de pouvoir trianguler, autour de laquelle il effectue des déplacements. De plus, cette méthode nécessite une étape d'initialisation peu triviale durant laquelle les caméras se déplacent suivant des translations pures.

3.2.3 Conclusion

Dans notre projet, nous recherchons une autonomie totale du système robot-caméra et donc le calibrage doit être réalisé sans mire. En plus des motivations classiques, zone observée non accessible ou des raisons évidentes de sécurité, en infrarouge s'ajoute la difficulté de fabriquer des mires d'une grande précision dans la bande spectrale $8-12\mu m$. Les méthodes présentées supposant un calibrage préalable de la caméra sont donc exclues. Pour les méthodes d'auto-calibrage (paramètres caméras et paramètres Pince-Caméra), il existe deux types de méthodes.

Les premières, Schmidt [Schmidt 2005] ou Andreff [Andreff 1999], sont basées sur un auto-calibrage préalable des caméras à un facteur d'échelle près, à l'aide d'un algorithme de "structure from motion", puis d'un calcul classique des paramètres Pince-Caméra. Or l'utilisation d'un algorithme de "structure from motion" dans un contexte bruité et avec peu de points extraits, qui est celui des caméras infrarouges, n'est pas suffisamment précis et favorise ainsi la propagation d'erreurs dans le reste du calcul. De plus, ces méthodes comportent de multiples étapes, et dans un contexte

avec un faible nombre de points extraits, nous avons intérêt à minimiser le nombre d'étapes de calibrage pour minimiser la propagation d'erreurs.

Les autres méthodes, Jordt [Jordt 2009] et Wei [qing Wei 1998], ont une approche très intéressante car elles permettent la résolution de l'ensemble des paramètres en une seule étape. Ces méthodes sont basées sur un algorithme de "bundle adjustment" et minimisent les reprojections entre les points $3D$ observés. Cependant, cette méthode nécessite l'estimation de la position $3D$ de l'ensemble des points, soit trois paramètres supplémentaires par point observé. Wei procède donc à une étape d'initialisation peu triviale afin de pouvoir estimer ces paramètres, ce qui l'oblige à n'utiliser qu'une unique vue de référence. Jordt, quant à lui simplifie ce problème en n'observant qu'un seul et unique point, impliquant donc un point de référence présent dans toutes les vues. Dans notre cas, l'utilisation d'une seule vue, ou d'un point de référence est impossible car d'une part, l'extraction de points sur des images infrarouges est très difficile, et d'autre part, le champ de vision de ces caméras étant très faible, le suivi de points est limité à de très faibles déplacements.

Nous cherchons donc à mettre en place un système disposant d'une grande autonomie basé sur une méthode d'auto-calibrage qui permet d'obtenir, en une seule étape, les paramètres de la transformation pince-caméra ainsi que les paramètres des caméras, le tout sans mire de calibrage et sans connaissance à priori de sa valeur. Compte tenu de la faible résolution spatiale et de la qualité des points extraits, nous supposons que dans l'application finale le robot est équipé d'un capteur de stéréovision infrarouge. Deux approches peuvent être envisagées, l'une en se basant sur une distance à minimiser dans l'espace $3D$ entre deux déplacements et l'autre dans l'espace $2D$ en se basant sur la distance d'un point d'une image à sa droite épipolaire associée dans cette même image après déplacement.

Tout d'abord, l'approche la plus intuitive, comme dans l'article [qing Wei 1998], est de minimiser les distances entre les reprojections des points $3D$ observés comme dans une approche classique de calibrage fort d'un banc stéréovision. Or, l'estimation de la position des points $3D$ ne nous apporte aucune information supplémentaire et augmente considérablement le nombre de paramètres à estimer (3 paramètres supplémentaires par point extrait). Contrairement à Wei [qing Wei 1998], nous proposons donc de minimiser non pas l'erreur de reprojection entre les points $3D$ mais directement l'erreur de reconstruction des points $3D$ entre deux positions successives. Cette minimisation n'est bien évidemment réalisable que dans le cas de l'utilisation d'un banc de stéréovision. L'idée est d'exprimer, pour deux prises de vue, chaque point $3D$ en fonction des paramètres intrinsèques et extrinsèques du banc de stéréovision. Puis, la position $3D$ de ces points est exprimé dans un même repère pour introduire les paramètres Pince-Caméra et les paramètres de déplacement du robot. Finalement, la méthode consiste à minimiser la distance entre chaque points $3D$ correspondants exprimés dans un même repère. L'estimation de l'ensemble des paramètres intrinsèques, extrinsèques et Pince-Caméra est accomplie en une seule étape. Les avantages supplémentaires de cette méthode résident dans le fait de ne pas estimer la position $3D$ des points extraits et de n'avoir aucune prise de vue considérée comme référence. De plus, l'initialisation peut se faire à partir de don-

nées constructeurs approximatives.

Ensuite, la seconde approche est basée sur une décomposition de la matrice Fondamentale. En effet la matrice Fondamentale est la seule information dont nous disposons dans un contexte entièrement non calibré. L'idée est de décomposer cette matrice Fondamentale en matrices de paramètres intrinsèques et en matrice Essentielle pour minimiser ainsi en une seule étape l'ensemble des paramètres. Dans un cadre mono caméra, la matrice Fondamentale considérée est issue de deux vues consécutives de cette même caméra. La matrice Essentielle s'exprime alors à partir des paramètres Pince-Caméra et des paramètres de déplacement du robot. Dans un cadre stéréovision, trois matrices Fondamentales sont considérées. La première est issue de la relation entre les deux caméras du banc stéréo. Elle permet d'exprimer la matrice Essentielle à partir des paramètres extrinsèques. Ensuite, deux autres sont issues du déplacement de chacune des caméras. Elles servent à minimiser l'ensemble des paramètres en une seule étape comme pour le cas mono caméra.

Pour chacune des méthodes proposées et dans le cadre de l'utilisation d'un banc de stéréovision, nous démontrons l'importance de simplifier au maximum le nombre de paramètres à déterminer grâce à une étape préalable de rectification non calibrée.

3.3 Proposition d'une méthode d'auto-calibrage basée sur la minimisation de distances dans l'espace 3D

A l'aide d'un nuage de points 3D acquis lors de chaque déplacement du robot, l'idée est de minimiser la distance entre les points 3D correspondants sur les différentes prises de vues, pour retrouver les paramètres intrinsèques de chaque caméra, la transformation rigide entre chaque caméra et les paramètres de la transformation pince-caméra.

3.3.1 Principe

Le principe de la méthode est de minimiser la distance entre les points 3D, noté P et exprimé dans le repère monde, noté P_r , entre deux positions successives du banc de stéréovision et donc en considérant un déplacement du robot de la position j à la position i . La fonction à minimiser est donc la suivante :

$$\sum_{ij} \|P_{r_i} - P_{r_j}\|^2 \quad (3.16)$$

Le repère monde pris en compte ici, est la position initiale de l'organe terminal, donc la position 0. Comme le montre la figure 3.2, le point P_r exprimé dans le repère monde s'écrit en fonction du déplacement du robot entre la position courante i et la position initiale, noté M_{P_0i} et en fonction de son expression dans le repère de la caméra C , noté P_c , suivant la relation suivante :

$$P_{r_i} = M_{P_0i} M_{PC} P_{c_i} \quad (3.17)$$

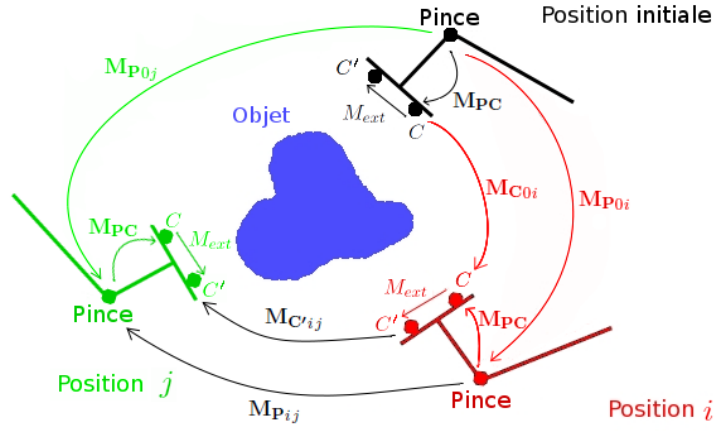


FIG. 3.2 – Géométrie du système par rapport à la position initiale

avec \mathbf{M}_{PC} la transformation pince-caméra.

En remplaçant le point P_r exprimé dans le repère monde par l'équation (3.17) dans l'équation (3.16), nous obtenons la fonction à minimiser qui fait apparaître la transformation pince-caméra :

$$\sum_{ij} \|\mathbf{M}_{P0i} \mathbf{M}_{PC} P_{c_i} - \mathbf{M}_{P0j} \mathbf{M}_{PC} P_{c_j}\|^2 \quad (3.18)$$

La généralisation à tous les déplacements $i j$ et à tous les points extraits k nous donne la fonction suivante :

$$\sum_{ij} \sum_k \|\mathbf{M}_{P0i} \mathbf{M}_{PC} P_{c_{i_k}} - \mathbf{M}_{P0j} \mathbf{M}_{PC} P_{c_{j_k}}\|^2 \quad (3.19)$$

Dans le cas d'un système non préalablement calibré, il faut exprimer les coordonnées du point P_c dans le repère de la caméra C en fonction des paramètres intrinsèques de chaque caméra et de la transformation rigide entre les deux caméras. Cela est l'objet de la sous-section suivante.

3.3.2 Expression du point P_c dans le repère de la caméra C

3.3.2.1 Dans le cas d'un banc de stéréovision classique

Lorsqu'un banc de stéréovision est calibré, il est possible d'exprimer chaque points $3D$ de l'espace comme une fonction des paramètres intrinsèques, extrinsèques et des projetés de ces points dans les images. La définition de ces paramètres est rappelée en annexe A.

Le point 3D en coordonnées normalisées P_c (resp. $P_{c'}$) est exprimé dans le repère de la caméra C (resp. C') avec les notations suivantes : $P_c = (X_c, Y_c, Z_c, 1)$ (resp. $P_{c'} = (X_{c'}, Y_{c'}, Z_{c'}, 1)$). Nous écrivons alors la relation suivante :

$$P_c = M_{\text{ext}} P_{c'} \quad (3.20)$$

Nous rappelons que M_{ext} la matrice de passage du repère de la caméra C' à la caméra C .

Or, les coordonnées des projections perspectives de P_c sur les plans image des deux caméras sont données par : $p_c = P_c/Z_c = (x_c \ y_c \ 1)^T$. L'équation (3.20) peut être alors réécrite sous la forme suivante :

$$Z_c p_c = Z_{c'} \mathbf{R}_{\text{ext}} p_{c'} + \mathbf{t}_{\text{ext}} \quad (3.21)$$

donc

$$\begin{pmatrix} p_c & -\mathbf{R}_{\text{ext}} p_{c'} \end{pmatrix} \begin{pmatrix} Z_c \\ Z_{c'} \end{pmatrix} = \mathbf{t}_{\text{ext}} \quad (3.22)$$

La résolution du système (3.22) en utilisant les moindres carrés nous donne :

$$\begin{pmatrix} Z_c \\ Z_{c'} \end{pmatrix} = \begin{pmatrix} \frac{\|\mathbf{R}_{\text{ext}} p_{c'}\|^2 \langle p_c | \mathbf{t}_{\text{ext}} \rangle - \langle p_c | \mathbf{R}_{\text{ext}} p_{c'} \rangle \langle \mathbf{R}_{\text{ext}} p_{c'} | \mathbf{t}_{\text{ext}} \rangle}{\|p_c\|^2 \|\mathbf{R}_{\text{ext}} p_{c'}\|^2 - \langle p_c | \mathbf{R}_{\text{ext}} p_{c'} \rangle^2} \\ \frac{\langle \mathbf{R}_{\text{ext}} p_{c'} | p_c \rangle \langle p_c | \mathbf{t}_{\text{ext}} \rangle - \|p_c\|^2 \langle \mathbf{R}_{\text{ext}} p_{c'} | \mathbf{t}_{\text{ext}} \rangle}{\|p_c\|^2 \|\mathbf{R}_{\text{ext}} p_{c'}\|^2 - \langle p_c | \mathbf{R}_{\text{ext}} p_{c'} \rangle^2} \end{pmatrix} \quad (3.23)$$

Avec $\langle . | . \rangle$ représentant le produit scalaire standard. De plus en inversant l'équation A.1 pour chaque caméra, nous avons :

$$p_c = \begin{pmatrix} \frac{X_c}{Z_c} \\ \frac{Y_c}{Z_c} \\ 1 \end{pmatrix} = \mathbf{K}^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{u-u_0}{\alpha_u} \\ \frac{v-v_0}{\alpha_v} \\ 1 \end{pmatrix} \quad (3.24)$$

$$p_{c'} = \begin{pmatrix} \frac{X_{c'}}{Z_{c'}} \\ \frac{Y_{c'}}{Z_{c'}} \\ 1 \end{pmatrix} = \mathbf{K}'^{-1} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{u'-u'_0}{\alpha_{u'}} \\ \frac{v'-v'_0}{\alpha_{v'}} \\ 1 \end{pmatrix} \quad (3.25)$$

En intégrant (3.24) dans (3.23), on peut exprimer les coordonnées du point 3D dans le repère de la caméra C :

$$P_c = \begin{pmatrix} \frac{Z_c(u-u_0)}{\alpha_u} \\ \frac{Z_c(v-v_0)}{\alpha_v} \\ Z_c \end{pmatrix} \quad (3.26)$$

3.3.2.2 Dans le cas d'un banc de stéréovision rectifié

Dans le cadre d'un banc de stéréovision rectifié, la relation entre P_c et $P_{c'}$ est triviale car les paramètres extrinsèques ne sont composés que d'une translation t_{ext_x}

selon l'axe X . Nous avons donc $X_{c'} = X_c - t_{ext_x}$, $Y_{c'} = Y_c$ et $Z_{c'} = Z_c$ et par conséquent :

$$\frac{X_c}{x_c} = \frac{Y_c}{y_c} = \frac{Z_c}{f} \quad ; \quad \frac{X_c - t_{ext_x}}{x_{c'}} = \frac{Y_c}{y_{c'}} = \frac{Z_c}{f'} \quad (3.27)$$

avec cette fois ci :

$$x_c = \frac{u - u_0}{k_u} \quad ; \quad x_{c'} = \frac{u' - u'_0}{k_{u'}} \quad ; \quad y_c = y_{c'} = \frac{v - v_0}{k_v} \quad (3.28)$$

Sachant que $k_u f = \alpha_u$, $k_v f = \alpha_v$ et en intégrant (3.28) dans (3.27), on obtient le système d'équation suivant :

$$\frac{\alpha_u X_c}{u - u_0} = \frac{\alpha_{u'}(X_c - t_{ext_x})}{u' - u'_0} = \frac{\alpha_v Y_c}{v - v_0} = Z_c \quad (3.29)$$

Finalement :

$$\mathbf{P}_c = \begin{pmatrix} \frac{\alpha_{u'} t_{ext_x} (u - u_0)}{\alpha_{u'}(u - u_0) - \alpha_u(u' - u'_0)} \\ \frac{\alpha_u \alpha_{u'} t_{ext_x} (v - v_0)}{\alpha_v(\alpha_{u'}(u - u_0) - \alpha_u(u' - u'_0))} \\ \frac{\alpha_u \alpha_{u'} t_{ext_x}}{\alpha_{u'}(u - u_0) - \alpha_u(u' - u'_0)} \end{pmatrix} \quad (3.30)$$

3.3.3 Problème à minimiser basé sur la distance entre les points 3D

En intégrant les équations (3.26), (3.24) et (3.23) dans le problème (3.19), l'expression de du problème à minimiser est donnée dans l'équation (3.31).

$$\min_P \sum_{ij} \sum_k \|\mathbf{M}_{\mathbf{P}_{0i}} \mathbf{M}_{\mathbf{P}_C} \begin{pmatrix} \frac{(u_{i_k} - u_0)(\|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 \langle \mathbf{p}_c | \mathbf{t}_{\text{ext}} \rangle - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle \langle \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} | \mathbf{t}_{\text{ext}} \rangle)}{\alpha_u (\|\mathbf{p}_c\|^2 \|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle^2)} \\ \frac{(v_{i_k} - v_0)(\|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 \langle \mathbf{p}_c | \mathbf{t}_{\text{ext}} \rangle - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle \langle \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} | \mathbf{t}_{\text{ext}} \rangle)}{\alpha_v (\|\mathbf{p}_c\|^2 \|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle^2)} \\ \frac{\|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 \langle \mathbf{p}_c | \mathbf{t}_{\text{ext}} \rangle - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle \langle \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} | \mathbf{t}_{\text{ext}} \rangle}{\|\mathbf{p}_c\|^2 \|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle^2} \end{pmatrix} - \mathbf{M}_{\mathbf{P}_{0j}} \mathbf{M}_{\mathbf{P}_C} \begin{pmatrix} \frac{(u_{j_k} - u_0)(\|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 \langle \mathbf{p}_c | \mathbf{t}_{\text{ext}} \rangle - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle \langle \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} | \mathbf{t}_{\text{ext}} \rangle)}{\alpha_u (\|\mathbf{p}_c\|^2 \|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle^2)} \\ \frac{(v_{j_k} - v_0)(\|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 \langle \mathbf{p}_c | \mathbf{t}_{\text{ext}} \rangle - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle \langle \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} | \mathbf{t}_{\text{ext}} \rangle)}{\alpha_v (\|\mathbf{p}_c\|^2 \|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle^2)} \\ \frac{\|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 \langle \mathbf{p}_c | \mathbf{t}_{\text{ext}} \rangle - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle \langle \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} | \mathbf{t}_{\text{ext}} \rangle}{\|\mathbf{p}_c\|^2 \|\mathbf{R}_{\text{ext}} \mathbf{p}_{c'}\|^2 - \langle \mathbf{p}_c | \mathbf{R}_{\text{ext}} \mathbf{p}_{c'} \rangle^2} \end{pmatrix} \|^2 \quad (3.31)$$

Avec $\mathbf{p}_c = \begin{pmatrix} u - u_0 & v - v_0 & 1 \\ \alpha_u & \alpha_v & \end{pmatrix}$ et $\mathbf{p}_{c'} = \begin{pmatrix} u' - u'_0 & v' - v'_0 & 1 \\ \alpha_{u'} & \alpha_{v'} & \end{pmatrix}$. Le vecteur P contient 20 paramètres (8 paramètres intrinsèques, 6 paramètres extrinsèques et 6 paramètres Pince-Caméra). La minimisation de cette fonction avec un nombre aussi important de paramètres n'a pas donné de solutions acceptables.

Nous proposons donc de diminuer le nombre de paramètres de la fonction en procédant à une étape préalable de rectification non calibrée présentée dans le cha-

pitre 2. Le nouveau problème de minimisation s'obtient alors en intégrant l'équation (3.30) dans la relation (3.19) et son expression est la suivante :

$$\min_P \sum_{ij} \sum_k \left\| \begin{matrix} \mathbf{M}_{\mathbf{P}0i} \mathbf{M}_{\mathbf{P}C} \\ -\mathbf{M}_{\mathbf{P}0j} \mathbf{M}_{\mathbf{P}C} \end{matrix} \begin{pmatrix} \frac{\alpha_{u'} t_{extx}(u_{i_k} - u_0)}{\alpha_{u'}(u_{i_k} - u_0) - \alpha_u(u'_{i_k} - u'_0)} \\ \frac{\alpha_u \alpha_{u'} t_{extx}(v_{i_k} - v_0)}{\alpha_v(\alpha_{u'}(u_{i_k} - u_0) - \alpha_u(u'_{i_k} - u'_0))} \\ \frac{\alpha_u \alpha_{u'} t_{extx}}{\alpha_{u'}(u_{i_k} - u_0) - \alpha_u(u'_{i_k} - u'_0)} \\ 1 \\ \frac{\alpha_{u'} t_{extx}(u_{j_k} - u_0)}{\alpha_{u'}(u_{j_k} - u_0) - \alpha_u(u'_{j_k} - u'_0)} \\ \frac{\alpha_u \alpha_{u'} t_{extx}(v_{j_k} - v_0)}{\alpha_v(\alpha_{u'}(u_{j_k} - u_0) - \alpha_u(u'_{j_k} - u'_0))} \\ \frac{\alpha_u \alpha_{u'} t_{extx}}{\alpha_{u'}(u_{j_k} - u_0) - \alpha_u(u'_{j_k} - u'_0)} \\ 1 \end{pmatrix} \right\|^2 \quad (3.32)$$

Le vecteur P ne comporte alors que 13 paramètres (6 paramètres intrinsèques, 1 paramètre extrinsèque et 6 paramètres Pince-Caméra).

3.3.4 Conclusion

Grâce à une étape préalable de rectification non calibrée, nous sommes en mesure de proposer dans l'espace 3D une fonction de minimisation dont le vecteur paramètre englobe à la fois les paramètres intrinsèques, extrinsèques et de la transformation pince-caméra. Au total, ce sont 13 paramètres qui sont estimés par minimisation de la fonction (3.32). Cette méthode est dénommée **Distance P3D** dans les tableaux de résultats de la section 3.5.

3.4 Proposition d'une méthode d'auto-calibrage basée sur la minimisation de distances dans l'espace 2D

L'idée est de minimiser la distance d'un point d'une image à sa droite épipolaire associée dans cette même image après déplacement. Nous utilisons la distance de Sampson en décomposant la matrice Fondamentale en matrice Essentielle en y intégrant les paramètres de la transformation pince-caméra.

3.4.1 Principe

La seule information géométrique disponible pour des caméras non calibrées, est la connaissance de la matrice Fondamentale. Sa détermination repose usuellement sur la minimisation de la distance approximée de Sampson présentée dans le chapitre 2 que nous rappelons ci-dessous :

$$\min_P \sum_k \frac{|\mathbf{m}'_k{}^T \mathbf{F} \mathbf{m}_k|^2}{|\pi(\mathbf{F} \mathbf{m}_k)|_2^2 + |\pi(\mathbf{F}^T \mathbf{m}'_k)|_2^2} \quad (3.33)$$

avec $\pi : (x_1, x_2, x_3) \rightarrow (x_1, x_2)$ étant la projection canonique orthogonale dans l'espace vectoriel, \mathbf{m} (resp. \mathbf{m}') représente les stéréo correspondants de la première image (resp. la seconde image) et k est le nombre de points extraits communs aux deux images.

Ensuite, l'idée est donc de décomposer la matrice Fondamentale pour obtenir les paramètres intrinsèques des caméras. Cette décomposition de la matrice Fondamentale repose sur la matrice Essentielle comme le montre la relation (A.5). La distance à minimiser (3.33) se ré-écrit alors :

$$\min_P \sum_k \frac{|\mathbf{m}'_k{}^T \mathbf{K}'^{-T} \mathbf{E} \mathbf{K}^{-1} \mathbf{m}_k|^2}{|\pi(\mathbf{K}'^{-T} \mathbf{E} \mathbf{K}^{-1} \mathbf{m}_k)|_2^2 + |\pi((\mathbf{K}'^{-T} \mathbf{E} \mathbf{K}^{-1})^T \mathbf{m}'_k)|_2^2} \quad (3.34)$$

Enfin, il faut décomposer la matrice Essentielle pour faire apparaître les paramètres de la transformation pince-caméra. Il faut donc prendre en compte le déplacement du robot et considérer non plus les matrices Fondamentales et Essentielles comme des relations entre une image gauche et droite d'un banc stéréo mais comme des transformations entre des images de la même caméra à différentes positions. Comme le montre la relation (A.6), lors d'un déplacement entre la position j et la position i la matrice Essentielle contient les informations du déplacement de la caméra $\mathbf{M}_{\mathbf{C}ij}$. Ainsi, nous avons l'égalité suivante :

$$\mathbf{E}_{ij} = [\mathbf{t}_{\mathbf{C}ij}]_{\times} \mathbf{R}_{\mathbf{C}ij} \quad (3.35)$$

où \mathbf{E}_{ij} est la matrice Essentielle entre les images aux positions j et i d'une même caméra et avec $\mathbf{t}_{\mathbf{C}ij}$ et $\mathbf{R}_{\mathbf{C}ij}$ les composantes en rotation et translation du déplacement de la caméra $\mathbf{M}_{\mathbf{C}ij}$.

De plus, selon l'équation (3.5), le déplacement de la caméra s'exprime en fonction du déplacement du robot $\mathbf{M}_{\mathbf{P}ij}$ et de la transformation pince-caméra $\mathbf{M}_{\mathbf{P}C}$ de la manière suivante :

$$\mathbf{M}_{\mathbf{C}ij} = \mathbf{M}_{\mathbf{P}C}^{-1} \mathbf{M}_{\mathbf{P}ij} \mathbf{M}_{\mathbf{P}C} \quad (3.36)$$

Ainsi les paramètres de la transformation pince-caméra sont intégrés dans la distance à minimiser. Cette fonction s'écrit différemment suivant le système de vision envisagé : système monoculaire, stéréovision ou stéréovision rectifiée.

3.4.2 Système monoculaire

Le cas du système monoculaire est présenté ici car il représente une étape clé pour la compréhension des fonctions basées sur un système de stéréovision. Il sera évalué uniquement avec des données de synthèse en comparaison aux critères basés sur un capteur de stéréovision.

3.4.2.1 Écriture de la matrice Essentielle

D'après la relation (A.5), la matrice Fondamentale entre deux positions j et i d'une même caméra s'écrit alors sous la forme suivante :

$$\mathbf{F}_{ij} = \mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} \quad (3.37)$$

Avec \mathbf{K} la matrice de paramètres intrinsèques de la caméra et \mathbf{E}_{ij} la matrice Essentielle contenant les informations de déplacement de la caméra de la position j à la position i .

D'après l'équation (3.13), nous exprimons la matrice Essentielle pour les positions j et i .

$$\mathbf{E}_{ij} = [\mathbf{t}_{C_{ij}}]_{\times} \mathbf{R}_{C_{ij}} \quad (3.38)$$

Avec $\mathbf{R}_{C_{ij}}$ et $\mathbf{t}_{C_{ij}}$ les composantes (rotation et translation) du déplacement de la caméra $M_{C_{ij}}$. Et comme décrit sur la figure 3.1 le déplacement de la caméra C ($M_{C_{ij}}$) est composé du déplacement du robot $\mathbf{M}_{P_{ij}}$ et de la matrice Pince-Caméra \mathbf{M}_{PC} :

$$\begin{aligned} \mathbf{M}_{C_{ij}} &= \begin{pmatrix} \mathbf{R}_{C_{ij}} & \mathbf{t}_{C_{ij}} \\ 0 & 1 \end{pmatrix} = \mathbf{M}_{PC}^{-1} \mathbf{M}_{P_{ij}} \mathbf{M}_{PC} \\ &= \begin{pmatrix} \mathbf{R}_{PC}^T & -\mathbf{R}_{PC}^T \mathbf{t}_{PC} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_{P_{ij}} & \mathbf{t}_{P_{ij}} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_{PC} & \mathbf{t}_{PC} \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{R}_{PC}^T \mathbf{R}_{P_{ij}} \mathbf{R}_{PC} & \mathbf{R}_{PC}^T \mathbf{R}_{P_{ij}} \mathbf{t}_{PC} + \mathbf{R}_{PC}^T \mathbf{t}_{P_{ij}} - \mathbf{R}_{PC}^T \mathbf{t}_{PC} \\ 0 & 1 \end{pmatrix} \end{aligned} \quad (3.39)$$

On peut donc en déduire l'écriture de la matrice Essentielle \mathbf{E}_{ij} :

$$\mathbf{E}_{ij} = [\mathbf{R}_{PC}^T \mathbf{R}_{P_{ij}} \mathbf{t}_{PC} + \mathbf{R}_{PC}^T \mathbf{t}_{P_{ij}} - \mathbf{R}_{PC}^T \mathbf{t}_{PC}]_{\times} \mathbf{R}_{PC}^T \mathbf{R}_{P_{ij}} \mathbf{R}_{PC} \quad (3.40)$$

3.4.2.2 Ecriture de la distance à minimiser

En écrivant l'équation (3.33) pour un déplacement entre la position j et la position i et le nombre de points extraits k communs lors de chaque déplacement, nous obtenons :

$$\min_P \sum_{ij} \sum_k \frac{|m_{jk}^T \mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} m_{ik}|^2}{|\pi(\mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} m_{ik})|_2^2 + |\pi(\mathbf{K}^{-1} \mathbf{E}_{ij}^T \mathbf{K}^{-T} m_{jk})|_2^2} \quad (3.41)$$

En remplaçant \mathbf{E}_{ij} par l'équation (3.40), le vecteur de paramètres P contient les paramètres intrinsèques de la caméra et ceux de la transformation Pince-Caméra. Au total 10 paramètres : 4 paramètres intrinsèques et 6 paramètres de la transformation Pince-Caméra sont à identifier. Cette méthode est dénommée **Mono** dans les tableaux de résultats de la section 3.5.

3.4.3 Système de stéréovision

Nous disposons maintenant sur l'organe terminal du robot d'un banc de stéréovision avec deux caméras (caméra C et caméra C').

3.4.3.1 Écriture des matrices Essentielles

Avec les conventions de notations précédentes, les matrices Fondamentales entre les positions j et i des caméras C d'une part et caméra C' d'autre part s'écrivent :

$$\mathbf{F}_{ij} = \mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} \quad ; \quad \mathbf{F}'_{ij} = \mathbf{K}'^{-T} \mathbf{E}'_{ij} \mathbf{K}'^{-1} \quad (3.42)$$

avec et \mathbf{K} (resp. \mathbf{K}') la matrice des paramètres intrinsèques de la caméra C (resp. C') et l'expression des matrices Essentielles suivantes :

$$\mathbf{E}_{ij} = [\mathbf{t}_{Cij}]_{\times} \mathbf{R}_{Cij} \quad ; \quad \mathbf{E}'_{ij} = [\mathbf{t}'_{Cij}]_{\times} \mathbf{R}'_{Cij} \quad (3.43)$$

avec \mathbf{t}_{Cij} (resp. \mathbf{t}'_{Cij}) et \mathbf{R}_{Cij} (resp. \mathbf{R}'_{Cij}) les translations et les rotations du déplacement de la caméra C : \mathbf{M}_{Cij} (resp. de la caméra C' : \mathbf{M}'_{Cij}) (voir figure 3.1).

