SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UNIVERSITÉ DE BOURGOGNE

Système de vision hybride à fovéation pour la vidéo-surveillance et la navigation robotique

FRANÇOIS RAMEAU

SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UNIVERSITÉ DE BOURGOGNE

THÈSE présentée par

FRANÇOIS RAMEAU

pour obtenir le

Grade de Docteur de
l'Université de Bourgogne

Spécialité : **Instrumentation et Informatique de l'Image**

Système de vision hybride à fovéation pour la vidéo-surveillance et la navigation robotique

Soutenue publiquement le 02 Décembre 2014 devant le Jury composé de :

MICHEL DHOME	Rapporteur	Directeur de recherche à l'Institut Pascal
MICHEL DEVY	Rapporteur	Directeur de recherche au LAAS
HELDER ARAUJO	Examinateur	Professeur à l'Université de Coimbra
CÉDRIC DEMONCEAUX	Co-Encadrant	Professeur à l'Université de Bourgogne
DÉSIRÉ SIDIBÉ	Co-Encadrant	Maître de conférence à l'Université de Bourgogne
DAVID FOFI	Directeur de thèse	Professeur à l'Université de Bourgogne

REMERCIEMENTS

Tout d'abord, je tiens à exprimer ma profonde gratitude à David Fofi le directeur de cette thèse, mais également à Cédric Demonceaux et Désiré Sidibé pour leur disponibilité inconditionnelle, leur soutien constant, leur habileté à me faire comprendre lorsque je m'égarais -souvent- et la confiance qu'ils m'ont accordée. De plus, il semble impossible de ne pas souligner leurs qualités scientifiques mais également humaines ainsi que leur humour respectif.

Merci également aux rapporteurs de ce manuscrit de thèse, Michel Dhome et Michel Devy pour leur travail et leurs retours très détaillés, pertinents qui ont su donner une nouvelle perspective à mes travaux. Je souhaite plus globalement remercier l'ensemble des membres constituant mon jury et notamment Helder Araujo pour sa précieuse présence durant ma soutenance en tant qu'examinateur et président du jury.

J'exprime également tous mes remerciements à la Direction Générale de l'Armement et au Conseil Régional de Bourgogne. Les deux organismes, qui m'ont permis de réaliser ces travaux de thèse, par leurs financements et leur support.

Un grand merci à tous les collègues (et amis) du laboratoire Le2i du Creusot :

- Vincent pour sa jovialité candide
- Guillaume & Mojdeh pour leur amitié
- Mohamed pour tous nos échanges mêlant élégamment métaphysique et matrices fondamentales
- Eric pour ses calembours de haute voltige
- DP pour sa soif intarissable (de savoir, bien entendu) et toute l'aide qu'il m'a apporté durant ma thèse
- Cansen pour son imperturbable zen attitude
- QingLin pour sa célèbre générosité et sympathie
- Alban pour nous avoir préservé autant que possible de ses odeurs corporelles après ses entraînements quasi-quotidien
- Vineet pour être Vineet
- Sik pour son habileté à parler très fort (même les jours où il est fatigué)
- Fabrice (a.k.a Don Mériaudeau) pour son dynamisme et son talent à résoudre n'importe quel problème (instantanément !)
- Omar pour sa gourmandise le poussant à subtiliser toute nourriture trainant sur les bureaux du labo (me préservant ainsi de l'obésité)
- David pour son humour corrosif

– Désiré pour toutes les qualités qu'on lui connait (tout le monde sait que seul sa grand-mère incarne plus la perfection que lui)
– Cédric pour tout le support qu'il m'a apporté durant presque quatre ans au laboratoire
– Ralph pour toute sa gentillesse et son aide ininterrompue (support technique et scientifique, agence matrimoniale... etc)
– Olivier M. pour toujours porter haut les couleurs du labo lors de ses exploits sportifs
– Raphael pour son aide inestimable
– Nathalie pour son aide, sa dévotion journalière et sa bienveillance à notre égard (quand les ordres de mission sont rendus à temps)
– Alice Alhem pour son esprit ingénu et sa sagacité de tout instant
– Christophe L. pour ses blagues et ses interprétations des génériques de notre enfance (et d'avant)
– Olivier A. pour son indéfectible solidarité tabagique
– Christophe S. pour sa propension à attirer autant d'étudiants égarés au laboratoire ("Where is Professor Stolz ?")
– Ouadi pour être toujours à l'écoute des autres
– Mazen pour nos soirées en République Tchèque
– Satya pour les connaissances en Hindi que j'ai acquises au fur et à mesure des années, en restant après 18h au travail
– Frédéric parcequ'il a la classe !
– Yohan pour sa connaissance étendue en spiritueux de toutes sortes
– Olivier L. pour sa jovialité et son style
– Shabayek pour sa bonne humeur inch'allah
– Adlane pour tous ses conseils toujours avisés
– Hamid pour sa grande dextérité à relier les thèses

Je n'oublierais pas de remercier l'ensemble de la Vecteo Team :
– Thomas pour ses analyses socio-politiques pointues et toujours de bon aloi
– Souhaiel pour sa grande sagesse et son flegme
– Antoine pour être un ami et un cordon bleu (il faudrait penser à m'inviter au fait ! ?)
– Mickael pour ses bises matinales (sans la langue)
– Wil et sa famille pour leur invitation à manger des pizzas (et à déménager)

Une thèse c'est aussi beaucoup de rencontres et je dois admettre en avoir fait beaucoup au cours de ces trois dernières années, il serait ici fastidieux de nommer toutes ces personnes qui sont maintenant pour la plupart de très bon amis. En voici toutefois une liste non exhaustive : Alper, Ozan, Ozan (oui deux fois), Sharib, Mélanie, Kassem, Andrea, Iris, Sophia, Suman, Taman, Ajad, Youssef, Jeffrey, Guillaume, Reinier, Astie, Kristina, Jérémie, Abdoulaye, Jilliam, Juan Manuel, Pierre, Damien, Peter (one love), Peter (from Germany), Amir, Benjamin, Soumya, Jhimli, Deepak, Lija, Maya, Sang, Sam, Giullia, Tommaso, Kedir, Anastasia, Abilash, Abinash, Diogone, Luis, Alexandru, Isabel, Ashutosh, Dadhichi, Andru, Chali Konda, Sai, Carlos, Emilie, Jonathan ... Merci égale-

ment à tous ceux que j'ai pu oublier ici !

Merci aux "potes", James, Dams, Alex, Zul', Bouboule, Benj', Max, Tim, Jocelyn, Maxime, Sonia ... pour m'avoir toujours soutenu pendant ma thèse mais aussi avant et pour toutes les soirée et les moments de joie partagés et à venir !

Je ne peux pas me permettre de rédiger des remerciements sans évoquer les Florences de France et de Navarre, merci à elles !

Finalement, merci du fond du coeur à toute ma famille et tout particulièrement à ma soeur Claire, mon frère William ainsi qu'à mes parents pour leur soutien de toujours.

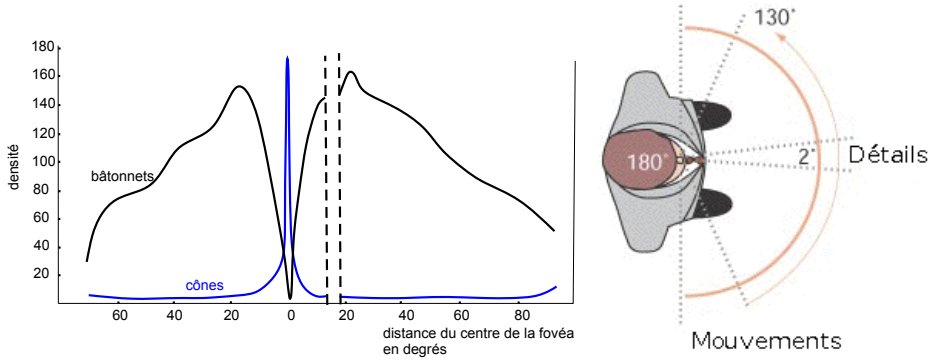
SOMMAIRE

1

INTRODUCTION

La vision est l'un des sens les plus développés chez l'Homme, elle nous permet de percevoir notre environnement, de reconnaître des objets, d'estimer des distances, d'effectuer des tâches complexes, etc ...
La rétine est l'organe sensible de la vision permettant la conversion des signaux lumineux perçus en signaux électriques transmis au cerveau par le nerf optique. Physiologiquement cette membrane photo-sensible est constituée de deux types de photo-récepteur, les cônes et les bâtonnets.
Les cônes sont particulièrement sensibles à la couleur et adaptés à la vision photopique (diurne) tandis que les bâtonnets, représentant 95% des cellules photo-sensibles de l'oeil, permettent la vision scotopique (vision en condition de faible luminosité). Les bâtonnets ne suffisent pas à eux seul à distinguer les couleurs mais sont 25 à 100 fois plus sensibles aux stimuli lumineux que les cônes. La répartition de ces récepteurs dans la rétine n'est pas uniforme (voir figure 1.1(a)), on distingue en effet deux zones appelées zone fovéale et zone périphérique.
La zone fovéale est une zone contenant un fort pourcentage de cônes et très peu de bâtonnets, elle est située dans le prolongement de l'axe optique et octroie une vision détaillée et tri-chromatiques de l'environnement mais ne couvre qu'un angle de vision de 2° environ (voir figure 1.1(b)). La zone périphérique, majoritairement constituée de bâtonnets, est quant à elle plus floue mais très sensible aux mouvements et aux faibles luminosités.
Ces deux zones ayants des propriétés distinctes livrent à la fois une vision globale mais où les détails sont mal perçus et une vision focalisée de haute résolution sur un point de fixation. Un évènement (*e.g.* un mouvement) détecté dans la vision périphérique du champ de vision peut donc être analysé finement en orientant le regard afin d'aligner la zone concernée avec le champ de vision fovéal [113].

Par analogie, cette thèse porte sur l'étude d'un système de vision artificielle reposant sur le même concept, où il est possible à la fois d'avoir une vision globale, mais de faible résolution, et une vision fovéale sur une zone d'intérêt. L'objectif est de faire collaborer deux caméras de types différents, nous appellerons cette association de caméras, un système de vision hybride. Le système qui nous concerne ici est constitué d'une caméra



(a) (b)

FIGURE 1.1 – (a) Densité des cônes et des bâtonnets dans la rétine (b) Champ de vue fovéal

de type omnidirectionnelle (vision périphérique) associée à une caméra PTZ (Pan-Tilt-Zoom pour panoramique-inclinaison-zoom) qui fera office de vision fovéale.

DÉFINITION DU PROBLÈME

Les caméras omnidirectionnelles permettent d'obtenir des images avec un grand champ de vue, cependant elles sont souvent dotées d'une résolution limitée et non-uniforme, ainsi que d'une focale fixe (pas de zoom optique). Elles entraînent de plus, une forte distorsion géométrique de l'image rendant complexe la plupart des traitements.
D'autre part, les caméras PTZ sont dites actives car elles peuvent être mécaniquement orientées dans de multiples directions. Malgré le champ de vision restreint d'une caméra perspective classique, la possibilité d'orienter la caméra dans une direction déterminée permet de couvrir l'ensemble de la scène (jusqu'à 360°). Le zoom offre quant à lui la possibilité d'obtenir une image de haute résolution sur une zone d'intérêt.
Le tandem formé par ces deux caméras permet de combiner les avantages offerts par chacune d'entre elles, à savoir la possibilité d'observer la scène dans sa globalité mais également de surveiller une zone désirée avec un niveau de détail ajustable.
Ce travail de thèse a donc pour objet, à la fois de permettre le suivi d'une cible à l'aide de notre banc de caméras mais également de permettre une reconstruction 3D par stéréoscopie hybride de l'environnement et de calculer le déplacement du robot équipé d'un tel capteur.

ORGANISATION DU DOCUMENT ET CONTRIBUTIONS

Comme l'illustre la figure 1.2, cette thèse s'articule autour de plusieurs problématiques réparties dans les chapitres suivants :

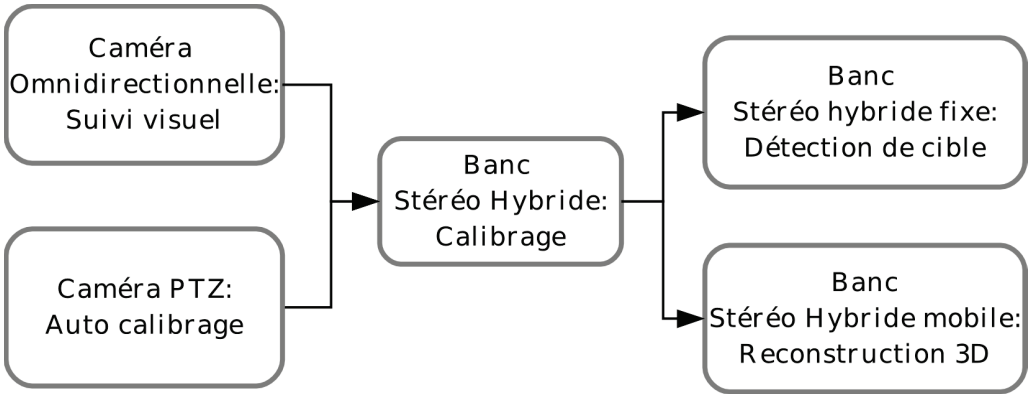


FIGURE 1.2 – Problématique globale.

Chapitre 2 : Géométrie des capteurs Dans ce chapitre nous présentons les modèles géométriques existants pour les caméras perspectives et omnidirectionnelles ainsi que la géométrie multi-vues allant de 2 à 4 poses.

Chapitre 3 : Suivi d'objet avec des caméras omnidirectionnelles Alors que le suivi de régions d'intérêt dans les images perspectives est un domaine désormais bien maitrisé, le problème reste encore très ouvert dans le cas d'images omnidirectionnelles. En effet, l'obtention d'une image panoramique à partir d'une seule prise de vue se fait au dépend de très fortes distorsions qui rendent inapplicables les méthodes usuelles. **Contribution :** La méthode développée adapte des algorithmes de suivi visuel usuels pour des images omnidirectionnelles. Cette méthode est basée sur une représentation sphérique de l'image qui permet de prendre en compte les distorsions et la résolution non-uniforme des images omnidirectionnelles.

Chapitre 4 : Auto-calibrage de caméra PTZ Le calibrage de caméra est une étape cruciale pour tout procédé de reconstruction 3D à partir d'images. En effet cette étape permet l'obtention des paramètres (internes ou externes à la caméra) décrivant la projection de l'espace tri-dimensionnels de la scène sur le plan image. De nombreuses méthodes permettent d'étalonner des caméras hors-ligne à l'aide de mires. Ces méthodes sont très efficaces lorsqu'il est question de caméra à paramètres intrinsèques fixes. Cependant, dans le cas d'une caméra PTZ, l'utilisation du zoom entraîne une modification des paramètres intrinsèques (distance focale et point principal). Notons de même que, une rotation de la caméra modifie les paramètres extrinsèques.
Contribution : Nous proposons dans cette thèse une approche faisant intervenir de nouvelles contraintes d'inégalité permettant d'intégrer des informations *a priori* sur la caméra afin d'améliorer la qualité de l'auto-calibrage pour ce type de caméra.

Chapitre 5 : Calibrage d'un système de vision hybride et contrôle d'une caméra PTZ au sein d'un banc de stéréo-vision hybride fixe Le calibrage d'un système de stéréo-vision est une tâche particulièrement importante lorsque l'on souhaite effectuer des reconstructions tri-dimensionnelles car elle facilite la mise en correspondance entre images et elle donne accès à une reconstruction à l'échelle réelle. C'est une opération devenue triviale en vision par ordinateur car de nombreux outils ont été mis à disposition par la communauté. Cependant, le calibrage géométrique de système de vision combinant différents types de caméra nécessite une adaptation des approches conventionnelles.
Contribution : Nous développons dans ce manuscrit une méthode permettant l'obtention des paramètres extrinsèques dans le cas d'un système de vision hétérogène.

Les systèmes de surveillance mettant en œuvre des caméras PTZ sont nombreux, cependant la question d'un capteur alliant une caméra rotative avec une caméra omnidirectionnelle n'est encore que peu étudiée. Dans les travaux portant sur ce cas spécifique, les auteurs proposent des méthodes très dépendantes de l'environnement.
Contribution : Nous proposons ici une approche plus générique en prenant avantage des connaissances géométriques du capteur obtenu par calibrage. Cette approche permet d'orienter la caméra rotative dans la direction d'une cible visible sur l'image panoramique sans utiliser de contraintes liées à la scène.