En utilisant un raisonnement analogue au paragraphe précédent, on peut déduire que la valeur de la matrice Essentielle \mathbf{E}_{ij} est la même que dans l'équation (3.40). Cependant comme le montre de nombreux travaux, lors du calibrage d'un banc de stéréovision avec de multiples vues, nous avons intérêt pour réduire l'erreur commise lors de l'estimation à considérer la transformation existante entre les deux caméras (cf. figure 3.1). Nous intégrons donc cette transformation \mathbf{M}_{ext} composée de la rotation \mathbf{R}_{ext} et de la translation \mathbf{t}_{ext} , dans l'expression du déplacement entre la position j et la position i de la caméra C' (\mathbf{M}'_{Cij}) par rapport à celui de la caméra C (\mathbf{M}_{Cij}). Le déplacement de la caméra C' est donc le suivant :

$$\begin{aligned} \mathbf{M}'_{Cij} &= \mathbf{M}_{\text{ext}}^{-1} \mathbf{M}_{Cij} \mathbf{M}_{\text{ext}} = \begin{pmatrix} \mathbf{R}'_{Cij} & \mathbf{t}'_{Cij} \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{R}_{\text{ext}}^T \mathbf{R}_{Cij} \mathbf{R}_{\text{ext}} & \mathbf{R}_{\text{ext}}^T \mathbf{R}_{Cij} \mathbf{t}_{\text{ext}} + \mathbf{R}_{\text{ext}}^T \mathbf{t}_{Cij} - \mathbf{R}_{\text{ext}}^T \mathbf{t}_{\text{ext}} \\ 0 & 1 \end{pmatrix} \end{aligned} \quad (3.44)$$

En remplaçant dans l'équation (3.44), les termes \mathbf{R}_{Cij} et \mathbf{t}_{Cij} par leur expression de l'équation (3.40), nous obtenons pour l'expression de la rotation \mathbf{R}'_{Cij} et de la translation \mathbf{t}'_{Cij} les formulations suivantes :

$$\mathbf{R}'_{Cij} = \mathbf{R}_{\text{ext}}^T \mathbf{R}_{\text{PC}}^T \mathbf{R}_{\text{P}ij} \mathbf{R}_{\text{PC}} \mathbf{R}_{\text{ext}} \quad (3.45)$$

$$\begin{aligned} \mathbf{t}'_{Cij} &= \mathbf{R}_{\text{ext}}^T \mathbf{R}_{\text{PC}}^T \mathbf{R}_{\text{P}ij} \mathbf{R}_{\text{PC}} \mathbf{t}_{\text{ext}} + \mathbf{R}_{\text{ext}}^T \mathbf{R}_{\text{PC}}^T \mathbf{R}_{\text{P}ij} \mathbf{t}_{\text{PC}} + \mathbf{R}_{\text{ext}}^T \mathbf{R}_{\text{PC}}^T \mathbf{t}_{\text{PC}} \\ &\quad - \mathbf{R}_{\text{ext}}^T \mathbf{R}_{\text{PC}}^T \mathbf{t}_{\text{PC}} - \mathbf{R}_{\text{ext}}^T \mathbf{t}_{\text{ext}} \end{aligned} \quad (3.46)$$

On obtient finalement la matrice Essentielle \mathbf{E}'_{ij} en intégrant les équations (3.45) dans l'équation (3.43).

3.4.3.2 Écriture de la distance à minimiser

En appliquant un raisonnement analogue au paragraphe précédent dans le cas d'une caméra (cf. section 3.4.2), la distance à minimiser pour un système de stéréo-

vision s'écrit :

$$\begin{aligned} \min_P \sum_{ij} \sum_k & \frac{|m_{jk}^T \mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} m_{ik}|^2}{|\pi(\mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} m_{ik})|_2^2 + |\pi(\mathbf{K}^{-1} \mathbf{E}_{ij}^T \mathbf{K}^{-T} m_{jk})|_2^2} \\ & + \frac{|m_{jk}'^T \mathbf{K}'^{-T} \mathbf{E}'_{ij} \mathbf{K}'^{-1} m'_{ik}|^2}{|\pi(\mathbf{K}'^{-T} \mathbf{E}'_{ij} \mathbf{K}'^{-1} m'_{ik})|_2^2 + |\pi(\mathbf{K}'^{-1} \mathbf{E}'_{ij}^T \mathbf{K}'^{-T} m'_{jk})|_2^2} \end{aligned} \quad (3.47)$$

Avec un raisonnement analogue au paragraphe précédent sur l'introduction de la transformation entre les deux caméras, la fonction peut être contrainte plus fortement en considérant la relation Fondamentale entre les deux caméras à la position i et à la position j .

$$\mathbf{F}_{ii} = \mathbf{F}_{jj} = \mathbf{K}'^{-T} \mathbf{E}_{\text{ext}} \mathbf{K}^{-1} \quad (3.48)$$

avec

$$\mathbf{E}_{\text{ext}} = [\mathbf{t}_{\text{ext}}]_{\times} \mathbf{R}_{\text{ext}} \quad (3.49)$$

L'équation (3.47) est donc enrichie de deux relations supplémentaires et elle s'écrit :

$$\begin{aligned} \min_P \sum_{ij} \sum_k & \frac{|m_{jk}^T \mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} m_{ik}|^2}{|\pi(\mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} m_{ik})|_2^2 + |\pi(\mathbf{K}^{-1} \mathbf{E}_{ij}^T \mathbf{K}^{-T} m_{jk})|_2^2} \\ & + \frac{|m_{jk}'^T \mathbf{K}'^{-T} \mathbf{E}'_{ij} \mathbf{K}'^{-1} m'_{ik}|^2}{|\pi(\mathbf{K}'^{-T} \mathbf{E}'_{ij} \mathbf{K}'^{-1} m'_{ik})|_2^2 + |\pi(\mathbf{K}'^{-1} \mathbf{E}'_{ij}^T \mathbf{K}'^{-T} m'_{jk})|_2^2} \\ & + \frac{|m_{ik}'^T \mathbf{K}'^{-T} \mathbf{E}_{\text{ext}} \mathbf{K}^{-1} m_{ik}|^2}{|\pi(\mathbf{K}'^{-T} \mathbf{E}_{\text{ext}} \mathbf{K}^{-1} m_{ik})|_2^2 + |\pi(\mathbf{K}'^{-1} \mathbf{E}_{\text{ext}}^T \mathbf{K}^{-T} m'_{ik})|_2^2} \\ & + \frac{|m_{jk}'^T \mathbf{K}'^{-T} \mathbf{E}_{\text{ext}} \mathbf{K}^{-1} m_{jk}|^2}{|\pi(\mathbf{K}'^{-T} \mathbf{E}_{\text{ext}} \mathbf{K}^{-1} m_{jk})|_2^2 + |\pi(\mathbf{K}'^{-1} \mathbf{E}_{\text{ext}}^T \mathbf{K}^{-T} m'_{jk})|_2^2} \end{aligned} \quad (3.50)$$

En remplaçant \mathbf{E}_{ij} par l'équation (3.40) et \mathbf{E}'_{ij} par l'équation (3.44), le vecteur de paramètres P contient les paramètres intrinsèques de la caméra et ceux de la transformation pince-caméra. Il comporte 20 paramètres inconnus : 8 paramètres des matrices intrinsèques \mathbf{K} et \mathbf{K}' , 6 paramètres de la matrice de transformation rigide \mathbf{M}_{ext} et 6 paramètres pour la transformation Pince-Caméra \mathbf{M}_{PC} . Cette méthode est dénommée **Stéréo** dans les tableaux de résultats de la section 3.5.

3.4.4 Système de stéréovision rectifié

Le système de stéréovision rectifié est un cas simplifié du précédent dont l'intérêt est de réduire le nombre de paramètres à estimer.

3.4.4.1 Écriture des matrices Essentielles

La matrice Essentielle entre la position j et i de la caméra C est identique à la précédente (cf. équation (3.40)). En revanche, la matrice essentielle entre la position

j et i de la caméra C' se simplifie du fait de la forme de la transformation rigide entre deux plans d'images rectifiées (cf. équation (A.8)). Nous pouvons donc écrire :

$$\begin{aligned}\mathbf{R}'_{Cij} &= \mathbf{R}_{PC}^T \mathbf{R}_{Pij} \mathbf{R}_{PC} & (3.51) \\ \mathbf{t}'_{Cij} &= \mathbf{R}_{PC}^T \mathbf{R}_{Pij} \mathbf{R}_{PC} \mathbf{t}_{\text{ext}} + \mathbf{R}_{PC}^T \mathbf{R}_{Pij} \mathbf{t}_{PC} + \mathbf{R}_{PC}^T \mathbf{t}_{Pij} - \mathbf{R}_{PC}^T \mathbf{t}_{PC} - \mathbf{t}_{\text{ext}}\end{aligned}$$

Là encore, la matrice Essentielle \mathbf{E}'_{ij} s'obtient en intégrant les équations (3.51) dans l'équation (3.43).

3.4.4.2 Écriture de la distance à minimiser

La contrainte sur les matrices Fondamentales $\mathbf{F}_{\text{rectii}}$ et $\mathbf{F}_{\text{rectjj}}$ devient obsolète, car la distance de Sampson (3.33) est par définition nul sur des images rectifiées. Nous pouvons donc écrire la fonction finale suivante :

$$\begin{aligned}\min_P \sum_{ij} \sum_k & \frac{|\mathbf{m}'_{jk}{}^T \mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} \mathbf{m}_{ik}|^2}{|\pi(\mathbf{K}^{-T} \mathbf{E}_{ij} \mathbf{K}^{-1} \mathbf{m}_{ik})|_2^2 + |\pi(\mathbf{K}^{-1} \mathbf{E}_{ij}^T \mathbf{K}^{-T} \mathbf{m}_{jk})|_2^2} \\ & + \frac{|\mathbf{m}'_{jk}{}^T \mathbf{K}'^{-T} \mathbf{E}'_{ij} \mathbf{K}'^{-1} \mathbf{m}'_{ik}|^2}{|\pi(\mathbf{K}'^{-T} \mathbf{E}'_{ij} \mathbf{K}'^{-1} \mathbf{m}'_{ik})|_2^2 + |\pi(\mathbf{K}'^{-1} \mathbf{E}'_{ij}{}^T \mathbf{K}'^{-T} \mathbf{m}'_{jk})|_2^2}\end{aligned} \quad (3.52)$$

En remplaçant \mathbf{E}_{ij} par l'équation (3.40) et \mathbf{E}'_{ij} par l'équation (3.51), le vecteur P composé de 13 paramètres : 6 intrinsèques, 1 extrinsèque, 6 paramètres de la transformation Pince-Caméra. Cette méthode est dénommée **Stéréo rect** dans les tableaux de résultats de la section 3.5.

3.4.5 Conclusion

Tout comme pour la méthode **Distance P3D**, une étape préalable de rectification non calibrée permet de réduire le nombre de paramètre à estimer (20 contre 13 paramètres), ainsi que la complexité de la fonction à minimiser.

3.5 Evaluation des méthodes d'auto-calibrage sur des données de synthèse

Cette section est consacrée aux tests sur des données de synthèse des méthodes de calibrage proposées, basées sur l'observation de points quelconques, qui fournissent en une seule étape l'ensemble des paramètres recherchés (paramètres intrinsèques et paramètres de la transformation Pince-Caméra).

Cette approche permet une étude approfondie des différentes méthodes en comparant les valeurs des paramètres estimés lors du calibrage à celles utilisées pour générer les données de synthèse.

Les sous-sections 3.5.1 et 3.5.2 présentent la démarche et les valeurs utilisées pour générer les données de synthèse. La sous-section 3.5.4 décrit les critères adoptés pour évaluer les performances des différentes méthodes. Enfin, les sous-sections

3.5.5, 3.5.8, 3.5.6 et 3.5.7 donnent les résultats dans les différentes conditions de la simulation.

3.5.1 Méthodologie

Les méthodes proposées ont été évaluées dans différents cas afin d'analyser la sensibilité aux conditions de calibrage. Les conditions envisagées pour la génération des données de synthèse sont :

- des données de synthèse non bruitées,
- une variation du nombre de points extraits,
- une introduction de bruit dans l'extraction des points,
- l'observation d'un plan : cas dégénéré où les points extraits ne sont plus uniformément répartis dans un espace $3D$ mais dans un espace $2D$.

Ces conditions ont été choisies car elles mettent en évidence les difficultés rencontrées avec des images infrarouges comme l'extraction d'un faible nombre de points, des images bruitées ou encore un faible champ de vision ne permettant pas d'observer la totalité de l'objet mais qu'une partie parfois plane.

3.5.2 Contexte expérimental

Nous avons simulé un système de stéréovision observant un nuage de 100 points $3D$ aléatoirement réparti en 6 vues différentes. Ensuite nous avons changé la répartition des points en observant des points sur un plan, fait varier le nombre de points de 4 à 400 et finalement ajouté un bruit aléatoire uniformément distribué entre 0 et 1 pixel, lors de l'extraction des points.

Les paramètres intrinsèques, extrinsèques et pince-caméra exacts utilisés sont décrits ci dessous. Leurs valeurs proviennent d'anciens calibrages forts de caméras infrarouges montées sur un robot six axes et d'expériences d'acquisitions sur ce robot. Ces données permettent de tester les méthodes dans des conditions proches de la réalité.

Dans cette section, les translations sont exprimées en mm et les angles sont exprimés en degrés et correspondent à la séquence suivante : rotations successives autour de l'axe z d'un angle r_z , puis autour du nouvel axe y' d'un angle r_y et enfin autour du nouvel axe x'' d'un angle r_x .

3.5.2.1 Paramètres de la transformation pince-caméra

Le tableau 3.1 présente les paramètres de la transformation pince-caméra.

t_{PC_x} (mm)	t_{PC_y} (mm)	t_{PC_z} (mm)	r_{PC_x} ($^\circ$)	r_{PC_y} ($^\circ$)	r_{PC_z} ($^\circ$)
152	36	-19	6.53	4.89	-3.98

TAB. 3.1 – Paramètres de la transformation pince-caméra.

3.5.2.2 Paramètres intrinsèques et extrinsèques monoculaire et stéréovision

Les paramètres intrinsèques sont résumés dans le tableau 3.2.

α_u	$\alpha_{u'}$	α_v	$\alpha_{v'}$	u_0	u'_0	v_0	v'_0
211.5742	213.5703	211.5957	213.4323	82.1637	85.4981	63.9174	63.7494

TAB. 3.2 – Paramètres intrinsèques des deux caméras.

Les paramètres de distorsions ainsi que le paramètre de skew sont considérés comme nuls.

Le tableau 3.3 décrit les paramètres extrinsèques.

t_{ext_x} (mm)	t_{ext_y} (mm)	t_{ext_z} (mm)	r_{ext_x} (°)	r_{ext_y} (°)	r_{ext_z} (°)
-154.2365	-1.1042	18.7118	-0.0062	15.8950	0.9942

TAB. 3.3 – Paramètres extrinsèques, déplacement rigide entre les deux caméras.

3.5.2.3 Paramètres intrinsèques et extrinsèques d'un système de stéréovision rectifié

Les données proviennent aussi d'anciens calibrages réalisés sur un système composé de deux caméras infrarouges préalablement rectifiées montées sur un robot six axes. Ces paramètres ainsi que la base (seul paramètre extrinsèque dans ce cas, il est noté t_{ext_x}) sont résumés dans le tableau 3.4.

α_u	$\alpha_{u'}$	α_v	u_0	u'_0	v_0	t_{ext_x}
195.8445	191.6099	201.5821	137.3928	26.9461	58.5826	-156.1435

TAB. 3.4 – Paramètres intrinsèques des deux caméras rectifiées.

On remarque que dans ce cas, les paramètres u_0 et u'_0 sont très différents des données constructeur ($u_0 = 82$) du cas précédent. Comme pour le cas d'un système stéréo classique, les paramètres de distorsions ainsi que le paramètre de skew sont considérés comme nuls.

3.5.2.4 Paramètres orientations et positions de la pince

Pour chaque vue (i), l'orientation et la position de la pince par rapport au repère robot (situé à l'origine de la chaîne cinématique du robot), notées $\mathbf{R}_{\mathbf{P}0i}$ et $\mathbf{t}_{\mathbf{P}0i}$, sont données dans le tableau 3.5.

vue	rotation en °			translation en mm		
	$r_{P_{0i_x}}$	$r_{P_{0i_y}}$	$r_{P_{0i_z}}$	$t_{P_{0i_x}}$	$t_{P_{0i_y}}$	$t_{P_{0i_z}}$
1	0	180	0	0	0	500
2	45	180	14	10	200	356
3	-43	-167	22	-318	-158	424
4	49	175	4	12	189	832
5	60	167	34	-321	376	294
6	-34	-163	-45	531	-269	756

TAB. 3.5 – Positions et orientations du robot lors des simulations.

3.5.2.5 Génération des données

Afin d'obtenir un set de points extraits, pour chaque vue nous appliquons, comme le montre l'équation (3.53), l'ensemble des transformations suivantes : déplacement de la pince, relation pince-caméra, matrice de paramètres extrinsèques et projection dans l'image à l'aide des paramètres intrinsèques.

$$\begin{pmatrix} u_k \\ v_k \end{pmatrix} = \mathbf{K} \begin{pmatrix} \mathbf{R}_{PC} & \mathbf{t}_{PC} \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{R}_{P_{0i}} & \mathbf{t}_{P_{0i}} \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} X_k \\ Y_k \\ Z_k \end{pmatrix}$$

$$\begin{pmatrix} u'_k \\ v'_k \end{pmatrix} = \mathbf{K}' \begin{pmatrix} \mathbf{R}_{ext} & \mathbf{t}_{ext} \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{R}_{PC} & \mathbf{t}_{PC} \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{R}_{P_{0i}} & \mathbf{t}_{P_{0i}} \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} X_k \\ Y_k \\ Z_k \end{pmatrix} \quad (3.53)$$

Avec \mathbf{K} la matrice de paramètres intrinsèques définie à l'équation (A.2). Les transformations \mathbf{M}_{PC} , \mathbf{M}_{ext} et $\mathbf{M}_{P_{0i}}$ sont inversées car nous avons gardé les mêmes notations que la figure 3.1.

Pour d'obtenir un ensemble de points extraits rectifié, nous appliquons la même équation (3.53) que pour un système stéréo classique. La structure utilisée dans les paramètres intrinsèques rectifiés est issue du tableau 3.4). Elle permet de générer directement des images rectifiées et donc de supprimer l'erreur apportée par l'utilisation des homographies de rectification.

3.5.3 Méthode de minimisation et initialisation

Le processus de minimisation est la fonction de minimisation classique sans contrainte de Matlab (la fonction "fminunc").

L'initialisation des paramètres est réalisée avec les données constructeurs de la caméra (distance focale et taille de la matrice) et avec des mesures grossières de la transformation pince-caméra (précision de l'ordre du centimètre pour les translations et de 10° pour l'orientation). Ces valeurs sont résumées dans le tableau 3.6.

α_u	$\alpha_{u'}$	α_v	$\alpha_{v'}$	u_0	u'_0	v_0	v'_0
200	200	200	200	82	82	64	64
t_{ext_x}	t_{ext_y}	t_{ext_z}	r_{ext_x}	r_{ext_y}	r_{ext_z}		
(mm)	(mm)	(mm)	(°)	(°)	(°)		
-150	0	20	0	0	0		
t_{PC_x}	t_{PC_y}	t_{PC_z}	r_{PC_x}	r_{PC_y}	r_{PC_z}		
(mm)	(mm)	(mm)	(°)	(°)	(°)		
150	40	-20	0	0	0		

TAB. 3.6 – Paramètres d'initialisation pour l'optimisation.

3.5.4 Critères d'évaluation

Afin d'évaluer les performances et de pouvoir comparer chacune des méthodes implémentées, différents critères ont été mis en oeuvre en fonction des tests effectués.

3.5.4.1 Critères uniquement valables avec des données de synthèse

Avec les données de synthèse, les paramètres sont préalablement fixés et par conséquent sont connus de manière exacte. La comparaison utilise donc ces vrais paramètres et nous établissons :

- l'erreur sur les paramètres intrinsèques exprimée en pourcentage (noté ε_{α_u} , $\varepsilon_{\alpha_{u'}}$, ε_{α_v} , ...).
- l'erreur sur les paramètres extrinsèques exprimée en millimètres ou en degrés (noté $\varepsilon_{t_{ext}}$ ou $\varepsilon_{r_{ext}}$).
- l'erreur sur les paramètres de la transformation pince-caméra exprimée en millimètres ou en degrés (notée $\varepsilon_{t_{PC}}$ ou $\varepsilon_{r_{PC}}$).
- La moyenne des erreurs sur les paramètres : intrinsèques σ_{int} (en %), extrinsèques σ_{ext} (en mm ou °) et les paramètres de la transformation pince-caméra σ_{PC} (en mm ou °).

Pour chaque test, dix calibrages avec un nuage de points différent sont réalisés et les résultats présentés sont basés sur le calibrage moyen.

3.5.4.2 Critères généraux

Sur les données réelles, une valeur approximative des paramètres peut être estimée mais leurs valeurs exactes sont inconnues. Nous utilisons pour l'évaluation le calcul de la distance euclidienne entre les points reconstruits par triangulation lors de chacune des vues. Ce critère est appelé **critère de reconstruction 3D** et il est caractérisé par trois grandeurs :

- la somme des distances des points 3D reconstruits notée Σ_d définie par l'équation (3.18).
- la moyenne de ces distances \bar{d} qui représente en réalité l'erreur de recalage entre les vues.
- l'écart type des distances σ_d .

De plus, un système multivues obéit à certaines contraintes, donc la contrainte épipolaire. La distance d'un point d'une image à sa droite épipolaire associée dans cette même image après déplacement est donc un critère pertinent, il est appelé ici **critère épipolaire** et il est qualifié par :

- la somme des distances des points d'une image à leurs droites épipolaires respectives dans cette même image après déplacement Σ_e défini par l'équation (3.52)
- la moyenne de ces distances \bar{e} .
- l'écart type des distances σ_e .

Enfin, le résidu (valeur finale de la fonction coût utilisée pour l'optimisation), noté *Res.*, est analysé pour étudier la convergence finale de chaque algorithme. En remarque, cette valeur n'est pas utilisable pour la comparaison entre critères car chaque fonction coût est différente.

Afin de pouvoir comparer les résultats obtenus avec des données réelles et des données de synthèse, ces critères sont appliqués lors de chacune des simulations.

3.5.5 Évaluation sur des données de synthèse non bruitées

Les données de synthèse sont obtenues à partir de six déplacements (6 vues) autour de 100 points *3D* aléatoirement répartis.

3.5.5.1 Evaluation des paramètres intrinsèques

Le tableau 3.7 propose l'erreur entre les paramètres intrinsèques estimés et les paramètres réels, exprimée en pourcentage, pour chaque méthode testée.

Erreur Méthode	ε_{α_u} (%)	$\varepsilon_{\alpha_{u'}}$ (%)	ε_{α_v} (%)	$\varepsilon_{\alpha_{v'}}$ (%)	ε_{u_0} (%)	$\varepsilon_{u'_0}$ (%)	ε_{v_0} (%)	$\varepsilon_{v'_0}$ (%)	σ_{int} (%)
Mono	0.28	–	0.23	–	0.73	–	0.25	–	0.37
Stéréo	3.53	2.85	5.67	5.25	0.97	3.57	3.78	5.13	3.84
Stéréo rect	0.09	0.32	0.20	0.20	2.15	1.95	0.31	0.31	0.84
Distance <i>P3D</i>	0.04	0.01	0.03	0.03	0.23	0.28	0.01	0.01	0.10

TAB. 3.7 – Comparaison des résultats sur les paramètres intrinsèques des caméras.

Nous constatons que l'erreur en pourcentage sur les paramètres intrinsèques est la plus élevée avec la méthode **Stéréo** avec une erreur moyenne autour de 4%. En comparaison, l'erreur moyenne sur chacun des paramètres est de 0.4% pour la méthode **Mono**. En effet le grand nombre de paramètres permet de répartir l'erreur durant l'optimisation sur la totalité des paramètres et favorise ainsi la convergence vers un minimum local éloigné de la solution réelle.

Dans le cas rectifié, pour les méthodes **Stéréo rect** et **Distance P3D**, on remarque que l'algorithme converge vers une solution très proche de la réalité avec une moyenne d'erreur inférieure à 0.3% pour l'ensemble des paramètres, excepté pour les abscisses à l'origine U_0 (avec près de 2% d'erreur pour la méthode **Stéréo**

rect). L'erreur sur ce paramètre est en grande partie due à l'initialisation. En effet, la valeur de ce paramètre ne peut être estimée précisément et est, comme le montre le tableau 3.4, très éloignée des données constructeur (cf. tableau 3.6).

Pour l'ensemble des paramètres intrinsèques, la méthode introduisant la plus faible erreur sur ces paramètres est la méthode **Distance P3D** avec une moyenne d'erreur de 0.1% sur l'ensemble des paramètres.

3.5.5.2 Évaluation des paramètres de la transformation rigide entre les deux caméras

Le tableau 3.8 représente les erreurs en millimètres et en degrés obtenues sur les paramètres extrinsèques.

Erreur Méthode	$\varepsilon_{t_{ext_x}}$ (mm)	$\varepsilon_{t_{ext_y}}$ (mm)	$\varepsilon_{t_{ext_z}}$ (mm)	$\varepsilon_{r_{ext_x}}$ (°)	$\varepsilon_{r_{ext_y}}$ (°)	$\varepsilon_{r_{ext_z}}$ (°)	$\sigma_{t_{ext}}/\sigma_{r_{ext}}$ (mm/°)
Stéréo	4.17	3.13	0.24	0.07	0.80	0.49	2.51/0.45
Stéréo rect	1.04	0	0	0	0	0	1.04/0
Distance P3D	0.28	0	0	0	0	0	0.28/0

TAB. 3.8 – Comparaison des résultats sur les paramètres extrinsèques des caméras.

L'utilisation d'un système de stéréovision rectifié (méthodes **Stéréo rect** ou **Distance P3D**) permet d'obtenir les erreurs les plus faibles sur l'estimation des paramètres extrinsèques. Ceci est due aux nombres de paramètres à estimer, car dans le cas rectifié ce paramètre est unique contre 6 pour le cas non-rectifié (méthode **Stéréo**). En effet, l'erreur sur la translation est de 1mm pour la méthode **Stéréo rect** et 0.2mm pour **Distance P3D** contre 5.2mm la méthode **Stéréo**.

On remarque cependant pour l'ensemble des méthodes, que l'erreur sur l'estimation des paramètres extrinsèques est très faible par rapport à la taille de la base, environ 3% pour la méthode **Stéréo**. Pour cette méthode, l'erreur sur les paramètres des rotations est inférieure au degré sur chacun des axes.

En remarque, la méthode **Mono** est sans objet dans ce tableau car la liaison rigide entre deux caméras n'existe pas.

3.5.5.3 Évaluation des paramètres de la transformation pince-caméra

Le tableau 3.9 nous montre les erreurs en millimètres et en degrés sur les paramètres de la transformation pince-caméra.

Ces résultats montrent une faible erreur sur l'estimation des paramètres pour la méthode **Distance P3D** avec une erreur moyenne de 0.4mm et une erreur sur les paramètres de rotation très faible avec moins de 0.1° d'erreur. Les méthodes basées sur une décomposition de la matrice Fondamentale présentent des erreurs plus élevées de l'ordre de 2mm soit environ 2% de la valeur des paramètres. La méthode **Stéréo** affiche une erreur de 6.32mm sur l'un des paramètres, ce qui montre une convergence de l'algorithme vers une solution éloignée des données réelles.

Erreur Méthode	$\varepsilon_{t_{PC_x}}$ (mm)	$\varepsilon_{t_{PC_y}}$ (mm)	$\varepsilon_{t_{PC_z}}$ (mm)	$\varepsilon_{r_{PC_x}}$ (°)	$\varepsilon_{r_{PC_y}}$ (°)	$\varepsilon_{r_{PC_z}}$ (°)	$\sigma_{t_{PC}}/\sigma_{r_{PC}}$ (mm/°)
Mono	1.63	1.98	1.54	0.11	0.02	0.03	1.72/0.05
Stéréo	1.37	6.32	0.28	0.06	0.03	0.13	2.66/0.07
Stéréo rect	3.56	1.33	1.93	0.11	0.27	0.02	2.28/0.13
Distance <i>P3D</i>	0.23	0.14	0.70	0.01	0.04	0.01	0.36/0.02

TAB. 3.9 – Comparaison des résultats sur les paramètres de transformation pince-caméra.

3.5.5.4 Évaluation des méthodes avec les critères généraux

Le tableau 3.10 nous montre les valeurs des critères généraux. De plus, ces critères sont directement issus des fonctions coût utilisées pour chacune des méthodes présentées.

Erreur Méthode	Σ_d (mm)	\bar{d} (mm)	σ_d (mm)	Σ_e (px)	\bar{e} (px)	σ_e (px)	Res.
Stéréo	12833	25.67	12.30	1384.2	1.38	2.75	1455.6
Stéréo rect	2873.8	5.75	4.30	8.23	$8.2e^{-3}$	0.01	8.23
Distance <i>P3D</i>	1.27	$2.5e^{-3}$	$1.4e^{-3}$	$6.8e^{-4}$	$6.8e^{-7}$	$1.7e^{-6}$	1.27

TAB. 3.10 – Comparaison des résultats avec les critères généraux.

Ces résultats sont le reflet de ceux obtenus précédemment. En effet, on constate que pour le cas **Stéréo** la distance moyenne entre les points 3D reconstruits (\bar{d}), qui représente en réalité l'erreur de recalage entre les vues est énorme (2.5cm) et l'erreur moyenne sur les distances épipolaires est supérieur au pixel. Le critère **Stéréo rect** présente quant à lui des erreurs faibles, tout à fait acceptables, et donc une solution proche des paramètres réels. Finalement, le critère **Distance P3D** présente des erreurs négligeables et montre ainsi une très bonne convergence vers la solution réelle.

En remarque, la méthode **Mono** est sans objet dans ce tableau car deux caméras sont nécessaires pour le calcul de ces critères.

3.5.5.5 Conclusion

Cette première simulation met en avant quelques défaillances sur la méthode **Stéréo**. Ces défaillances proviennent d'un trop grand nombre de paramètres à estimer, ce qui a pour conséquence d'augmenter le nombre de minima locaux lors de l'optimisation. Néanmoins les résultats obtenus sur les méthodes utilisant un banc de stéréovision rectifié sont très encourageants et montrent qu'il est possible, avec une initialisation grossière, de converger vers une solution proche de la réalité.

3.5.6 Influence du nombre de points

L'utilisation de caméras infrarouges implique certaines contraintes comme travailler avec un faible nombre de points extraits. Nous avons donc choisi d'étudier l'influence de la quantité de points observés sur la précision de chacun des paramètres. Les déplacements du robot (6 vues), les paramètres et la répartition des points restent identiques à la première simulation.

Tout d'abord, nous présentons dans les tableaux (3.11, 3.12, 3.13 et 3.14) les cas extrêmes avec un nuage de 9 et 400 points 3D aléatoires. Ils montrent que quelque soit la méthode, le nombre de points a très peu d'influence sur les erreurs des paramètres estimés.

Erreur Méthode	ε_{α_u} (%)	$\varepsilon_{\alpha_{u'}}$ (%)	ε_{α_v} (%)	$\varepsilon_{\alpha_{v'}}$ (%)	ε_{u_0} (%)	$\varepsilon_{u'_0}$ (%)	ε_{v_0} (%)	$\varepsilon_{v'_0}$ (%)	σ_{int} (%)
Mono (9 points)	0.10	–	0.05	–	0.23	–	0.91	–	0.32
Mono (400 points)	0.20	–	0.17	–	0.55	–	0.35	–	0.32
Stéréo (9 points)	4.53	4.32	5.94	6.00	0.99	2.80	1.68	4.54	3.85
Stéréo (400 points)	3.44	2.84	5.69	5.23	0.93	3.33	3.69	4.87	3.75
Stéréo rect (9 points)	0.07	0.24	1.50	1.50	1.05	0.44	3.06	3.06	1.06
Stéréo rect (400 points)	0.14	0.35	0.24	0.24	2.14	1.91	0.06	0.06	0.81
Distance <i>P3D</i> (9 points)	0.03	0.03	0.04	0.04	0.11	0.13	0.01	0.01	0.06
Distance <i>P3D</i> (400 points)	$3.9e^{-3}$	$3.4e^{-3}$	$4.4e^{-3}$	$4.4e^{-3}$	0.01	0.01	$6.5e^{-3}$	$6.5e^{-3}$	$7.2e^{-3}$

TAB. 3.11 – Comparaison des résultats sur les paramètres intrinsèques des caméras.

Erreur Méthode	$\varepsilon_{t_{ext_x}}$ (mm)	$\varepsilon_{t_{ext_y}}$ (mm)	$\varepsilon_{t_{ext_z}}$ (mm)	$\varepsilon_{r_{ext_x}}$ (°)	$\varepsilon_{r_{ext_y}}$ (°)	$\varepsilon_{r_{ext_z}}$ (°)	$\sigma_{t_{ext}}/\sigma_{r_{ext}}$ (mm/°)
Stéréo (9 points)	4.14	3.47	0.11	0.07	0.69	0.47	2.57/0.41
Stéréo (400 points)	4.09	3.14	0.25	0.11	0.75	0.47	2.49/0.44
Stéréo rect (9 points)	1.87	0	0	0	0	0	1.87/0
Stéréo rect (400 points)	1.07	0	0	0	0	0	1.07/0
Distance <i>P3D</i> (9 points)	0.10	0	0	0	0	0	0.10/0
Distance <i>P3D</i> (400 points)	0.01	0	0	0	0	0	0.01/0

TAB. 3.12 – Comparaison des résultats sur les paramètres extrinsèques des caméras.

Ensuite, les figures 3.3, 3.4 et 3.5 présentent une analyse plus détaillée en fonction d'un nombre de points croissant.

La figure 3.3 représente l'évolution de l'erreur d'estimation des paramètres intrinsèques et extrinsèques en fonction du nombre de points. Cette figure met en évidence que, quel que soit le nombre de point, l'estimation des paramètres intrinsèques est extrêmement stable et précise pour les méthodes Mono, Stéréo rect et Distance *P3D*. Cependant, pour l'estimation des paramètres extrinsèques, nous observons une légère variation de l'erreur (de l'ordre de 1%) en fonction du nombre de points utilisé pour la méthode Stéréo rect.

Erreur Méthode	$\varepsilon_{t_{PCx}}$ (mm)	$\varepsilon_{t_{PCy}}$ (mm)	$\varepsilon_{t_{PCz}}$ (mm)	$\varepsilon_{r_{PCx}}$ (°)	$\varepsilon_{r_{PCy}}$ (°)	$\varepsilon_{r_{PCz}}$ (°)	$\sigma_{t_{PC}}/\sigma_{r_{PC}}$ (mm/°)
Mono (9 points)	2.00	1.55	1.56	0.23	0.02	0.03	1.70/0.09
Mono (400 points)	1.76	1.99	1.55	0.19	0.00	0.01	1.77/0.07
Stéréo (9 points)	1.42	6.42	0.28	0.10	0.13	0.20	2.71/0.14
Stéréo (400 points)	1.40	6.30	0.20	0.04	0.06	0.15	2.63/0.08
Stéréo rect (9 points)	4.11	4.28	3.48	0.71	0.13	0.08	3.96/0.31
Stéréo rect (400 points)	3.58	1.27	1.88	0.13	0.26	0.01	2.24/0.13
Distance P3D (9 points)	0.20	0.16	0.35	0.01	0.01	0.00	0.24/0.01
Distance P3D (400 points)	0.01	0.01	0.03	0.00	$3.2e^{-3}$	0.00	0.02/1.6e ⁻³

TAB. 3.13 – Comparaison des résultats sur les paramètres du Hand-eye.

Erreur Méthode	Σ_d (mm)	\bar{d} (mm)	σ_d (mm)	Σ_e (px)	\bar{e} (px)	σ_e (px)	Res.
Stéréo (9 points)	933.3	20.73	7.66	101.55	1.12	1.95	107.70
Stéréo (400 points)	61056	30.52	11.84	4490.9	1.12	2.12	6275.5
Stéréo rect (9 points)	267.8	5.95	3.22	8.51	0.09	0.12	8.51
Stéréo rect (400 points)	12022	6.01	4.68	38.63	$9.7e^{-3}$	0.01	38.63
Distance P3D (9 points)	1.39	0.03	0.01	$3.6e^{-3}$	$3.9e^{-5}$	$9.3e^{-5}$	1.39
Distance P3D (400 points)	6.26	$3.1e^{-3}$	$1.4e^{-3}$	$1.2e^{-3}$	$3.1e^{-7}$	$5.1e^{-7}$	6.26

TAB. 3.14 – Comparaison des résultats avec les critères généraux.

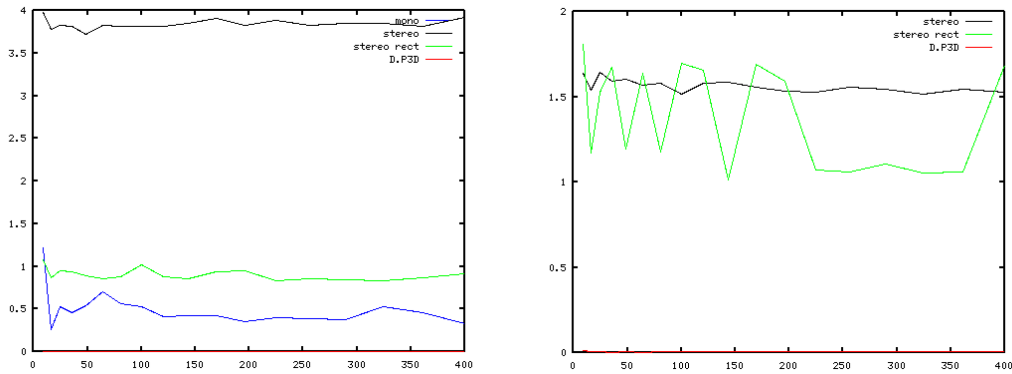


FIG. 3.3 – Évolution de la moyenne de l'erreur sur les paramètres intrinsèques σ_{int} (à gauche) et sur la translation des paramètres extrinsèques σ_{ext} (à droite) en fonction du nombre de points.

Les figures 3.4 et 3.5 représentent l'influence du nombre de points sur la qualité d'estimation des paramètres Pince-Caméra représentée par les critères définis ci-dessus. Sur ces graphes, une très légère diminution de l'erreur est observée avec l'augmen-

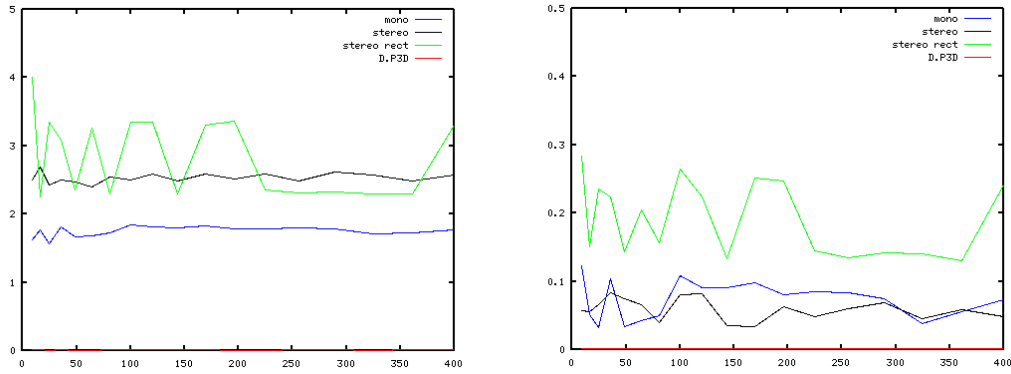


FIG. 3.4 – Évolution de la moyenne de l'erreur sur la translation Pince-Caméra $\sigma_{t_{PC}}$ (à gauche) et sur la rotation Pince-Caméra $\sigma_{r_{PC}}$ (à droite) en fonction du nombre de points.

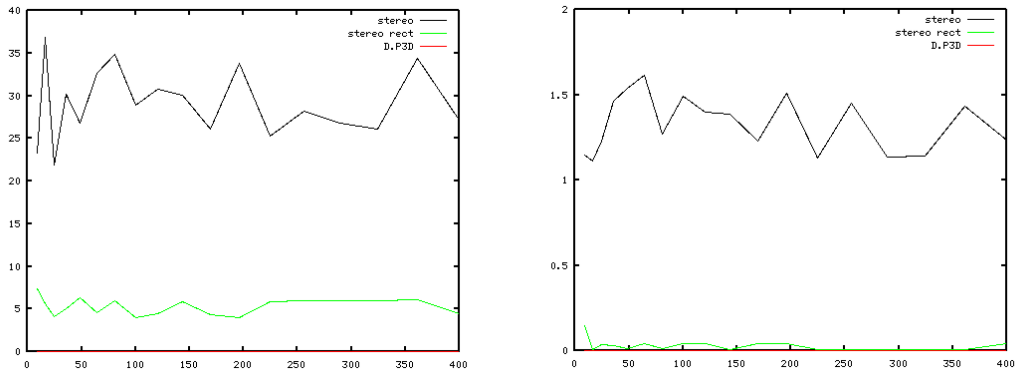


FIG. 3.5 – Évolution de la valeur moyenne de la distance des points 3D reconstruits \bar{d} (à gauche) et de la valeur moyenne de la distance des points à leur droite épipolaire \bar{e} (à droite) en fonction du nombre de points.

tation du nombre de points. Là encore, une légère variation de l'erreur pour la méthode Stéréo rect est notée en fonction du set de point utilisé. Cette tendance semble montrer une plus grande sensibilité à la répartition des points pour cette méthode.

L'ensemble des résultats montre une grande stabilité des résultats en fonction du nombre de points avec une variation comprise entre 1% et 2% d'erreur, variation mettant plus en évidence un petit manque de stabilité venant de l'optimisation qu'une influence du nombre points.

En conclusion, dans la totalité des cas, l'influence du nombre de points sur la précision de notre algorithme, quelque soit la méthode, est très faible. Ceci est d'une grande importance car l'extraction d'un grand nombre de points sur des images infrarouges reste un problème majeur.

3.5.7 Influence du bruit sur les points des images

L'utilisation d'images infrarouges implique généralement de travailler avec des images contenant du bruit, ceci est due à la technologie utilisée ou encore à certaines perturbations thermiques. Afin d'évaluer l'influence du bruit lors de l'extraction des points, nous avons ajouté un bruit aléatoire à chacun des points extraits et nous avons fait varier ce bruit de 0 à +/-2 pixels. Pour cette simulation, nous prenons un nuage de 100 points 3D aléatoires avec 6 mouvements du robot.

Les tableaux 3.15, 3.16, 3.17 et 3.18 représentent les erreurs sur chaque paramètre en fonction de la méthode pour un bruit de +/-1 pixel.

Erreur Méthode	ε_{α_u} (%)	$\varepsilon_{\alpha_{u'}}$ (%)	ε_{α_v} (%)	$\varepsilon_{\alpha_{v'}}$ (%)	ε_{u_0} (%)	$\varepsilon_{u'_0}$ (%)	ε_{v_0} (%)	$\varepsilon_{v'_0}$ (%)	σ_{int} (%)
Mono	0.16	-	0.15	-	0.43	-	1.06	-	0.45
Stéréo	3.89	2.74	5.64	5.07	0.19	3.84	2.99	4.76	3.64
Stéréo rect	0.29	0.20	1.12	1.12	0.44	0.70	2.54	2.54	0.88
Distance <i>P3D</i>	3.22	3.02	0.44	0.44	4.18	5.15	0.62	0.62	2.77

TAB. 3.15 – Comparaison des résultats sur les paramètres intrinsèques des caméras.

Erreur Méthode	$\varepsilon_{t_{extx}}$ (mm)	$\varepsilon_{t_{exty}}$ (mm)	$\varepsilon_{t_{extz}}$ (mm)	$\varepsilon_{r_{extx}}$ (°)	$\varepsilon_{r_{exty}}$ (°)	$\varepsilon_{r_{extz}}$ (°)	$\sigma_{t_{ext}}/\sigma_{r_{ext}}$ (mm/°)
Stéréo	4.08	4.11	0.08	0.19	0.86	0.37	2.76/0.48
Stéréo rect	1.38	0	0	0	0	0	1.38/0
Distance <i>P3D</i>	2.62	0	0	0	0	0	2.62/0

TAB. 3.16 – Comparaison des résultats sur les paramètres extrinsèques des caméras.

Erreur Méthode	$\varepsilon_{t_{PCx}}$ (mm)	$\varepsilon_{t_{PCy}}$ (mm)	$\varepsilon_{t_{PCz}}$ (mm)	$\varepsilon_{r_{PCx}}$ (°)	$\varepsilon_{r_{PCy}}$ (°)	$\varepsilon_{r_{PCz}}$ (°)	$\sigma_{t_{PC}}/\sigma_{r_{PC}}$ (mm/°)
Mono	1.68	2.29	1.35	0.18	0.01	0.9	1.77/0.09
Stéréo	1.57	6.29	0.61	0.03	0.17	0.10	2.83/0.10
Stéréo rect	4.34	3.03	3.43	0.47	0.21	0.16	3.60/0.28
Distance <i>P3D</i>	1.47	5.12	32.30	0.26	1.03	0.12	12.96/0.47

TAB. 3.17 – Comparaison des résultats sur les paramètres du Hand-eye.

En comparant ces résultats avec ceux des tableaux 3.7, 3.8, 3.9 et 3.10, deux comportements différents sont observés. Le premier constat est que l'erreur sur l'ensemble des paramètres augmente considérablement avec la méthode **Distance P3D**

Erreur Méthode	Σ_d (mm)	\bar{d} (mm)	σ_d (mm)	Σ_e (px)	\bar{e} (px)	σ_e (px)	Res.
Stéréo	19349	38.69	18.41	1411	1.411	2.66	1534
Stéréo rect	9061	18.12	6.56	78.41	0.07	0.11	78.41
Distance <i>P3D</i>	2194.7	4.38	4.68	239.92	0.23	0.36	2194.7

TAB. 3.18 – Comparaison des résultats avec les critères généraux.

lorsque le bruit augmente. Le second est que l'influence de ce bruit sur l'erreur est très faible avec les méthodes basées sur une décomposition de la matrice Fondamentale.

Une analyse plus complète a été réalisée sur les figures 3.6, 3.7 et 3.8 en fonction d'une évolution linéaire du bruit sur les points extraits entre 0 et ± 2 pixels.

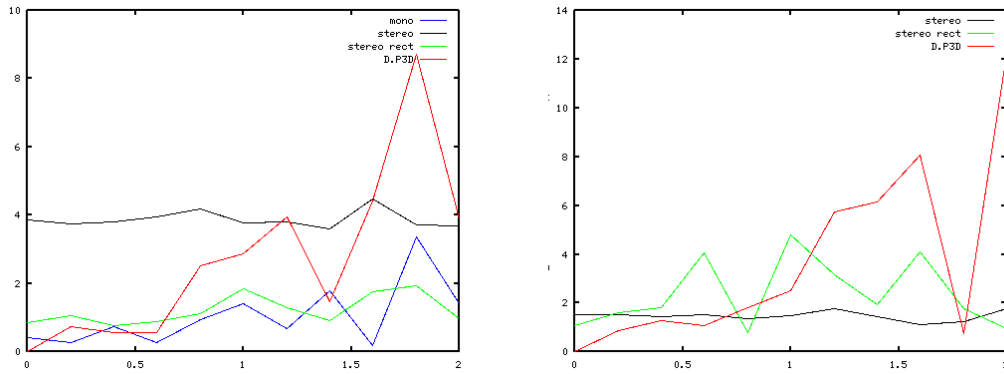


FIG. 3.6 – Évolution de la moyenne de l'erreur sur les paramètres intrinsèques σ_{int} (à gauche) et sur la translation des paramètres extrinsèques σ_{text} (à droite) en fonction de l'écart type du bruit.

La figure 3.6 représente la valeur moyenne de l'erreur des paramètres intrinsèques, σ_{int} , et l'erreur de translation des paramètres extrinsèques, σ_{text} , pour une variation du bruit de 0 à ± 2 pixels. Cette figure montre une augmentation plus importante de l'erreur lors de l'estimation des paramètres intrinsèques avec la méthode Distance *P3D* lorsque le niveau de bruit augmente. En revanche, l'erreur reste constante avec les méthodes basées sur une décomposition de la matrice Fondamentale. Avec l'estimation paramètres extrinsèques, la tendance est confirmée et la méthode Stéréo rect offre la précision la plus élevée lorsque le niveau de bruit est supérieur à 1.2 pixel.

L'erreur de translation et de rotation sur les paramètres Pince-Caméra, $\sigma_{t_{PC}}$ et $\sigma_{r_{PC}}$, représentée sur la figure 3.7, est minimale avec la méthode Distance *P3D* lorsque la valeur du bruit est inférieur à 0.6 pixel. Cependant, pour des valeurs de bruit supérieures, la tendance s'inverse et la méthode Stéréo rect est significa-

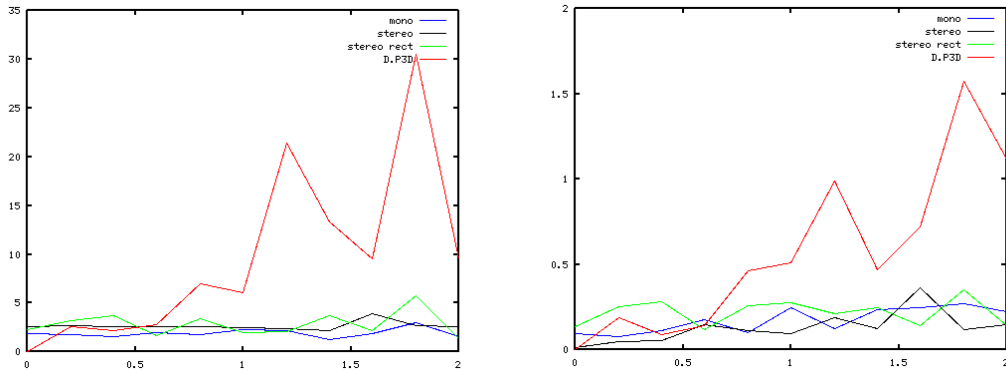


FIG. 3.7 – Évolution de la moyenne de l'erreur sur la translation Pince-Caméra $\sigma_{t_{PC}}$ (à gauche) et sur la rotation Pince-Caméra $\sigma_{r_{PC}}$ (à droite) en fonction de l'écart type du bruit.

tivement plus performante avec une erreur maximale de $4mm$ (resp. 0.2°) pour la translation (resp. la rotation).

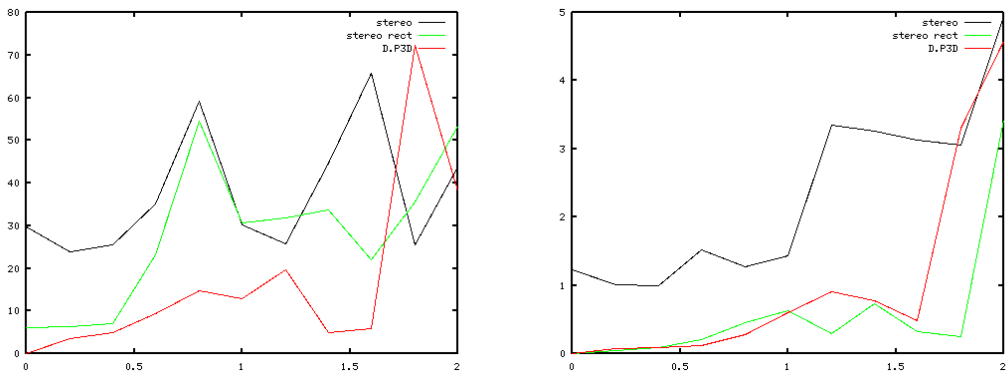


FIG. 3.8 – Évolution de la valeur moyenne de la distance des points $3D$ reconstruits \bar{d} (à gauche) et de la valeur moyenne de la distance des points à leur droite épipolaire \bar{e} (à droite) en fonction de l'écart type du bruit.

La figure 3.8 montre la valeur moyenne \bar{d} de la distance entre les points $3D$ reconstruits et la valeur moyenne \bar{e} de la distance des points à leur droite épipolaire en fonction du bruit. Là encore, on observe un meilleur comportement de la méthode Stéréo rect quand la valeur du bruit dépasse 0.6 pixel.

En conclusion, on constate que la méthode Stéréo rect est moins sensible à l'influence du bruit. Dans un contexte d'extraction avec un niveau de bruit faible

(moins de 0,5 pixel), la méthode Distance *P3D* est plus précise. Toutefois, lorsque le niveau de bruit augmente, la méthode Stéréo rect est moins sensible et elle est alors plus efficace. Il est à noter qu'avec les techniques d'extraction actuelles, telles que Harris[Harris 1988] ou SIFT[Lowe 2004], le niveau de bruit est rarement au dessus de 0.5 pixels.

3.5.8 Influence de la répartition des points

Nous avons remarqué lors d'expérimentations que, comme le champ de vision des caméras infrarouges est très faible, la majorité des points extraits se situent sur des plans ou des zones quasi-planes. L'objectif est donc d'étudier l'influence de cette répartition des points sur l'erreur obtenue sur les paramètres estimés.

Le nombre de points (100 points), les déplacements du robot (6 vues) et l'ensemble des paramètres restent identiques aux simulations précédentes. Seule la position des points est différente, ils ne sont plus disposés dans l'espace en 3D mais aléatoirement sur un plan en 2D.

Les tableaux 3.19, 3.20, 3.21 et 3.22 illustrent le pourcentage d'erreurs, pour les paramètres intrinsèques, et l'erreur en millimètres et degrés, pour les paramètres extrinsèques et de la transformation pince-caméra en fonction des méthodes.

Erreur Méthode	ε_{α_u} (%)	$\varepsilon_{\alpha_{u'}}$ (%)	ε_{α_v} (%)	$\varepsilon_{\alpha_{v'}}$ (%)	ε_{u_0} (%)	$\varepsilon_{u'_0}$ (%)	ε_{v_0} (%)	$\varepsilon_{v'_0}$ (%)	σ_{int} (%)
Mono	0.90	–	0.32	–	1.44	–	4.41	–	1.77
Stéréo	2.91	2.87	5.23	4.88	0.02	5.25	5.20	5.67	4.00
Stéréo rect	0.26	0.36	1.10	1.10	1.07	0.61	4.33	4.33	1.29
Distance <i>P3D</i>	0.06	0.05	0.07	0.07	0.19	0.23	0.01	0.01	0.10

TAB. 3.19 – Comparaison des résultats sur les paramètres intrinsèques des caméras.

Erreur Méthode	$\varepsilon_{t_{ext_x}}$ (mm)	$\varepsilon_{t_{ext_y}}$ (mm)	$\varepsilon_{t_{ext_z}}$ (mm)	$\varepsilon_{r_{ext_x}}$ (°)	$\varepsilon_{r_{ext_y}}$ (°)	$\varepsilon_{r_{ext_z}}$ (°)	$\sigma_{t_{ext}}/\sigma_{r_{ext}}$ (mm/°)
Stéréo	4.21	3.31	0.05	0.22	0.73	0.48	2.52/0.48
Stéréo rect	1.82	0	0	0	0	0	1.82/0
Distance <i>P3D</i>	0.18	0	0	0	0	0	0.18/0

TAB. 3.20 – Comparaison des résultats sur les paramètres extrinsèques des caméras.

Ces résultats montrent que l'erreur sur l'estimation des paramètres intrinsèques pour les méthodes **Mono** et **Stéréo rect** augmente. Une répartition des points sur un plan *2D* supprime implicitement une contrainte d'espace. Les méthodes basées sur une décomposition de la matrice Fondamentale sont donc plus influencées que la méthode **Distance P3D**.

Erreur Méthode	ε_{tPC_x} (mm)	ε_{tPC_y} (mm)	ε_{tPC_z} (mm)	ε_{rPC_x} (°)	ε_{rPC_y} (°)	ε_{rPC_z} (°)	$\sigma_{tPC}/\sigma_{rPC}$ (mm/°)
Mono	1.32	1.80	1.57	0.42	0.13	0.10	1.56/0.22
Stéréo	1.26	7.03	0.72	0.23	0.09	0.28	3.00/0.20
Stéréo rect	4.19	3.36	3.12	0.79	0.12	0.02	3.56/0.31
Distance <i>P3D</i>	0.25	0.26	0.65	0.02	0.03	0.00	0.39/0.02

TAB. 3.21 – Comparaison des résultats sur les paramètres du Hand-eye.

Erreur Méthode	Σ_d (mm)	\bar{d} (mm)	σ_d (mm)	Σ_e (px)	\bar{e} (px)	σ_e (px)	Res.
Stéréo	21795	43.59	21.49	1665.9	1.66	3.34	1859.5
Stéréo rect	1790.6	3.58	2.10	31.39	0.03	0.04	31.39
Distance <i>P3D</i>	1.95	$3.9e^{-3}$	$1.9e^{-3}$	$6.1e^{-4}$	$6.1e^{-7}$	$12.5e^{-7}$	1.95

TAB. 3.22 – Comparaison des résultats avec les critères généraux.

3.5.9 Conclusion

Pour conclure, l'ensemble de ces simulations ont montré les qualités et les défauts de chaque méthode. La méthode **Stéréo**, due à un trop grand nombre de paramètres à estimer, ne converge jamais vers un résultat acceptable. Elle ne sera pas utilisée lors des tests sur des images réelles. La méthode **Mono** est plus précise et très stable et pourra être utilisée dans des systèmes ne nécessitant qu'une caméra. Les méthodes **Stéréo rect** et **Distance P3D** sont les méthodes les plus précises. La méthode **Stéréo rect** est plus stable et plus précise dans les contextes bruités, la méthode **Distance P3D** est plus précise dans tous les autres cas. De plus nous notons qu'afin d'obtenir le meilleur résultat, l'important n'est pas d'extraire un maximum de points lors de chaque vue, mais plutôt de favoriser la qualité et la précision de l'extraction.

3.6 Performances des méthodes de calibrage sélectionnées sur des données réelles

Cette section est consacrée à l'évaluation des deux méthodes de calibrage sélectionnées, la méthode **Stéréo rect** et **distance 3D** sur des images réelles.

Cette évaluation est menée d'une part avec des images visibles qui représentent un contexte habituel et favorable car les images sont de meilleure résolution et plus texturées. L'extraction des points est facilité et nous sommes donc dans des conditions d'un nombre de points extraits élevés, repartis uniformément et avec un faible niveau de bruit. Et d'autre part, avec des images infrarouges qui représentent le contexte final de notre application, et donc dans des conditions moins favorables présentant un faible nombre de points extraits avec un niveau de bruit élevé.

En remarque, l'algorithme d'optimisation et l'initialisation restent identiques à ceux de la section précédente.

3.6.1 Contexte expérimental

Le contexte expérimental est décrit ici de façon succincte, l'ensemble de l'expérimentation et le matériel utilisé sont décrits de manière plus détaillée dans le chapitre 5.

L'objet $3D$ utilisé est composé de deux surfaces planes de 20cm formant un angle de 90° . Pour les images infrarouges, cet objet est chauffé à $35^\circ C$ à l'aide de résistances chauffantes. Les caméras utilisées sont :

- pour les images visibles issues de caméras CCD de résolution 800×600 pixels.
- pour les images infrarouges issues de caméras infrarouges microbolométriques ($8 - 12\mu m$) de résolution 164×128 pixels.

Les données, qui sont les points extraits des images pour 8 déplacements du robot, sont de l'ordre de 1300 points par déplacement pour les images visibles et de 13 points par déplacement pour les images infrarouges.

3.6.2 Données issues d'images visibles

La figure 3.9 montre l'objet $3D$ dans l'image gauche et droite dans le domaine visible.

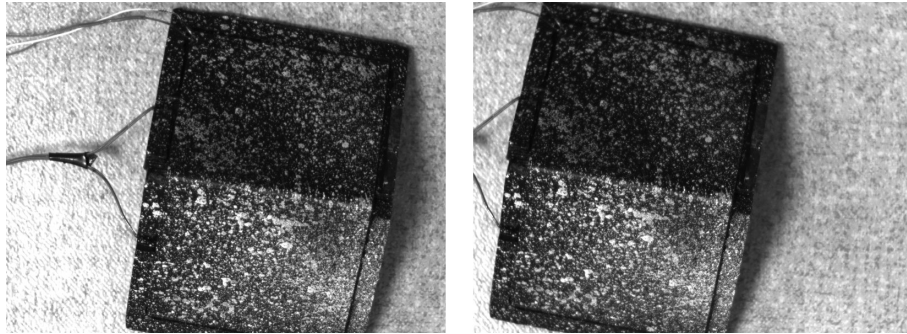


FIG. 3.9 – Exemple d'images visibles utilisées pour l'évaluation.

Le tableau 3.23 représente la somme, la moyenne et l'écart type des distances de chacun des points $3D$, Σ_d , \bar{d} , σ_d . Il illustre également la somme, la moyenne, et l'écart type des distances croisées entre chaque point et la droite épipolaire qui lui est associée Σ_e , \bar{e} , σ_e .

Erreur Méthode	Σ_d (mm)	\bar{d} (mm)	σ_d (mm)	Σ_e (px)	\bar{e} (px)	σ_e (px)
Stéréo rect	5931.40	2.36	2.16	3568.99	0.71	1.40
Distance $P3D$	2061.63	0.82	0.73	3042.58	0.60	1.45

TAB. 3.23 – Comparaison des résultats sur des images visibles.

Le tableau montre que les deux méthodes donnent des résultats acceptables

pour notre application. Cependant, la méthode **Distance 3D** reste globalement plus précise.

Ce test reflète les résultats obtenus lors des simulations en contexte très légèrement bruité (environs 0.1 pixel de bruit) et permet de valider la méthode dans un contexte réel avec des images contenant un grand nombre de points extraits et avec un faible niveau de bruit.

3.6.3 Données issues d'images infrarouges

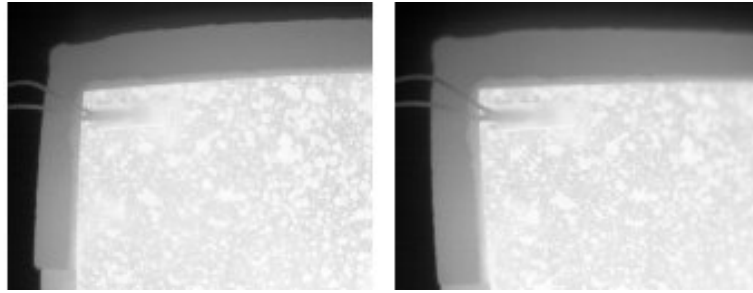


FIG. 3.10 – Exemple d'images infrarouges utilisées pour l'évaluation.