Chapitre 6 : Reconstruction 3D à l'aide d'un système de vision hétérogène monté sur un robot mobile L'emploi de deux caméras offre plusieurs avantages. Tout d'abord ce dispositif permet une reconstruction à une échelle métrique et par conséquent la possibilité d'estimer le véritable déplacement du banc de stéréo-vision. De plus le calibrage préalable du système peut faciliter la mise en correspondance entre images. Ces systèmes sont couramment utilisés pour la navigation robotique pour leur capacité à estimer le déplacement à l'échelle réelle tout en évitant les dérives inhérentes à d'autre instruments de mesure tel que l'odométrie.
Cependant l'utilisation de systèmes de vision hybride pour la navigation n'a soulevé que peu d'intérêt jusqu'à présent. Ce type d'approche peut pourtant constituer un grand avantage. Une vue globale de l'environnement assure par exemple l'existence de points de correspondance. D'autre part, une reconstruction 3D fine peut être obtenue à l'échelle réelle avec la caméra perspective. Nous verrons dans cette thèse qu'il est possible à l'aide d'un système de stéréo-vision hybride d'estimer le déplacement d'un robot mobile.
Contribution : Nous proposons dans cette thèse une approche de "structure from motion" sans recouvrement adaptée aux particularités de notre capteur. Malgré l'existence d'un champ de vue commun entre les deux caméras de notre système de vision hybride, la forte différence entre les images rend l'étape de mise en correspondance inter-caméra particulièrement difficile. C'est pourquoi l'usage d'approches généralement réservées aux bancs de caméras à champs de vue disjoints est particulièrement intéressant.

PUBLICATIONS DE L'AUTEUR

Revues internationales :

1. François RAMEAU, Désiré SIDIBE, Cédric DEMONCEAUX, David FOFI, "Visual Tracking with Omnidirectional Cameras : An Efficient Approach", IET Electronics Letters, 47(21), pp. 1183-1184, October 2011.

Conférences internationales :

1. Francois RAMEAU, Cédric DEMONCEAUX, Désiré SIDIBE, David FOFI, "Control of a PTZ Camera in a Hybrid Vision System", 9th International Conference on Computer Vision Theory and Applications, Lisbon (VISAPP), Portugal, January 2014.
2. François RAMEAU, Adlane HABED, Cédric DEMONCEAUX, Désiré SIDIBE, David FOFI, "Self-Calibration of PTZ Camera using New LMI Constraints", 11th Asian Conference on Computer Vision (ACCV), Daejeon, South Korea, November 2012.

Workshops internationaux :

1. François RAMEAU, Désiré SIDIBE, Cédric DEMONCEAUX, David FOFI, "Tracking Moving Objects With a Catadioptric Sensor Using Particle Filter", 11th Workshop on Omnidirectional Vision and Camera Networks (OMNIVIS'11), Barcelona (Spain), November 2011.

Conférences nationales :

1. François RAMEAU, Cédric DEMONCEAUX, Désiré SIDIBÉ, David FOFI, "Étude d'un système de stéréo-vision hybride", Congrès des jeunes chercheurs en vision par ordinateur (ORASIS), Cluny, France, 2013.
2. François RAMEAU, Désiré SIDIBÉ, Cédric DEMONCEAUX, David FOFI, "Une approche performante de suivi visuel pour les caméras catadioptriques", Reconnaissance des Formes et Intelligence Artificielle (RFIA), Lyon, France, 2012.

2

GÉOMÉTRIE DES CAPTEURS

Dans ce chapitre nous nous concentrerons sur les bases nécessaires à la compréhension des différentes approches proposées dans cette thèse. Nous aborderons d'abord la modélisation géométrique des caméras perspectives. Les différentes modalités d'acquisition d'images panoramiques ainsi que les modèles géométriques associés seront également étudiés. Nous nous intéresserons ensuite à la géométrie multi-vues, où nous couvrirons les cas allant de deux à quatre vues.

2.1/ GÉOMÉTRIE PROJECTIVE ET NOTATIONS

La géométrie projective peut modéliser la façon dont nous percevons visuellement le monde qui nous entoure (et plus généralement la projection d'un espace à n dimensions sur un espace de dimension inférieur), c'est pourquoi on la retrouve parfois sous le nom de "géométrie descriptive". Les effets de cette géométrie sont particulièrement perceptibles sur des clichés photographiques tels que présentés figure 2.1(a) où les propriétés de la géométrie euclidienne ne sont pas préservées. En effet, sur les images ainsi obtenues les objets circulaires deviennent elliptiques tandis que les parallélogrammes apparaissent sous formes de quadrilatères quelconques. Bien que le théorème de Pappus (3$^{\text{ème}}$ siècle av. JC) soit souvent considéré comme la prémisse de cette géométrie, elle a pourtant émergé assez tardivement dans l'histoire des Mathématiques sous l'impulsion de l'ingénieur Français Desargues au 17$^{\text{ème}}$ siècle. Ses propriétés ont également été utilisées auparavant par de nombreux artistes de la Renaissance, soucieux d'offrir un rendu plus réaliste de leurs oeuvres (voir fig. 2.1(b)). Pour plus de détails concernant l'histoire de la géométrie projective, nous conseillons la lecture de [45].
La géométrie euclidienne offre une description des objets avec des dimensions et des angles fixes et mesurables. La géométrie projective quant à elle ne conserve ni les longueurs, ni les angles ni même le parallélisme des lignes et des plans ce qui en fait un outil essentiel pour décrire la formation des images sur la surface photosensible d'une caméra.

Les coordonnées homogènes introduites par Möbius ont révolutionnés l'approche de la

FIGURE 2.1 – (a) Photographie des colonnes de Buren (b) tableau intitulé La Remise des clefs à Saint Pierre de Vannucci.

géométrie projective en facilitant grandement les calculs dans l'espace projectif. Le principe de ces coordonnées est de représenter un point dans un espace de dimension n par un vecteur de $n+1$ coordonnées : $\mathbf{x} \simeq [x_1, ..., x_{n+1}]^T$. De cette manière il est possible de représenter les points existants à une distance finie autant que les points situés à l'infini, de plus cela permet de prendre en compte le fait que l'échelle n'est pas préservée en géométrie projective.
Plus explicitement, dans le cas d'un point localisé à une distance finie la coordonnée supplémentaire peut être considérée comme le facteur d'échelle (d'où l'utilisation du symbole $\simeq$ représentant l'égalité à l'échelle près), les coordonnées cartésiennes de ce point sont alors : $[x_1/x_{n+1}, ..., x_n/x_{n+1}]^T$. Dans le cas d'un point situé à l'infini la valeur du "facteur d'échelle" x_{n+1} sera égal ou très proche de zéros $\mathbf{x} \simeq [x_1, ..., 0]^T$. Nous verrons par la suite que cette manière de représenter l'espace offre de nombreux avantages.

NOTATIONS

x	Un scalaire.
$\simeq$	Égalité à un facteur multiplicatif près.
$\times$	Produit vectoriel.
$\mathbf{p}$	Point image.
$\mathbf{P}$	Point 3D.
$\mathbf{P}_s$	Point sur la sphère unité.
$\mathbf{H}$	Matrice d'homographie.
$det(\mathbf{H})$	Déterminant de la matrice $\mathbf{H}$.
$\mathbf{R}$	Matrice de rotation.
$\mathbf{t}$	Vecteur de translation.

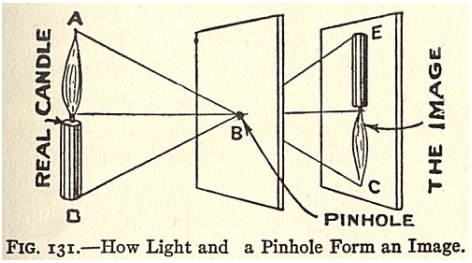


FIG. 131.—How Light and a Pinhole Form an Image.

FIGURE 2.2 – Représentation grossière du principe de fonctionnement d'une caméra sténopé illustré en 1925 dans "*The Boy Scientist*".

2.2/ LE MODÈLE STÉNOPÉ

Le principe de toute caméra "conventionnelle" consiste en la projection d'un espace tridimensionnel (le monde réel) sur un plan 2D (l'image). Cette projection peut être caractérisée géométriquement à l'aide de différents modèles de projection. Chaque modèle utilise un ensemble de paramètres permettant de décrire au mieux le fonctionnement physique du capteur. Nous présentons ici le plus simple mais également le plus utilisé pour décrire le fonctionnement d'une caméra perspective, le modèle sténopé. La traduction anglo-saxonne de sténopé est "*pinhole*", littéralement "trou d'épingle" ce qui représente bien le dispositif optique responsable de la formation de l'image (voir figure 2.2). Ce trou infiniment petit est en effet le point de convergence des rayons lumineux. Dans ce modèle la caméra peut être représentée par deux éléments, le plan image $\pi_\mathbf{i}$ - aussi appelé plan rétinien - et la position du centre optique $\mathbf{O}$ le "trou d'épingle" (que nous appelerons également point focal) au travers duquel les rayons de lumières reflétés par la scène se propagent. Dans un dispositif physique, ce point focal sera toujours situé devant la surface photosensible. L'image résultante sera donc la projection inversée de la scène. En vision par ordinateur, -par convenance- on considère en général ce point comme étant localisé derrière le plan image, de cette manière l'image obtenue ne subira pas cette inversion. Dans cette configuration, un point 3D $\mathbf{P}$ situé dans la scène aux coordonnées (X, Y, Z) dans le repère monde se projette en un point $\mathbf{p}$ de coordonnées image (x, y) exprimé généralement en pixels dans le cas d'une caméra numérique (figure 2.3). Le modèle sténopé met donc en oeuvre un ensemble de paramètres permettant de modéliser cette projection perspective. On peut décomposer ces paramètres en deux catégories distinctes : les paramètres intrinsèques ($\mathbf{K}$) exprimés sous forme d'une matrice de taille 3×3 et les paramètres extrinsèques composés d'une matrice de rotation 3×3 $\mathbf{R}$ et d'un vecteur de translation 3×1 $\mathbf{t}$. Nous verrons également que la projection d'un point 3D sur le capteur peut s'exprimer de manière linéaire à l'aide d'une matrice de transformation de taille 3×4 appelée matrice de projection. Cependant le modèle sténopé n'est pas parfaitement respecté, aujourd'hui ce n'est plus un simple trou qui permet la formation de l'image mais un ensemble complexe de lentilles permettant de focaliser la lumière sur la surface photosensible. L'utilisation de ces systèmes optiques entraîne des

imperfections dans le modèle linéaire brièvement évoqué précédemment. Nous montrerons qu'il est également possible de prendre en considération les distorsions induites par l'utilisation de lentilles à l'aide de relations mathématiques non-linéaires incorporées au modèle. Notons que ce modèle n'est qu'une approximation de la géométrie réelle d'une

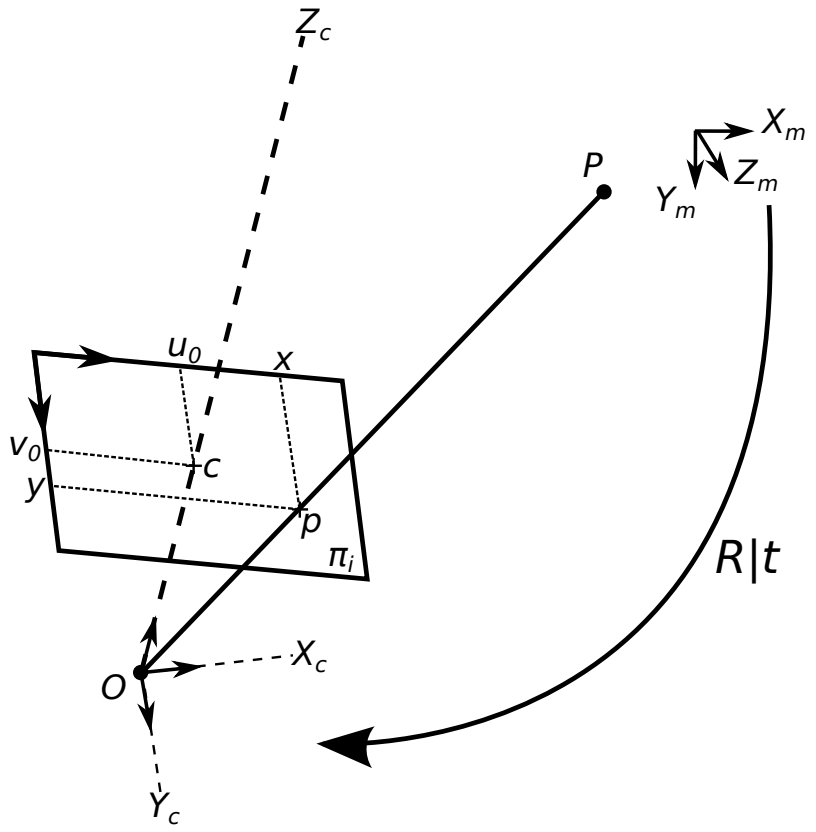


FIGURE 2.3 – Le modèle sténopé

caméra et que de nombreux paramètres ne sont pas pris en compte comme par exemple le flou. Cependant ce modèle n'admet qu'un nombre limité de paramètres. On évite ainsi les "erreurs numériques" liées à la résolution de grands systèmes paramétriques.

2.2.1/ LES PARAMÈTRES INTRINSÈQUES

Les paramètres intrinsèques sont les paramètres internes à la caméra et modélisent les caractéristiques optiques du capteur. La ligne perpendiculaire à $\pi_\mathbf{i}$ traversant le centre optique est appelé l'axe principal, l'intersection de l'axe principal avec le plan image est connu sous le nom de point principal $\mathbf{c}$ localisé dans l'image aux coordonnées (u_0, v_0) exprimé en pixels. Rappelons également que la distance entre le plan caméra - plan parallèle au plan image et passant par $\mathbf{O}$ - et le plan image correspond à la distance focale f. Les paramètres présentés ici ne tiennent pas compte des paramètres extrinsèques ce qui signifie que le point 3D est déjà exprimé dans le repère caméra.

Projection d'un point à l'aide des coordonnées homogènes La projection de $\mathbf{P}$ (exprimé dans le repère caméra) sur $\pi_\mathbf{i}$ peut s'exprimer de la manière suivante :

$$(X, Y, Z)^T \mapsto (fX/Z, fY/Z)^T \tag{2.1}$$

Cette transformation dans l'espace euclidien est non-linéaire, toutefois sa réécriture à l'aide des coordonnées homogènes fournit la relation linéaire suivante :

$$\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{2.2}$$

La matrice de projection homogène $diag(f, f, 1)[\mathbf{I} \mid \mathbf{0}]$ permet donc le passage entre le repère caméra et le repère image.

Passage aux coordonnées pixeliques L'équation (2.1) exprime la projection d'un point 3D dans le plan image normalisé avec comme origine le point principal $\mathbf{c}$. Cependant, en pratique l'origine des coordonnées image (exprimé en pixels) se situe dans le coin supérieur gauche de l'image. Il est possible d'exprimer ce changement de repère de la manière suivante :

$$\mathbf{P} \mapsto \mathbf{p}, \tag{2.3}$$

$$(X, Y, Z)^T \mapsto (fX/Z + u_0, fY/Z + v_0)^T, \tag{2.4}$$

cette fois encore on peut simplifier cette relation à l'aide des coordonnées homogènes :

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \tag{2.5}$$

La dérivation des paramètres internes à la caméra présentée jusqu'à présent considère que les pixels sont orthogonaux et possèdent la même échelle sur les deux axes qui constituent le capteur, en d'autre termes que les pixels sont parfaitement carrés. Pourtant ce n'est pas toujours le cas en pratique pour les capteurs CCD et CMOS utilisés (voir figure 2.4). Il est possible de modéliser ces imperfections à l'aide de deux paramètres : λ qui n'est autre qu'un facteur d'échelle entre la largeur et la hauteur d'un pixel, et s qui correspond à la non-orthogonalité d'un pixel $s = tan(\alpha)f$. La matrice intrinsèque complète

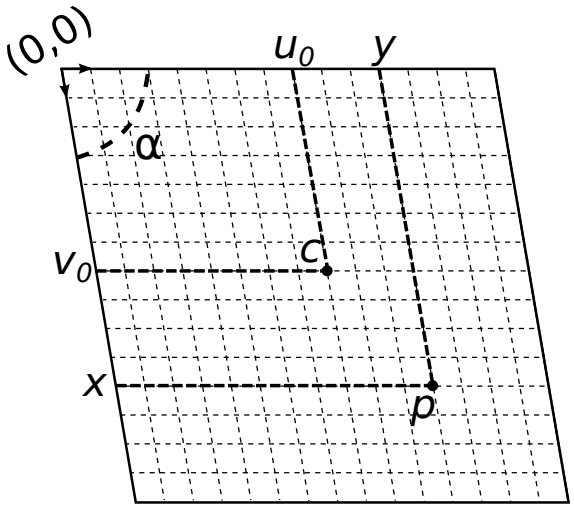


FIGURE 2.4 – Représentation du capteur

peut donc s'écrire :

$$\mathbf{K} = \begin{bmatrix} f & s & u_0 \\ 0 & \lambda f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.6}$$

Ainsi, la matrice $\mathbf{K}$ contient 5 paramètres propres à la caméra :
- λ est le rapport d'aspect de pixel. Ce paramètre est représentatif du rapport largeur/hauteur des pixels qui constituent l'image. Dans le cas idéal (c'est-à-dire si les pixels sont carrés), on aura λ=1.
- f correspond à la distance focale. Il s'agit de la distance entre le centre optique et le plan image. Ce facteur est responsable du grossissement de l'image (zoom).
- u_0 et v_0 sont les coordonnées du point principal sur l'image. Ce point correspond à la projection orthogonale du centre optique sur le plan image.
- s correspond à la non-orthogonalité des pixels (qu'on appelle aussi "obliquité").

2.2.2/ LES PARAMÈTRES EXTRINSÈQUES

Les paramètres extrinsèques déterminent la position et l'orientation de la caméra par rapport à un référentiel monde donné. Par exemple, un point 3D $\mathbf{P_m}$ est initialement exprimé dans le repère monde $\mathbf{O_m}$ suivant les coordonnées (X_m, Y_m, Z_m). Afin d'exprimer ce point dans le repère caméra il faut donc lui appliquer une transformation rigide, constituée d'une rotation $\mathbf{R}$ et d'une translation $\mathbf{t} = [t_x, t_y, t_z]^T$.

Rotation La rotation permet d'ajuster l'orientation de la caméra afin de l'aligner sur le repère monde, de cette manière l'axe Z_m sera perpendiculaire au plan image $\pi_\mathbf{i}$. Pour ce faire, on passera par l'utilisation d'une matrice de rotation $\mathbf{R}$ de taille 3×3 permettant une

rotation composée autour des trois axes. $\mathbf{R}$ est une matrice orthogonale de déterminant 1. $\mathbf{R}$ peut être obtenue par le produit de trois autres matrices de rotation autour d'un seul axe X, Y et Z respectivement :

$$\mathbf{R_x}(\phi_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\phi_x) & -sin(\phi_x) \\ 0 & sin(\phi_x) & cos(\phi_x) \end{bmatrix} \tag{2.7}$$

$$\mathbf{R_y}(\phi_\mathbf{y}) = \begin{bmatrix} cos(\phi_y) & 0 & sin(\phi_y) \\ 0 & 1 & 0 \\ -sin(\phi_y) & 0 & cos(\phi_y) \end{bmatrix} \tag{2.8}$$

$$\mathbf{R_z}(\phi_z) = \begin{bmatrix} cos(\phi_z) & -sin(\phi_y) & 0 \\ sin(\phi_z) & cos(\phi_z) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.9}$$

Avec ϕ l'angle de rotation exprimé en radians.
Il existe différents types de représentation permettant d'obtenir une matrice de rotation sur les trois axes :
Avec les angles de Cardan (lacet, tangage, roulis)

$$\mathbf{R}(\phi_z, \phi_y, \phi_x) = \mathbf{R_z}(\phi_z)\mathbf{R_y}(\phi_y)\mathbf{R_x}(\phi_x) \tag{2.10}$$

Avec la représentation d'Euler (z-y-z)

$$\mathbf{R}(\phi_z, \phi_y, \phi_x) = \mathbf{R_z}(\phi_z)\mathbf{R_y}(\phi_y)\mathbf{R_z}(\phi_x) \tag{2.11}$$

La rotation permettant de passer le point $\mathbf{P_m}$ dans un autre repère est donnée par :

$$\widehat{\mathbf{P_c}} = \mathbf{R}\mathbf{P_m} \tag{2.12}$$

On obtient ainsi le point $\widehat{\mathbf{P_c}}(\widehat{X_c}, \widehat{Y_c}, \widehat{Z_c})$.

Translation On peut ensuite appliquer une translation $\mathbf{t}$ au point $\widehat{\mathbf{P_c}}$ de manière à "superposer" la position du centre optique de la caméra avec le repère monde. D'un point de vue algébrique, il s'agit simplement d'une addition sur les trois axes du vecteur $\widehat{\mathbf{P_c}}$:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = \begin{pmatrix} \widehat{X_c} \\ \widehat{Y_c} \\ \widehat{Z_c} \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \tag{2.13}$$

On obtient alors le point $\mathbf{P_c}$ exprimé dans le repère caméra.

2.2.3/ LA MATRICE DE PROJECTION

Finalement il est possible de combiner les paramètres intrinsèques et extrinsèques afin d'obtenir une expression linéaire permettant la projection d'un point de la scène dans l'image. La projection globale peut s'écrire en coordonnées homogènes :

$$\mathbf{p} \simeq \mathbf{K}.[\mathbf{R}|\mathbf{t}]\mathbf{P_m} \tag{2.14}$$

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} f & s & u_0 \\ 0 & \lambda f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} & \mathbf{R}_{3\times3} & & \begin{matrix} t_x \\ t_y \\ t_z \end{matrix} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{pmatrix} \tag{2.15}$$

On appellera $\mathbf{M} = \mathbf{KR}[\mathbf{I} \mid \mathbf{t}]$ la matrice de projection de taille 3×4 modélisant les transformations géométriques permettant la projection. Pour résumer la matrice de projection admet en tout : 5 degrés de liberté pour la matrice intrinsèque (f,λ,s,u_0,v_0) et 6 dégrées de liberté pour les paramètres extrinsèques (3 pour la rotation sur chaque axe et 3 pour le vecteur de translation $\mathbf{t}$), soit au total 11 degrés de liberté.

2.2.4/ LA MODÉLISATION DES DISTORSIONS

L'utilisation d'un système optique composé de lentilles peut entraîner des distorsions de l'image invalidant la relation linéaire offerte par la matrice de projection. Ce type d'abbération géométrique entraîne une anamorphose notoirement visible sur les lignes droites présentes dans l'image. En effet, comme cela est mentionné la section 2.1, la transformation perspective conserve normalement les lignes droites, ce constat n'est plus vrai si le système optique induit des distorsions. Dans ce cas de figure, les lignes apparaissant sur l'image seront courbées.
Le modèle sténopé reste cependant acceptable lorsque le ratio entre l'épaisseur de la lentille et le rayon de courbure de ses faces est faible, on est alors en présence d'une lentille mince ne modifiant que peu la fiabilité du modèle. Les lentilles n'entrainant pas de distorsions sont désignées comme rectilinéaires. Ce n'est cependant pas le cas pour toutes les caméras, notamment en ce qui concerne celles équipées d'objectif à courte focale (par exemple une optique *fisheye*). Ces distorsions doivent être prises en compte si l'on souhaite réaliser des panoramas/reconstructions précises et réalistes. Ces imperfections peuvent être modélisées par une approximation radiale et tangentielle afin de corriger l'erreur de positionnement par rapport au modèle parfait. Si la projection idéale d'un point 3D dans le plan image $\mathbf{p}(x,y)$ ne correspond pas exactement à son projeté réel $\widehat{\mathbf{p}}(\widehat{x},\widehat{y})$, alors cette différence peut être compensée à l'aide d'un modèle de distorsion $D(\mathbf{p})$:

$$\widehat{\mathbf{p}} = \mathbf{p} + D(\mathbf{p}). \tag{2.16}$$

L'impact de la distorsion s'exercera différement en fonction de la position du pixel sur l'image.

Distorsion radiale Comme l'illustre la figure 2.5, la distorsion radiale provoque un déplacement de la position idéale des pixels vers l'intérieur ou l'extérieur de l'image depuis son centre. Les points localisés au centre de l'image étant par conséquent moins affectés. Cette altération est directement causée par la courbure de la lentille.
Une distorsion radiale entraînant le déplacement des points en direction de son centre est appelé distorsion en barillet arrondissant ainsi les bords de l'image et modifiant l'échelle des objets. C'est le type de distorsion que l'on retrouvera sur les caméras *fisheye* ou encore sur un judas de porte.
Dans le cas inverse si la distorsion "étire" l'image sur ses côtés on parlera de distorsion en cousinet. Ces distorsions sont symétriques par rapport au point principal de l'image et peuvent être modélisées à l'aide d'un modèle polynomial. Si $r = \sqrt{(\mathbf{c} - \mathbf{p})^2}$ correspond à la distance entre le point principal $\mathbf{c}(u_0, v_0)$ et la projection observée du point $\widehat{\mathbf{p}}$, la distorsion D peut s'exprimer de la manière suivante :

$$D = \begin{pmatrix} (x - u_0)(k_1 r^2 + k_2 r^4 + \cdots) \\ (y - v_0)(k_1 r^2 + k_2 r^4 + \cdots) \end{pmatrix}, \tag{2.17}$$

avec k_n les coefficients de distorsions. Généralement un polynôme de degré 4 est utilisé ($n = 2$) pour modéliser la distorsion toutefois un polynôme de degré 2 (un seul coefficient) est souvent suffisant [147]. De la même manière il est possible d'utiliser un polynôme de degré supérieur afin de rectifier l'image. La correction des images revient à redresser les lignes droites dans l'image (voir figure 2.6).

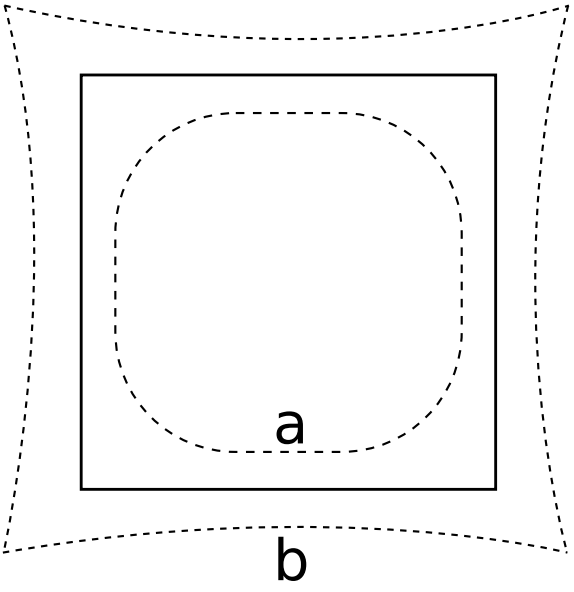


FIGURE 2.5 – Distorsion radiale (a) négative/en barillet (b) positive/en cousinet

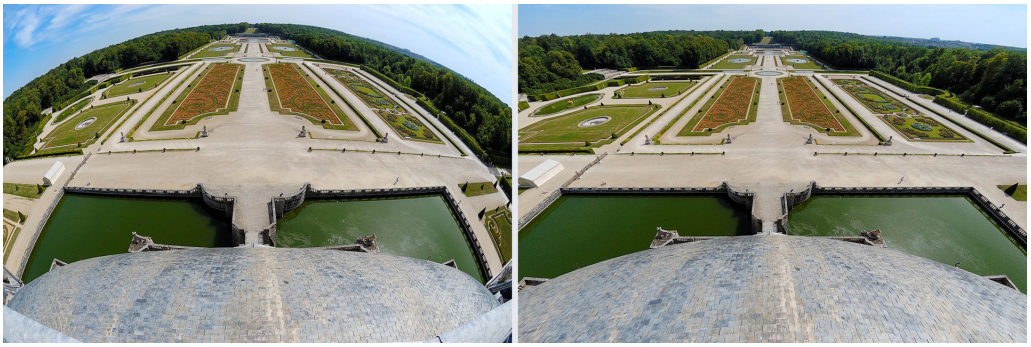


FIGURE 2.6 – Correction d'une image *fisheye*. A gauche l'image originale, à droite l'image rectifiée (Image de J.P. Roche)

Distorsion tangentielle La distorsion tangentielle est liée à un mauvais alignement des lentilles dans l'objectif, et peut, par exemple, se traduire par un positionnement non perpendiculaire à l'axe optique. Ce type de défaut est donc directement lié à la fabrication de l'objectif et apparaît souvent dans les caméras de mauvaise qualité. Elle peut également se modéliser sous forme d'un polynôme de degré n, pourtant cette imperfection est bien souvent considérée comme négligeable [109, 160].

2.3/ LA VISION OMNIDIRECTIONNELLE

Une caméra "conventionnelle" possède un champ de vue relativement limité, en général de l'ordre de 40° à 60°, ce qui constitue une contrainte dans bon nombre d'utilisations. Un capteur omnidirectionnel permet de pallier cet inconvénient en fournissant une vision panoramique de la scène allant jusqu'à 360°. Les capteurs omnidirectionnels sont d'ores et déjà largement utilisés dans différents domaines, comme la vidéo surveillance [137], la reconstruction 3D [22], la navigation robotique [151], la capture d'événements sportifs [8] ou encore la réalisation d'œuvres artistiques ...
Dans cette section nous aborderons à la fois la manière d'obtenir de telles images mais également les modèles de projections correspondants.

2.3.1/ LES DIFFÉRENTES MODALITÉS D'ACQUISITION D'IMAGES OMNIDIRECTIONNELLES

Les dispositifs existants permettant des acquisitions de ce type d'images peuvent être classés selon leur appartenance à trois principaux groupes : la vision panoramique à partir d'un ensemble d'images (polydioptrique), l'utilisation de lentilles spécifiques (*fisheye*) ou encore l'emploi d'un miroir convexe (catadioptrique). L'avantage principal offert par les caméras de type omnidirectionnel est de fournir une vue d'ensemble de la scène ce qui a pour effet, une réduction du coût du dispositif (moins de caméras mises en œuvre pour la

FIGURE 2.7 – Image de l'opéra de Paris acquisse avec un cylindrographe

surveillance d'une zone déterminée sauf dans le cas d'un système polydioptrique), une réduction du temps d'acquisition (une seule acquisition est nécessaire pour les caméras *fisheye* et catadioptriques). Cependant, le compromis temps d'acquisition/résolution doit être en concordance avec l'application souhaitée. De plus, l'inconvénient majeur découlant de l'utilisation de ces systèmes est souvent l'anamorphose (distorsion de l'image) induite par le procédé mis en oeuvre pour l'acquisition de l'image, voir figure 2.8(a). Cette forte distorsion radiale rend impossible l'utilisation du modèle sténopé généralement admis pour des caméras perspectives conventionnelles. On notera également une résolution spatiale non-uniforme sur l'ensemble de l'image et une prédisposition aux aberrations chromatiques avec les dispositifs évoqués.

2.3.1.1/ LES CAMÉRAS ROTATIVES

Il est possible d'obtenir des images panoramiques à partir d'une caméra en mouvement. Cette approche est d'ailleurs la première à avoir permis la capture d'une prise de vue panoramique avec le cylindrographe [132] puis l'invention brevetée des frères Lumières, le "Photorama" en 1900 (voir figure 2.7).
L'emploi d'un capteur CCD linéaire monté sur un axe de rotation vertical (pan) permet une acquisition relativement simple d'une image omnidirectionnelle [20, 76], la résolution spatiale de l'image sera ici déterminée par la vitesse de rotation de la caméra sur son axe. Le panorama peut également être obtenu à partir d'une caméra matricielle, les premières méthodes disponibles dans la littérature se basent sur une caméra admettant une rotation pure sur l'axe vertical, la mise en correspondance (*mozaicing*) permet alors l'obtention d'une image cylindrique de la scène. Il est possible de généraliser ce principe à un système possédant un nombre de degrés de liberté plus important comme une caméra PTZ [156, 164] permettant la création d'une image sphérique de l'ensemble de l'espace environnant avec une résolution très élevée (*mozaicing* utilisant plusieurs niveaux de zoom). L'utilisation des méthodes décrites ci-dessus a l'avantage de créer des images de très haute résolution. De nombreuses approches permettant une acquisition de ce type de cliché sont décrites dans l'ouvrage de Capel [37]. Ces différentes stratégies soulèvent cependant un grand nombre de problèmes liés à la mise en correspondance

FIGURE 2.8 – (a) Image obtenue à l'aide d'un objectif *fisheye* (b) lentille *fisheye*

d'images qui limitent l'utilisation de ce type de système à des acquisitions de scènes fixes et peuvent également nécessiter un temps d'acquisition important.

2.3.1.2/ LES CAMÉRAS *fisheye*

On appellera dans ce document caméra *fisheye* toute caméra équipée d'une lentille de type *fisheye* (ou œil de poisson en français, voir figure 2.8(b)). Ce type d'optique permet d'imposer une distance focale très courte et par conséquent un champ de vue très large pouvant atteindre 185°. L'acquisition de ce type d'image hémisphérique se fait une fois encore au détriment d'une forte distorsion radiale (dit en barillet) déformant les lignes droites vers l'extérieur de l'image (phénomène clairement visible sur la figure 2.8(a)). De plus, à l'instar des caméras catadioptriques, la résolution de l'image est plus forte en son centre qu'à sa périphérie. On notera cependant que l'usage de ce genre de capteur permet de se soustraire du "blind spot" central caractéristique des caméras catadioptriques (figure 2.10(a)).
Les optiques *fisheye* sont composées d'un ensemble de lentilles ne permettant pas de respecter un centre de projection unique, ce qui a pour conséquence de rendre théoriquement impossible la rectification de l'image [77]. Il existe cependant différents modèles permettant de caractériser ces distorsions [53]. Généralement il s'agit de modéliser la distorsion radiale par différentes transformations (polynomiales, logarithmiques, rationnelles).