Le tableau 3.24 donne les six paramètres : Σ_d , \bar{d} , σ_d (qui représentent respectivement la somme, la moyenne et l'écart type des distances entre les points 3D) et Σ_e , \bar{e} , σ_e (qui représentent respectivement la somme, la moyenne et l'écart type des distances croisées entre chaque point et la droite épipolaire qui lui est associée).

Erreur Méthode	Σ_d (mm)	\bar{d} (mm)	σ_d (mm)	Σ_e (px)	\bar{e} (px)	σ_e (px)
Stéréo rect	$2.66e + 05$	4661.76	$1.36e + 04$	2636.89	23.13	29.74
Distance P3D	113.82	2.14	5.29	630.55	5.94	12.72

TAB. 3.24 – Comparaison des résultats sur des images infrarouges.

Le tableau 3.24 montre que dans ce cas, avec un faible nombre de points et une mauvaise précision d'extraction, les deux critères ont plus de mal à converger. Le critère **Stéréo rect** diverge complètement avec en moyenne 4.6m d'erreur de recalage entre deux nuages de points 3D. Le critère **Distance P3D** quant à lui converge vers une solution comparable à une simulation avec un niveau de bruit de 0.4 pixel. Cependant cette solution reste tout à fait acceptable pour notre application car l'erreur moyenne sur le recalage est de 2mm, ce qui est inférieur à la précision de reconstruction.

3.7 Conclusion

Ce chapitre décrit une méthode d'auto-calibrage prenant en compte les paramètres des caméras ainsi que les paramètres Pince-Caméra. Cette méthode estime l'ensemble des paramètres en une seule étape d'optimisation et fonctionne avec des caméras infrarouges et par conséquent avec un faible nombre de points. Plusieurs fonctions coûts ont été proposées et ont été testées dans de multiples conditions.

Les résultats obtenus pour la première méthode sont correct, celle ci permet l'auto-calibrage d'une unique caméra montée sur un robot et pourra être utilisé sur des robots autonomes qui ne nécessitent pas une grande précision de reconstruction ou qui évoluent dans un milieu hostile où l'utilisation de mires est impossible.

La deuxième fonction coût a été présentée car elle représente une étape clé à la compréhension des méthodes **Stéréo rect** et **Distance P3D**. Elle permet aussi de montrer la difficulté de trouver une solution acceptable dès lors que l'on utilise un banc de stéréovision. En effet, le nombre de paramètres à estimer étant trop grand, cette solution ne converge jamais vers une solution acceptable quelque soit la situation.

Les méthodes **Stéréo rect** et **Distance P3D** sont des méthodes basées sur l'utilisation d'un banc de stéréovision préalablement rectifié. La rectification permet de diminuer le nombre de paramètres à estimer et de converger vers un résultat correct. Les résultats obtenus, dans la majorité des cas testés, sont tout à fait acceptables pour une utilisation dans notre projet. Nous avons montré que la méthode **Distance P3D** est plus précise, hormis lors d'un contexte très bruité, où l'utilisation de la méthode **Stéréo rect** est plus judicieuse.

Au final, nous avons appliqué ces méthodes à des cas réels, permettant de valider l'approche avec deux caméras infrarouges montées sur un robot six axes, étape nécessaire au projet R3T. Il est à noter que dans ce cas seul la méthode **Distance P3D** fonctionne.

Stéréovision infrarouge

Sommaire

4.1	Extraction et appariement	80
4.1.1	Détecteur "Phase Congruency"	80
4.1.2	Détecteur de "Harris"	81
4.1.3	Détecteur "SIFT"	82
4.1.4	Comparaison des détecteurs sur des images infrarouges	84
4.2	Filtrage des faux appariements	86
4.2.1	La méthode des "M-Estimeurs"	88
4.2.2	La méthode de "RANSAC"	88
4.2.3	La méthode "LTS" ou "LMS"	89
4.2.4	Comparaison	89
4.3	Corrélation	91
4.3.1	Principe de la corrélation	91
4.3.2	Méthodes locales	91
4.3.3	Méthodes globales	92
4.3.4	Conclusion	93
4.4	Reconstruction 3D avec des images infrarouges	93
4.5	Conclusion	94

Dans ce chapitre nous présentons l'ensemble des étapes nécessaires à la reconstruction tridimensionnelle ($3D$) d'objets à partir d'un banc de stéréovision composé de deux caméras thermiques fonctionnant dans une bande de 8 à $12\mu m$ (cf. figure 4.1). La stéréovision permet à partir de deux images, prises sous angle de vue différent, d'un même objet, de déterminer sa forme, ses dimensions et sa position si les caméras sont calibrées.



FIG. 4.1 – Présentation du système d'acquisition (à gauche de la figure) et de l'objet à reconstruire (à droite de la figure).

La méthode de stéréovision est basée sur des caméras infrarouges dont la résolution est de 164×128 pixels (cette valeur est faible en comparaison de celle d'une matrice CCD ordinaire (1024×768 pixels)). De plus, les images obtenues possèdent généralement très peu de texture et de détails. Comme le montre la transformée de Laplace sur l'image infrarouge de la figure 4.2, les variations d'intensité sont très faibles.

Les étapes de la reconstruction $3D$ sont analysées en prenant en compte les spécificités de ces images infrarouges. Il est donc nécessaire pour chaque étape de déterminer les algorithmes les plus adaptés à la configuration retenue dans le projet basée sur deux caméras infrarouges non calibrées. La figure 4.3 illustre les six étapes de la reconstruction : acquisition, appariement de points d'intérêts, rectification, autocalibrage, corrélation et reconstruction $3D$.

La méthode débute donc par l'acquisition d'une paire d'images (dite ci-après images droite et gauche) depuis le banc stéréoscopique. Ces images acquises étant peu texturées, la première difficulté est de détecter un premier ensemble de points communs aux deux images, appelés points appariés ou stéréo correspondants. Cet

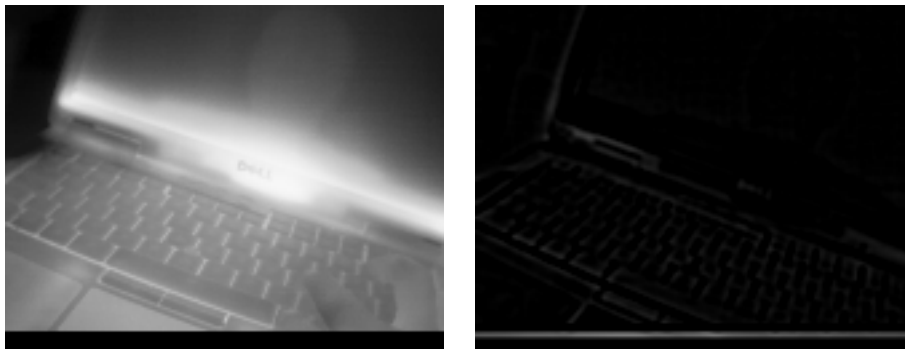


FIG. 4.2 – Image thermique d'un clavier d'ordinateur et sa transformée de Laplace.

ensemble de points peut ne représenter que 0.001% de l'image. Il doit cependant comporter un nombre de points suffisants pour réaliser l'étape suivante, la rectification, étape déterminante dans notre approche (décrite au chapitre 2), et l'étape d'autocalibrage (décrite au chapitre 3). Pour cela, nous avons évalué trois détecteurs de points proposés dans la littérature.

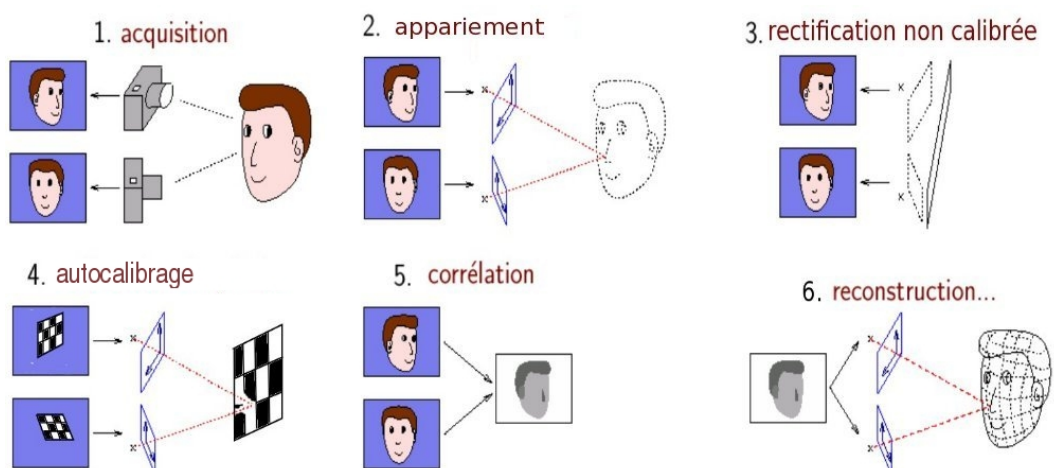


FIG. 4.3 – Approche proposée pour la reconstruction 3D infrarouge.

L'étape de corrélation ou mise en correspondance de tous les points entre chaque image consiste à localiser, dans l'image gauche et droite, les projections d'un même point 3D de la scène observée. La qualité de la reconstruction dépend de la qualité et de la densité de l'appariement. Il existe de nombreuses méthodes de mise en correspondance mais nous nous focalisons uniquement sur les méthodes de corrélation dense. En infrarouge, cette étape s'avère délicate et nous avons testé différentes méthodes de la littérature.

Sur la base de cette mise en correspondance dense, la position de chaque point

est alors calculée en fonction de la distance observée entre les points stéréo correspondants, nommée distance de disparité. Une valeur élevée de cette distance est représentative d'un point proche et inversement. La reconstruction euclidienne est ensuite réalisée dans le repère des caméras rectifiées. La scène ne peut être reconstruite en 3D dans l'espace euclidien que si l'étape d'autocalibrage du système rectifié a été effectué.

Le chapitre s'articule autour des différentes étapes de la reconstruction 3D : l'extraction d'un premier ensemble de points, la mise en correspondance, la corrélation et la triangulation. Pour chaque étape, les méthodes existantes sont analysées et comparées afin de déterminer la plus adaptée aux images infrarouges.

4.1 Extraction et appariement

L'étape d'extraction et appariement de points d'intérêts est l'une des étapes les plus importantes dans la stéréovision. En effet, elle fournit les points de bases qui doivent être précis et adaptés aux étapes comme la rectification et l'autocalibrage. Dans sa thèse Hajebi [Hajebi 2007] montre que lors d'observation de scènes quelconques à partir de caméras infrarouges, la méthode la plus adaptée pour l'extraction de points est une méthode basée sur la "congruence de phase". Cependant les images de notre application sont obtenues avec des prises de vues proches de l'objet et présentent moins de contours marqués. Les résultats étant moins concluants que ceux de la thèse, nous avons décidé de le comparer, sur nos images, avec les deux détecteurs de points d'intérêts les plus utilisés en vision classique : Harris[Harris 1988] et SIFT[Lowe 2004].

4.1.1 Détecteur "Phase Congruency"

La méthode d'extraction "Phase congruency" a pour particularité d'être robuste aux changements d'illumination et de contraste. Cette méthode a été introduite par Morroni [Morroni 1987], son principe est basé sur la décomposition de l'image en séries de Fourier. Les points d'intérêt sont localisés aux endroits où les composantes de Fourier sont les plus en phase. Afin de détecter ces points d'intérêt, il propose le critère suivant :

$$PC_1(x) = \frac{\sum_n A_n \cos(\phi_n(x) - \bar{\phi}(x))}{\sum_n A_n(x)} \quad (4.1)$$

où A_n représente l'amplitude de la n^{ime} composante en cosinus, $\phi_n(x)$ la phase à la position x et $\bar{\phi}(x)$ la moyenne des phases des différentes composantes de la décomposition de Fourier.

Cependant ce critère étant assez imprécis et très sensible au bruit, Kovési [Kovési 1999] a mis au point une mesure différente qui consiste à soustraire au cosinus l'amplitude des sinus de la déviation de phase, ce qui produit une réponse plus localisée. Cette nouvelle mesure intègre également la compensation du

bruit :

$$PC_2(x) = \frac{\sum_n W(x) [A_n(x) (\cos(\phi_n(x) - \phi(\bar{x})) - |\sin(\phi_n(x) - \phi(\bar{x}))|) - T]}{\sum_n A_n(x) + \varepsilon} \quad (4.2)$$

Le terme $W(x)$ est un facteur adapté au nombre de fréquences (la congruence sur grand nombre de fréquences est plus importante que la congruence sur peu de fréquences). Une petite constante ε est incorporé pour éviter la division par zéro. Les symboles [...] indiquent que la quantité contenue est égal à elle-même lorsque sa valeur est positive, et à zéro autrement. Nous pouvons donc noter que seules les valeurs énergétiques qui dépassent T , l'influence du bruit estimé, sont comptabilisés dans le résultat.

Afin de ne détecter que les coins et les contours, Kovesi [Kovesi 2003] propose un critère basé sur le détecteur 4.2 où les mesures suivantes sont effectuées sur chaque points de l'image :

$$a = \sum_{\theta} (PC_2(\theta) \cos(\theta))^2 \quad (4.3)$$

$$b = 2 \sum_{\theta} (PC_2(\theta) \cos(\theta))(PC_2(\theta) \sin(\theta)) \quad (4.4)$$

$$c = \sum_{\theta} (PC_2(\theta) \sin(\theta))^2 \quad (4.5)$$

Il définit ensuite, sur chaque pixel, un coefficient basé sur la covariance du moment maximum de la congruence de phase :

$$M = \frac{1}{2}(c + a + \sqrt{b^2 + (a - c)^2}) \quad (4.6)$$

Un ensemble de points d'intérêts est ensuite déterminé en appliquant un seuillage sur l'image de coefficients.

Pour la mise en correspondance des points entre deux images, Hajebi [Hajebi 2007] caractérise chacun des points à l'aide de coefficients "Log-Gabor", calculés à différentes échelles et différentes directions lors de l'extraction des points. Ensuite une simple comparaison est effectuée entre les points et seuls les points similaires sont retenus. Cette méthode de corrélation avait été proposée par Lades [Lades 1993].

4.1.2 Détecteur de "Harris"

L'idée du détecteur de Harris [Harris 1988] est de considérer le voisinage d'un pixel et de déterminer les changements moyens de l'intensité dans le voisinage considéré lorsque la fenêtre se déplace dans différentes directions. Plus précisément, ce changement d'intensité en fonction des déplacements (x, y) autour de la position (u, v) , est décrite par une fonction E définie comme suit :

$$E(x, y) = \sum_{u, v} w(u, v) \left[x \frac{\delta I}{\delta x} + y \frac{\delta I}{\delta y} + o(x^2, y^2) \right]^2 \quad (4.7)$$

avec :

- w spécifie la fenêtre considérée autour de (u, v) : chaque point est pondéré avec une fonction gaussienne $w(u, v) = e^{-\frac{u^2+v^2}{2\sigma^2}}$ avec σ^2 la variance,
- $I(u, v)$ représente l'intensité du pixel (u, v) ,
- $E(x, y)$ représente la moyenne du changement d'intensité lorsque la fenêtre est déplacée de (x, y) .

Afin de détecter des caractéristiques différentes dans l'image, la fonction E est décomposée de la manière suivante, en négligeant les termes au dessus de second ordre :

$$E(x, y) = Ax^2 + 2Cxy + By^2 = (x, y)M(x, y)^t \quad (4.8)$$

avec M la matrice qui caractérise le comportement local de la fonction E et exprimée par la relation :

$$M = \begin{pmatrix} A & C \\ C & B \end{pmatrix} \quad (4.9)$$

Les valeurs propres de cette matrice correspondent aux courbures principales associées à E :

- Si les deux courbures sont de faibles valeurs, alors la région considérée a une intensité approximativement constante.
- Si une des courbures est de forte valeur alors que l'autre est de faible valeur alors la région contient un contour.
- Si les deux courbures sont de fortes valeurs alors l'intensité varie fortement dans toutes les directions, ce qui caractérise un coin.

En conséquence, Harris et Stephen propose l'opérateur suivant pour détecter les coins et les contours dans une image :

$$R = Det(M) - kTrace(M)^2 \quad (4.10)$$

avec $Det(M) = AB - C^2$ et $Trace(M) = A + B$

Les valeurs de R sont positives au voisinage d'un coin, négatives au voisinage d'un contour et faibles dans une région d'intensité constante.

4.1.3 Détecteur "SIFT"

L'algorithme SIFT [Lowe 2004](Scale Invariant Feature Transform) sert non seulement à détecter des zones d'intérêts dans une image, mais il permet aussi de caractériser chacun des points d'intérêts afin de faciliter la mise en correspondance. La particularité de cet algorithme est qu'il est invariant aux rotations, aux changements d'échelle et d'illumination.

La première étape, permettant la détection de points d'intérêts, consiste à décomposer l'image en un espace échelle Gaussien(voir figure4.4). Cette décomposition consiste à appliquer récursivement à l'image $I(x, y)$ un filtre Gaussien. Elle est caractérisée par la fonction :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (4.11)$$

où $*$ est le produit de convolution sur x et y , σ le facteur d'échelle représentatif du nombre d'octaves (cf fig4.4) et

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (4.12)$$

La recherche de points d'intérêts s'effectue ensuite en recherchant les extrema dans l'espace de différence des Gaussiennes $D(x, y, \sigma)$:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (4.13)$$

avec k un facteur représentatif du nombre d'images par octave.

Les extrema sont calculés par rapport à leur voisins directs mais aussi par rapport à l'espace échelle (voir figure 4.4). Finalement, les extrema sont filtrés à l'aide d'un détecteur de contours afin d'obtenir un set de points d'intérêts.

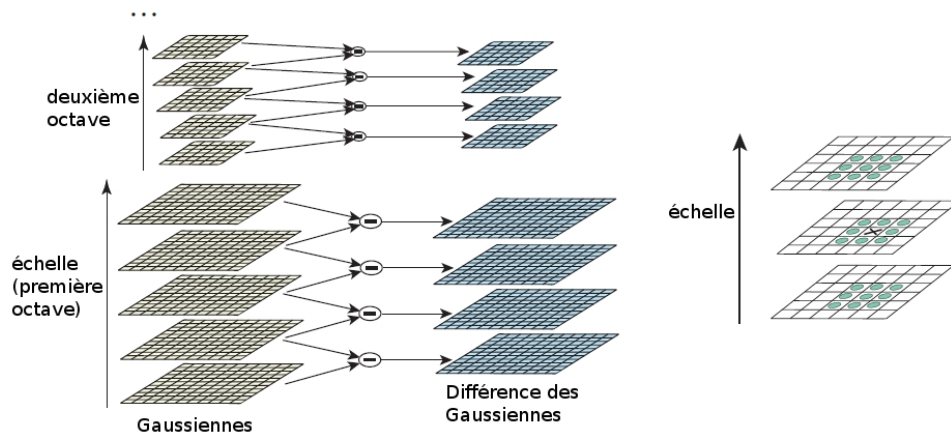


FIG. 4.4 – Représentation de la décomposition d'une image en espace échelle (à gauche) ; Recherche d'extréma dans l'espace de différence des gaussiennes (à droite)

Dans le but de faciliter la mise en correspondance de ces points entre différentes images, un descripteur est associé à chacun d'eux. Ce descripteur est obtenu en calculant, dans un premier temps, l'orientation θ et la magnitude m du gradient de l'image autour de chacun des points.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (4.14)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (4.15)$$

Ces données sont ensuite pondérées par une gaussienne centrée sur le point d'intérêt (représenté par un cercle sur la figure 4.5), puis regroupées dans plusieurs histogrammes (4 dans la figure 4.5) permettant de décrire l'entourage du point d'intérêt.

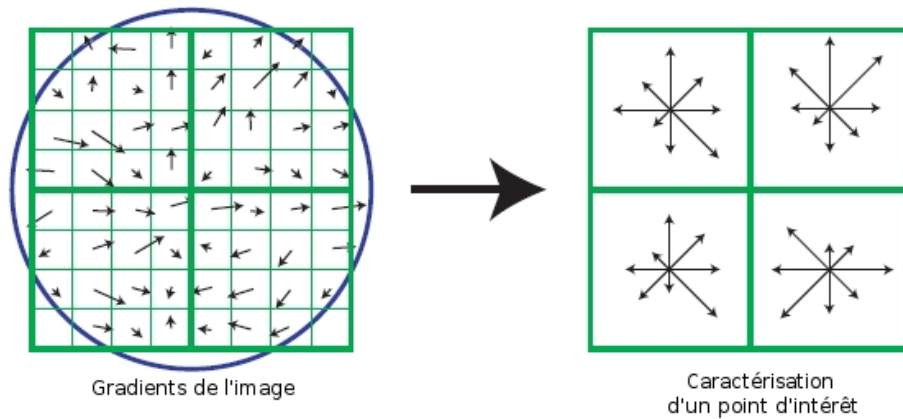


FIG. 4.5 – Orientation et magnitude des gradients de l'image autour d'un point d'intérêt (à gauche) ; caractérisation d'un point d'intérêt (à droite)

4.1.4 Comparaison des détecteurs sur des images infrarouges

Le but de ce paragraphe est de déterminer, pour chacune des étapes de la reconstruction, quel est le détecteur de points d'intérêts le plus adapté. Nous n'abordons pas ici l'étape de filtrage des faux appariements qui est traitée dans le paragraphe 4.2.

La figure 4.6 nous montre le nombre de points d'intérêts extraits grâce aux détecteurs "Phase Congruency", "Harris" ou "SIFT" appliqués sur une image infrarouge. Le détecteur "Phase congruency" extrait le plus grand nombre de points qui est de 833 points extraits. Le détecteur "Harris" (respectivement "SIFT") extrait 655 points d'intérêts (respectivement 144 points d'intérêts). Cependant, les points extraits avec le détecteur "Phase congruency" sont regroupés dans des zones de l'image présentant un fort contraste et la répartition n'est donc pas homogène.

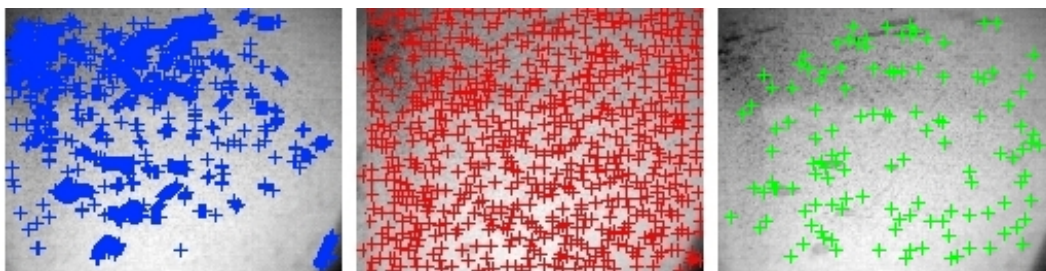


FIG. 4.6 – Extraction de points d'intérêts : Phase congruency (en bleu), Harris (en rouge) et SIFT (en vert)

Pour l'étape de rectification proposée (cf. chapitre 2), l'ensemble de points idéal

pour obtenir une grande précision et une faible distorsion devrait être composé d'un grand nombre de points uniformément répartis. Dans l'approche actuelle, cette étape est réalisée à l'initialisation sur la première prise de vue. La figure 4.7 illustre le résultat en appliquant les détecteurs "Phase Congruency+Log-Gabor" présenté au paragraphe précédent , "Harris+ZNCC" et "SIFT" avec un banc stéréo infrarouge statique lors d'une unique prise de vue.

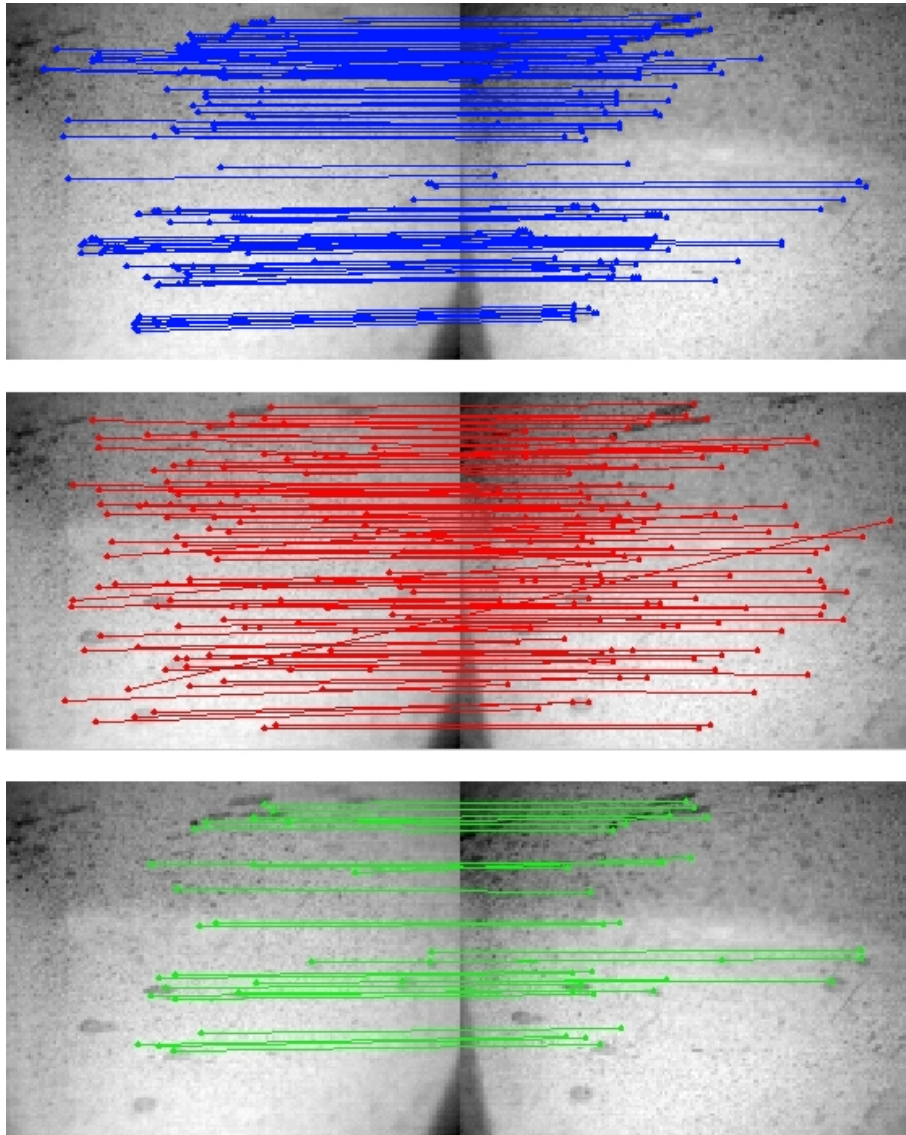


FIG. 4.7 – Corrélation de points d'intérêts : Phase Congruency+Log-Gabor (en bleu), Harris+ZNCC (en rouge), SIFT (en vert)

Le nombre de points appariés est de 153 avec le critère "Phase Congruency+Log-Gabor", de 113 pour le critère "ZNCC" appliqué aux points de Harris et de 40 pour le

détecteur "SIFT". La proportion de points corrélés par rapport au nombre de points extraits est largement inférieure pour les détecteurs de "Phase Congruency+Log-Gabor" et "Harris+ZNCC" (18 et 17%). Cette proportion est de 27% avec le détecteur "SIFT". Cependant, le nombre de points appariés est deux ou trois fois supérieur avec les détecteurs de "Phase Congruency+Log-Gabor" et "Harris+ZNCC". En revanche, la répartition de points la plus homogène est obtenue avec le détecteur de "Harris". Cette propriété étant essentielle pour l'étape de rectification, nous utiliserons le détecteur de "Harris" pour cette étape. On notera néanmoins la présence de quelques faux appariements malgré leur filtrage (cf. paragraphe 4.2).

L'étape d'autocalibrage (cf. chapitre 3) nécessite quant à elle un set de points précis corrélés au minimum entre quatre images issues d'un banc stéréo en déplacement. La deuxième comparaison illustrée sur la figure 4.8 est réalisée lors d'un déplacement du banc stéréo infrarouge. Elle montre les points extraits et corrélés sur les quatre différentes images. On remarque que seuls les détecteurs "Phase congruency+Log-Gabor" et "SIFT" sont capables de corrélérer correctement plusieurs points (resp. 28 et 8) dans les quatre images. Sur les 4 points corrélés avec le détecteur de "Harris+ZNCC", seul deux ne sont pas des faux appariements. En effet, les caractérisations des points par les méthodes "Log-Gabor" ou par "SIFT" sont beaucoup plus robustes. Le suivi de plusieurs points lors du déplacement des caméras est mieux assuré. Pour cette étape d'autocalibrage, nous choisissons le détecteur "Phase congruency" associé à une caractérisation des points par coefficient "Log-Gabor".

Pour conclure, ces tests montrent qu'un seul algorithme ne satisfait pas toutes les situations. Cependant quelque soit la méthode utilisée, un grand nombre de faux appariements est présent et un filtrage est nécessaire. Le paragraphe suivant compare les différentes méthodes de calculs robustes existantes afin d'obtenir un ensemble de points sans faux appariements.

4.2 Filtrage des faux appariements

Compte tenu de la faible résolution et du manque de texture des images infrarouge, l'ensemble des points extraits comporte généralement un nombre non négligeable de faux appariements. En contexte non calibré, le filtrage de ces points ne repose que sur la vérification de la relation de la matrice fondamentale (Cf. equation 2.3). Dans ce cas, les neuf coefficients de la matrice fondamentale sont calculés à partir de huit appariements et en résolvant un système linéaire composé de huit équations. Cette approche repose sur l'algorithme des huit points proposé par Longuet-Higgins.

Plusieurs méthodes disponibles dans la littérature ont été testées afin de détecter les faux appariements : les "M-Estimeurs"[Beaton 1974][Huber 1981], "RANSAC"[Fischler 1981], "Least Median Square"[Rousseeuw 1987] et "Least Trimmed Square"[Rousseeuw 1987].