2.3.1.3/ LES CAMÉRAS POLYDIOPTRIQUES

Les caméras polydioptriques sont quant à elles constituées d'un ensemble de caméras permettant de couvrir un champ de vue plus large. Les transformations inter-caméras sont généralement connues à l'aide d'un calibrage préalable du capteur permettant d'associer les images provenant de toutes les caméras afin de former l'image omnidirectionnelle. L'avantage avec ces systèmes de vision est qu'ils permettent l'acquisition d'images

panoramiques de très haute résolution et contrairement aux systèmes faisant intervenir des caméras rotatives ils sont tout à fait exploitables sur des robots mobiles. On peut toutefois nuancer la praticité de ces capteurs qui nécessitent une synchronisation entre les caméras suffisamment précise, une phase de calibrage plus contraignante ainsi qu'un coût élevé.
Les systèmes de vision polydioptriques peuvent être constitués de caméras perspectives, mais aussi d'un ensemble de caméras omnidirectionnelles, on parle alors de système poly-omnidirectionnelle [148]. De nombreux systèmes commerciaux tel que la Ladybug (voir figure 2.9), le Panono ou encore la Bublcam démocratisent l'utilisation de ces caméras non seulement pour les entreprises ("Google street view" reposait sur l'utilisation de la ladybug jusqu'en 2008) mais aussi pour les particuliers. De nombreux travaux portent actuellement sur la navigation de véhicules autonomes équipés de caméras polydioptriques [128, 107].

FIGURE 2.9 – (a)Image acquise avec un système multi-caméras (b) Le capteur LadyBug

2.3.1.4/ LES CAMÉRAS CATADIOPTRIQUES

Une caméra catadioptrique consiste en l'association d'une caméra et d'un miroir permettant ainsi d'élargir ou de modifier le champ de vue. Étymologiquement, on retrouve d'ailleurs le terme "dioptrique" qui est l'étude de la réfraction de la lumière (lentille) tandis que "catoptrique" concerne la réflexion des rayons (miroir). L'utilisation de miroirs convexes afin d'obtenir une vision étendue d'une scène n'est pas un concept nouveau et était déjà employé au 16ème siècle afin de surveiller les établissements d'orfèvreries ou bancaires. Cependant la première utilisation de ce type de capteur permettant d'acquérir une image omnidirectionnelle est très récente avec le dépôt de brevet de Rees en 1970 [145]. Celui-ci suggérait l'utilisation d'un miroir hyperboloïde afin d'obtenir une image à 360°. Les caméras catadioptriques se sont par la suite démocratisées en robotique pour leur capacité à visualiser une large zone avec une seule caméra et par conséquent sans nécessiter de synchronisation (contrairement à un système multi-caméra). On notera tout de même quelques inconvénients, notamment l'encombrement, le coût, ainsi que la résolution faible et non uniforme. Dans les sections suivantes, nous verrons que différentes formes de miroir peuvent être utilisées et que ce choix est prépondérant lorsqu'il est

question de modéliser la projection du capteur.

FIGURE 2.10 – (a) Image obtenue à l'aide d'une caméra catadioptrique (b) Différent type de caméras catadioptriques

2.3.2/ LES CAMÉRAS CATADIOPTRIQUES À POINT DE VUE UNIQUE

Les caméras centrales également désignées comme caméras à point de vue unique (PVU) regroupent toutes les caméras où les rayons lumineux convergent vers un seul et même point. C'est par exemple le cas d'une caméra perspective sans distorsion et respectant ainsi le modèle sténopé (voir figure 2.3). Cette contrainte est également respectée pour différents types de caméras omnidirectionnelles spécifiquement pensées pour faire converger les rayons de lumière en un point unique. C'est le cas pour certaines caméras catadioptriques où le choix de la caméra et la forme du miroir qui y est associé ont été spécialement conçus pour respecter cette caractéristique. Théoriquement toutes les formes de miroir dérivées de section de conique permettent de satisfaire le PVU, cependant elles ne sont pas toutes applicables en pratique. Notons également que pour respecter le PVU le foyer du miroir doit être confondu avec le centre optique de la caméra.
En 1997 Nayar et Baker[133] référence 4 configurations dont le dispositif est physiquement réalisable :

Parmi ces différents couples miroir/caméra, seules les deux configurations impliquant un miroir de forme convexe présentent une véritable utilité dans le cadre de la vision omnidirectionnelle. D'un point de vue pratique la configuration utilisant une caméra perspective est souvent préférée pour des raisons liées au coût et à l'encombrement du capteur. D'autres configurations de caméra catadioptrique à PVU sont théoriquement valides mais ne peuvent pas être mise en œuvre physiquement, nous ne les traiterons donc pas dans ce manuscrit de thèse. Cependant ces différentes configurations sont exposées dans l'article [133]. L'avantage de pouvoir modéliser la projection d'une caméra par un PVU est de permettre la rectification de l'image. De cette façon il est possible d'obtenir facilement une

Tableau 2.1 – Configurations de capteur à point de vue unique réalisables

Caméra	Miroir
Perspective	Convexe hyperbolique
Perspective	Plan
Perspective	Concave ellipsoïdale
Orthographique	Convexe parabolique

image perspective à partir d'une image complètement anamorphosée et par conséquent d'appliquer des méthodes initialement destinées aux caméras perspectives.

2.3.3/ MODÈLES DE PROJECTION DES CAMÉRAS CATADIOPTRIQUES CENTRALES

Nayar & Baker [133] se sont penchés sur la formation des images catadioptriques pour tous les cas respectant le PVU. Nous détaillerons ici les deux possibilités permettant à la fois un centre de projection unique et une vision panoramique. Les formules présentées ici sont reprises de [53, 177, 133].

Caméra hypercatadioptrique Il est ici question de la configuration comprenant un miroir hyperbolique et une caméra perspective possédant une distance focale f, comme représenté sur la figure 2.11. Nous considérons également que l'axe principal de la caméra passant par son centre optique $\mathbf{O}$ est aligné avec le foyer du miroir $\mathbf{O_m}$ et espacé d'une distance d. Le miroir quant à lui est caractérisé par l'équation polaire suivante :

$$\rho = \frac{p}{1 + ecos(\theta)}, \tag{2.18}$$

avec e l'exentricité et p le paramètre de l'hyperbole.
La projection d'un point dans le monde $\mathbf{P}$ peut donc se décrire en deux étapes, la première étant la projection de $\mathbf{P}$ sur la surface du mirroir ρ :

$$\mathbf{P_m} = \frac{\rho \mathbf{P}}{\| \mathbf{P} \|}. \tag{2.19}$$

La deuxième étape consiste en la projection du point miroir $\mathbf{P_m}(X_m, Y_m, Z_m)$ sur le plan image $\mathbf{p}(x, y)$:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d \end{bmatrix} \begin{pmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{pmatrix}. \tag{2.20}$$

Il est également essentiel que le centre optique de la caméra $\mathbf{O}$ corresponde au deuxième foyer de l'hyperbole, forçant donc la distance $d = \frac{2ep}{1-e^2}$. De plus nous savons que $cos(\theta) =$

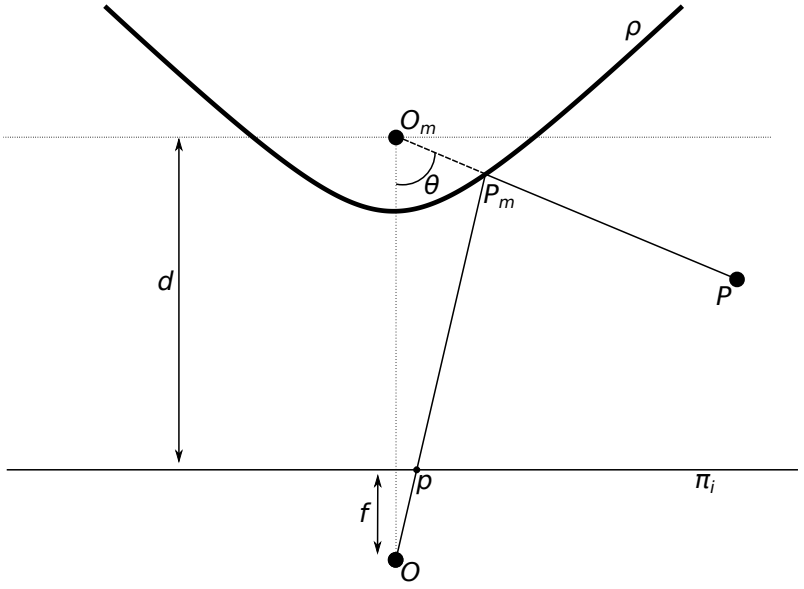


FIGURE 2.11 – Formation d'une image hypercatadioptrique

$\frac{Z}{\|\mathbf{P}\|}$, la projection peut donc s'exprimer (dans le repère pixellique) :

$$(x, y) = \left(\frac{\frac{1-e^2}{1+e^2} fX}{\frac{2e}{1+e^2}\sqrt{X^2+Y^2+Z^2}+Z} + u_0, \frac{\frac{1-e^2}{1+e^2} fY}{\frac{2e}{1+e^2}\sqrt{X^2+Y^2+Z^2}+Z} + v_0 \right). \tag{2.21}$$

Caméra paracatadioptique Cette configuration est assez différente de la précédente dans le sens où une caméra orthographique est utilisée. Ce type de caméra admet un point central situé à l'infini et par conséquent une focale $f = \infty$ et $d = \infty$ (voir figure 2.12). De plus la forme du miroir admet une exentricité $e = 1$, l'équation (2.21) peut en conséquence être reformulée :

$$(x, y) = \left(\frac{pX}{\sqrt{X^2+Y^2+Z^2}+Z} + u_0, \frac{pY}{\sqrt{X^2+Y^2+Z^2}+Z} + v_0 \right). \tag{2.22}$$

2.3.4/ MODÈLE DE PROJECTION DES CAMÉRAS *fisheye*

D'un point de vue théorique les caméras *fisheye* ne respectent pas le PVU, mais il a été prouvé par différents travaux tels que [177, 48], que le modèle de projection sphérique unifié constitue une très bonne approximation pour des applications telles que l'asservissement visuel [165] ou la navigation robotique. Généralement, les caméras légèrement non-centrales sont mieux décrites par le modèle sphérique que par tout autre modèle. D'autres modèles basés sur différents types de transformations permettent également de modéliser la projection d'une optique *fisheye*, l'ensemble de ces approches est résumé

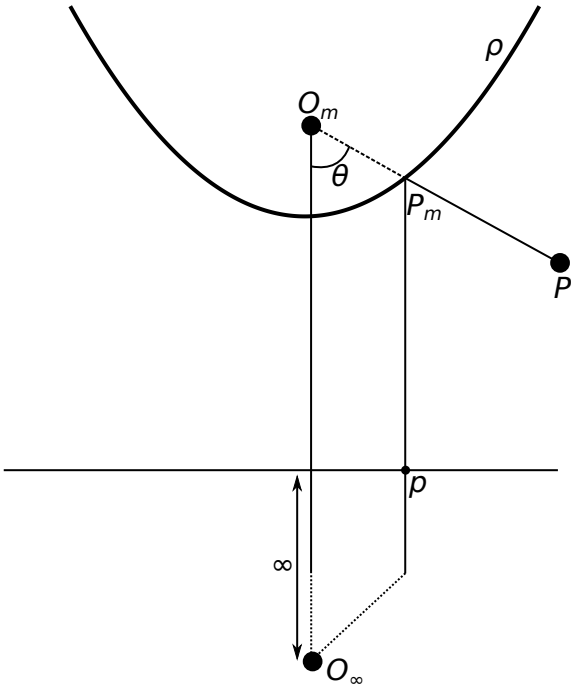


FIGURE 2.12 – Formation d'une image paracatadioptique

dans [48]. La plupart de ces modèles consistent à trouver la relation existante entre la distance euclidienne d'un point $\mathbf{p}(x, y)$ avec le centre optique $r = \sqrt{x^2 + y^2}$ et son équivalent dans le modèle sténopé $r' = \sqrt{x'^2 + y'^2}$ (voir figure 2.13). Fitzgibbon [67] propose par exemple le modèle suivant :

$$r' = k_1 \frac{r}{1 - k_2 r^2}. \tag{2.23}$$

L'approche la plus simple permet quant à elle d'établir une relation entre l'angle d'incidence θ et r :

$$r = f\theta. \tag{2.24}$$

L'article [92] fournit également de nombreuses informations sur le fonctionnement optique d'une caméra *fisheye*.

2.3.5/ LE MODÈLE SPHÉRIQUE UNIFIÉ

Nous avons montré que les caméras à point de vue unique forment un large panel comprenant les caméras perspectives, les caméras catadioptriques centrales et -dans une certaine mesure- les caméras équipées de lentille *fisheye*. Certaines de ces caméras offrent des images sujettes à de forte distorsions ce qui rend difficile, voire impossible, l'utilisation de méthodes développées pour des caméras perspectives conventionnelles ; prenons par exemple la géométrie épipolaire [149], la détection de contour [56] ou encore la mise en correspondance de points caractéristiques [49].
Comme cela a été évoqué dans les sections 2.3.3 et 2.3.4 des modèles spécifiques ont

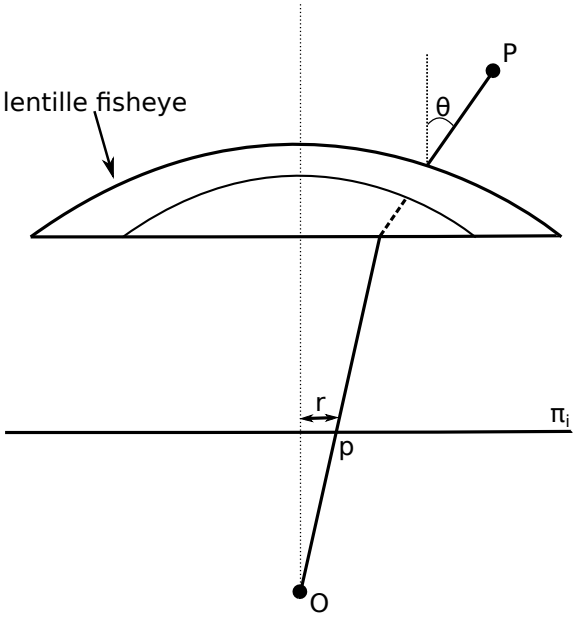


FIGURE 2.13 – Formation d'une image *fisheye*

été développés indépendamment pour chaque système. Il est cependant possible de représenter tous ces types de caméra à l'aide d'un modèle de projection stéréographique unique : le modèle sphérique unifié [72, 18]. Le processus de formation des images peut se traduire pour toute caméra centrale par une double projection sur une sphère gaussienne. Tout d'abord un point dans la scène $\mathbf{P}$ est projeté sur la sphère en $\mathbf{P_s}$. Cette première projection est suivie d'une seconde sur le plan image $\pi_\mathbf{i}$ formant ainsi le pixel $\mathbf{p_i}$. Cette projection part d'un point $\mathbf{O_c}$ situé au dessus du centre de la sphère. La distance l séparant ce point $\mathbf{O_c}$ et le centre de la sphère $\mathbf{O}$ modélise la distorsion radiale inhérente à la caméra utilisée, pour une caméra perspective sans distorsion cette distance est nulle. L'ensemble de cette projection est résumé dans la figure 2.14 et développé plus en détails dans les sous-sections suivantes.