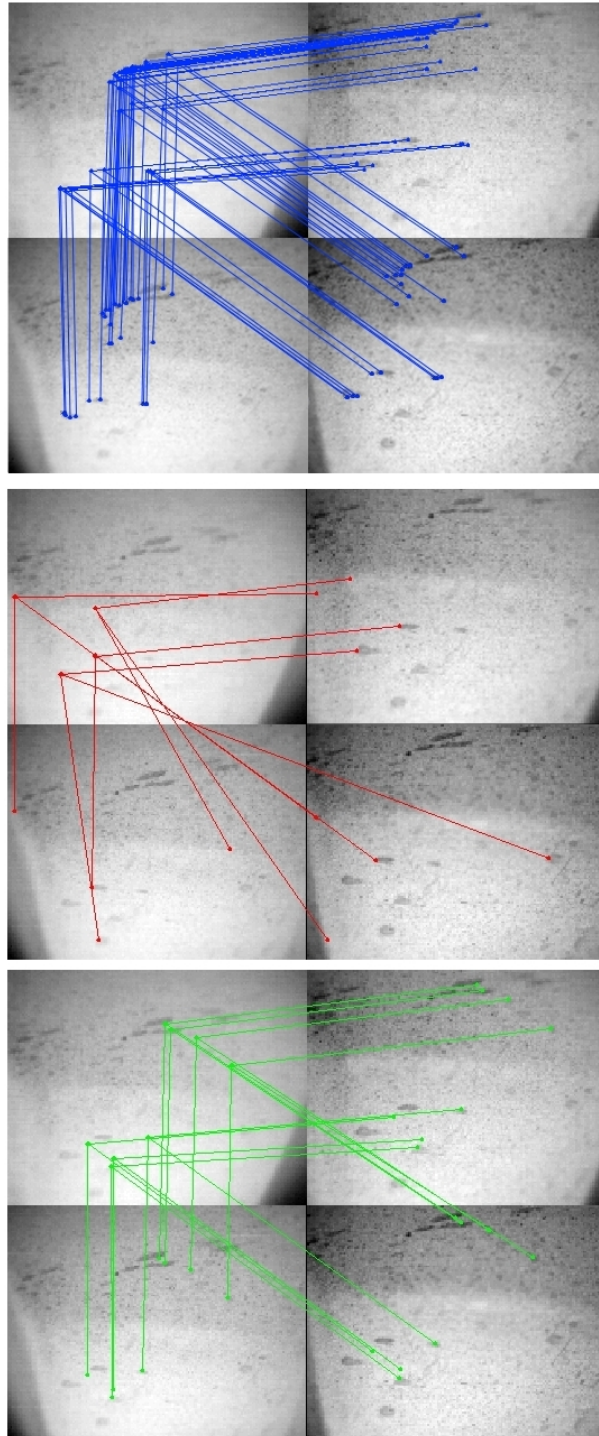


FIG. 4.8 – Corrélation entre quatre images : Phase congruency+Log-Gabor (en bleu), Harris+ZNCC (en rouge), SIFT (en vert)

4.2.1 La méthode des "M-Estimeurs"

Les algorithmes de type "M-Estimeur" réduisent l'effet de fausses mesures lors de la minimisation de l'erreur résiduelle en utilisant une fonction de poids. L'idée est d'estimer le modèle (ici la matrice fondamentale) de façon récursive en attribuant un poids à tous les points lors de chaque itération. Ainsi les "M-Estimeurs" sont basés sur la résolution de la relation suivante :

$$\min_F \sum_{i=1}^n w_i(r_i) \quad \text{où} \quad r_i = m_i'^T F m_i \quad (4.16)$$

avec F la matrice Fondamentale et $w_i()$ la fonction de poids associée au couple de points $(m, m')_i$. Plusieurs fonctions de poids existent dans la littérature parmi lesquelles on peut citer la fonction de poids proposée par Tukey [Beaton 1974] :

$$w_i(r_i) = \begin{cases} \frac{c^2}{6} [1 - (1 - (\frac{r_i}{c})^2)^3] & \text{si } |r_i| \leq c \\ \frac{c^2}{6} & \text{si } |r_i| > c \end{cases} \quad (4.17)$$

avec $c = 4.6851\sigma$, où σ est une estimation robuste de l'écart type :

$$\sigma = 1,4826 [1 + \frac{5}{n-1}] \text{median}_i |r_i| \quad (4.18)$$

où n représente le nombre total de points appariés.

Une autre fonction de poids intéressante a été proposée par Huber [Huber 1981] :

$$w_i(r_i) = \begin{cases} \frac{r_i^2}{2} & \text{si } |r_i| \leq c \\ c(|r_i| - \frac{c}{2}) & \text{si } |r_i| > c \end{cases} \quad (4.19)$$

avec $c = 1.345\sigma$.

La principale différence entre ces deux fonctions est qu'avec celle de Tukey les mauvais appariements sont éliminés lors de la résolution alors qu'avec la fonction de Hubert ils sont seulement pénalisés. La robustesse de cette méthode réside dans l'affectation sélective de poids aux mesures afin de donner moins d'importance aux faux appariement lors de la résolution du système linéaire. Au bout d'un certain nombre d'itérations, le modèle converge vers une solution unique et tous les points ayant un poids trop faible sont considérés comme de faux appariements.

4.2.2 La méthode de "RANSAC"

L'algorithme "RANSAC" a été proposé par Fischler et Bolles [Fischler 1981] et contrairement aux techniques de régression classiques, cette méthode repose sur un vote probabiliste. L'algorithme "RANSAC" repose sur l'estimation d'un modèle à partir d'un nombre minimal de points tirés aléatoirement. Cette estimation est répétée plusieurs fois. Le modèle retenu est celui dont le support est le plus grand. Le support est défini comme étant le nombre de points dont la distance au modèle est inférieure à un certain seuil. Il s'agit donc d'une méthode de vote ; chaque point

vote pour une solution obtenue après un tirage aléatoire ; la solution qui obtient le plus de vote l'emporte, si ce nombre de votes est assez important.

De plus, il est possible de connaître le nombre d'itération minimum itt_{min} nécessaire afin d'avoir la probabilité p de converger vers la bonne solution :

$$itt_{min} = \frac{\log(1 - p)}{\log(1 - (1 - nb_{in})^{n_{min}})} \quad (4.20)$$

avec nb_{in} le pourcentage de points correctement appariés et n_{min} le nombre minimal de points nécessaires pour estimer le modèle.

4.2.3 La méthode "LTS" ou "LMS"

La méthode dite "Least Trimmed Square" a été proposée par Rousseeuw [Rousseeuw 1987]. Elle consiste à minimiser uniquement la somme des carrés des q premiers résidus. q est généralement choisi comme étant la moitié du nombre total de points, ce nombre est optimisé à partir de la connaissance du pourcentage de faux appariements présents.

L'algorithme "Least Median Square" a été proposé aussi par Rousseeuw [Rousseeuw 1987]. Son principe consiste à résoudre le problème de minimisation non linéaire suivant :

$$C(x) = \text{median}(r_1^2(x), r_2^2(x), \dots, r_n^2(x)) \quad (4.21)$$

où le terme r_i^2 est l'erreur quadratique résiduelle associée à la correspondance entre les points m_{1i} et m_{2i} .

L'estimation du modèle se fait à partir d'un nombre minimal de points (ex : $s = 2$ points pour une ligne, $s = 4$ points pour une homographie, $s = 7$ points une matrice fondamentale, etc.) tirés aléatoirement. Cette estimation est répétée plusieurs fois, le modèle retenu est celui dont la valeur médiane des erreurs quadratiques résiduelles est la plus petite. En effet, l'idée intuitive de base consiste à dire que si un faux point a été utilisé pour l'estimation du modèle alors ce dernier aura une valeur médiane des erreurs quadratiques résiduelles bien supérieure à un modèle ne comportant aucun faux point. Une fois le modèle obtenu, les faux points sont éliminés en fixant un seuil sur les erreurs quadratiques de chacun des points. Le modèle final est alors calculé en ne prenant en compte que les bons appariements.

4.2.4 Comparaison

Le tableau 4.1 résume les résultats obtenus lors de la comparaison des différentes méthodes. Les différents algorithmes ont été évalués sur trois critères lors d'un calcul de matrice fondamentale :

- Max : erreur quadratique résiduelle la plus grande.
- M_{err} : moyenne des erreurs quadratiques résiduelles sur tous les points appariés.
- M_{ec} : écart type des erreurs à la moyenne.

Erreur calculée	"LTS"	"RANSAC"	"LMS"	"M-Estimateur"
Max	1.44	1.55	1.40	1.75
M_{err}	0.28	0.29	0.28	0.34
M_{ec}	0.07	0.15	0.05	0.11

TAB. 4.1 – Comparaison des méthodes de détection et d'élimination des faux appariements.

Les méthodes de vote ("RANSAC") ou de pondération itérative ("M-Estimateur") présentent des valeurs d'erreur max et d'écart type supérieures aux autres algorithmes. Elles ne convergent donc pas aussi bien que les méthodes de tris ("LMS" et "LTS"). Nous avons donc fait notre choix sur la petite différence d'écart type entre ces deux méthodes. Nous avons choisi l'algorithme "Least Median Square" pour éliminer les faux appariements.

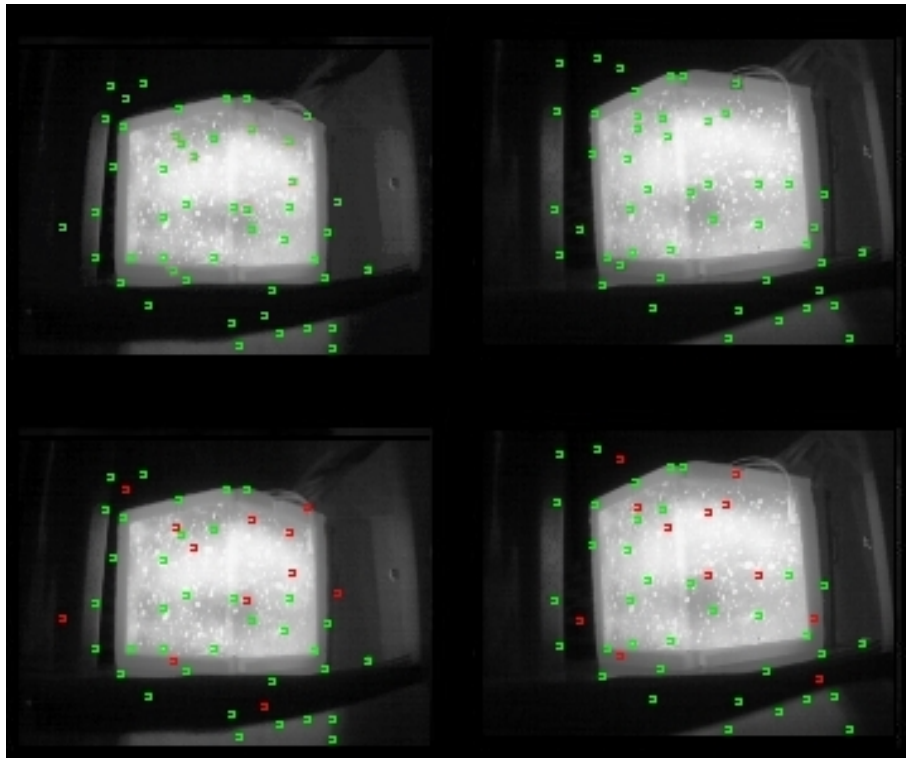


FIG. 4.9 – Résultats du détecteur de Harris sur des images infrarouges, puis résultats du filtrage par "LMS" (les points rouges représentent les faux appariements).

Comme le montre la figure 4.9, une fois les étapes d'acquisition, d'appariement et de filtrage faites, nous obtenons un faible nombre de points appariés.

4.3 Corrélation

Une fois obtenue la paire d'images rectifiées, l'étape suivante consiste à reconstruire l'objet observé de façon dense afin d'obtenir le maximum d'informations sur sa géométrie. La reconstruction dense par stéréovision, se fait en deux étapes : d'abord construire une image de disparité en appariant chacun des pixels de l'image de gauche à un pixel de l'image de droite, puis reconstruire par triangulation un point $3D$ à partir des pixels appariés. Comme les images infrarouges utilisées ne possèdent presque pas de texture, l'algorithme d'appariement devra être suffisamment robuste. Dans cette section, nous décrivons comment nous avons choisi la méthode la plus adaptée pour ces images.

4.3.1 Principe de la corrélation

Afin de pouvoir tester différentes méthodes et différents types de corrélation, nous avons exploité les algorithmes disponibles sur le site de l'Université de Middlebury [Scharstein 2002].

Il existe deux types de méthodes pour construire l'image de disparité : les méthodes locales et globales. Toutes exploitent une mesure de ressemblance entre pixels des deux images, mesure obtenue par corrélation entre voisinages de ces pixels. La première étape est de choisir la fonction de corrélation qui permet de comparer les pixels entre les deux images, et la taille du voisinage sur lequel elle est appliquée. De nombreux scores ont été proposés : "le carré de la différence d'intensité"(SSD)[Hannah 1974], "la valeur absolue de la différence d'intensité"(SAD)[Kanade 1995], le score de corrélation normalisé (ZNCC)...

Pour notre évaluation nous ne disposons pas d'images de disparité de référence. Nous avons donc pris comme critère d'évaluation des méthodes d'appariement, la densité de l'image de disparité puis la précision de la reconstruction estimée à l'aide du logiciel Geomagic. Cette évaluation est détaillée dans le paragraphe suivant(4.4).

4.3.2 Méthodes locales

Les méthodes locales recherchent des appariements indépendamment pour chaque pixel de l'image gauche, sans contrainte de cohérence avec les appariements trouvés pour les voisins. Pour chaque pixel à gauche, on calcule d'abord la fonction de corrélation entre ce pixel à gauche, et un ensemble de pixels de l'image droite, typiquement ceux pris sur la même ligne, compris dans un écart admissible de disparité. Ensuite, on recherche le score optimal parmi tous ces scores de corrélation, ce qui permet d'obtenir la disparité pour le pixel étudié. Une interpolation entre scores proches de l'optimum permet de calculer une disparité subpixelique ; une vérification en partant de l'image droite permet de filtrer quelques erreurs.

Les méthodes locales sont des méthodes souvent rapides mais qui ne tiennent pas compte de l'ensemble de l'image. Les erreurs générées sont généralement dues au manque de texture ou à des textures récurrentes. Ces erreurs se traduisent par des pixels non corrélés ou des disparités incohérentes.

Nous avons évalué les deux méthodes locales [Kanade 1995] [Hannah 1974] sur des images infrarouges. Les résultats présentés sur la figure 4.10, nous montrent que les zones les moins texturées comportent un grand nombre de faux appariements quelques soit la méthode de corrélation locale utilisée. Nous pouvons donc conclure que ces méthodes ne sont pas adaptées à des scènes très peu texturées et donc ne conviennent pas pour des images infrarouges.

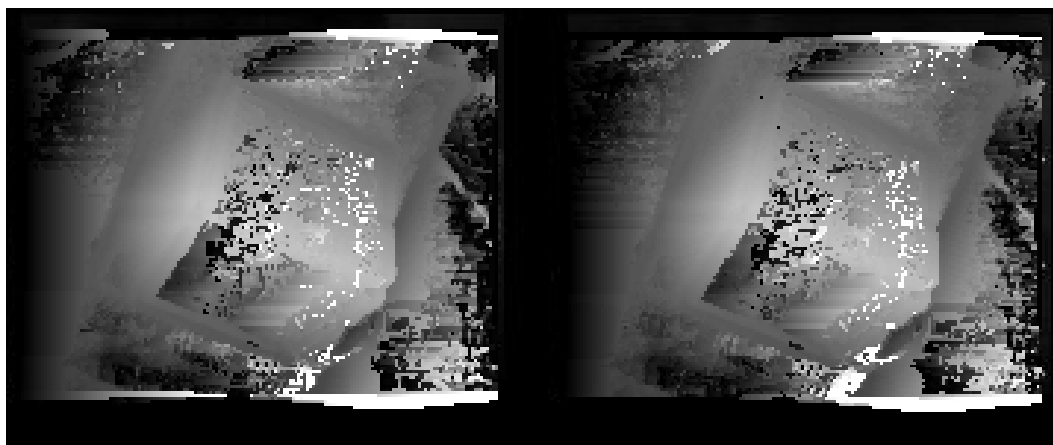


FIG. 4.10 – Application de méthode de corrélation locale, "SD" à gauche et "AD" à droite.

4.3.3 Méthodes globales

Les méthodes globales recherchent un ensemble d'appariements qui satisfont d'une part des critères de ressemblance entre pixels appariés, mais aussi des contraintes de cohérence, liées à la continuité de la surface observés. Le problème est traduit sous une forme adaptée à un traitement global : un graphe reliant les possibles pixels appariés dans lequel il faut chercher un chemin ou une coupure optimal, ou une fonction d'énergie à minimiser. Une fois construit ce graphe ou cette fonction d'énergie, il existe de multiples algorithmes pour trouver la solution optimale : programmation dynamique, coupure de graphe, relaxation, propagation de contraintes...

Les méthodes globales nécessitent plus de temps de calcul, car elles optimisent la disparité en la recherchant simultanément pour tout un ensemble de pixels. Ces méthodes sont paramétrables : on peut définir le coefficient de lissage de l'algorithme ainsi que pénaliser plus ou moins les discontinuités en fonction de la scène observée.

Nous avons évalué deux méthodes globales de types différents, le *Graph Cut* ("GC") [Boykov 2001] qui recherche les appariements pour tous les pixels de l'image, et la *Scanline Optimisation* ("SO") qui traite successivement chaque ligne, donc sans appliquer de contrainte inter-lignes [Scharstein 2002]. "GC" est très lourd, mais vu la faible taille des images infrarouges, ce défaut est ici limité ; il pourrait s'appliquer

sur un graphe réduit, comme proposé en [Zureiki 2007]; comme tous les appariements sont recherchés ensemble, cette méthode peut se satisfaire d'une rectification approchée, et applique des contraintes inter-lignes.

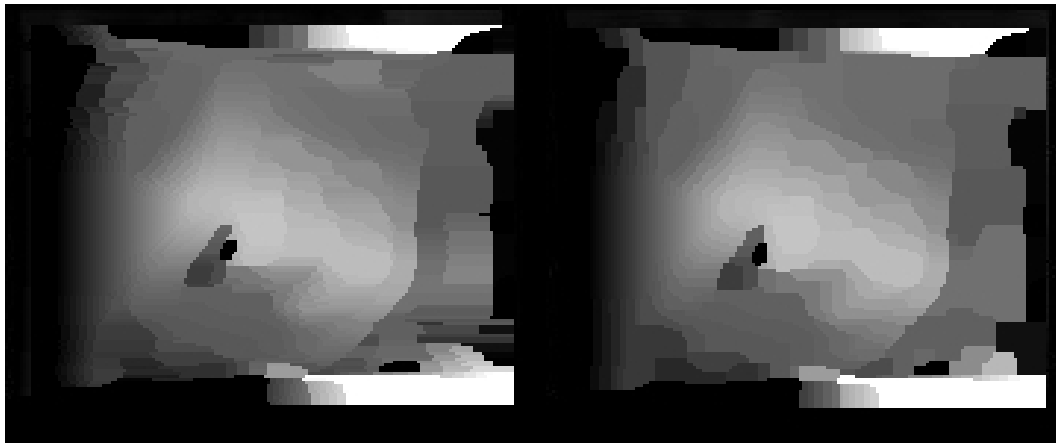


FIG. 4.11 – Application des méthodes de corrélation globales, "SO" à gauche et "GC" à droite.

Les résultats de la figure 4.11 nous montrent des images de disparités similaires pour des paramètres d'initialisation similaires. Cependant l'algorithme "SO" s'exécute en 2, 23s contre 105, 48s pour la méthode "GC". Comme attendu, ces approches globales permettent de supprimer les incohérences constatées lors de l'application des méthodes locales.

4.3.4 Conclusion

Pour conclure, nous avons testé quatre méthodes de corrélation différentes et nous avons observé une grande différence de résultats entre les algorithmes locaux et globaux. En effet, après filtrage des erreurs les plus grossières, les méthodes locales donnent une image de disparité beaucoup moins dense que les méthodes globales.

4.4 Reconstruction 3D avec des images infrarouges

La reconstruction 3D s'opère en plusieurs étapes. Tout d'abord, à partir de l'image de disparité et des paramètres de calibrage des caméras qui ont été obtenus dans le chapitre 3, la reconstruction est effectuée dans le repère de la caméra gauche. Ensuite, le maillage et la projection de texture ont été réalisés grâce au logiciel Geomagic. Le résultat de la reconstruction ainsi réalisé est illustré sur la figure 4.12.

Cette figure valide notre approche de reconstruction 3D. En effet, l'erreur maximale mesurée est de 8mm sur le modèle reconstruit de notre objet. A titre de comparaison, nous procédons à un calcul approché de la résolution spatiale, *res*, de

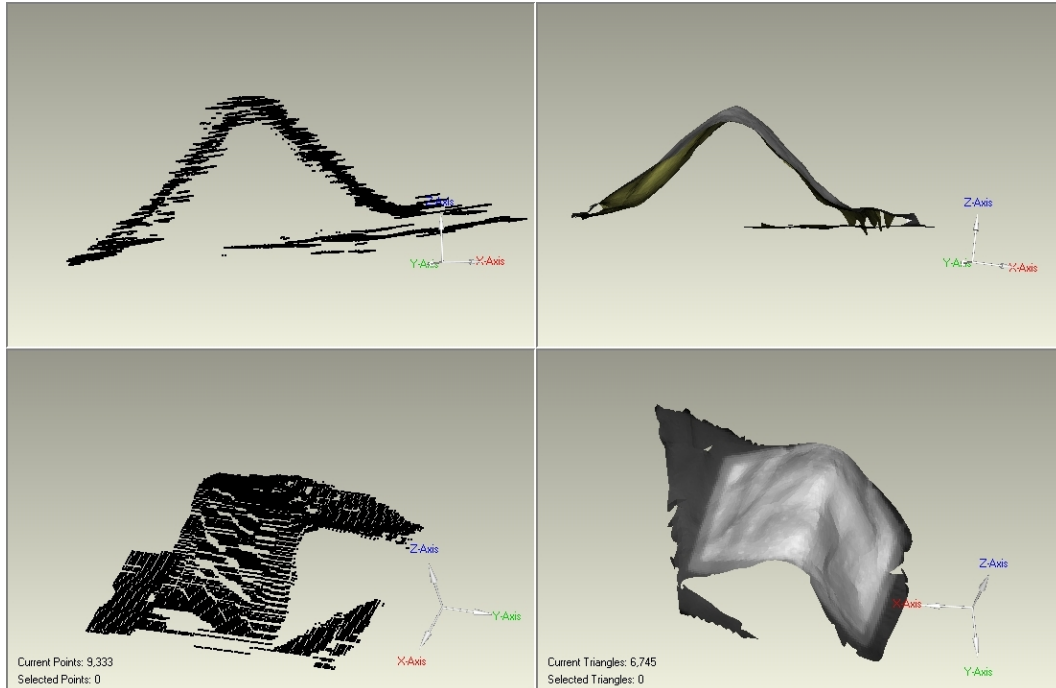


FIG. 4.12 – Nuage de points filtrés après reconstruction puis maillage et projection de texture.

notre système selon l'équation 4.22.

$$res = \frac{d \times t_{corr} \times t_{pix}}{f} \quad (4.22)$$

avec d la distance caméra-objet qui est de $30cm$, t_{corr} la taille de la demie fenêtre de corrélation qui est de 4.5 pixels, f la focale qui est de $11mm$ et t_{pix} la taille de pixel qui est de $51\mu m$. La résolution spatiale, res , est égale à $6,25mm$.

En comparant l'erreur maximale et la résolution, nous pouvons conclure que la reconstruction $3D$ est acceptable. Même dans un cas très défavorable avec des caméras de très faible résolution, une reconstruction $3D$ dense est donc possible avec des caméras infrarouges. Les coûts diminuant régulièrement, des caméras avec des tailles de matrice de 320×240 pixels et des tailles de pixel de $25 \mu m$ seront bientôt disponibles à des coûts compatibles avec des applications robotiques. De plus, des tailles de matrice de 1024×960 pixels sont annoncées pour l'année prochaine.

4.5 Conclusion

Dans ce chapitre nous présentons les différentes étapes de la reconstruction $3D$ à partir d'un banc de stéréovision composé de caméras infrarouges non calibrées.

Nous nous sommes attachés à monter les spécificités de cette reconstruction qui est basée sur des images peu texturées. Dans ces conditions particulières, nous

avons montré la nécessité d'adapter chacune des étapes car la plupart des algorithmes fonctionnant en vision classique entraîneraient de larges erreurs. En effet, comme le montre le calcul d'erreur de la reconstruction, le manque de résolution spatiale associé à de trop grandes erreurs de calcul conduiraient à une erreur qui serait supérieure au centimètre. Nous avons évalué différentes méthodes d'extraction, de filtrage et de corrélation, afin de déterminer la plus adaptée. Nous avons ainsi montré que les méthodes de corrélation locale n'étaient pas assez robustes pour une reconstruction dense. Finalement nous avons validé notre approche à l'aide du modèle *3D* de l'objet reconstruit grâce à une étape d'autocalibrage réalisée sur des images rectifiées.

Le prochain chapitre traite d'une méthode de reconstruction *3D* entièrement non calibrée en incluant de la stéréovision multivues possibles à partir d'un robot cartésien six axes présenté sur la figure 4.1.

Application

Sommaire

5.1	Système de perception actif robotisé	98
5.1.1	Robot cartésien six axes	99
5.1.2	Caméras	99
5.1.3	Objets chauffants	101
5.2	Vérification du modèle géométrique des caméras infrarouges	102
5.2.1	Problème d'extraction des points de la mire	102
5.2.2	Validation du modèle géométrique de la camera infrarouge . .	103
5.3	Reconstruction 3D infrarouge active	105
5.3.1	Étape de rectification	106
5.3.2	Étape d'autocalibrage	107
5.3.3	Corrélation, Triangulation et Recalage	110
5.4	Comparaison avec le modèle de référence	112
5.5	Conclusion	115

L'objectif de ce chapitre est d'intégrer les résultats des chapitres précédents (rectification infrarouge non calibrée, autocalibrage et stéréovision infrarouge) sur un système de perception actif robotisé composé d'un robot six axes et d'un banc de stéréovision pour modéliser un objet en $3D$ de manière automatique dans le domaine spectral infrarouge. Ce chapitre évalue de manière expérimentale les performances de la reconstruction $3D$ infrarouge active.

Nous avons montré que l'approche idéale pour obtenir ce modèle $3D$ est de procéder à une rectification non calibrée préalable du banc de stéréovision pour minimiser le nombre de paramètres à estimer lors de l'autocalibrage global du banc. Les différentes étapes du processus de reconstruction $3D$ infrarouge actif automatisé, sont alors les suivantes : dans un premier temps, la méthode de rectification non calibrée simplifie le modèle du banc de stéréovision et les différentes étapes de la reconstruction. Dans un deuxième temps, la méthode d'autocalibrage du système complet fournit les paramètres intrinsèques et extrinsèques du banc de stéréovision et ceux de la transformation pince-caméra. Ces paramètres sont indispensables pour obtenir une reconstruction $3D$ euclidienne. Dans un troisième temps, les étapes de corrélation et de triangulation permettent d'obtenir le modèle $3D$ de l'objet observé à partir de différentes vues acquises avec le robot. Ces étapes ont été réalisées sur deux objets de tailles différentes avec des caméras CCD visibles classiques pour valider l'approche dans une configuration optimale de référence (haute résolution et images texturées) et avec des caméras infrarouges pour évaluer les performances de la reconstruction $3D$ infrarouge et active. Enfin, les modèles obtenus des deux objets, avec les deux configurations de capteur, sont comparés aux modèles de référence déterminés par une méthode de numérisation basée sur un scanner laser $3D$ disponible sur étagères, dont la précision est de l'ordre de $50\mu m$.

Tout d'abord, ce dernier chapitre présente le système de perception actif robotisé composé d'un robot cartésien six axes et des caméras visibles et infrarouges et il décrit les objets qui seront ensuite numérisés. Ensuite, notre modèle géométrique étant basé sur le modèle projectif de caméra pin-hole, nous vérifions que les caméras de technologie microbolométrique respectent cette hypothèse. Enfin, les résultats de l'ensemble du processus de reconstruction multivues sont présentés avec des caméras CCD visibles et infrarouges. Les modèles $3D$ obtenus dans les deux configurations sont comparés au modèle de référence obtenu par un système de numérisation précis.

5.1 Système de perception actif robotisé

Le système de perception actif robotisé assure deux fonctions : le positionnement et l'orientation du système de vision en un ensemble de configurations pré-sélectionnées, sont assurés par un robot cartésien six axes ; la mesure de la forme $3D$ de l'objet, est fournie par deux caméras montées en banc de stéréovision.

5.1.1 Robot cartésien six axes

Le robot cartésien de la figure 5.1 a pour fonction de positionner le système de mesure dans un volume de travail de $1000 \times 1000 \times 800\text{mm}$ avec la possibilité de l'orienter dans toutes les directions avec trois rotations autour des axes X, Y et Z (cf. figure 5.2).



FIG. 5.1 – Robot cartésien festo.

La position cartésienne de l'organe terminal est donnée par une transformation (translation et rotation) entre le repère monde et le repère de l'organe terminal (repère "hand" en figure 5.2). Cette transformation est obtenue à partir des positions articulaires et des dimensions précises du robot, supposées connues.

Notons que les configurations de l'organe terminal pour modéliser nos objets ont été calculées hors ligne ; elle se situe sur une demi-sphère englobant l'objet à modéliser, de sorte que les axes optiques des caméras passent au plus près du centre de cette demi-sphère.

5.1.2 Caméras

A des fins de comparaison, deux types de caméras ont été utilisées : des caméras infrarouges nécessaires au projet et des caméras CCD visibles classiques en référence.

5.1.2.1 Caméra infrarouge A10

Les caméras thermiques A10 sont basées sur des matrices de microbolomètres (détecteurs thermiques à résistance variable). Elles fournissent donc l'image thermique de l'environnement comme le montre la figure 5.3.

La caméra thermique A10 est non refroidie et elle a été choisie en raison de son faible encombrement et de son faible poids qui sont des qualités requises pour

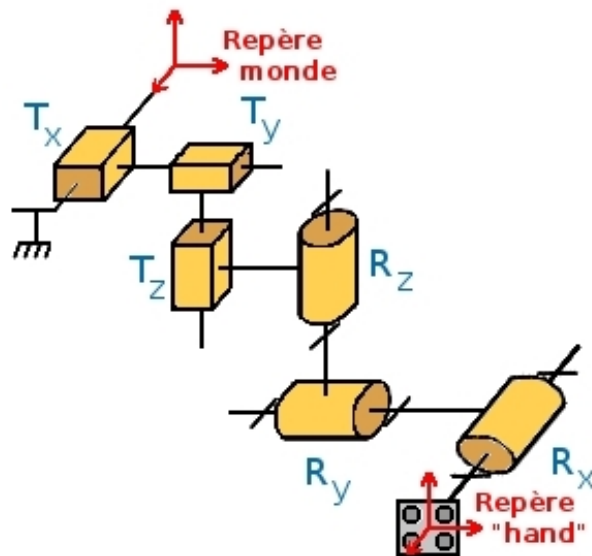


FIG. 5.2 – Chaîne cinématique du robot cartésien.

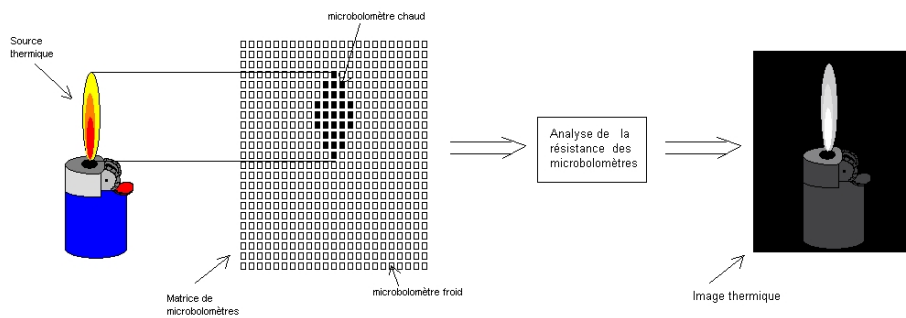


FIG. 5.3 – Schéma explicatif des caméras microbolométriques

un système embarqué sur robot (cf. figure 5.4). Mais, elle possède une résolution spatiale faible de $(164 \times 128 \text{ pixels})$ et une résolution thermique de 0.08°C .

Cette caméra est donc idéale pour les applications embarquées d'imagerie thermique avec des contraintes d'espace, de poids et/ou de puissance. Sa résolution est suffisante pour la surveillance thermique des procédés industriels, la détection et la prévention des incendies et les applications de sécurité et de surveillance.



FIG. 5.4 – Caméra A10

5.1.2.2 Caméra CCD

Les caméras visibles utilisées ici sont des caméras fonctionnant avec des matrices CCD classiques (cf. figure 5.5). Ces caméras, des Sony XCD-X700, sont équipées de capteurs 1/2", sont monochromatiques et délivrent une image haute résolution (1024×768) avec une résolution de 8 bits.



FIG. 5.5 – Caméra Sony XCD-X700

5.1.3 Objets chauffants

Nous avons choisi d'évaluer notre méthode sur deux types d'objets comme le montre la figure 5.6. Le premier objet est composé de deux plaques d'une dizaine de centimètres formant un angle, il est noté "**Angle droit**" dans le tableau des résultats. Le second objet, plus volumineux, représente une pyramide de hauteur une vingtaine de centimètres dont la base mesure une trentaine de centimètres avec des cotés d'une dizaine de centimètre. Cet objet est noté "**Pyramide**" dans le paragraphe résultat. Ces objets sont équipés de résistances chauffantes pour élever les surfaces à des températures jusqu'à $40^{\circ}C$, afin de les discriminer de l'environnement dans les images infrarouges.



FIG. 5.6 – Objets utilisés pour la reconstruction 3D : "Angle droit" à gauche et "Pyramide" à droite.

5.2 Vérification du modèle géométrique des caméras infrarouges

Cette section a pour objectif de vérifier que le modèle de caméra pin-hole est applicable aux caméras infrarouges. Cette validation est réalisée par analyse de la cohérence des valeurs des paramètres intrinsèques du modèle pinhole par un calibrage fort réalisé à partir d'une mire de calibrage et de la "Camera Calibration Toolbox" de Matlab.

La mire de calibrage est composée d'un motif connu. Dans notre cas, ce motif est un damier composé de carrés d'un centimètre de côté. Un repère est choisi dans un coin de la mire puis les intersections des carrés du damier sont extraits dans de multiples positions de la mire. Les valeurs des paramètres intrinsèques sont identifiées à partir des positions des points d'intérêt extraits de la mire et en prenant comme initialisation les données constructeur de la caméra. Enfin, une fois les valeurs des paramètres intrinsèques des deux caméras obtenues, les valeurs des paramètres extrinsèques du banc stéréo sont identifiées par corrélation des points des images de mires communs aux deux caméras.

5.2.1 Problème d'extraction des points de la mire

Le premier problème rencontré est l'extraction des points de la mire. En effet, les motifs des mires utilisées dans le visible, à température ambiante, ne sont pas détectés par la caméra infrarouge. Une mire chauffante reproduisant les motifs a donc été conçue. La figure 5.7 donne son schéma de principe qui repose sur l'exploitation de la différence d'émissivité entre deux matériaux.

La mire est constituée d'une plaque représentant les motifs dessinés avec des matériaux aux propriétés émissives différentes. Cette plaque est un circuit imprimé dont la différence d'émissivité est créée entre de l'époxy ($\epsilon_{poxy} \approx 0.8$) et de l'or poli ($\epsilon_{or} \approx 0.1$). Comme le montre la figure 5.8, cette différence est assez grande pour qu'une fois chauffé le motif apparaisse très contrasté.

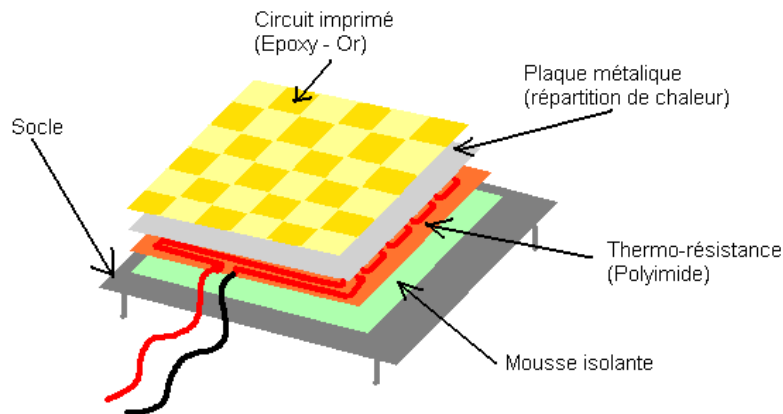


FIG. 5.7 – Schéma de la mire chauffante

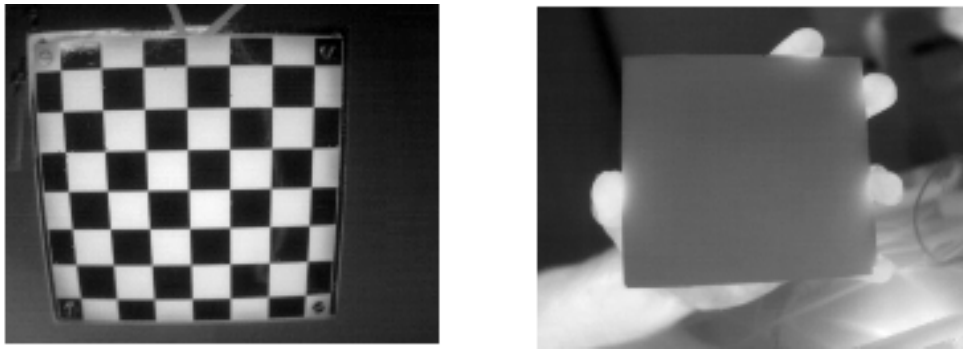


FIG. 5.8 – Image de mires avec une caméra thermique. À gauche, image d'une mire chauffée et à droite image d'une mire "classique" utilisée dans le visible.

5.2.2 Validation du modèle géométrique de la camera infrarouge

La figure 5.9 illustre l'extraction des points d'intérêt de la mire dans des positions différentes. La multiplicité des prises de vues permet de prendre en compte des données 3D lors du calibrage, même si la mire est plane.

Les résultats obtenus pour les paramètres intrinsèques de l'une des caméras infrarouges sont exprimés dans le tableau 5.1. Les valeurs estimées des paramètres intrinsèques sont très proches des données constructeur. Les déformations introduites sur les images sont donc faibles. De plus, en reprojétant les points extraits sur les images d'origine, l'erreur de reprojection est de 0.10550 pixel sur les lignes (u) et de 0.09841 pixel sur les colonnes (v). L'erreur de reprojection des points de mires extraits sur les images est de l'ordre du dixième de pixel. Nous concluons que le modèle "pin-hole" convient parfaitement aux caméras infrarouges utilisées.

Ensuite, il s'agit de procéder à l'estimation des paramètres extrinsèques du banc

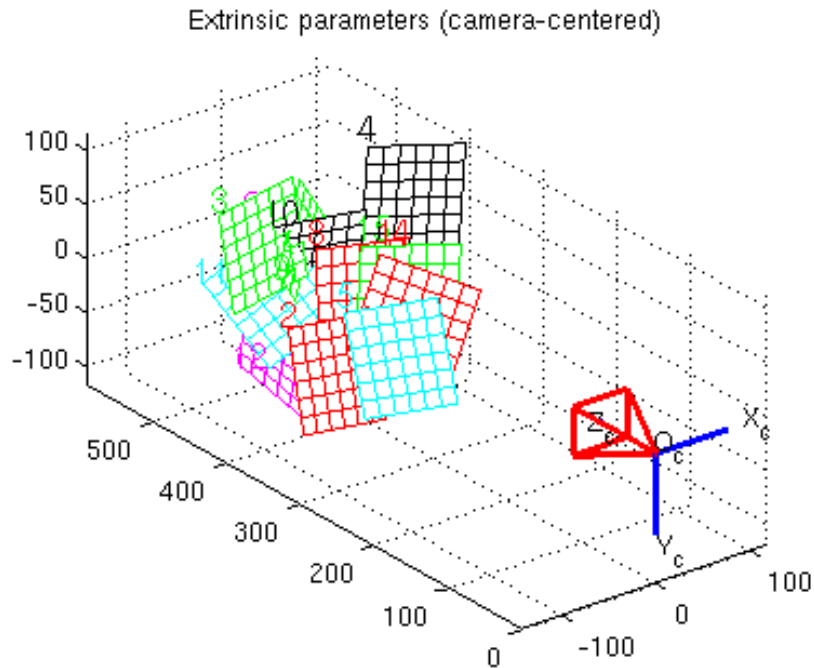


FIG. 5.9 – Extraction des mires.

	α_u	α_v	u_0	v_0
Résultats de calibrage	212.57412	212.47842	84.89107	65.87503
Données constructeur	210	210	82	64

TAB. 5.1 – Comparaison entre les paramètres intrinsèques obtenus par calibrage et les données constructeurs.

stéréo illustré sur la figure 5.10.

Les résultats obtenus montrent la position de la caméra de gauche dans le repère de la caméra de droite :

$$R = \begin{pmatrix} 0.9616 & -0.0167 & 0.2739 \\ 0.0173 & 0.9999 & 0.0001 \\ -0.2738 & 0.0046 & 0.9618 \end{pmatrix} ; \quad T = \begin{pmatrix} -154.23651 \\ -1.10422 \\ 18.71175 \end{pmatrix} \quad (5.1)$$

Ces résultats sont cohérents avec le montage physique observé. En effet, la distance entre les deux caméras est de 15.5cm et leur angle de convergence est de 17° .

En conclusion, le modèle de caméra adopté convient avec le matériel utilisé, car l'ensemble des résultats reste cohérent avec la réalité. De plus l'erreur de reprojection en pixel est très faible.

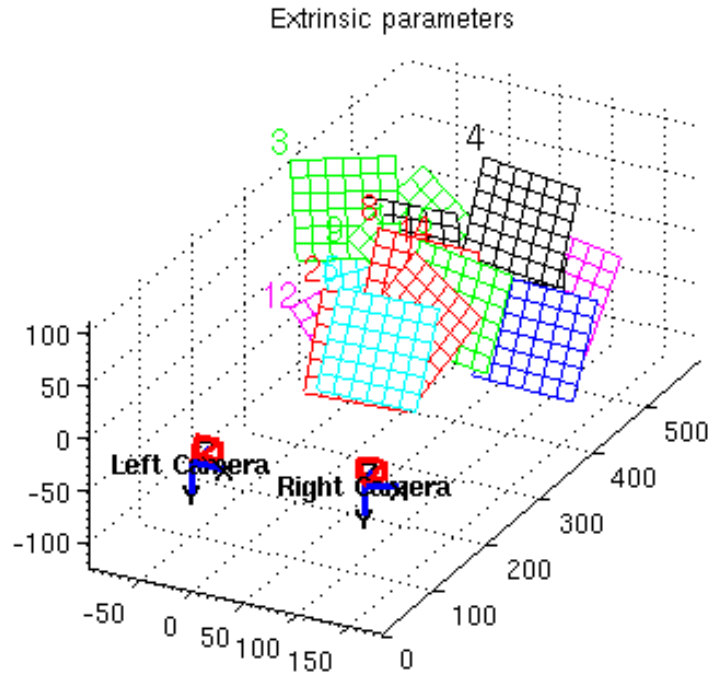


FIG. 5.10 – Mires vu par les deux caméras

5.3 Reconstruction 3D infrarouge active

Afin d'obtenir une reconstruction multivues de l'objet observé, l'ensemble des étapes des chapitres précédents ont été intégrées dans un seul et même processus. La première étape est d'acquérir un certain nombre de vues à partir du déplacement du robot pour extraire des points sur la totalité de l'objet ou la partie que l'on désire reconstruire. Le nombre de vue dépend généralement de la résolution des caméras utilisées, de leur angle d'ouverture, de la distance de prise de vue et de l'angle de convergence entre les deux caméras du banc de stéréovision.

Une fois les acquisitions réalisées, les points d'intérêt sont extraits sur chacune des paires d'images pour calculer la paire d'homographies de rectification du banc de stéréovision tout en minimisant les déformations projectives (cf chap. 2). Après cette étape de rectification, l'ensemble des déplacements et des points extraits sont utilisés dans une étape d'autocalibrage du banc de stéréovision (cf chap. 3). Ensuite, tous les pixels dans chaque paire d'images acquises sont corrélés afin d'obtenir plusieurs vues 3D de l'objet (cf chap. 4). Finalement, l'ensemble des images 3D partielles sont recalées dans un même repère à l'aide des paramètres obtenus lors de l'étape d'autocalibrage et des positions de l'organe terminal du robot données avec précision par son système de contrôle. Ce recalage permet de construire une vue 3D complète de l'objet observé.

Afin de valider l'approche dans un cas de référence, des premiers tests ont été

effectués sur des images visibles issues de caméras CCD classiques. Ensuite, les tests avec des images infrarouges issues de caméras microbolométriques ont été menés. Afin de pouvoir comparer les résultats entre les reconstructions obtenues, les objets utilisés sont les mêmes, à l'exception qu'ils sont chauffés à une température de 30°C pour les images infrarouges (voir paragraphe 5.1.3).

Nous présentons les différentes étapes nécessaires à la reconstruction multivues d'objets. Pour chaque étape, la méthode utilisée est définie, puis les résultats obtenus pour chacun des objets dans le visible et l'infrarouge sont analysés et comparés.

5.3.1 Étape de rectification

La rectification est une étape clé dans notre processus de reconstruction car elle permet de simplifier le modèle du banc stéréoscopique. Une description détaillée est proposée au chapitre 2.

5.3.1.1 Extraction d'un premier set de points

Contrairement à l'étape d'autocalibrage, l'étape de rectification ne nécessite pas forcément de déplacement pour obtenir l'ensemble de points nécessaires à sa réalisation. Cependant, plus le nombre de points extraits est grand plus la rectification est précise. Nous avons donc utilisé un détecteur de Harris car, comme nous l'avons montré au chapitre 4, c'est le plus adapté à ce genre de situation.

5.3.1.2 Méthode

Nous avons montré dans le chapitre 2 que les deux méthodes de rectification proposées avaient des performances similaires. Nous avons choisi d'utiliser celle basée sur des contraintes algébriques car elle permet de fixer directement des valeurs numériques des contraintes qui sont de $\pm 2^{\circ}$ pour l'orthogonalité et de $\pm 5\%$ d'erreur sur la taille et le rapport d'aspect.

5.3.1.3 Résultats

Le tableau 5.2 montre les résultats obtenus sur chaque objet avec des caméras visibles et des caméras infrarouges. Les critères d'évaluation ont été présentés au chapitre 2. La précision sur la matrice fondamentale n'est pas présentée car elle n'est pas calculée avec cette méthode.

Ces résultats montrent que la déformation d'image due à la rectification est très légèrement supérieure sur les images infrarouges. Cependant sa précision est nettement meilleure. Dans l'ensemble, ces résultats sont très satisfaisants car, quelques soit l'objet ou les caméras utilisées, la précision est très inférieure au pixel et les déformations sont minimales. Les figures 5.11 montrent pour chacun des cas, une paire d'images originales et une paire d'images rectifiées.

Échantillon	Caméras	Orthogonalité E_o		Rapport d'aspect E_a		Erreur Rect. E_r	
		H'	H	H'	H	moy	σ
Angle droit	Visibles	90.04	89.89	1.0007	0.9982	0.55	0.30
	Infrarouges	89.66	90.10	0.9954	0.9988	0.10	0.07
Pyramide	Visibles	90.46	91.62	1.0077	1.0275	0.29	0.25
	Infrarouges	89.29	90.07	0.9882	0.9991	0.09	0.06

TAB. 5.2 – Comparaison des résultats de rectification obtenus entre les caméras visibles et infrarouges.

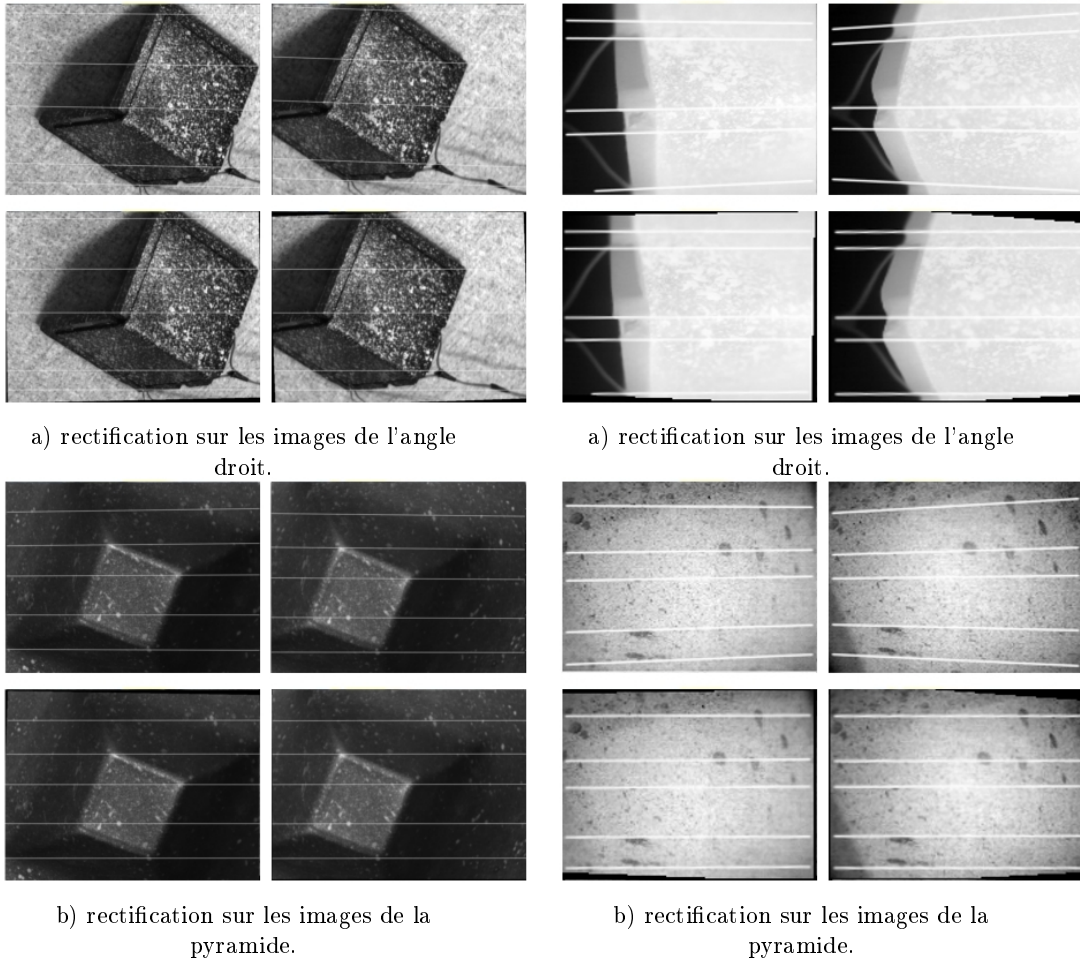


FIG. 5.11 – Rectification utilisant les caméras visibles (à gauche) et infrarouges (à droite).

5.3.2 Étape d'autocalibrage

La reconstruction 3D dans un espace euclidien n'est possible que si les paramètres intrinsèques et extrinsèques des caméras sont connus. L'autocalibrage décrit dans le chapitre 3 permet d'obtenir ces paramètres.

5.3.2.1 Extraction d'un ensemble de points entre quatre images

Nous avons montré au chapitre 3 que l'étape d'autocalibrage nécessite l'extraction d'un ensemble de points communs aux deux images issues d'un banc stéréoscopique et aux deux images issues de ce même banc stéréoscopique après déplacement. Nous avons utilisé un détecteur "Phase Congruency" qui est le plus adapté dans ce cas comme cela a été démontré au chapitre 4.

Le tableau 5.3 présente pour chacun des cas le nombre de déplacements réalisés ainsi que le nombre total de points extraits.

Objet	Caméras	Déplacements	Points extraits
Angle droit	Visibles	8	10616
	Infrarouges	8	106
Pyramide	Visibles	10	3864
	Infrarouges	15	57

TAB. 5.3 – Nombre de points extraits avec des caméras visibles et infrarouges.

Pour un même nombre de vues avec le même objet "Angle droit", le nombre de points extraits en infrarouge est environ 100 fois moins important que le nombre de points extraits avec des caméras visibles. De plus, le nombre de points extraits pour l'objet "Angle droit" est bien supérieur à celui de l'objet "Pyramide" car l'objet "Angle droit" est plus petit que l'objet "Pyramide". Pour chaque déplacement, la zone de recouvrement est donc plus importante. Pour l'objet "Pyramide", avec les caméras infrarouges et une zone de recouvrement faible, le nombre de points extraits est très faible. Nous avons ici seulement une moyenne de 3.8 points par vue.

5.3.2.2 Méthode

Nous avons montré dans le chapitre 3 que la méthode d'autocalibrage la plus adaptée avec un système équipé d'un banc de stéréovision infrarouge est celle basée sur la minimisation des distances entre les points $3D$ reconstruits (cf. section 3.3). Cette méthode a donc été appliquée à tous les cas traités.

5.3.2.3 Résultats

Le tableau 5.4 présente les résultats obtenus lors du calcul d'autocalibrage. Les critères utilisés pour évaluer les performances ont été définis au chapitre 3.

Ces résultats montrent que, comme nous l'avons montré au chapitre 3, malgré un nombre de points faible (57 points pour 15 déplacements dans le cas de l'objet "Pyramide" avec les caméras infrarouges) l'erreur moyenne entre les points $3D$ ou l'erreur sur la distance épipolaire restent faibles (1.32 et 3.25mm). Les autocalibrages réalisés avec les caméras visibles sont globalement plus précis avec une moyenne d'erreur inférieure à 1mm.

La figure 5.12 représente, pour chaque cas, l'ensemble des points extraits qui sont reconstruits à l'aide d'une part des paramètres constructeurs et d'autre part

Échantillon	Caméras	Σ_d (mm)	\bar{d} (mm)	σ_d (mm)	Σ_e (mm)	\bar{e} (mm)	σ_e (mm)
Angle droit	Visibles	4392.06	0.82	0.73	6386.18	0.60	1.45
	Infrarouge	113.82	2.14	5.29	630.55	5.94	12.72
Pyramide	Visibles	1852.11	0.95	0.66	3970.74	1.02	1.84
	Infrarouge	37.74	1.32	0.86	185.74	3.25	7.96

TAB. 5.4 – Comparaison des résultats d'autocalibrage obtenus entre caméras visibles et infrarouges.

des paramètres estimés.

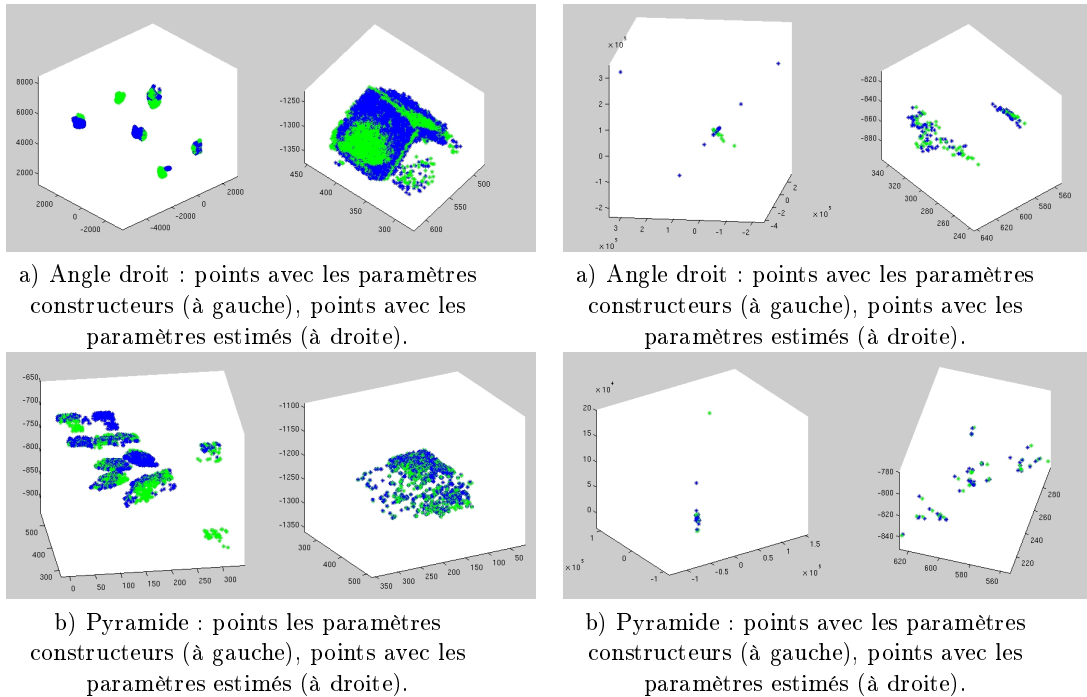


FIG. 5.12 – Autocalibrage avec les caméras visibles (à gauche) et infrarouges (à droite).

Avec les caméras visibles, l'objet est reconstruit dans sa totalité de manière quasi dense sans même procéder à une étape de corrélation. En effet lors de chacun des déplacements du robot l'objet est quasiment visible dans sa totalité, le nombre de points appariés entre les vues est donc très important. Cependant, ce grand nombre de point pénalise le temps de calcul qui est, pour les caméras visibles, de 50 min pour l'objet "Pyramide" et de 4 min pour l'objet "Angle droit" contre quelques secondes pour les caméras infrarouges.

5.3.3 Corrélation, Triangulation et Recalage

Les étapes précédentes ayant fourni l'ensemble des paramètres de calibrage pour le système robotique, nous procédons la reconstruction 3D complète de l'objet observé.

5.3.3.1 Méthode

Pour l'étape de corrélation, et comme nous l'avons montré dans le chapitre 4, les méthodes globales sont plus adaptées aux images infrarouges. Nous appliquons ici la méthode du *Scanline Optimisation* disponible sur le site de l'université de Middlebury [Scharstein 2002].

Étant donné que pour un banc de stéréovision rectifié, les cartes de disparités sont obtenues et que les paramètres de calibrage sont connus, il est donc triviale de trianguler l'ensemble des points dans le repère d'une des caméras rectifiées.

De plus l'étape précédente fournit aussi les paramètres de changement de repère de la caméra rectifiée au robot et par conséquent au monde. Le recalage est donc obtenu simplement en transformant tous les points triangulés dans un repère commun, le repère monde, cela en exploitant le mouvement connu des capteurs et la transformée *Hand-Eye* estimée précédemment dans la procédure d'auto-calibrage.

5.3.3.2 Résultats

Pour la reconstruction complète de l'objet "Angle droit", nous avons utilisé uniquement deux vues avec les caméras visibles et huit vues avec les caméras infrarouges. La figure 5.13 représente, pour le cas visible, les deux vues de manière indépendante puis affichées dans un même repère.

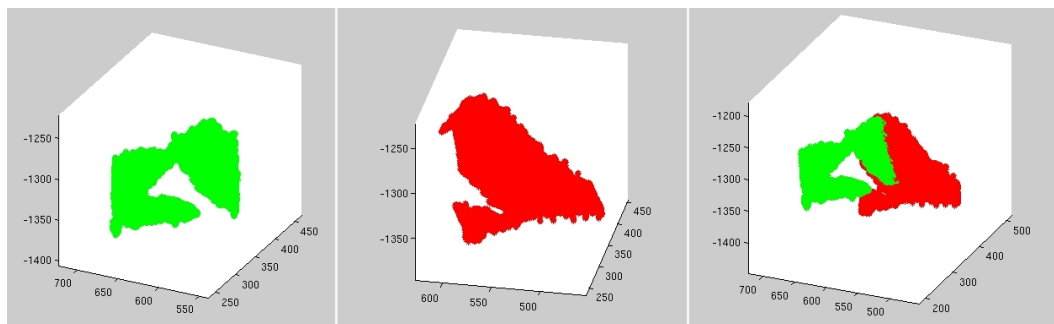


FIG. 5.13 – Exemple de deux images 3D partielles puis d'une vue 3D complète avec les images recalées.

La figure 5.14 montre la reconstruction complète de ce même objet, pour le cas infrarouge, sous différents angles.

Pour la reconstruction complète de l'objet "Pyramide", seules trois vues sont utilisées pour le cas visible. La figure 5.15 représente ces trois vues de manière

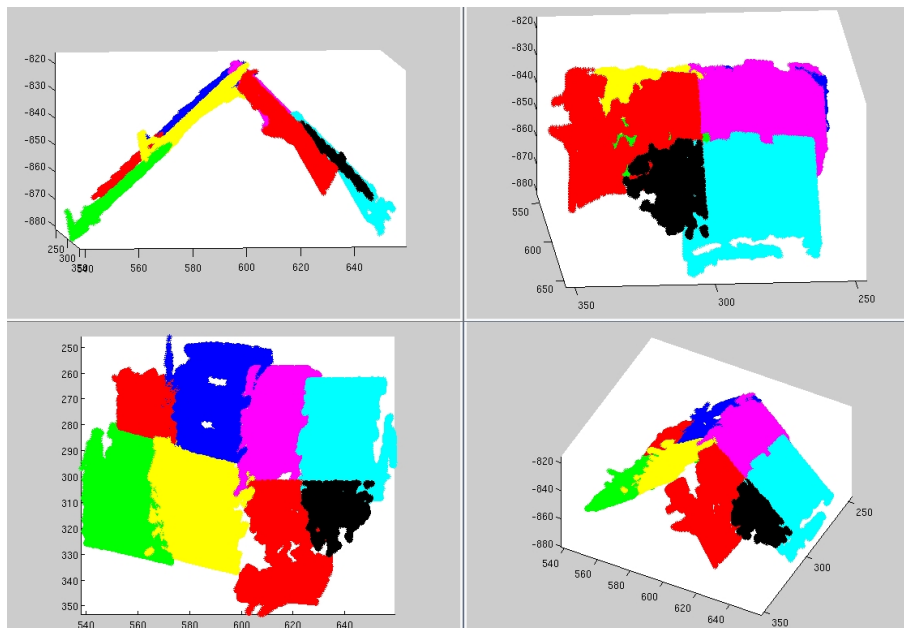


FIG. 5.14 – Exemple d'une vue 3D complète avec 8 images 3D partielles recalées.

indépendante puis affichées dans un même repère.