Projection d'un point dans le monde sur le plan image Étape 1 : Un point 3D $\mathbf{P}$ est projeté sur la sphère :

$$\mathbf{P_s} = \frac{\mathbf{P}}{\| \mathbf{P} \|} = \begin{pmatrix} X_s \\ Y_s \\ Z_s \end{pmatrix} \tag{2.25}$$

Etape 2 : Projection stéréographique du point sphérique sur le plan image :

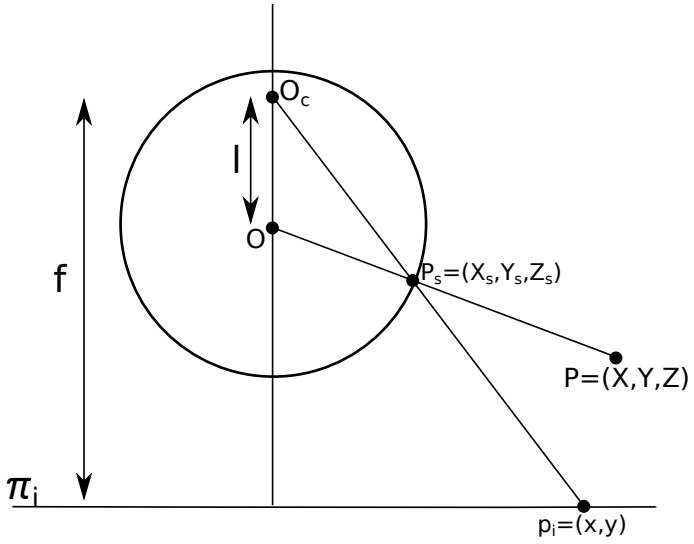


FIGURE 2.14 – Modèle sphérique unifié

$$x = \frac{X_s.f_x}{l + Z_s} + u_0 \tag{2.26}$$

$$y = \frac{Y_s.f_y}{l + Z_s} + v_0 \tag{2.27}$$

(2.28)

La projection globale se caractérise donc de la manière suivante :

$$\mathbf{p} = (x, y) = \left(\frac{X.f_x}{l\sqrt{X^2 + Y^2 + Z^2} + Z} + u_0, \frac{Y.f_y}{l\sqrt{X^2 + Y^2 + Z^2} + Z} + v_0 \right) \tag{2.29}$$

Re-projection d'un point image sur la sphère Afin de permettre un travail directement sur la sphère il est également important de maîtriser le modèle de projection inverse, c'est-à-dire la projection de l'image sur la sphère :

$$\begin{cases} Z_s = \frac{-2.l.\omega + \sqrt{(2.l.\omega)^2 - 4(\omega+1).(l^2.\omega - 1)}}{2(\omega+1)} \\ X_s = x_t(Z_s + l) \\ Y_s = y_t(Z_s + l) \end{cases}$$

, avec $\begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} \simeq \mathbf{K}^{-1}\mathbf{p}_i$ et $\omega = x_t^2 + y_t^2$.

Récapitulatif Toutes les caméras centrales permettent l'utilisation du modèle sphérique unifié, la différence majeure résidant dans la valeur du paramètre l. Les équations (2.21) modélisant les caméras hypercatadioptriques et (2.29) sont équivalentes si l'on pose : $l = \frac{2e}{1+e^2} < 1$ et $f_x = f_y = \frac{1-e^2}{1+e^2} f$.
Cette équivalence est également vraie pour le modèle de projection paracatadioptrique, puisque l'équation (2.22) est strictement similaire au modèle sphérique avec $l = 1$ et $f_x = f_y = p$.
En effet, pour les caméras catadioptriques les points projetés se situent derrière le plan image, ce qui force l a être compris entre 0 et 1 pour le cas paracatadioptrique tandis que $l = 1$ pour une caméra hypercatadioptrique. Cette distance sera nulle dans le cas d'une caméra perspective tandis que $l > 1$ pour les caméras *fisheye* [177, 48]. Ce modèle constitue également une très bonne approximation lorsqu'il est question de système multi-caméras où les centres optiques des caméras constituant le banc de vision sont suffisamment proches relativement à la profondeur de la scène. Dans ce cas de figure, la translation existante entre les caméras peut être négligée et le modèle sphérique s'appliquer [128, 104]. De la même manière il est possible d'appliquer ce modèle dans le cas de caméra rotative telles que les caméras de type PTZ. Le modèle sphérique unifié nous fournit donc un outil très polyvalent permettant d'homogénéiser le modèle de projection même dans le cas qui nous concerne, c'est-à-dire les systèmes de vision hétérogène.

2.3.6/ MODÈLE DE PROJECTION GÉNÉRIQUE

Dans le cas de caméras non-centrales (figure 2.15), ou pour des réseaux de caméras les modèles présentés précédemment ne permettent pas de caractériser la géométrie responsable de la formation de l'image. C'est d'ailleurs pour les systèmes multi-caméras que Pless [139] a introduit un modèle de caméra généralisé permettant de modéliser les projections ne respectant pas le point de vue unique. Dans cet article il définit de quelle manière un réseau de caméras centrales peut être exprimé comme une seule caméra non-centrale.
Ce modèle est particulièrement utile car il n'impose pas de contrainte dans la conception physique des capteurs catadioptriques. De plus, il permet de représenter des systèmes de vision hybride composés de n caméras centrales/non-centrales [89]. Dans ce modèle générique chaque pixel $\mathbf{p}$ se rapporte à un rayon exprimé sous forme de ligne de Plücker composée de 6 coordonnées afin de décrire la direction du rayon dans l'espace [140].
La mise en pratique de cette approche reste cependant plus complexe que les modèles décrits précédemment. Pour le calibrage d'une caméra non-centrale, nous renvoyons à [163].

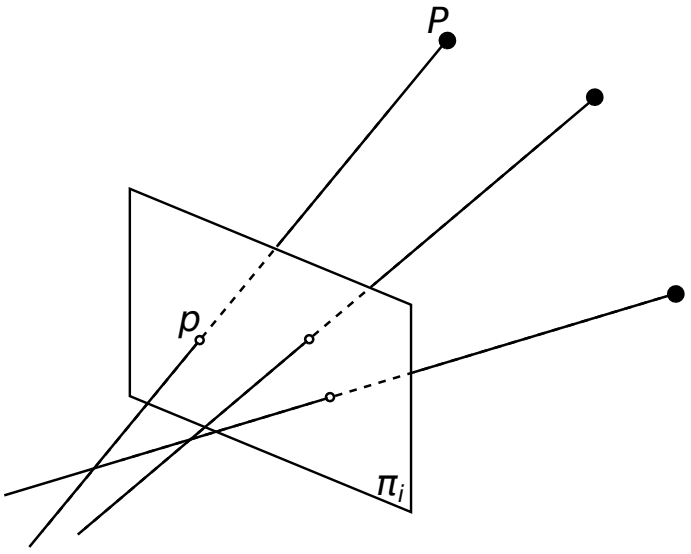


FIGURE 2.15 – Projection non centrale

2.4/ LA PROJECTION PLANAIRE

Lorsque la scène observée contient un plan, il est possible de caractériser la projection des points appartenant à ce plan sur $\pi_\mathbf{i}$ par une transformation homographique. Cette homographie s'exprimant sous forme d'une matrice $\mathbf{H}$ de taille 3×3 contient l'ensemble des informations concernant le plan et la pose de la caméra l'observant.

Plan à une distance finie Tout plan π dans la scène peut être caractérisé par son vecteur normal $\mathbf{n}$ et par la distance orthogonale d entre le plan et le repère de la scène. Un plan π est donc représenté par un vecteur de 4 éléments $\pi \simeq (\mathbf{n}, d)^T$. Un point 3D $\mathbf{P} = (X, Y, Z, 1)$ appartenant au plan π dans la scène respecte les deux relations suivantes :

$$\mathbf{n}^T \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = -d \quad \text{et} \quad \mathbf{p} \simeq \mathbf{KR} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \mathbf{Kt}. \tag{2.30}$$

L'une étant la projection de ce point dans l'image $\mathbf{p} \simeq \mathbf{MP}$, l'autre validant l'existence de ce point sur le plan $\pi^T \mathbf{P} = 0$ (voir figure 2.16). Si $d \neq 0$ il est possible de combiner les deux équations précédentes :

$$\mathbf{p} \simeq \mathbf{KR} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - \mathbf{Kt} \frac{\mathbf{n}^T}{d} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \tag{2.31}$$

ce qui revient à la relation compacte suivante : $p \simeq \mathbf{H}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ où l'homographie $\mathbf{H}$ est définie de la manière suivante :

$$\mathbf{H} = \mathbf{K}\left(\mathbf{R} - \mathbf{t}\frac{\mathbf{n}^T}{d}\right). \tag{2.32}$$

Cette matrice $\mathbf{H}$ constitue donc une transformation directe entre les points du plan et les points images. Nous verrons par la suite que cette propriété est très souvent exploitée surtout lorsqu'il est question de calibrer une caméra (c'est-à-dire de calculer ses paramètres intrinsèques).

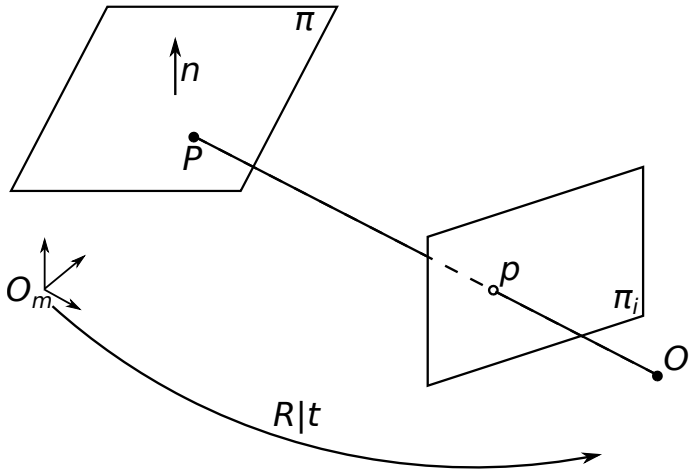


FIGURE 2.16 – Projection homographique

Plan à l'infini Cette relation existe également entre les points images et les points appartenant au plan à l'infini. Par exemple un point de fuite $\mathbf{P}_\infty = (X_\infty, Y_\infty, Z_\infty, 0)$ - projeté sur l'image en un point $\mathbf{p}$ - sera situé à une distance $d = \infty$ donc $\lim_{d\to\infty}(\mathbf{K}\mathbf{t}\frac{\mathbf{n}^T}{d}) = 0$, l'homographie à l'infinie $\mathbf{H}_\infty$ peut donc s'écrire :

$$\mathbf{H}_\infty \simeq \mathbf{K}\mathbf{R}. \tag{2.33}$$

Cette particularité est très intéressante puisque cette homographie ne dépend plus de la position de la caméra mais seulement de son orientation et de ses paramètres intrinsèques. Nous verrons dans cette thèse une méthode d'auto-calibrage prenant avantage de l'homographie à l'infini.

2.5/ LA GÉOMÉTRIE MULTI-VUES

Dans cette section nous traiterons de la géométrie multi-vues, c'est-à-dire des relations qui existent entre la projection des points de la scène dans plusieurs vues. Tout d'abord nous aborderons le cas faisant intervenir deux vues - aussi appelé géométrie épipolaire

- puis nous développerons les approches basées sur l'utilisation de tenseurs permettant de lier des amères visibles sur trois ou quatre vues. De façon à être le plus générique possible les explications ci-dessous concerneront le cas de caméra perspectives mais également de toutes les caméras à PVU ou assimilables (catadioptrique, *fisheye*, ...).

2.5.1/ GÉOMÉTRIE BI-FOCALE

La projection d'un point 3D $\mathbf{P}(X, Y, Z)$ sur les deux images π_1 et π_2 respectivement en $\mathbf{p_1}$ et $\mathbf{p_2}$ peut s'exprimer à l'aide des matrices de projection associées à chacune des caméras. Si le repère lié à la scène correspond au repère de la première caméra, on obtient les relations suivantes :

$$\begin{aligned} \mathbf{p_1} = \mathbf{M_1}[\mathbf{P}\ 1]^T \rightarrow \mathbf{p_1} = \mathbf{K_1}[\mathbf{I} \mid \mathbf{0}][\mathbf{P}\ 1]^T \\ \mathbf{p_2} = \mathbf{M_2}[\mathbf{P}\ 1]^T \rightarrow \mathbf{p_2} = \mathbf{K_2}[R_{12} \mid \mathbf{t_{12}}][\mathbf{P}\ 1]^T \end{aligned} \tag{2.34}$$

où $\mathbf{R_{12}}$ et $\mathbf{t_{12}}$ sont la rotation et la translation entre la première et la seconde caméra. Les matrices $\mathbf{K_1}$ et $\mathbf{K_2}$ étant quant à elles les matrices intrinsèques de la première et de la seconde caméra respectivement.
Remarquons que la figure 2.18 fait état de deux épipoles $\mathbf{e_{12}}$ et $\mathbf{e_{21}}$, ils correspondent respectivement à la projection du centre optique de la seconde caméra $\mathbf{O_2}$ sur le plan image de la première (π_1) et inversement. On obtient donc :

$$\mathbf{e_{12}} = \mathbf{K_1}\mathbf{R_{12}^T}\mathbf{t_{12}} \tag{2.35}$$

$$\mathbf{e_{21}} = \mathbf{K_2}\mathbf{t_{12}} \tag{2.36}$$

A noter que les épipoles ne vivent pas nécessairement dans l'image visible.
Les épipoles sont importants dans la géométrie multi-vue car ils caractérisent la pose relative de deux caméras entre elles.

2.5.1.1/ HOMOGRAPHIE ENTRE DEUX VUES

Au même titre qu'il existe une relation directe entre un plan et sa projection dans l'image, il existe une homographie liant la projection d'un plan sur deux prises de vue (voir figure 2.17). Prenons un point $\mathbf{P}$ appartenant au plan π, ce point est projeté sur les deux plans image respectivement aux points $\mathbf{p_1}$ et $\mathbf{p_2}$. Il existe une homographie $\mathbf{H_{12}}$ respectant :

$$\mathbf{p_2} \simeq \mathbf{H_{12}}\mathbf{p_1}, \tag{2.37}$$

avec cette fois une matrice d'homographie dépendante des paramètres intrinsèques des deux caméras $\mathbf{K_1}$ et $\mathbf{K_2}$:

$$\mathbf{H_{12}} \simeq \mathbf{K_2}(\mathbf{R} - \mathbf{t}\frac{\mathbf{n}^T}{d})\mathbf{K_1}^{-1}. \tag{2.38}$$

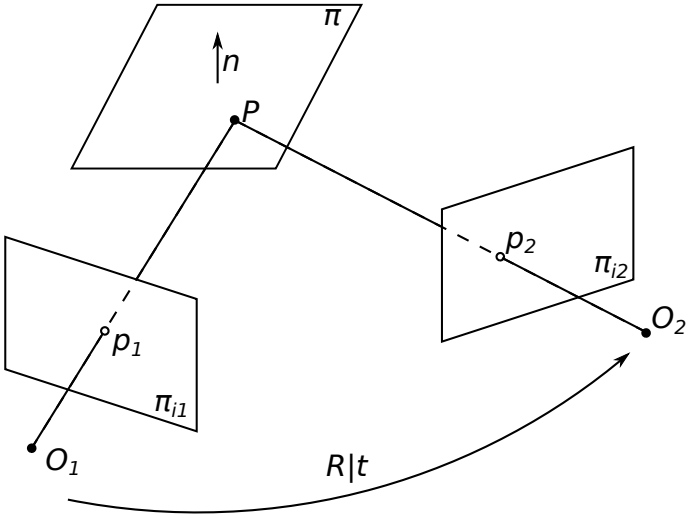


FIGURE 2.17 – Homographie entre deux vues

2.5.1.2/ LA GÉOMÉTRIE ÉPIPOLAIRE

La géométrie épipolaire correspond au modèle mathématique liant deux images d'une même scène capturée sous deux points de vue différents. La géométrie épipolaire se base sur l'intersection des plans images avec le plan épipolaire $\pi_{\mathbf{e}}$ formé par les centres focaux des caméras (la *baseline*) et le point 3D $\mathbf{P}$ projeté sur les deux images aux points de correspondance $\mathbf{p_1}$ et $\mathbf{p_2}$. En absence d'*a priori* sur la scène la position du point 3D $\mathbf{P}$ est inconnu entraînant une ambiguïté sur la position du point de correspondance de $\mathbf{p_1}$ dans l'autre image. On peut toutefois affirmer que celui-ci sera nécessairement localisé sur la ligne épipolaire $\mathbf{l_2}$ résultante, $\mathbf{p_1} \mapsto l_2$ (phénomène illustré figure 2.18). Tous les plans épipolaires possibles passent par les épipoles $\mathbf{e_{12}}$ et $\mathbf{e_{21}}$ formés par l'intersection de la baseline et des plans images. Cette contrainte est particulièrement utile lorsqu'on

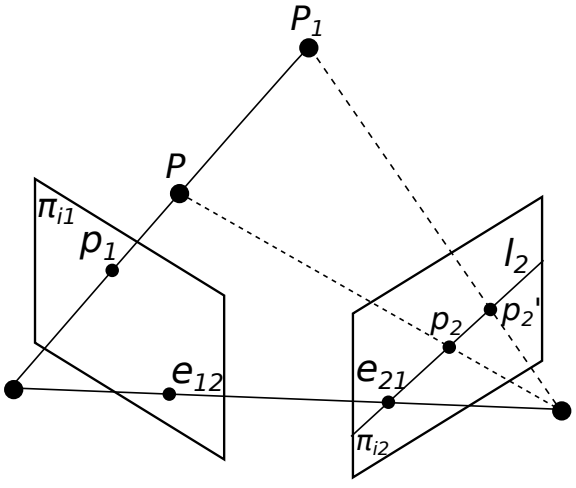


FIGURE 2.18 – Géométrie épipolaire

doit par exemple trouver des points de correspondance entre images [180]. Mais nous verrons également qu'elle est essentielle dans le calcul de pose des caméras ou pour la rectification d'image [83].

2.5.1.3/ LA MATRICE FONDAMENTALE

La géométrie épipolaire peut être formalisée de manière mathématique sous forme d'une matrice 3x3, appelée matrice fondamentale. Nous allons en reprendre le développement afin de mettre en avant ses propriétés et souligner l'intérêt d'un tel formalisme. Les équations (2.34) représentent la projection d'un point 3D sur deux caméras, elles peuvent également s'exprimer sous la forme suivante :

$$\mathbf{K_1}^{-1}\mathbf{p_1} \simeq \mathbf{P}, \tag{2.39}$$

$$\mathbf{K_2}^{-1}\mathbf{p_2} \simeq \mathbf{R_{12}P} + \mathbf{t_{12}}, \tag{2.40}$$

avec $\mathbf{p_1}$ en coordonnées homogènes $(x_1, y_1, 1)^T$ et $\mathbf{p_2} = (x_2, y_2, 1)^T$. En substituant l'équation (2.39) dans l'équation (2.40), on obtient :

$$\mathbf{K_2}^{-1}\mathbf{p}_2 \simeq \mathbf{R_{12}K_1}^{-1}\mathbf{p}_2 + \mathbf{t_{12}}. \tag{2.41}$$

On notera qu'à cette étape la relation entre les deux points de correspondance ne dépend plus du tout de la position du point 3D mais simplement du positionnement des caméras. L'équation précédente est simplifiable à l'aide de la matrice anti-symétrique de $\mathbf{t_{12}}$ ($\mathbf{t}_{[\times]} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$ pour tout vecteur $\mathbf{t} = [t_x\ t_y\ t_z]^T$) :

$$\mathbf{t_{12[\times]}K_2}^{-1}\mathbf{p2} \simeq \mathbf{t_{12[\times]}R_{12}K_1}^{-1}\mathbf{p1}, \tag{2.42}$$

une multiplication par $\mathbf{p_2}^T\mathbf{K_2}^{-T}$ donne la relation appelée "contrainte épipolaire" :

$$\mathbf{p_2}^T\mathbf{K_2}^{-T}\mathbf{t_{12[\times]}R_{12}K_1}^{-1}\mathbf{p_1} = 0, \tag{2.43}$$

$$\mathbf{p_2}^T\mathbf{F_{12}p_1} = 0, \tag{2.44}$$

avec $\mathbf{F_{12}} = \mathbf{K_2}^{-T}\mathbf{t_{12[\times]}R_{12}K_1}^{-1}$ de taille 3×3. Cette matrice de rang 2 a de nombreuses propriétés :

Correspondances Si deux points $\mathbf{p_1}$ et $\mathbf{p_2}$ sont homologues alors la contrainte épipolaire $\mathbf{p_2}^T\mathbf{F_{12}p_1} = 0$ est respectée. Cela revient à dire que ces deux points sont coplanaires sur le même plan épipolaire.

Transposition Si $\mathbf{F_{12}}$ est la matrice fondamentale exprimant la contrainte épipolaire existante entre la caméra 1 et la caméra 2, alors sa transposée définit la relation liant la caméra 2 à la caméra 1 : $\mathbf{F_{12}}^T = \mathbf{F_{21}}$.

Ligne épipolaire Il est très simple de calculer l'équation d'une ligne épipolaire $\mathbf{l}_2$ -dans la seconde image- formée par $\mathbf{p_1}$ à l'aide de la matrice fondamentale : $\mathbf{l_2} = \mathbf{F_{12}p_1}$. De la même manière $\mathbf{l_1} = \mathbf{F_{12}}^T\mathbf{p_2}$ et $\mathbf{l_1} = \mathbf{p_2F_{12}}$.

Les épipoles Toutes les lignes épipolaires passent par les épipoles, ce qui signifie que pour tout point $\mathbf{p_1}$ la condition $\mathbf{e_{21}}^T\mathbf{F_{12}p_1} = 0$ est toujours satisfaite. Donc $\mathbf{e_{21}}^T\mathbf{F_{12}} = 0$, en d'autres termes l'épipole dans la seconde image $\mathbf{e_{21}}$ est le vecteur nul gauche de $\mathbf{F_{12}}$.

CALCUL DE LA MATRICE FONDAMENTALE

Il est possible à partir de la contrainte épipolaire $\mathbf{p_2}^T\mathbf{F_{12}p_1} = 0$ de résoudre linéairement les entrées de la matrice fondamentale. Si on considère les éléments composants $\mathbf{F}$ de la manière suivante :

$$\mathbf{F} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \tag{2.45}$$

et en développant la contrainte épipolaire, on obtient l'équation ci-dessous :

$$x_2x_1f_{11} + x_2y_1f_{12} + x_2f_{13} + y_2x_1f_{21} + y_2y_1f_{22} + y_2f_{23} + x_1f_{31} + y_1f_{32} + f33 = 0 \tag{2.46}$$

sous forme vectorielle on peut réécrire :

$$(x_2x_1, x_2y_1, x_2, y_2x_1, y_2y_1, y_2, x_1, y_1, 1)\mathbf{f} = 0 \tag{2.47}$$

avec $\mathbf{f} = (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33})^T$. On peut ainsi résoudre $\mathbf{F}$ sous forme d'un problème linéaire de type $\mathbf{Af} = 0$ à l'aide d'un ensemble de n points :

$$\mathbf{Af} = \begin{bmatrix} x_2^1x_1^1 & x_2^1y_1^1 & x_2^1 & y_2^1x_1^1 & y_2^1y_1^1 & y_2^1 & x_1^1 & y_1^1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_2^nx_1^n & x_2^ny_1^n & x_2^n & y_2^nx_1^n & y_2^ny_1^n & y_2^n & x_1^n & y_1^n & 1 \end{bmatrix}\mathbf{f} = 0. \tag{2.48}$$

Chaque point de correspondance fournissant une équation, la méthode la plus élémentaire nécessite un ensemble minimum de huit points afin de résoudre $\mathbf{f}$ car $\mathbf{f}$ est définit à l'échelle près.
La résolution de la matrice fondamentale se révèle cependant plus complexe en pratique puisqu'une normalisation des coordonnées des points de correspondance est nécessaire afin d'assurer la stabilité numérique de la méthode. Il faut en outre tenir compte des cas

dégénérés tel qu'un choix de points coplanaires ou encore une rotation pure de la caméra. L'algorithme le plus basique est l'algorithme des huit points normalisés de Longuet [117]. Hartley y a par la suite ajouté une contrainte en forçant la singularité de $\mathbf{F}$ après la résolution linéaire ayant pour effet de garantir la "validité" de la matrice fondamentale [81]. Le calcul de la matrice fondamentale a été très largement étudié, on retrouve donc un grand nombre d'approches différentes dans la littérature telles que des estimations robustes à l'aide LMedS (least median of square) [180], RANSAC [66, 87, 64] ou ses variantes [129, 178]. On notera également l'existence d'algorithmes forçant la contrainte $det(\mathbf{F}) = 0$ réduisant le nombre de points nécessaire à 7 [95, 80]. Ces approches fournissent cependant trois résultats possibles, et leurs implémentations reste plus complexe que l'algorithme des 8 points.

2.5.1.4/ LA MATRICE ESSENTIELLE

La matrice essentielle est l'équivalent de la matrice fondamentale dans le cas calibré, c'est-à-dire que les points de correspondance utilisés ne sont plus exprimés dans le plan image $\pi_\mathbf{i}$ mais dans le plan rétinien. Si l'on considère un point $\mathbf{p}$ vivant dans le plan image alors son expression dans le plan rétinien $\widehat{\mathbf{p}}$ s'exprimera de la manière suivante :

$$\widehat{\mathbf{p}} = \mathbf{K}^{-1}\mathbf{p}. \tag{2.49}$$

La matrice essentielle notée $\mathbf{E_{12}}$ permet de décrire la géométrie entre deux images provenant de deux caméras calibrées et dont les points de correspondance sont respectivement $\widehat{\mathbf{p_1}}$ et $\widehat{\mathbf{p_2}}$. La contrainte épipolaire peut alors être écrite de la manière suivante :

$$\widehat{\mathbf{p_2}}^T\mathbf{E_{12}}\widehat{\mathbf{p_1}} = 0. \tag{2.50}$$

Si l'on reprend l'équation (2.43), on remarque une relation simple et directe entre la matrice fondamentale et la matrice essentielle :

$$\mathbf{F_{12}} = {\mathbf{K_2}}^{-T}\mathbf{E_{12}}{\mathbf{K_1}}^{-1}. \tag{2.51}$$

On en déduit :

$$\mathbf{E_{12}} = \mathbf{t_{12[\times]}}\mathbf{R_{12}}. \tag{2.52}$$

La matrice essentielle contient donc la transformation rigide entre deux caméras, ce qui à l'échelle près lui confère un total de 5 degrés de liberté. Cela en fait un outil indispensable pour estimer le déplacement d'une caméra. L'approche permettant l'estimation de la matrice essentielle la plus basique repose sur l'utilisation de six points mis en correspondance, une approche minimale utilisant 5 points existe également [114] et permet notamment de gérer certains cas dégénérés tel que celui où les points mis en correspondance sont coplanaires.

CALCUL DE LA MATRICE ESSENTIELLE

Nous présentons ici la résolution linéaire des entrées d'une matrice essentielle à l'aide de 6 points de correspondance respectant la contrainte épipolaire $\widehat{\mathbf{p_2}}^T \mathbf{E_{12}} \widehat{\mathbf{p_1}} = 0$. Avec les points $\widehat{\mathbf{p_1}} = (x_1, y_1, z_1)^T$ et $\widehat{\mathbf{p_2}} = (x_2, y_2, z_2)^T$. La résolution du système peut s'écrire :

$$\mathbf{Ae} = \begin{bmatrix} x_2^1 x_1^1 & x_2^1 y_1^1 & x_2^1 z_1^1 & y_2^1 x_1^1 & y_2^1 y_1^1 & y_2^1 z_1^1 & x_1^1 z_2^1 & y_1^1 z_2^1 & z_2^1 z_1^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_2^n x_1^n & x_2^n y_1^n & x_2^n z_1^n & y_2^n x_1^n & y_2^n y_1^n & y_2^1 n z_1^n & x_1^n z_2^n & y_1^n z_2^n & z_2^n z_1^n \end{bmatrix} \mathbf{e} = 0 \tag{2.53}$$

EXTRACTION DES PARAMÈTRES EXTRINSÈQUES $\mathbf{R}$ ET $\mathbf{t}$

Le calcul de $\mathbf{E_{12}}$ n'est pas suffisant pour déterminer le déplacement de la caméra, pour cela il est nécessaire d'en extraire les paramètres extrinsèques. La méthode que nous présentons ici est issue de [87]. Tout d'abord, considérons la décomposition en valeurs singulières de $\mathbf{E_{12}}$:

$$\mathbf{E_{12}} = \mathbf{U\Sigma V}^T. \tag{2.54}$$

Avec la matrice diagonale $\mathbf{\Sigma}$ égale à $\begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{bmatrix}$, puisque la matrice essentielle est de rang 2 (ce qui sous-entend l'existence d'une valeur singulière nulle et deux valeurs singulières non nulles).

Extraction de la translation $\mathbf{t}_{12}$:
Sachant que :

$$\mathbf{E_{12}t_{12}} = 0, \tag{2.55}$$

la translation correspond à la troisième colonne de la matrice orthogonale $\mathbf{U}$ (position de la valeur singulière nulle) :

$$\mathbf{t_{12}} \simeq \pm \mathbf{U}(0, 0, 1)^T. \tag{2.56}$$

Cette translation est cependant estimée à l'échelle près avec un signe indéterminé.

Extraction de la rotation :
Pour la résolution de la rotation, deux résultats sont possibles $\mathbf{R_{12}}^1$ et $\mathbf{R_{12}}^2$:

$$\mathbf{R_{12}}^1 = \mathbf{UWV}^T, \tag{2.57}$$

$$\mathbf{R_{12}}^2 = \mathbf{UW}^T\mathbf{V}^T, \tag{2.58}$$

avec une matrice $\mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Par conséquent, il existe quatre matrices de projection

possible pour la caméra 2 (considérant la caméra 1 comme référence), cependant une seule correspond réellement au déplacement de la caméra (à l'échelle). Pour pouvoir effectuer une reconstruction de l'environnement il est donc indispensable de déterminer quelle configuration est correcte parmi ces quatre possibilités :

$$\mathbf{M_2^1} = [\mathbf{R_{12}}^1 \mid \mathbf{t_{12}}], \tag{2.59}$$

$$\mathbf{M_2^2} = [\mathbf{R_{12}}^1 \mid -\mathbf{t_{12}}], \tag{2.60}$$

$$\mathbf{M_2^3} = [\mathbf{R_{12}}^2 \mid \mathbf{t_{12}}], \tag{2.61}$$

$$\mathbf{M_2^4} = [\mathbf{R_{12}}^2 \mid -\mathbf{t_{12}}]. \tag{2.62}$$

Pour ce faire, il est essentiel d'employer la contrainte dite de chiralité. Cette contrainte assure que les points reconstruits sont bien une solution physiquement possible par triangulation, en d'autres termes que les points 3D reconstruits existent devant les deux caméras. On peut ainsi rejeter les solutions physiquement impossibles [87].

2.5.1.5/ MATRICE FONDAMENTALE OMNIDIRECTIONNELLE/HYBRIDE

Nous avons montré dans la section 2.3.5 que le modèle de projection sphérique est particulièrement adapté pour tous les capteurs respectant le PVU, en revanche on notera que la projection sphérique ne peut pas directement être exprimée sous forme linéaire comme cela est le cas pour une caméra perspective décrite par le modèle sténopé. L'absence de matrice de projection linéaire pose un problème dans le sens où la matrice fondamentale classique n'est plus valide et doit être adaptée. En effet, sur des images omnidirectionnelles les lignes épipolaires deviennent des coniques épipolaires.
Pour résoudre ce problème, la plupart des approches se basent sur l'utilisation de coordonnées augmentées (*lifted coordinates*) permettant d'exprimer la matrice fondamentale de manière linéaire à l'aide des surfaces de Veronese, plus adaptées à l'étude des coniques. Cette approche a d'abord été employée par Geyer et Daniilidis afin de définir une matrice fondamentale adaptée à la géométrie des capteurs paracatadioptriques [73]. Par la suite Claus et Fitzgibbon [41] ont également proposé une méthode permettant de décrire la géométrie épipolaire sur des caméras *fisheye* à l'aide d'un nouveau modèle de projection, le calcul de la "matrice fondamentale augmentée" proposée s'appuie sur l'appariement de 36 points de correspondance. Des travaux plus récents menées par Micusik *et al*. [130] ont permis de généraliser cette approche à toutes les caméras respectant le PVU. Par la suite, d'autre travaux initiés par Sturm [162] concernant les systèmes de vision hybride sont apparus afin de permettre la mise en correspondance d'images omnidirectionnelles et d'images perspectives à l'aide d'une matrice fondamentale adaptée à la géométrie de chacune des caméras [143, 21, 17]. L'utilisation des surfaces de Veronese permettant d'obtenir une matrice de projection linéaire pour le modèle sphérique est également utilisée pour le calibrage des systèmes de vision omndirectionnels à PVU.

2.5.1.6/ STÉRÉO OMNIDIRECTIONNELLE/HYBRIDE CALIBRÉ

Dans cette thèse, nous traiterons essentiellement du cas de caméras omnidirectionnelles calibrées (dont les paramètres intrinsèques sont connus), dans ces circonstances la géométrie épipolaire - et toute la géométrie projective en générale- est préservée par l'utilisation du modèle sphérique unifié. La matrice essentielle garde donc sa structure, c'est-à-dire une matrice $\mathbf{E}$ de taille 3×3 composée de la rotation et de la matrice anti-symétrique issue du vecteur de translation. Les lignes épipolaires sur les plans images (comme cela est illustré dans la figure 2.18) deviennent des cercles épipolaires (C_1 et C_2) sur les sphères représentant les caméras car il s'agit de l'intersection du plan épipolaire avec les sphères susmentionnées (voir figure 2.19). Les points de correspondance détectés sur chaque image peuvent alors être reprojetées sur leur sphère respective $\mathbf{p_1} \rightarrow \mathbf{P_{S1}}$ et $\mathbf{p_2} \rightarrow \mathbf{P_{S2}}$ afin de calculer linéairement les composants de $\mathbf{E_{12}}$ à l'aide de la méthode présentée dans la section 2.5.1.4.
Cette représentation est particulièrement avantageuse dans le sens où l'emploi des méthodes initialement destinées à la géométrie multi-vues conventionnelles sont applicables sans autres modifications particulières. De nombreux travaux ont déjà tiré avantage de cette géométrie, pour la rectification d'images omnidirectionnelles, la navigation robotique, la vidéo surveillance ...