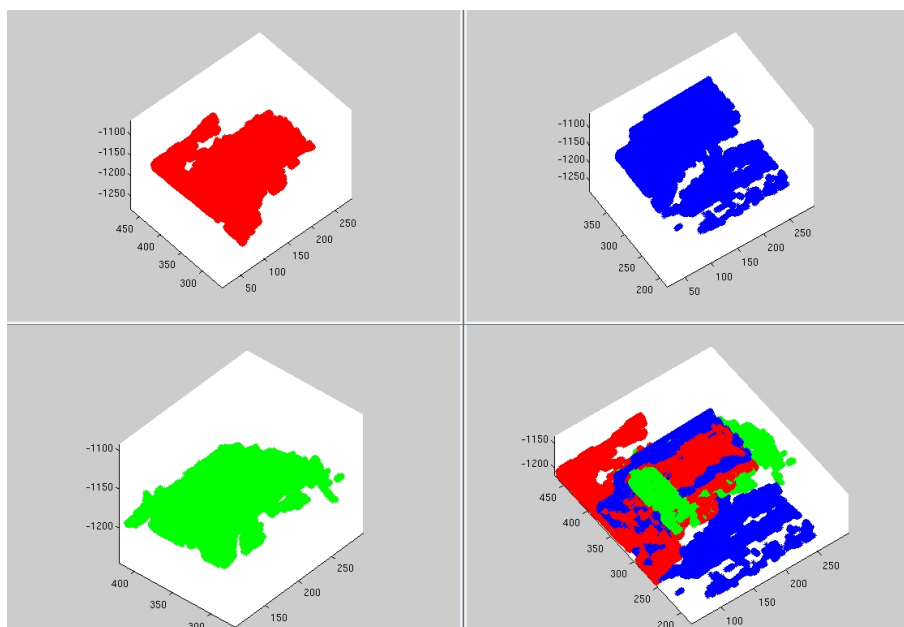


FIG. 5.15 – Exemple de trois images 3D partielles puis d'une vue 3D complète avec les images recalées

Pour le cas infrarouge, le champ de vision des caméras infrarouges est beaucoup plus faible et cet objet est relativement grand. Onze prises de vue différentes n'ont permis de reconstruire que partiellement l'objet. La figure 5.16 montre cette reconstruction partielle de la pyramide sous différents angles.

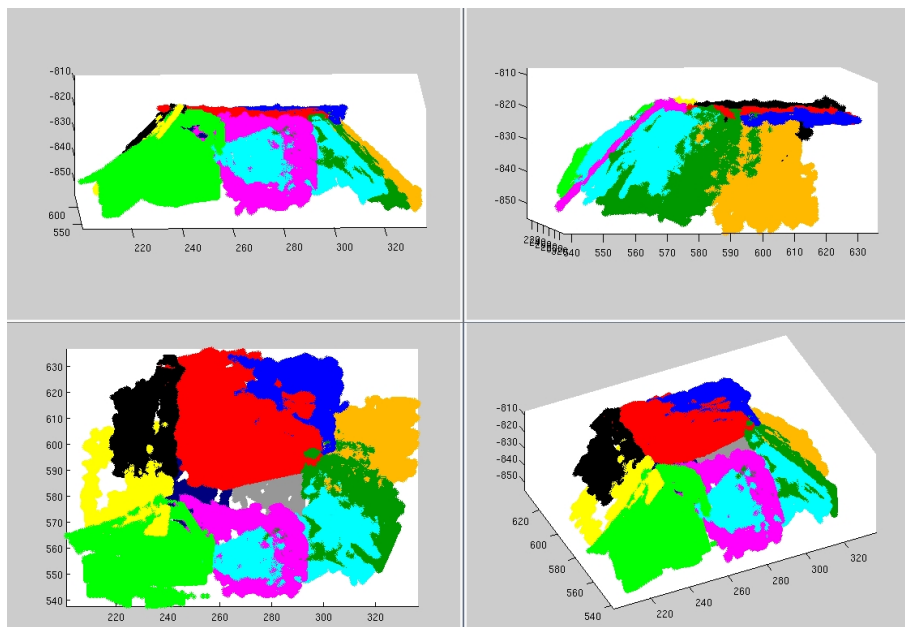


FIG. 5.16 – Exemple d'une vue 3D où 11 images 3D partielles sont recalées

Finalement, les figures 5.17, 5.18, 5.19, 5.20, illustrent pour chaque cas, l'objet reconstruit avec le maillage des points et la projection de texture réalisés avec le logiciel géomagic.

5.4 Comparaison avec le modèle de référence

Afin de connaître la précision exacte de notre processus de reconstruction, nous avons comparé les deux reconstructions (visible et infrarouge) avec le modèle numérique 3D de la pyramide. Ce modèle est issu d'une numérisation par un scanner laser dont la précision est de $50\mu m$. Ce modèle est donc considéré comme le modèle de référence. La comparaison est effectuée avec le logiciel Géomagic Qualify.

La figure 5.21 montre la différence entre la reconstruction 3D réalisée à partir des caméras visibles et la reconstruction de référence. La première figure illustre l'échelle totale de l'erreur de reconstruction qui est comprise entre $4.2mm$ et $-14.3mm$. La valeur maximale de $14mm$ est observée sur les bords de l'objet et sur très peu de points. L'échelle des erreurs est ensuite resserrée et les deux autres figures montrent les zones de l'objet où l'erreur est inférieure à $\pm 2mm$ et à $\pm 1mm$. L'erreur moyenne sur l'objet lui-même est seulement de $0.5mm$ et l'écart type est de $0.6mm$. L'écart

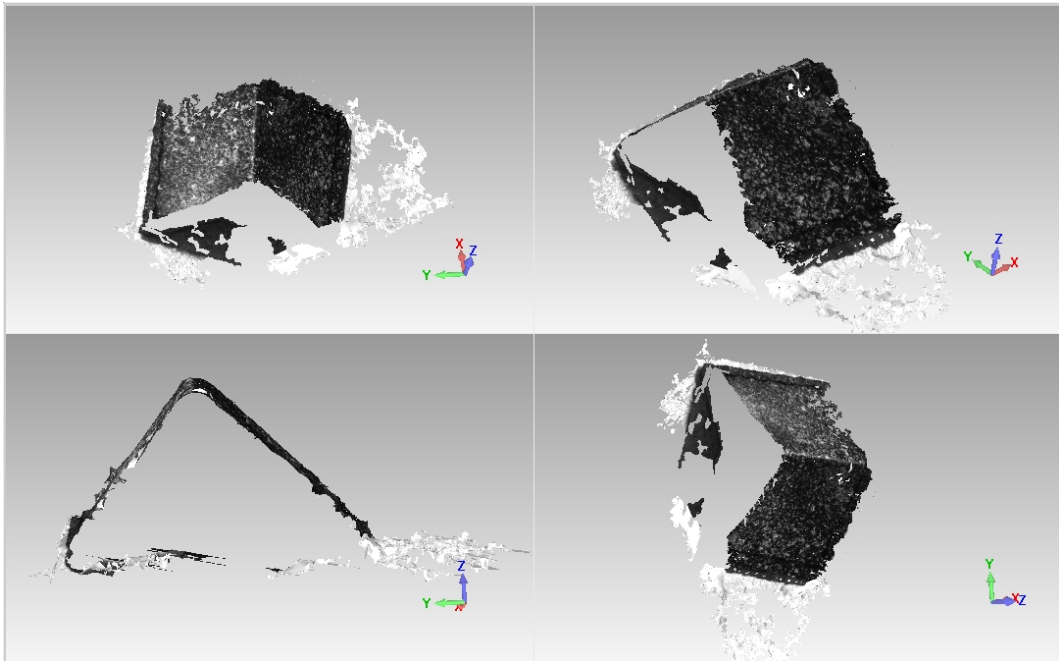


FIG. 5.17 – Angle droit reconstruit avec les caméras visibles.

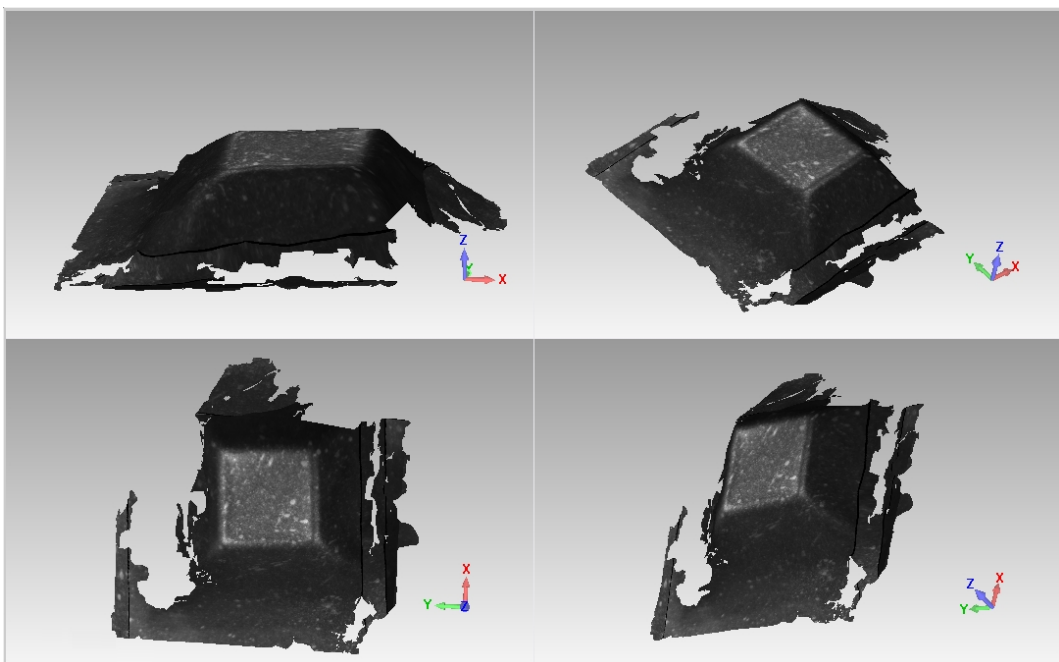


FIG. 5.18 – Pyramide reconstruit avec les caméras visibles.

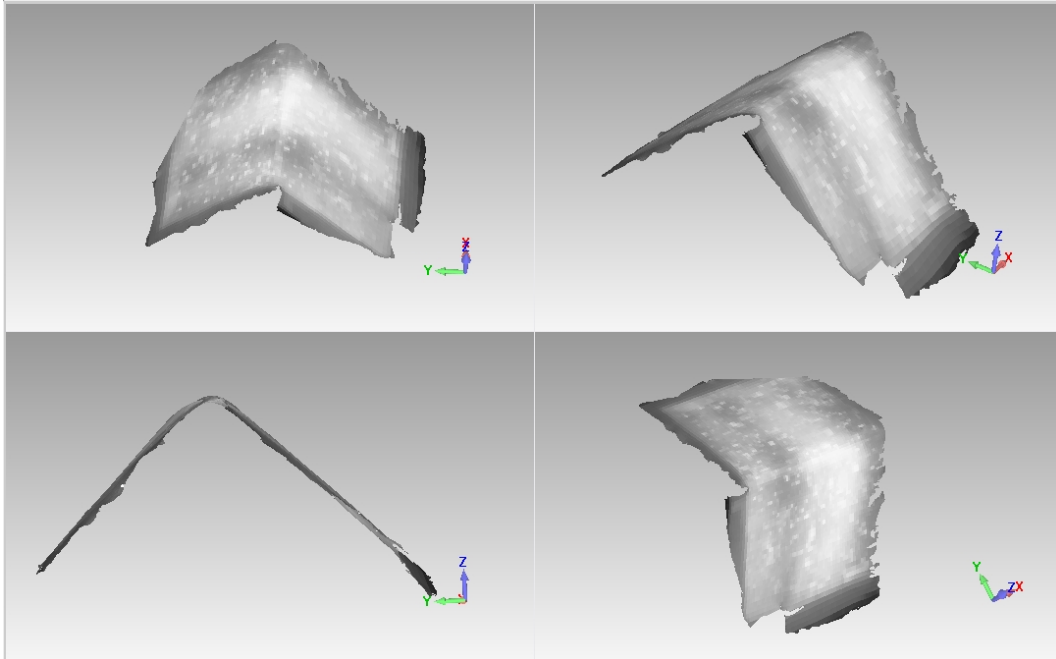


FIG. 5.19 – Angle droit reconstruit avec les caméras infrarouges.

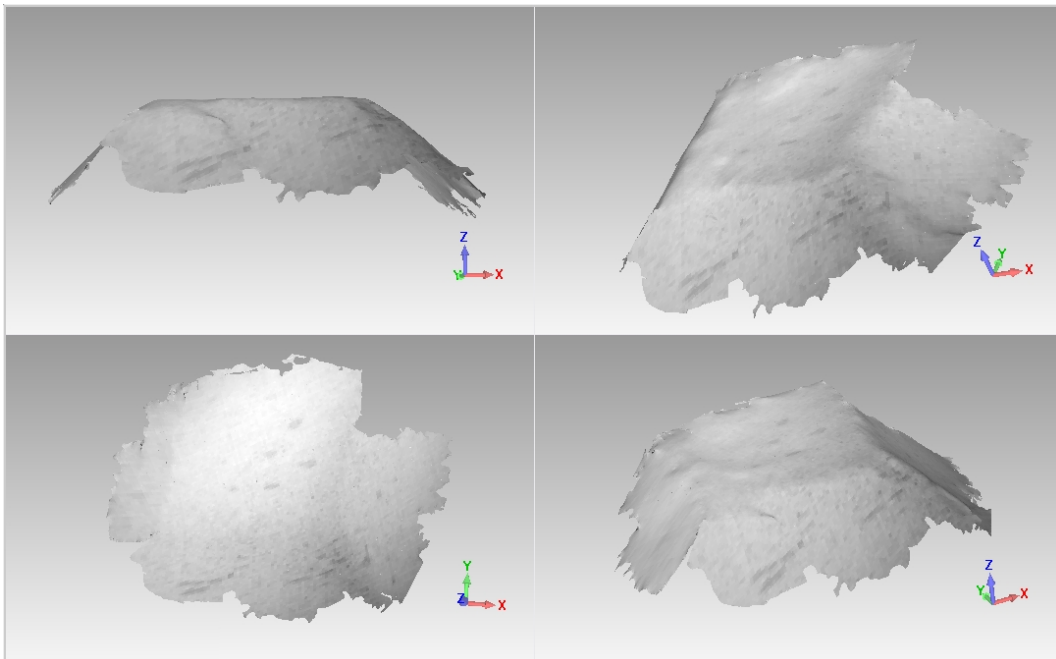


FIG. 5.20 – Pyramide reconstruit avec les caméras infrarouges.

maximal est observé sur le haut de la pyramide qui concentre le maximum d'erreur du recalage.

Un calcul de résolution spatiale (cf. chapitre 4) nous montre que pour une distance de mesure d'environ 30cm , une demi fenêtre de corrélation de 2.5 pixels, une taille de pixel de $12\mu\text{m}$ et une distance focale de 25mm , la précision de reconstruction est de 0.3mm dans un cas idéal. L'erreur obtenue sur notre reconstruction est donc tout à fait acceptable.

La figure 5.22 montre l'erreur entre la reconstruction $3D$ réalisée à partir des caméras infrarouge et la reconstruction de référence. Sur la première image avec une échelle d'erreur complète, l'erreur de la reconstruction est comprise entre 4.1mm et -12.8mm . L'erreur maximale est également observée sur les bords de l'objet. En reserrant l'échelle, $\pm 4\text{mm}$ et $\pm 1.2\text{mm}$ dans ce cas, cela permet de mieux observer la répartition d'erreur sur la pièce. L'erreur moyenne est ici de 1.7mm avec un écart type de 1.8mm .

Pour une distance d'observation d'environ 5cm , une demi fenêtre de corrélation de 4.5 pixels, une taille de pixel de $51\mu\text{m}$ et une distance focale de 11mm , la résolution spatiale est de 1mm . Ces résultats valident donc l'approche proposée et répondent à l'une des problématiques majeures de la thèse : il est donc possible d'obtenir le modèle $3D$ d'un objet, de manière entièrement automatique, à partir d'un banc de stéréovision infrarouge et avec une précision inférieure au centimètre.

5.5 Conclusion

Dans ce chapitre, nous avons tout d'abord présenté la technologie utilisée et détaillé le fonctionnement de chacun des systèmes. Il a ainsi été montré comment fonctionne une caméra microbolométrique et quelle était la chaîne cinématique du robot six axes utilisé.

Nous avons ensuite validé expérimentalement le modèle de caméra microbolométrique, le principal problème était de pouvoir fabriquer une mire de calibrage précise dans le domaine de l'infrarouge lointain ($8 - 12\mu\text{m}$). Ce problème a été réglé en utilisant des matériaux possédant des émissivités très différentes (Epoxy-Or), puis la validation du modèle de caméra a été réalisée à partir d'algorithmes classiques en vision par ordinateur.

Enfin, notre approche a été validée dans sa globalité en appliquant l'ensemble du processus afin de reconstruire deux objets différents. Dans un premier temps nous avons utilisé des images issues de caméras visibles. Le principal problème rencontré porte sur le nombre important de points extraits, en effet un trop grand nombre de points dégrade considérablement le temps de calcul de chaque algorithme. Dans un deuxième temps les tests ont porté sur des images issues de caméras infrarouges, où nous avons montré que la reconstruction $3D$ multivues fonctionne de manière précise, de l'ordre de 0.5mm avec les caméras visibles et 2mm avec les caméras infrarouges. Nous avons ainsi validé l'approche dans sa globalité et montré qu'il était possible d'obtenir un modèle $3D$ précis de l'objet observé à partir d'une construc-

tion stéréoscopique, multivue et infrarouge.

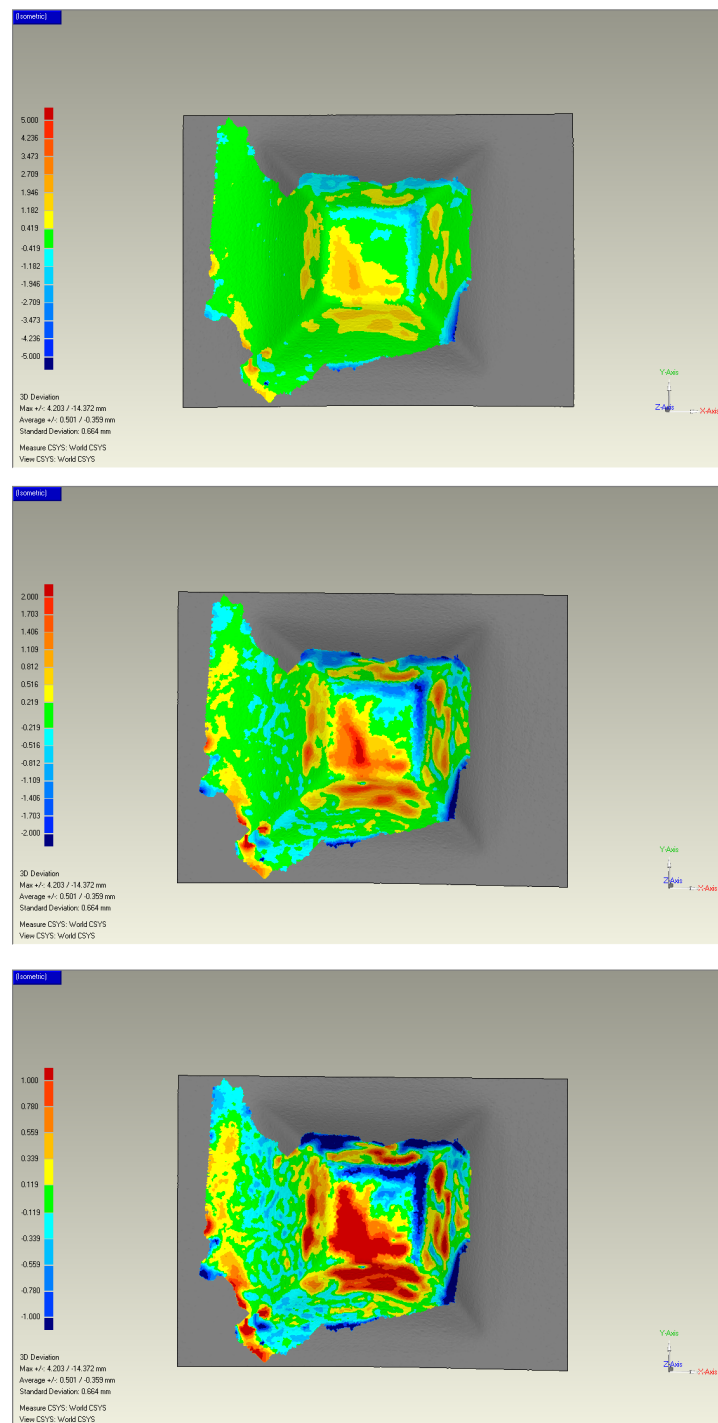


FIG. 5.21 – Erreur avec différentes échelles entre la reconstruction 3D avec des caméras visibles et la reconstruction de référence.

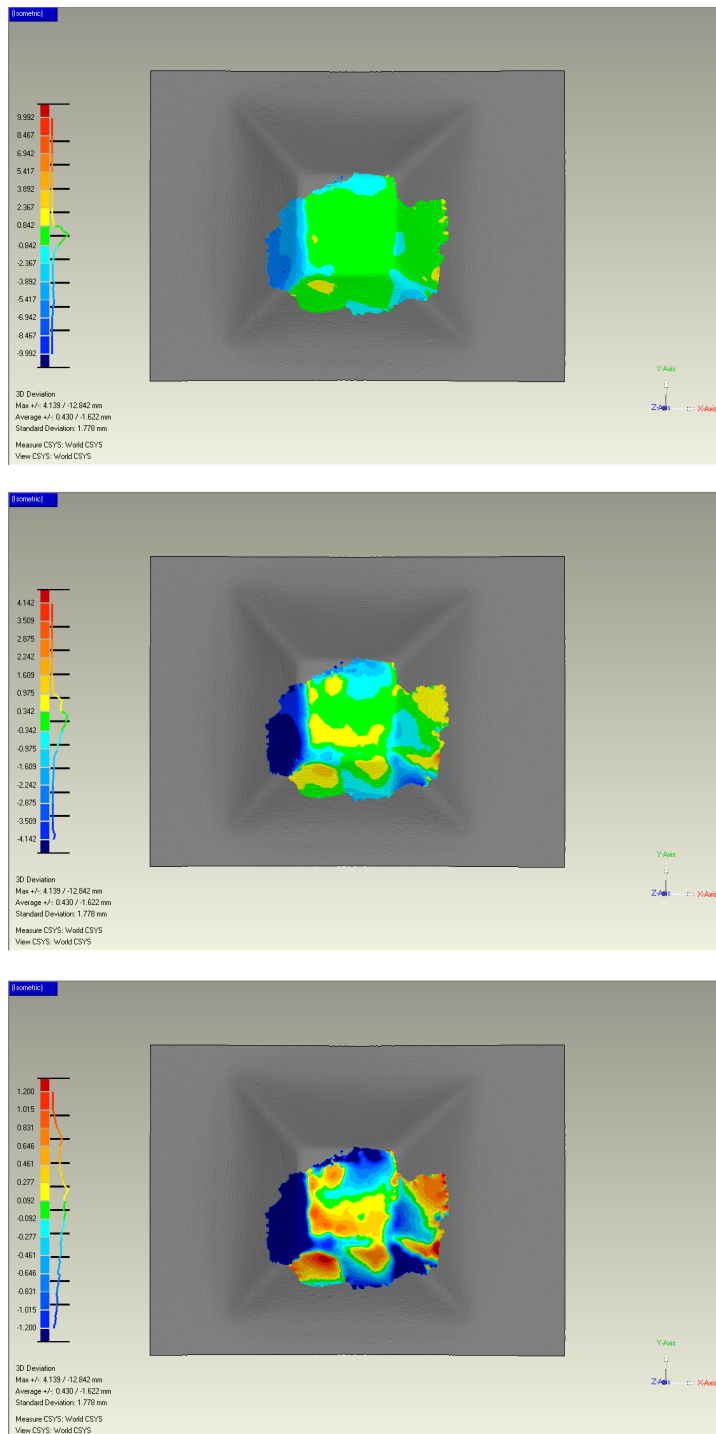


FIG. 5.22 – Erreur avec différentes échelles entre la reconstruction 3D avec des caméras infrarouges et la reconstruction de référence.

Conclusion

6.1 Bilan des travaux réalisés

Dans ce document, nous avons présenté l'ensemble des étapes pour réaliser la reconstruction *3D* infrarouge active complète d'un objet à partir d'un système composé d'un banc de stéréovision de caméras infrarouges monté sur un robot cartésien six axes. Le contexte applicatif est la mesure de champs de températures vraies, ce qui nécessite de connaître la forme de l'objet, mais aussi de contrôler localement la position et l'orientation du capteur vis-à-vis de la surface inspectée. Cette reconstruction s'opère donc dans un contexte industriel où une qualité essentielle est l'autonomie et la facilité de mise en oeuvre du système.

La démarche a alors consisté à analyser les méthodes existantes en vision classique pour chaque étape. Ensuite, nous avons déterminé ou proposé la méthode la plus adaptée aux caractéristiques des images infrarouges (faible résolution et peu de texture). Un effort particulier a porté sur l'étape d'autocalibrage pour laquelle nous avons proposé une approche originale qui repose sur une étape préalable de rectification elle-même spécifique. Finalement, nous avons validé le processus global de reconstruction *3D* infrarouge active en comparant les résultats obtenus avec un scanner *3D* laser précis à 50 micromètres. Ces résultats montrent que la reconstruction *3D* d'un objet simple avec plusieurs vues infrarouges de résolution uniquement de 164×128 pixels au pas de $50\mu m$ est obtenue avec une précision inférieure au centimètre.

Pour l'étape préalable de rectification, indispensable à l'étape d'autocalibrage, nous nous sommes intéressés aux méthodes de rectification d'images de stéréovision non calibrées en tenant compte des contraintes liées à l'imagerie infrarouge. Il a été montré que sur un petit nombre de points appariés avec une faible précision, les méthodes classiques qui reposent sur un calcul préalable de la matrice Fondamentale, génèrent d'importantes déformations projectives. Nous avons proposé une méthode dans laquelle les homographies nécessaires à la rectification sont estimées en une seule étape, sans calcul explicite de la matrice Fondamentale, limitant ainsi la propagation d'erreurs à travers les différents calculs. Cette méthode est basée sur un algorithme d'optimisation sous contraintes. La fonction coût à minimiser est la somme des distances croisées des points d'une image à la ligne épipolaire associée dans l'autre image. Cependant afin de limiter les déformations projectives, cette fonction coût est minimisée sous un ensemble de contraintes qui représentent des propriétés physiques à satisfaire entre images originales et images rectifiées,

afin d'assurer la préservation du rapport d'aspect, de la taille et de l'orthogonalité. Ces contraintes sont calculées pour les deux homographies soit par une expression analytique soit par une forme géométrique. Finalement, il a été montré que cette méthode permet d'obtenir une erreur de rectification finale inférieure à l'erreur de calcul de la matrice Fondamentale (si elle était calculée) tout en limitant les déformations projectives.

L'étape d'autocalibrage du système estime l'ensemble des paramètres qui modélisent à la fois le système de vision et le robot, en considérant que nous disposons d'une bonne estimation des mouvements du robot. Nous avons proposé une méthode d'autocalibrage en une seule étape pour minimiser là encore la propagation d'erreurs dans les calculs. Elle est particulièrement adaptée pour des images infrarouges qui disposent d'un faible nombre de points. Nous avons évalué différentes fonctions coût pour l'algorithme d'optimisation. La première fonction coût est fondée sur l'invariance de la position d'un point $3D$ reconstruit depuis différentes positions du capteur. Les autres fonctions sont basées sur la géométrie épipolaire et sur la décomposition des matrices Fondamentales associées à deux quelconques positions du capteur ; une telle transformation intègre les paramètres du système robot-caméra, ainsi que le mouvement, supposé connu, du capteur. Ces fonctions ont été comparées dans différentes configurations de caméras (caméra visible, caméra infrarouge, stéréovision rectifiée...) et différentes conditions d'acquisition d'images (points bruités, nombre et répartition des points...). Nous avons également réalisé une étude de sensibilité avec des images simulées, donc en comparant avec une vérité terrain exacte ; la conclusion de cette étude est que les fonctions coût basées sur l'utilisation d'un banc de stéréovision préalablement rectifié permet d'obtenir des résultats plus précis et une erreur de recalage plus faible. En effet, la rectification diminue le nombre de paramètres à estimer. Au final, nous avons appliqué ces méthodes à des cas réels, validant ainsi l'approche avec deux caméras infrarouges montées sur un robot six axes, étape nécessaire au projet R3T. Les résultats sur les images réelles ont confirmé cette évaluation en simulation.

Les différentes étapes de la reconstruction $3D$ à partir d'un banc de stéréovision composé de caméras infrarouges non calibrées ont ensuite été présentées. La première étape analysée a été l'étape d'extraction de points d'intérêt. Il existe dans la littérature un grand nombre de comparaisons sur les degrés d'invariance des différents points (Susan, Harris, SIFT, SURF, CensurE...), mais peu de travaux ont été consacrés aux images infrarouges. Nous avons comparé trois algorithmes, les "classiques" algorithmes "Harris" et "SIFT", et un algorithme spécifique proposé pour les images infrarouges, appelé "Phase Congruency". Nous avons montré qu'il n'existait pas une méthode universelle. Il est nécessaire d'adapter l'extracteur en fonction de la situation ; selon le cas, on peut avoir besoin d'un grand nombre de points bien répartis sur les images, ou de peu de points très caractéristiques. Pour notre application, le détecteur de Harris donne les meilleurs résultats pour l'étape de rectification alors que le détecteur "Phase Congruency" est plus adapté pour l'étape d'autocalibrage.

Nous avons en effet testé de multiples méthodes de filtrage pour supprimer les faux appariements lors de la phase d'optimisation qui exploite des paires de points d'intérêt appariés (par exemple pour évaluer la matrice fondamentale entre deux images). Nous avons montré que dans notre cas, les méthodes fondées sur le tri des erreurs de la fonction coût à minimiser (*Least Median Square* ou *Least Trimmed Square*) étaient plus robustes que les méthodes de vote (RANSAC) ou de pondération itérative (M-estimateur).

Ensuite, nous avons évalué différentes méthodes de stéréovision dense par corrélation entre images rectifiées, afin d'obtenir une carte de disparité la plus dense possible. De très nombreux travaux ont été consacrés à la stéréovision ; nous n'avons pas trouvé de raisons spécifiques pour développer un algorithme spécifique pour des images infrarouges. Par contre, nous avons montré que les méthodes de corrélation locale n'étaient pas assez robustes pour obtenir une reconstruction dense sur de telles images.

Nous avons consacré le dernier chapitre de ce mémoire à la partie expérimentale. Nous avons tout d'abord présenté l'ensemble des technologies utilisées, puis l'approche globale regroupant toutes les étapes proposées dans les chapitres précédents a été validée. Pour cette validation, nous avons utilisé deux objets de taille différente et effectué la reconstruction avec des caméras visibles et infrarouges. Ces différentes reconstructions ont montré que la reconstruction 3D multivues proposée fonctionne de manière précise et donne des résultats très satisfaisants.