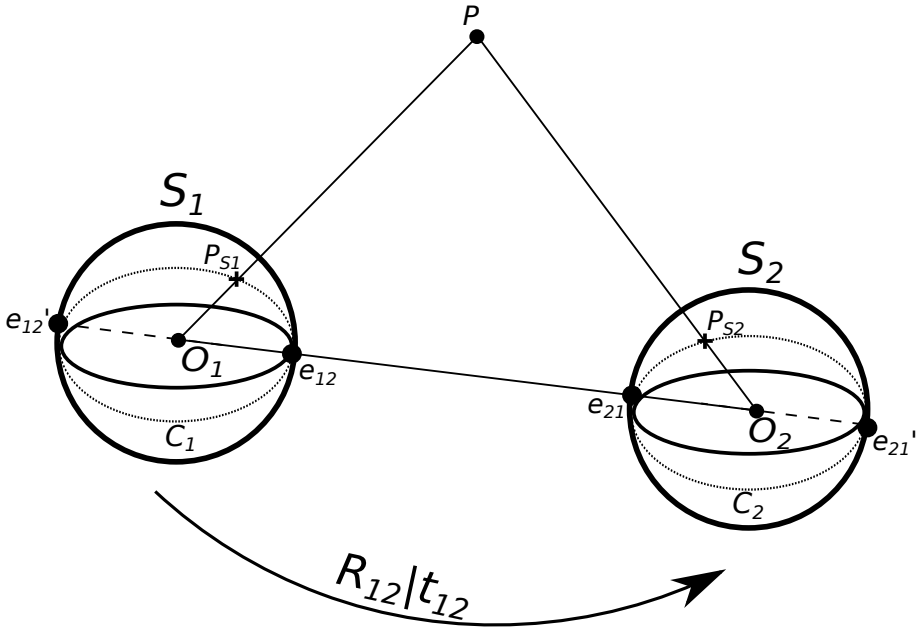


FIGURE 2.19 – Géométrie épipolaire avec le modèle spérique

2.5.1.7/ LA CONTRAINTE ÉPIPOLAIRE GÉNÉRALISÉE

Il est également possible d'étendre la contrainte épipolaire au modèle de vision généralisé déjà évoqué dans la section 2.3.6. Dans ce cas si un point image $\mathbf{p_1}$ associé à la ligne

de Plücker $\mathbf{L_1}$ est en correspondance sur une autre vue avec un point $\mathbf{p_2}$ dont le rayon est exprimé par $\mathbf{L_2}$, alors la contrainte épipolaire généralisée liant ces deux points peut s'écrire :

$$\mathbf{L_2}^T \begin{bmatrix} \mathbf{E_{12}} & \mathbf{R_{12}} \\ \mathbf{R_{12}} & \mathbf{0} \end{bmatrix} \mathbf{L_1}^T = 0. \tag{2.63}$$

Ce que nous appelons ici la matrice essentielle généralisée $\mathbf{E_{G12}}$ est la matrice de taille 6×6 liant les deux lignes de Plücker $\mathbf{L_1}$ et $\mathbf{L_2}$. Notons que la matrice essentielle "classique" décrite précédemment ne permet qu'une estimation du mouvement à l'échelle près, tandis que la matrice essentielle généralisée peut permettre une estimation du mouvement à l'échelle métrique. Il existe d'ailleurs plusieurs méthodes linéaires permettant le calcul de $\mathbf{E_{G12}}$ dont la principale nécessite la mise en correspondance de 17 points entre deux vues [139], ce qui est difficilement applicable en pratique si une estimation robuste est requise. Une autre approche ne nécessitant que 6 points de correspondance existe également, elle a néanmoins le désavantage d'offrir 64 solutions possibles. Toutefois, de nombreuses variantes nécessitant moins de points à l'aide d'informations additionnelles permettent de résoudre cette matrice essentielle. Par exemple, dans [89] les auteurs utilisent l'estimation de la rotation provenant de la centrale inertielle d'un drône afin de résoudre la translation métrique à l'aide de seulement 3 points.

2.5.1.8/ TRIANGULATION

La triangulation en vision artificielle est le processus permettant de déterminer la position d'un point 3D à partir de la projection de ce point sur deux images ou plus, on peut également parler de reconstruction. La triangulation permettant une reconstruction métrique ou euclidienne nécessite à la fois une connaissance des paramètres intrinsèques et extrinsèques des caméras. Le principe de la triangulation est trivial dans la mesure où l'on cherche simplement à obtenir l'intersection des lignes de vues passant par les points de correspondance. Cependant les rayons s'intersectent rarement en pratique particulièrement en présence de bruit, de distorsion ou simplement due à la limitation de la résolution du capteur. C'est pour ces différentes raisons que de nombreuses méthodes permettent de déterminer la position optimale du point reconstruit. Nous nous contenterons ici de présenter la méthode la plus employée, la triangulation linéaire [84].
Un point dans la scène $\mathbf{P}$ se projette sur deux images aux points $\mathbf{p_1} = \mathbf{M_1P}$ et $\mathbf{p_2} = \mathbf{M_2P}$. Le facteur d'échelle inhérent à la projection perspective peut être éliminé par un produit vectoriel. Par exemple pour le point $\mathbf{p_1}$, nous avons la relation $\mathbf{p_1} \times \mathbf{M_1P} = 0$, que nous pouvons également réécrire sous la forme suivante :

$$x_1(m_1^{3T}\mathbf{P}) - (m_1^{1T}\mathbf{P}) = 0 \tag{2.64}$$
$$y_1(m_1^{3T}\mathbf{P}) - (m_1^{2T}\mathbf{P}) = 0 \tag{2.65}$$
$$x_1(m_1^{2T}\mathbf{P}) - y_1(m_1^{1T}\mathbf{P}) = 0 \tag{2.66}$$

Avec m_1^n la $n^{\text{ème}}$ ligne de la matrice de projection $\mathbf{M}_1$. Cette série d'équations peut être réécrite sous forme de problème linéaire $\mathbf{AP} = 0$, avec :

$$A = \begin{bmatrix} x_1 m_1^{3T} - m_1^{1T} \\ y_1 m_1^{3T} - m_1^{2T} \\ x_2 m_2^{3T} - m_2^{1T} \\ y_2 m_2^{3T} - m_2^{2T} \end{bmatrix}. \tag{2.67}$$

Ce qui fait un total de deux équations par images pour chaque point homologue afin de résoudre les quatre inconnues de $\mathbf{P}$ exprimées en coordonnées homogènes.
Une comparaison des approches de triangulations les plus courantes est proposée par Hartley et Sturm dans [86].

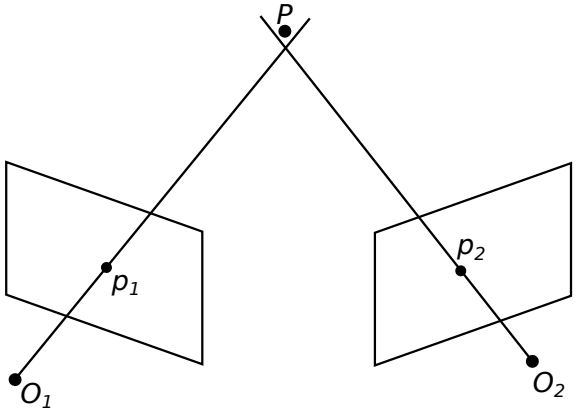


FIGURE 2.20 – Triangulation en situation réelle

2.5.2/ TENSEUR TRI/QUADRI-FOCAL

Il existe différentes manières de représenter les relations existantes entre plusieurs vues, la manière la plus directe peut s'exprimer à l'aide des matrices de projections, ici entre quatre vues :

$$\underbrace{\begin{bmatrix} \mathbf{M_1} & \mathbf{p}_1 & & & \\ \mathbf{M_2} & & \mathbf{p_2} & & \\ \mathbf{M_3} & & & \mathbf{p_3} & \\ \mathbf{M_4} & & & & \mathbf{p_4} \end{bmatrix}}_{\mathbf{C}} \begin{bmatrix} \mathbf{P_m} \\ -\lambda_1 \\ -\lambda_2 \\ -\lambda_3 \\ -\lambda_4 \end{bmatrix} = 0, \tag{2.68}$$

où $\lambda_{1\cdots4}$ sont des facteurs d'échelles. La matrice $\mathbf{C}_{12\times8}$ présente dans l'équation (2.68) admet une solution non nulle (rank($\mathbf{C}$)≤ 7) ce qui implique que tous les mineurs (déterminants de ses sous-matrices carrées) de taille 8×8 sont égaux à zéro. C'est ce déterminant qui définit les relations bi/tri/quadri-linéaires liant les points de correspondance entre les vues. Le choix de ce mineur - c'est-à-dire des 8 lignes au sein de la matrice

$\mathbf{C}$ le constituant- est donc prépondérant dans la construction de la contrainte, plusieurs configurations sont possibles :

– si l'on choisit seulement une ligne par caméra on aboutira à une relation trilinéaire
– si l'on choisit seulement une ligne pour deux caméras la contrainte résultante revient à une contrainte épipolaire (aussi appelée contrainte bifocale)
– finalement si deux lignes par caméras sont prises en compte, il s'agira alors d'une relation quadrilinéaire

Nous proposons ici un exemple de mineur respectant les contraintes quadrilinéaires :

$$det\begin{bmatrix} \mathbf{M}_1^1 & x_1 & & & \\ \mathbf{M}_1^2 & y_1 & & & \\ \mathbf{M}_2^1 & & x_2 & & \\ \mathbf{M}_2^2 & & y_2 & & \\ \mathbf{M}_3^1 & & & x_3 & \\ \mathbf{M}_3^2 & & & y_3 & \\ \mathbf{M}_4^1 & & & & x_4 \\ \mathbf{M}_4^2 & & & & y_4 \end{bmatrix} = 0, \tag{2.69}$$

où $\mathbf{M}^i$ correspond à la $i^{\text{ème}}$ ligne de $\mathbf{M}$ et x_j, y_j les coordonnées en x et y du point sur la $j^{\text{ème}}$ image.

Ces relations linéaires peuvent être reformulées sous forme de tenseurs, aussi appelés tenseurs bi/tri/quadri-focal. Ils permettent de mapper différentes primitives géométriques telles que les points ou les lignes entre chacune des vues de manière directe et élégante. Si les poses des caméras sont connues le calcul du tenseur tri/quadri-focal est trivial et peut s'exprimer de la manière suivante, pour les tenseur tri-focaux :

$$\mathbf{T}^{i,j,k} = det([\mathbf{M}_1^i\ \mathbf{M}_2^j\ \mathbf{M}_3^k]^T), \tag{2.70}$$

pour les tenseurs quadrifocaux :

$$\mathbf{Q}^{i,j,k,l} = det([\mathbf{M}_1^i\ \mathbf{M}_2^j\ \mathbf{M}_3^k\ \mathbf{M}_4^l]^T). \tag{2.71}$$

$\mathbf{T}$ correspond au tenseur tri-focal de taille $3 \times 3 \times 3$ (27 entrées) et $\mathbf{Q}$ au tenseur quadri-focal de taille $3 \times 3 \times 3 \times 3$ (81 entrées).

Ce type de tenseur fournit des relations géométriques intéressantes -entre lignes et/ou points de chaque caméra- appelés transfert. Le transfert ligne-ligne-ligne-ligne dans le cas d'un tenseur quadri-focal peut par exemple s'exprimer ainsi :

$$\mathbf{Q}^{i,j,k,l}\mathbf{l}_1^i\mathbf{l}_2^j\mathbf{l}_3^k\mathbf{l}_4^l = 0, \tag{2.72}$$

où l_i^v est la $i^{\text{ème}}$ ligne de correspondance dans la $v^{\text{ème}}$ vue. Cette relation peut être étendue aux transfert de points, rendant cependant la formulation plus complexe :

$$\mathbf{P_1}^i\mathbf{P_2}^j\mathbf{P_3}^k\mathbf{P_4}^l\varepsilon_{ipw}\varepsilon_{jqx}\varepsilon_{kry}\varepsilon_{lsz}\mathbf{Q}^{i,j,k,l} = 0_{wxyz} \tag{2.73}$$

en substituant les lignes $\mathbf{l}_i^1$ par $\varepsilon_{ipw}\mathbf{P}_1^i$, représentant ainsi toute les lignes passant par $\mathbf{P}_1$ et où ε_{ijk} est le tenseur de Levi-Civita :

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{si } (i,j,k) = (1,2,3), (2,3,1) \text{ ou } (3,1,2) \\ -1 & \text{si } (i,j,k) = (3,2,1), (1,3,2) \text{ ou } (2,1,3) \\ 0 & \text{autrement} \end{cases}$$

Il est possible d'estimer les tenseurs tri/quadri-focal à l'aide d'un simple appariement de points/lignes entre les images, au même titre que la matrice essentielle, les tenseurs sont définis à l'échelle près. Pour plus de détails quant à la résolution des tenseurs présentés ici il est conseillé de consulter le livre [87].

2.6/ CONCLUSION

Dans ce chapitre, nous avons présenté les fondamentaux de la géométrie de nos capteurs et de la géométrie multi-vues utilisées dans cette thèse.
Un descriptif exhaustif des dispositifs d'acquisition d'images panoramiques a également été proposé, nous verrons dans les chapitres suivants plusieurs applications mettant en œuvre différentes caméras de ce type.

3

SUIVI VISUEL POUR LES CAMÉRAS OMNIDIRECTIONNELLES

Ce chapitre s'intéresse à l'adaptation des méthodes de suivi visuel pour les caméras omnidirectionnelles à point de vue unique. Dans le cadre de cette thèse c'est une étape essentielle -comme le rappelle la figure 3.1- puisqu'elle doit par la suite permettre l'orientation de la caméra PTZ sur la cible en question.

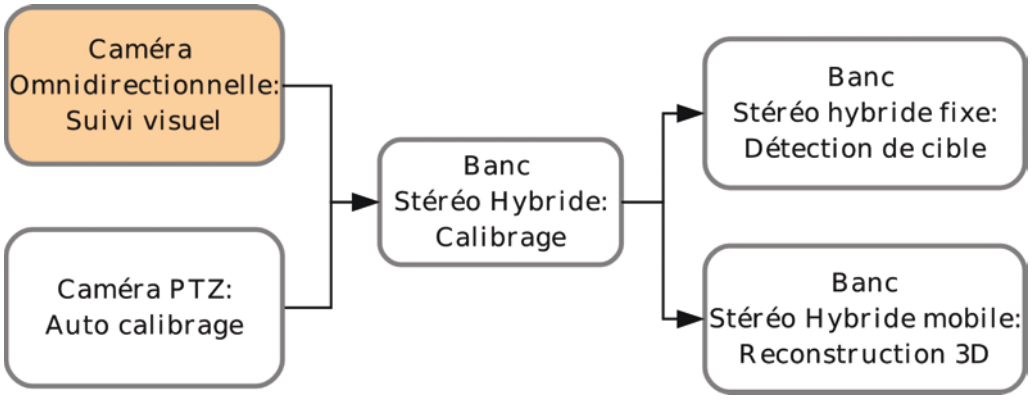
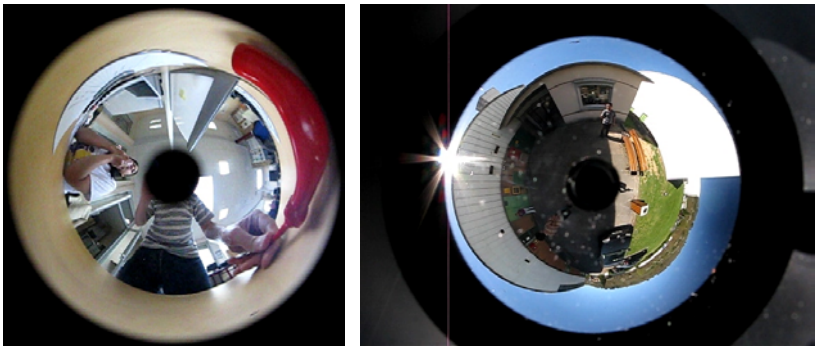


FIGURE 3.1 – Problématique globale

La littérature concernant le suivi visuel est très large et de nombreuses approches permettent d'effectuer le suivi d'un objet dans une suite d'images. Ces différentes méthodes sont couramment utilisées dans des séquences d'images perspectives, cependant nous ne pouvons pas directement les appliquer à des vidéos acquises à l'aide d'une caméra omnidirectionnelle. Cette incompatibilité est principalement liée à la distorsion géométrique induite par l'utilisation d'un miroir ou d'une optique particulière. Comme nous pouvons le constater sur les images de la figure 3.2, le miroir implique également une forte sensibilité à l'illumination de la scène ainsi qu'une non-uniformité de la résolution de l'image qui rendent le suivi plus difficile. Ici, nous proposons une adaptation d'algorithmes de suivi visuel (la méthode *mean-shift* et le filtrage particulaire) à la géométrie des capteurs omnidirectionnels. Ce chapitre est articulé en six parties : tout d'abord, nous proposons une présentation générale du suivi visuel, ses défis, ses caractéristiques. Dans

la section 3.2, une brève bibliographie à propos du suivi d'objet avec des capteurs omnidirectionnels est proposée. Dans la section suivante une méthode permettant une adaptation du voisinage à la géométrie du capteur est présentée. Une autre section rendra compte de la représentation d'une cible en utilisant un histogramme couleur. Ensuite, nous proposerons dans la section 3.5 une adaptation de deux méthodes, l'algorithme *mean-shift* et le filtrage particulaire, pour les images omnidirectionnelles. Enfin, la dernière section est dédiée aux résultats et aux conclusions de ce chapitre.