6.2 Quelques perspectives

Le système de reconstruction 3D infrarouge actif présenté est suffisamment précis pour les besoins du projet R3T, mais cette précision pourrait être encore largement améliorée. La reconstruction globale de l'objet s'opère en transformant chaque modèle local dans un même repère à partir des paramètres issus du calibrage et des codeurs du robot. Aucune méthode de recalage n'a été implémentée. Il serait donc simple d'ajouter un algorithme de type ICP permettant de corriger ce recalage. Cela pourrait se faire de manière indépendante par un ICP classique ; il serait préférable d'essayer plutôt une optimisation de type *Bundle Adjustment* et de réestimer à cette occasion tous les paramètres du système, le modèle de l'objet (points 3D), le mouvement du capteur (positions successives de la caméra), mais aussi paramètres intrinsèques du système stéréo rectifié et transformée *Hand-Eye*.

Plusieurs stratégies d'optimisation locale pour résoudre un tel système avec de si nombreux paramètres, pourraient être évaluées, afin d'affiner progressivement chaque inconnu, en considérant connus les autres. Un travail plus théorique pourrait être dédié à l'évaluation de méthodes d'optimisation globale, déjà validées pour certaines étapes.

Concernant le contexte applicatif, à savoir le projet R3T et ses possibles suites, il reste à intégrer nos travaux avec ceux des autres partenaires du projet pour arriver au but, à savoir construire un champ de températures vraies sur un objet de forme complexe et d'émissivité non uniforme.

Dans nos travaux nous nous sommes focalisés sur les fonctions visuelles de reconstructions 3D. Même si nous parlons de reconstruction "active", nous n'avons pas consacré de temps à optimiser les mouvements du capteur. Plusieurs approches pourraient être appliquées pour ce faire.

Des approches de planification traitent de la recherche du meilleur point de vue (problème connu sous l'appellation NBV, pour *Next Best View*), afin de compléter le modèle d'un objet inconnu a priori, ou afin de mieux détecter les défauts sur un objet de forme connu. Il s'agit encore de problèmes d'optimisation, souvent difficiles car les fonctions d'utilité des points de vue sont souvent assez "plates". De nouvelles pistes pourraient être proposées pour le capteur issu des travaux R3T, afin de prendre en compte des contraintes propres à ce capteur : points de vue avec axe optique orthogonal à la surface. . .

Un autre moyen pour acquérir successivement plusieurs vues sur un objet, consiste à acquérir en continu pendant l'exécution d'une trajectoire en mode asservi sur l'image. Plusieurs techniques d'asservissement ont été proposées ; le robot cartésien est un outil particulièrement adapté pour tester des méthodes innovantes d'asservissement, par exemple les méthodes dites photométriques exploitant toute l'information disponible dans les images.

Enfin, l'amélioration de ces résultats permettrait de nouvelles utilisations en métrologie comme l'inspection de défaut sur des objets transparents ou réfléchissants.

Publications

Conférence : ORASIS - Congrès des jeunes chercheurs en vision par ordinateur

Titre : Rectification Projective d'Images Stéréo non Calibrées Infrarouges avec prise en compte globale de la minimisation des distorsions

Auteurs : Benoit Ducarouge, Thierry Sentenac, Florian Bugarin, Michel Devy,

Année : 2009

Résumé : L'article expose une approche symétrique globale pour estimer les homographies de la rectification projective à partir d'images infrarouges non calibrées et de la géométrie épipolaire. Cette étape est importante dans le cadre de la stéréoscopie par corrélation car elle permet de se ramener à une géométrie épipolaire simple, dans laquelle les droites épipolaires sont parallèles aux lignes des images. Le processus de calcul des homographies est basé sur l'optimisation d'un critère non linéaire prenant en compte la géométrie épipolaire ; ce processus est réalisé sous des contraintes qui réduisent dans le même temps les distorsions des images. Le critère d'optimisation et l'espace de contraintes choisis nous assurent une meilleure précision et moins de déformations projectives que les méthodes couramment utilisées dans la littérature.

Conférence : RFIA - Conférence en Reconnaissance des Formes et Intelligence Artificielle

Titre : Modélisation d'un objet 3D à partir d'Images Stéréo Infrarouges avec calibrage sur images rectifiées

Auteurs : Benoit Ducarouge, Thierry Sentenac, Michel Devy,

Année : 2010

Résumé : L'article expose une méthode reconstruction du modèle 3D d'un objet basée sur un banc de stéréovision infrarouge. L'approche est centrée sur une rectification projective suivie d'un calibrage simplifié sur ces images. Le modèle 3D s'obtient ensuite par corrélation dense des images infrarouges rectifiées. A partir de l'extraction et de l'appariement d'un premier ensemble d'un faible nombre de points

d'intérêt entre les deux images, nous détaillons le calcul des homographies de la rectification projective à partir d'images infrarouges non calibrées. Cette phase est réalisée par l'optimisation d'un critère non linéaire qui intègre des contraintes pour réduire les déformations des images rectifiées. Cette méthode assure une meilleure précision et des déformations projectives moindres que les méthodes couramment utilisées dans la littérature tout en minimisant la propagation d'erreurs. Ensuite, nous exposons le calibrage sur ces images rectifiées qui présente l'avantage de ne nécessiter qu'un nombre réduit de paramètres du banc de stéréovision. Enfin, afin d'obtenir une image 3D dense de l'objet observé, plusieurs méthodes de corrélation exploitant des critères locaux ou globaux ont été évaluées sur des images infrarouges rectifiées.

Conférence : 3DPVT - International Symposium on 3D Data Processing, Visualization and Transmission

Titre : A global projective rectification approach for uncalibrated infrared stereo images

Auteurs : Benoit Ducarouge, Thierry Sentenac, Florian Bugarin, Michel Devy,

Année : 2010

Résumé : This paper describes a global and symmetric approach proposed to estimate homographies required for the projective rectification of uncalibrated infrared images. Rectification is an important step in 3D reconstruction from stereo images, since it allows making less complex the epipolar geometry, with epipolar lines parallel to the images lines. The most of known methods proposed for rectification of uncalibrated stereo images are based on the preliminary estimation of the fundamental matrix and two steps for rectifying images. These methods applied to infrared images induce large error. The proposed method is performed in only one step, without a Fundamental matrix estimation. This is done by an optimization method under constraints that allows both optimizing precision while minimizing images distortions. Optimization criteria and constraints space are selected in order to give a rectification accuracy similar to the estimation error of the fundamental matrix. It is shown that residual projective distortions on the rectified images, evaluated from independant measurements, are very weak.

Annexe 1 : Modélisation d'une caméra et d'un capteur stéréovision

A.1 Définition et modélisation du système mono caméra

Afin de modéliser la caméra de façon simple, le modèle utilisé est le modèle sténopé qui caractérise une caméra par son centre optique et son plan image. Ce modèle est une approximation au premier ordre du processus de formation d'image et ne tient pas compte des distorsions dues à la lentille des caméras. Les images sont formées par projection sur le plan image à travers le centre optique comme illustré à la figure A.1.

Soit $\mathbf{P} = (x, y, z, 1)^T$ un point 3D exprimé dans le repère de la scène en coordonnées homogènes, alors $\mathbf{p} = (u, v, 1)$ sa projection dans le plan image est donnée par :

$$\mathbf{p} = \mathbf{KTP} \quad (\text{A.1})$$

avec $\mathbf{T} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix}$ la matrice des paramètres extrinsèques composée de \mathbf{R} et \mathbf{t} qui définissent l'orientation et la position du repère caméra dans le repère monde. \mathbf{K} est la matrice des paramètres intrinsèques donnée par :

$$\mathbf{K} = \begin{pmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (\text{A.2})$$

où $\alpha_u = fk_u, \alpha_v = fk_v, \gamma = fk_u \cos(\theta), u_0, v_0$ sont les paramètres intrinsèques à identifier. f est la longueur focale en unités métriques, k_u et k_v sont le nombre de pixels par unité métrique le long des axes horizontaux et verticaux et θ est l'angle entre les axes u et v . Finalement (u_0, v_0) est la position du point principal, soit l'intersection de l'axe optique avec le plan image. Aussi, l'angle entre les axes u et v étant très près de 90° , nous admettons donc l'hypothèse : $\gamma = 0$.

Une caméra est dite calibrée lorsque les paramètres de sa matrice \mathbf{K} sont identifiés.

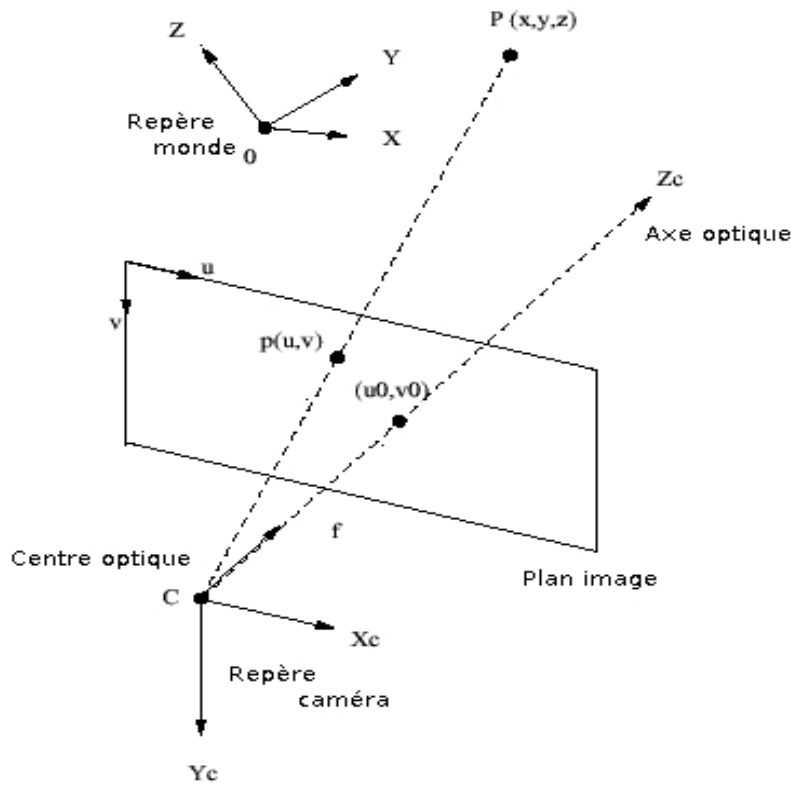


FIG. A.1 – *Modèle de projection sténopé.*

A.2 Définition et modélisation d'un système stéréovision

A.2.1 Modèle

Le système stéréovision est constitué de deux caméras (C et C') assemblées de manière rigide (nous n'aborderons pas les systèmes actifs de stéréovision), il est décrit par les modèles des deux caméras, et par la transformation rigide liant leurs repères. Si le repère monde est placé au centre optique de la première caméra C , les deux matrices de projection sont définies comme suit :

$$\begin{aligned}
 m &= \mathbf{K} \begin{pmatrix} \mathbf{I} & 0 \\ 0 & 1 \end{pmatrix} \mathbf{P} \\
 m' &= \mathbf{K}' \begin{pmatrix} \mathbf{R}_{\text{ext}} & \mathbf{t}_{\text{ext}} \\ 0 & 1 \end{pmatrix} \mathbf{P}
 \end{aligned}
 \tag{A.3}$$

Le système stéréo caméra est donc défini à partir des paramètres intrinsèques de chacune des caméras (\mathbf{K} et \mathbf{K}') et la transformation de la deuxième caméra par rapport à la première, notée \mathbf{M}_{ext} (\mathbf{R}_{ext} et \mathbf{t}_{ext}), appelés paramètres extrinsèques. Il y a donc 16 paramètres au total (5+5 = 10 paramètres intrinsèques, 3 paramètres

de rotation et 3 de translation).

A.2.2 Matrice essentielle

En éliminant \mathbf{P} dans l'équation (A.3), on obtient :

$$\mathbf{m}'^T \mathbf{K}'^{-T} [\mathbf{t}_{\text{ext}}]_{\times} \mathbf{R}_{\text{ext}} \mathbf{K}^{-1} \mathbf{m} = \mathbf{m}'^T \mathbf{K}'^{-T} \mathbf{E}_{\text{ext}} \mathbf{K}^{-1} \mathbf{m} = 0 \quad (\text{A.4})$$

avec \mathbf{E}_{ext} la matrice Essentielle qui décrit la transformation épipolaire gauche-droite. Le mouvement d'un objet inconnu entre les deux points de vues représentés dans ce cas par les deux images ou représentés par le déplacement d'une image entre deux instants peut être obtenu grâce à la connaissance de cette matrice. Elle est définie en fonction de la matrice Fondamentale \mathbf{F} et des paramètres intrinsèques (\mathbf{K} et \mathbf{K}') des caméras associées à chaque point de vue. Avec les notations précédentes, nous obtenons :

$$\mathbf{E}_{\text{ext}} = \mathbf{K}'^T \mathbf{F} \mathbf{K} \quad (\text{A.5})$$

Nous notons que entre deux points de vue à deux instants avec une seule caméra, cette relation est valable avec $\mathbf{K} = \mathbf{K}'$.

La matrice essentielle s'exprime donc en fonction de la transformation ($\mathbf{R}_{\text{ext}}, \mathbf{t}_{\text{ext}}$) entre les repères des deux points de vues (deux images stéréo ou deux images à deux instants) par la relation suivante :

$$\mathbf{E}_{\text{ext}} = [\mathbf{t}_{\text{ext}}]_{\times} \mathbf{R}_{\text{ext}} \quad (\text{A.6})$$

La matrice $[\mathbf{t}_{\text{ext}}]_{\times}$ est la matrice antisymétrique de rang deux défini comme suit :

$$[\mathbf{t}_{\text{ext}}]_{\times} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & t_x \\ -t_y & t_x & 0 \end{pmatrix} \quad (\text{A.7})$$

Cette représentation permet de réaliser une simple multiplication au lieu d'un produit en croix. La matrice essentielle comprend cinq degrés de liberté : 3 pour la rotation et 3 la translation et moins un en raison du facteur d'échelle.

A.2.3 Cas d'un système de stéréovision rectifié

Dans un système de stéréovision rectifié, les deux images rectifiées sont coplanaires et leurs abscisses sont colinéaires. La matrice de paramètres extrinsèques, \mathbf{M}_{ext} , relative aux deux caméras est donc réduite à un seul paramètre t_{ext_x} :

$$\mathbf{M}_{\text{ext}} = \begin{pmatrix} 1 & 0 & 0 & t_{ext_x} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.8})$$

De plus, pour un système rectifié, les lignes épipolaires étant alignées entre les deux images, certains paramètres intrinsèques des deux caméras sont identiques

($\alpha_v = \alpha_{v'} = \alpha_v$ et que $v_0 = v'_0 = v_0$). Les matrices de paramètres intrinsèques \mathbf{K} et \mathbf{K}' sont de la forme :

$$\mathbf{K} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} ; \quad \mathbf{K}' = \begin{pmatrix} \alpha_{u'} & 0 & u'_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (\text{A.9})$$

Bibliographie

- [Andreff 1999] Nicolas Andreff, Radu Horaud et Bernard Espiau. *On-Line Hand-Eye Calibration*. 3D Digital Imaging and Modeling, International Conference on, vol. 0, page 0430, 1999. 7, 45, 46, 47
- [Beaton 1974] Albert E. Beaton et John W. Tukey. *The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data*. Technometrics, vol. 16, no. 2, pages 147–185, 1974. 86, 88
- [Bertozzi 2007] M. Bertozzi, A. Broggi, C. Caraffi, M. Del Rose, M. Felisa et G. Vezzi. *Pedestrian detection by means of far-infrared stereo vision*. Computer Vision and Image Understanding, vol. 106, no. 2-3, pages 194 – 204, 2007. Special issue on Advances in Vision Algorithms and Systems beyond the Visible Spectrum. 1
- [Boykov 2001] Yuri Boykov, Olga Veksler et Ramin Zabih. *Fast Approximate Energy Minimization via Graph Cuts*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pages 1222–1239, 2001. 92
- [Bugarin 2010] F. Bugarin, D. Henrion, T. Sentenac, J.B. Lasserre et J.J. Orteu. *Optimisation globale polynomiale appliquée à la rectification projective d'images non calibrées*. RFIA 2010, Jan. 2010. 20, 29
- [Chou 1991] Jack C. K. Chou et M. Kamel. *Finding the position and orientation of a sensor on a robot manipulator using quaternions*. Int. J. Rob. Res., vol. 10, pages 240–254, June 1991. 44
- [Christy 1996] Stéphane Christy et Radu Horaud. *Euclidean Shape and Motion from Multiple Perspective Views by Affine Iterations*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, pages 1098–1104, November 1996. 46
- [C.T. Lawrence 1997] J.L. Zhou C.T. Lawrence et A.L. Tits. *User's guide for CF-SQP version 2.5*. Rapport technique, Electrical Engineering Department and Institute for Systems Research, University of Maryland, 1997. 29
- [Daniilidis 1996] K. Daniilidis et E. Bayro-Corrochano. *The dual quaternion approach to hand-eye calibration*. In Pattern Recognition, 1996., Proceedings of the 13th International Conference on, volume 1, 1996. 7, 44
- [Devernay 1997] F. Devernay. *Vision stéréoscopique et propriétés différentielles des surfaces*. PhD thesis, Ecole polytechnique, Février 1997. 6, 18
- [Eren 2009] Gonen Eren, Olivier Aubreton, Fabrice Meriaudeau, L.A. Sanchez Secades, David Fofi, A. Teoman Naskali, Frederic Truchetet et Aytul Ercil. *Scanning from heating : 3D shape estimation of transparent objects from local surface heating*. Opt. Express, vol. 17, no. 14, pages 11457–11468, 2009. 4
- [Eveland 2003] Christopher K. Eveland, Diego A. Socolinsky et Lawrence B. Wolff. *Tracking human faces in infrared video*. Image and Vision Computing, vol. 21,

- no. 7, pages 579 – 590, 2003. Computer Vision beyond the visible spectrum. 1
- [Fischler 1981] Martin A. Fischler et Robert C. Bolles. *Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography*. Commun. ACM, vol. 24, no. 6, pages 381–395, 1981. 86, 88
- [Gluckman 2001] Joshua Gluckman et Shree K. Nayar. *Rectifying Transformations That Minimize Resampling Effects*. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, vol. 1, page 111, 2001. 19, 20, 23
- [Hajebi 2007] K. Hajebi et J.S. Zelek. *Dense Surface from Infrared Stereo*. In Applications of Computer Vision, 2007. WACV '07. IEEE Workshop on, pages 21–21, Feb. 2007. 80, 81
- [Hajebi 2008] K. Hajebi et J.S. Zelek. *Structure from Infrared Stereo Images*. In Computer and Robot Vision, 2008. CRV '08. Canadian Conference on, pages 105–112, May 2008. 4
- [Hannah 1974] Marsha Jo Hannah. *Computer matching of areas in stereo images*. PhD thesis, Stanford, CA, USA, 1974. 91, 92
- [Harris 1988] C. Harris et M. Stephens. *A Combined Corner and Edge Detector*. Rapport technique, The Plessey Company plc, 1988. 72, 80, 81
- [Hartley 1994] R.I. Hartley. *An algorithm for self calibration from several views*. In In Proceedings of the Conference on Computer Vision and Pattern Recognition, pages 908–912, Seattle, Washington, USA, 1994. 6
- [Hartley 1999] R. Hartley. *Theorie and Practice of Projective Rectification*. International Journal of Computer Vision, vol. 35, pages 115–127, 1999. 16, 17, 20, 23, 30
- [Hartley 2003] Richard Hartley et Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003. 22
- [Heigl 2004] B Heigl. *Plenoptic Scene Modelling from Uncalibrated Image Sequences*. Rapport technique, ibidem-Verl, 2004. 46
- [Horaud 1998] Radu Horaud et Gabriella Csurka. *Self-Calibration and Euclidean Reconstruction Using Motions of a Stereo Rig*. Computer Vision, IEEE International Conference on, vol. 0, page 96, 1998. 6
- [Huber 1981] P.J. Huber. *Robust statistics*. Wiley, 1981. 86, 88
- [Isgro 1999] Francesco Isgro et Emanuele Trucco. *Projective Rectification Without Epipolar Geometry*. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, vol. 1, page 1094, 1999. 15, 20, 21, 22
- [Jordt 2009] Andreas Jordt, Nils T Siebel et Gerald Sommer. *Automatic High-Precision Self-Calibration of Camera-Robot Systems*, 2009. 7, 46, 48
- [Kanade 1995] T. Kanade, H. Kano, S. Kimura, A. Yoshida et K. Oda. *Development of a video-rate stereo machine*. In Intelligent Robots and Systems 95. 'Human

- Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on, volume 3, pages 95–100 vol.3, Aug 1995. 91, 92
- [Kovesi 1999] Peter Kovesi. *Image Features from Phase Congruency*, 1999. 80
- [Kovesi 2003] Peter Kovesi. *Phase congruency detects corners and edges*. 2003. 81
- [Lades 1993] Martin Lades, Jan C. Vorbrüggen, Joachim Buhmann, Jörg Lange, Christoph V. D. Malsburg, Rolf P. Würtz et Wolfgang Konen. *Distortion Invariant Object Recognition in the Dynamic Link Architecture*. IEEE Trans. Computers, vol. 42, pages 300–311, 1993. 81
- [Lasserre 2001] J.B. Lasserre. *Global Optimization with polynomials and the problem of moments*. SIAM Journal on Optimization, vol. 11, pages 796–817, 2001. 29
- [Lawrence 2001] C.T. Lawrence et A.L. Tits. *A computationally efficient feasible sequential quadratic programming algorithm*. SIAM Journal Optimization, pages 1092–1118, 2001. 29
- [Loop 1999] C. Loop et Z. Zhang. *Computing Rectifying Homographies for Stereo Vision*. Rapport technique, Microsoft Research, 1999. 17, 24, 30
- [Lowe 2004] David G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision, vol. 60, no. 2, pages 91–110, November 2004. 72, 80, 82
- [Luong 1992] Q-T. Luong. *Matrice Fondamentale et calibration visuelle sur l'environnement vers une plus grande autonomie des systemes robotiques*. PhD thesis, Université de Paris-sud, centre d'Orsay, Décembre 1992. 22
- [Maldague 2001] X. Maldague. *Theory and practice of infrared technology for non-destructive testing*. John Wiley Interscience, 2001. 1
- [Mallon 2005] J. Mallon et Whelan P. *Projective Rectification from the fundamental matrix*. Image and Vision Computing, vol. 23, pages 643–650, 2005. 18, 20, 23, 30
- [Maybank 1992] Stephen J. Maybank et Olivier D. Faugeras. *A theory of self-calibration of a moving camera*. International Journal of Computer Vision, vol. 8, pages 123–151, 1992. 10.1007/BF00127171. 6
- [Morrone 1987] M.C. Morrone et R.A. Owens. *Feature detection from local energy*. Pattern Recognition Letters, vol. 6, no. 5, pages 303 – 313, 1987. 80
- [Orteu 2008] J.J Orteu, Y. Rotrou, T. Sentenac et L. Robert. *An Innovative Method for 3-D Shape, Strain and Temperature Full-Field Measurement using a Single Type of Camera : Principle and Preliminary Results*. Experimental Mechanics, vol. 48(2), pages 63–179, April 2008. 4
- [Owens 1999] K. Owens et L. Matthies. *Passive night vision sensor comparison for unmanned ground vehicle stereo vision navigation*. In Computer Vision Beyond the Visible Spectrum : Methods and Applications, 1999. (CVBVS '99) Proceedings. IEEE Workshop on, pages 59–68, 1999. 1

- [Prakash 2006] Surya Prakash, Pei Yean Lee, Terry Caelli et Tim Raupach. *Robust thermal camera calibration and 3D mapping of object surface temperatures*. volume 6205, page 62050J. SPIE, 2006. 4
- [qing Wei 1998] Guo qing Wei, Klaus Arbter et Gerd Hirzinger. *Active Self-calibration of Robotic Eyes and Hand-eye Relationships with Model Identification*. IEEE Trans Robotics and Automation, 1998. 7, 46, 47, 48
- [R.Horaud 1995] R.Horaud et O.Monga. *Vision par ordinateur : outils fondamentaux*. HERMES, 1995. 7, 44
- [Rousseeuw 1987] P.J. Rousseeuw et A.M. Leroy. *Robust regression and outlier detection*. Wiley, 1987. 86, 89
- [Scharstein 2002] Daniel Scharstein et Richard Szeliski. *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*. Int. J. Comput. Vision, vol. 47, no. 1-3, pages 7–42, 2002. 91, 92, 110
- [Schmidt 2005] Jochen Schmidt, Florian Vogt et Heinrich Niemann. *Calibration Free Hand Eye Calibration : A Structure from Motion Approach*. In Walter G. Kropatsch, Robert Sablatnig et Allan Hanbury, editeurs, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 67–74. Springer Berlin / Heidelberg, 2005. 7, 46, 47
- [Shiu 1989] Yiu Cheung Shiu et Shaheen Ahmad. *Calibration of wrist mounted robotic sensors by solving homogeneous transform equations of the form $AX=XB$* . IEEE journal of robotics and automation, vol. 5, pages 16–29, 1989. 43
- [Sturm 1997] Peter Franz Sturm. *Vision 3D non calibrée : contributions à la reconstruction projective et étude des mouvements critiques pour l'auto-calibrage*. PhD thesis, 1997. 6
- [Tsai 1988] Roger Y. Tsai et Reimer K. Lenz. *A new technique for fully autonomous and efficient 3D robotics hand-eye calibration*. In *Proceedings of the 4th international symposium on Robotics Research*, pages 287–297, Cambridge, MA, USA, 1988. MIT Press. 7, 43
- [Wu 2007] H-H. Wu et C-C. Chen. *Scene reconstruction, pose estimation and tracking*. I-Tech, 2007. 24
- [Zhang 1993] Zhengyou Zhang, Quang-Tuan Luong et Olivier Faugeras. *Motion of an uncalibrated stereo rig : self-calibration and metric reconstruction*. Research Report RR-2079, INRIA, 1993. 6
- [Zhuang 1992] H. Zhuang et Y. C. Shiu. *A noise tolerant algorithm for wrist-mounted robotic sensor calibration with or without sensor orientation measurement*. International Conference on Intelligent Robots and Systems, vol. 2, pages 1095–1100, 1992. 44
- [Zureiki 2007] A. Zureiki, M. Devy et R. Chatila. *Stereo matching using reduced-graph cuts*. In *IEEE International Conference on Image Processing (ICIP 2007)*, pages 237–240, San Antonio (USA), 2007. 93

Reconstruction 3D infrarouge par perception active

Résumé : Ces travaux de thèse ont été menés dans le contexte du projet ANR blanc "Real Time and True Temperature measurement" (R3T), dédié à la métrologie thermique à partir de mesures dans l'infrarouge. L'estimation d'une température vraie à partir d'une mesure de température apparente par une caméra infrarouge, exploite un modèle radiométrique dans lequel apparaît des facteurs qui dépendent de la nature et de la forme de l'objet considéré. Ces travaux portent sur la construction d'un modèle géométrique de l'objet à partir de caméras infrarouges déplacées par un robot autour d'un objet.

Ces caméras, par rapport à des caméras standards, ont des caractéristiques spécifiques : faible résolution, peu de texture. Afin de faciliter la mise en oeuvre et de minimiser la complexité du système final, nous avons choisi une approche de stéréovision non calibrée. Nous avons donc un banc de stéréovision infrarouge embarqué sur un robot cartésien, pour acquérir plusieurs vues de l'objet d'intérêt ; les principales étapes concernent la rectification non calibrée des images acquises par le banc stéréo, puis le calibrage des caméras rectifiées et de la relation main-oeil sans utilisation de mire, puis la construction de modèles 3D locaux denses et le recalage de ces modèles partiels pour construire un modèle global de l'objet. Les contributions portent sur les deux premières étapes, rectification et calibrage pour la stéréovision.

Pour la rectification non calibrée, il est proposé une approche d'optimisation sous contraintes qui estime les homographies, à appliquer sur ces images pour les rectifier, sans calcul préalable de la matrice Fondamentale, tout en minimisant les déformations projectives entre images d'origine et images rectifiées. La fonction coût est calculée à partir de la distance de Sampson avec une décomposition de la matrice fondamentale. Deux types de contraintes, géométriques et algébriques, sont comparés pour minimiser les déformations projectives. L'approche proposée est comparée aux méthodes proposées par Loop et Zhang, Hartley, Mallon et al... sur des jeux de données classiques de la littérature. Il est montré que les résultats sont au moins équivalents sur des images classiques et meilleurs sur des images de faible qualité comme des images infrarouges.

Pour le calibrage sans mire, l'auteur propose de calibrer les caméras ainsi que la transformation main-oeil, indispensable dès lors que le banc stéréo est porté par un robot, en une seule étape ; l'une des originalités est que cette méthode permet de calibrer les caméras préalablement rectifiées et ainsi de minimiser le nombre de paramètres à estimer. De même plusieurs critères sont proposés et évalués par de nombreux résultats sur des données de synthèse et sur des données réelles.

Finalement, les méthodes de stéréovision testées pour ce contexte applicatif sont rapidement décrites ; des résultats expérimentaux acquis sur des objets sont présentés ainsi que des comparaisons vis-à-vis d'une vérité terrain connue.

Mots clés : infrarouge, stéréovision, rectification, autocalibrage, reconstruction 3D.

3D infrared reconstruction with active perception

Abstract : This dissertation was lead in the context of the R3T project (Real Time and True Temperature measurement), dedicated to metrology from thermal infrared measurements. The estimation of true temperature from apparent temperature measurement by an infrared camera uses a radiometric model which depends on nature and shape of the considered object. This work focuses on the construction of a geometric model from infrared cameras moved by a robot around an object.

Those cameras, in comparison with standard ones, have specific characteristics : low resolution, low texture. To minimize the complexity and easily implement the final system, we chose an stereo approach using uncalibrated cameras. So we have an infrared stereoring embeded on a Cartesian robot, to acquire multiple views of the object of interest. First, the main steps implemented concern uncalibrated images rectification and autocalibration of infrared stereoring and hand-eye transformation without use of a calibration pattern. Then, the reconstruction of locals 3D models and the merge of these models was done to reach a global model of the object. The contributions cover the first two stages, rectification and autocalibration, for the other stereo reconstruction steps, different algorithms were tested and the best was chosen for our application.

For the uncalibrated images rectification, an optimization approach under constraints is proposed. The estimation of rectification homographies is done, without the Fundamental matrix determination, while minimizing the distortion between original and corrected images. The cost function is based on the Sampson's distance with breakdown of the Fundamental matrix. Two constraints, geometrical and analytical, are compared to minimize distortion. The proposed approach is compared to methods proposed by Loop and Zhang, Hartley, Mallon et al ... on data sets from state of art. It is shown that the results are at least equivalent on conventional images and better on low quality images such as infrared images.

For the autocalibration, the author proposes to calibrate cameras and hand-eye transformation, essential whenever the stereoring is embeded on a robot, in one step. One of the originality is that this method allows to calibrate rectified cameras and so minimize the number of parameters to estimate. Similarly, several criteria are proposed and evaluated by numerous results on synthetic and real data.

Finally, all methods of stereovision tested for this application context are briefly described, the experimental results obtained on objects are presented and compared to ground truth.

Keywords : infrared, stereovision, rectification, autocalibration, 3D reconstruction.