(a) (b)

FIGURE 3.2 – (a) Exemple d'image catadioptrique (b) Image catadioptrique avec éblouissement

3.1/ LE SUIVI VISUEL

Le suivi visuel est un sujet prédominant dans le domaine de la vision par ordinateur. Son utilisation est souvent nécessaire pour des applications telles que la vidéo surveillance, la robotique mobile, la réalité augmentée ou encore la compression vidéo [120]. Cela en fait un sujet déjà très largement développé en vision par ordinateur.
Le suivi visuel est l'estimation dans le temps de la position d'un objet d'intérêt au sein d'une séquence d'images. Il est en effet question, à partir d'une initialisation de la cible sur la première image, de la localiser sur les images suivantes. Bien que tous les algorithmes existants ne suivent pas tous le même schéma, la plupart des approches peut clairement se décomposer en quatre étapes majeures :

1. Le premier point concerne l'initialisation de la cible, il s'agit parfois d'une détection de la cible dans une image à l'aide de détecteurs tels que ceux proposés par Viola et Jones [174]. Dans notre cas nous considérerons une initialisation manuelle de la cible, nous ne nous étendrons donc pas sur sa détection. Pour plus de détails à ce sujet, nous conseillons la lecture de l'ouvrage de Bishop [29].
2. Le second élément important concerne la caractérisation de la cible, c'est-à dire-qu'elles sont les caractéristiques discriminantes qui permettent de modéliser l'objet d'intérêt et donc de le suivre efficacement. Nous verrons plus en détails dans les

sections à venir qu'il en existe une grande variété.

3. La représentation de la forme de l'objet doit également être choisie avec soin en fonction de l'application désirée. Dans la plupart des approches existantes il s'agit d'un simple rectangle englobant, mais nous verrons que ce n'est pas toujours le plus approprié.

4. Enfin la dernière étape essentielle au suivi visuel concerne la stratégie de localisation de l'objet.

Ces grandes lignes sont cependant loin d'être exhaustives puisque de plus en plus de méthodes font également intervenir des mécanismes d'apprentissage afin de prendre en compte les possibles changements d'apparence de la cible dans le temps. De plus, il est important de savoir que les points développés sont fortement inter-dépendants.

3.1.1/ LES DIFFICULTÉS RENCONTRÉES

Le suivi visuel est entièrement basé sur l'apparence de la cible et de la représentation que l'on en fait. Cependant l'aspect initial de cette dernière peut être amené à changer au cours du temps. Afin de mieux cerner les défis que doivent surmonter les algorithmes de suivi nous listons ici les raisons qui peuvent conduire à ces changements :

– *L'illumination* : La direction, l'intensité mais aussi la couleur de la lumière ambiante peut fortement influencer l'apparence de l'objet. Par exemple dans une scène en extérieur un simple changement météorologique peut modifier l'image de la cible.
– *Les changements de poses* : Le déplacement de l'objet dans la scène peut entraîner de considérables changements d'apparences sur l'image. En effet son orientation, son échelle, mais également ses couleurs peuvent en être modifiés.
– *Le bruit* : Le bruit dans les images peut également altérer la représentation de l'objet et en conséquence troubler le suivi.
– *Les occultations* : Une occultation est une disparition totale ou partielle de l'objet derrière des éléments de la scène. Dans le cas où la cible est partiellement visible et si elle conserve ses caractéristiques principales alors le suivi peut toutefois se poursuivre. Dans le cas d'une occultation totale de l'objet des stratégies prédictives doivent être mises en œuvre afin d'estimer sa position.
– *"L'encombrement" de fond* : Ce phénomène est un problème récurrent pour toutes les approches de suivi. Ce terme signifie qu'il existe une forte similarité entre l'objet d'intérêt et les éléments constituants l'arrière plan.

3.1.2/ EXTRACTION DE PRIMITIVES

Le choix des caractéristiques à extraire afin de modéliser l'apparence de la cible est essentiel pour assurer une bonne robustesse du suivi. L'essentiel étant de trouver un

bon compromis entre l'invariance des primitives et la précision de suivi qu'elles permettent. La sélection des détecteurs doit être en conformité avec l'environnement auquel est confronté l'algorithme. Nous présentons ici les primitives les plus courantes.

La couleur La couleur présente un intérêt particulier pour le suivi de cible puisqu'elle permet une caractérisation à la fois assez discriminante mais aussi robuste aux changements mineurs d'illumination, de poses et peut également assurer un suivi efficace même en présence d'occultation partielle de l'objet.
Sur des images acquises numériquement les informations couleurs sont codées suivant différents espaces couleur, la plupart étant clairement définie par la CIE (Commission Internationale de l'Eclairage) tel que le codage RGB, XYZ, CIELab, HSV ... L'étude de ces différents espaces colorimétriques est très intéressante car certaines représentations assurent une meilleure invariance aux éléments évoqués précédemment. Van de Sande *et al.* proposent un comparatif de nombreux espaces couleurs pour la caractérisation d'objets dans [173].

Le gradient Le gradient d'une image peut fournir de précieuses informations sur l'aspect d'un objet, il correspond à la dérivée de l'image caractérisant ainsi les changements d'intensité locale. Il fournit plus exactement deux types d'informations, à la fois sur la direction du gradient mais aussi sur son intensité. Parmi les approches les plus populaires nous citerons Prewitt [141] et Sobel [157].

Les contours et les lignes Les contours sont des primitives usuelles en traitement d'image correspondant à la binarisation du gradient d'une image, réduisant ainsi la quantité de données à traiter. Ils sont également très utilisés pour le suivi visuel à la fois pour la simplicité de leur détection et pour les informations qu'ils apportent sur l'objet d'intérêt. Il existe de nombreuses approches permettant l'extraction de ce type de primitive tel que le détecteur de Canny [35]. Les contours souffrent toutefois de certaines limitations. Ils sont particulièrement sensibles aux changements d'illumination, aux bruits, changements de poses de l'objet d'intérêt... De plus, leur utilisation ignore certaines informations importantes tel que la couleur de la cible et ils sont inefficaces lorsque l'objet n'est pas ou peu texturé.
A partir des contours il est également possible d'extraire des lignes par exemple à l'aide de la transformée de Hough [94].

Les points d'intérêt La détection de points ou régions d'intérêt consiste à localiser des pixels ou ensemble de pixels possédant des caractéristiques locales remarquables. Depuis bientôt trente ans, nous avons pu assister au développement d'un grand nombre de détecteurs et de descripteurs de points d'intérêt de plus en plus performants. Concer-

nant le suivi de points d'intérêt le détecteur de Harris [79] reste un très bon standard, le suivi de points KLT (pour Kanade-Lucas-Tomasi [153]) employant un détecteur proche de celui proposé par Harris est d'ailleurs toujours largement utilisé. D'autres détecteurs tels que SIFT [118] ou SURF [24] sont toutefois très communs en suivi pour leur très grande robustesse aux changements d'échelle et d'orientation [112].

3.1.3/ REPRÉSENTATION DE LA CIBLE

Nous traiterons dans cette partie de la représentation de la cible. Cette représentation peut être scindée en deux parties, d'une part la représentation de la forme de l'objet et de l'autre la représentation de son apparence.

3.1.3.1/ REPRÉSENTATION DE LA FORME DE L'OBJET

La localisation et la représentation géométrique d'un objet sur l'image peuvent être codées sous forme d'un vecteur que nous appellerons vecteur d'état $\mathbf{S}_t$ modélisant l'état de la cible à l'instant t. La représentation la plus simple d'un objet peut, par exemple, être la position de son centroïde. Dans ce cas $\mathbf{S}_t$ contiendra au minimum les coordonnées de ce barycentre $\mathbf{S}_t = \{x_t, y_t\}$. Il est cependant courant de décrire la forme d'un objet à l'aide d'une zone autour de la cible, le plus fréquemment une boîte englobante est utilisée [146, 183] nécessitant alors un vecteur plus long $\mathbf{S}_t = \{x_t, y_t, l_t, h_t\}$ comprenant la largeur et la hauteur du rectangle. De la même manière, il est possible d'employer une ellipse afin d'ajouter une rotation à notre fenêtre de suivi [97] (voir figure 3.3).
Le suivi basé contour est également une méthode efficace et permet une description plus précise de la forme de l'objet. Cependant, cette représentation est plus complexe puisque l'approche classique consiste à prendre un échantillon de points de contrôle le long du contour de la cible [96]. Le vecteur d'état de la cible est alors la concaténation des coordonnées de tous ces points, ce qui - en fonction du nombre de points de contrôle utilisés - mène à une représentation de haute dimensionnalité. Si l'objet est connu il est également possible d'apprendre un modèle de contours valides comme dans [47] afin de suivre les contours d'une main.
Un grand éventail de modèles plus spécifiques mais nécessitant des *a priori* sur l'objet d'intérêt peuvent être mis en oeuvre. C'est par exemple le cas pour les cibles articulées, dans [4] ce type de modèle est utilisé pour suivre les mouvements humains. Les objets déformables peuvent également jouir d'une représentation particulière. Dans l'article [46] Cootes *et al.* se servent d'un modèle 3D de visage pour en suivre les déformations et ainsi reconnaître les expressions faciales.
Rappelons que les vecteurs d'état peuvent contenir d'autres informations concernant notamment la trajectoire de la cible telle que la vélocité bien souvent nécessaire pour l'utilisation de méthodes prédictives.

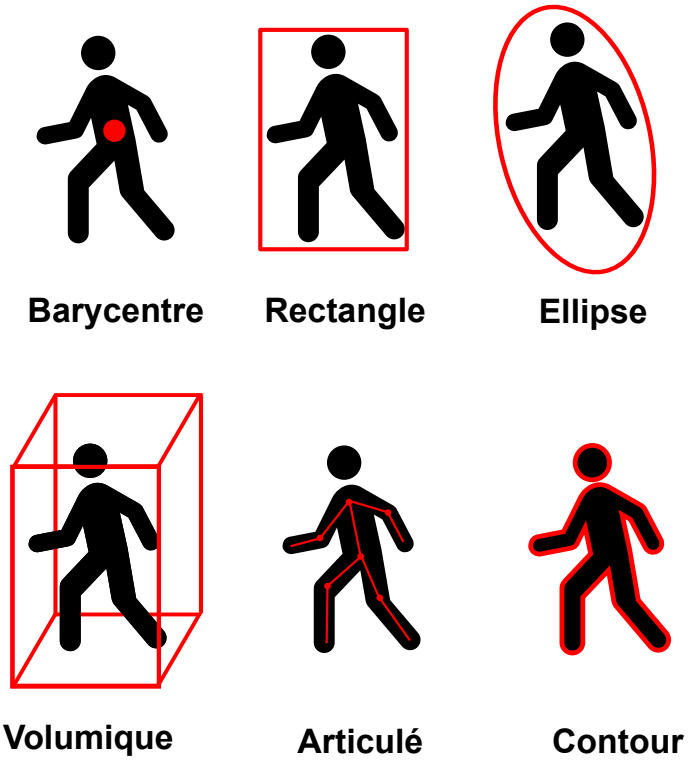


FIGURE 3.3 – Les différents types de représentation de forme.

3.1.3.2/ REPRÉSENTATION DE L'APPARENCE DE L'OBJET

Dans le cas où la caractérisation de l'objet passe par une fenêtre englobante permettant d'approximer la position de l'objet, alors l'ensemble des informations constituant la région d'intérêt peut être utilisé. Plusieurs techniques sont alors possibles pour permettre de caractériser l'objet d'intérêt.

Méthodes basées recalage De nombreuse approches basées sur des algorithmes de recalage existent. Cette stratégie consiste à optimiser une transformation appliquée au modèle afin de minimiser sa différence avec l'objet visible sur l'image courante. Cette différence pouvant être calculée par une mesure de similarité telle que la somme des différences absolues de l'intensité des pixels. L'avantage de ces méthodes dites "denses" est qu'elles prennent en compte l'ensemble des pixels constituant le modèle. Cependant la déformation à appliquer n'est pas toujours simple à définir. Cette transformation peut être affine (comprenant la translation et la rotation du modèle), ou encore dans le cas d'un plan, une transformation homographique [124]. Pour des déformations plus complexes d'autres modèles de transformation non rigide doivent être exploités [155]. Les résultats de suivi obtenus sont souvent très précis avec ces méthodes, elles souffrent cependant de quelques limitations. En effet, afin de garantir une convergence correcte, les mouvements entre les images doivent être suffisamment petits. De plus, une étape

d'interpolation est bien souvent nécessaire ce qui peut à la fois altérer la qualité du suivi mais également le ralentir.

Méthodes basées histogramme Une autre manière plus compacte de représenter l'apparence d'une cible est d'en extraire des histogrammes afin d'en créer un modèle statistique. Ces histogrammes peuvent être de plusieurs natures, les plus courants étant les histogrammes de couleurs ou de gradients. Nous verrons que ces représentations par histogramme offrent une représentation invariante aux transformations projectives.

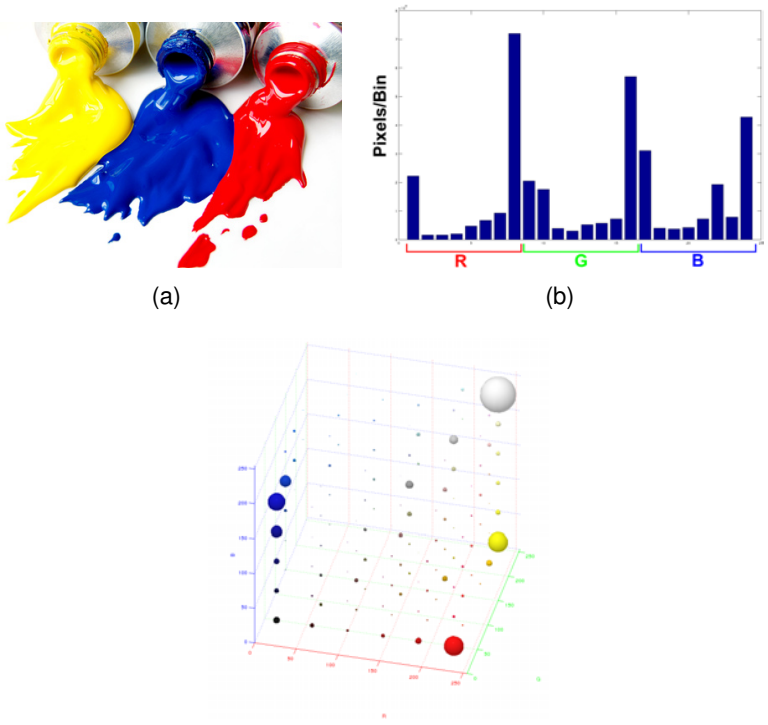
Les histogrammes couleurs

Un histogramme couleur permet de représenter la distribution des couleurs dans l'image. De cette manière on peut décrire les particularités colorimétriques de l'objet d'intérêt sans avoir à prendre en compte la localisation spatiale des pixels, c'est ce qui en fait un descripteur invariant aux changements d'échelle, à la rotation et aux occultations partielles de la cible. Afin de rendre la représentation par histogramme plus robuste on regroupe généralement les niveaux d'intensités en un nombre de cellules déterminé. Un noyau conférant une information spatiale supplémentaire peut également être utilisé afin de permettre une localisation plus précise de la cible.
Deux approches sont possibles pour construire des histogrammes couleurs, la première consistant simplement en la concaténation des histogrammes de chacun des canaux couleurs traités indépendamment. La seconde est la création d'un histogramme couleur 3D (voir fig. 3.4).

Les histogrammes de gradient

De la même manière que l'on construit des histogrammes couleurs il est possible de construire un histogramme des gradients orientés (HOG) de la cible [51]. Cette représentation a l'avantage d'être robuste aux changements d'illumination. C'est également un descripteur efficace lorsque les contours de l'objet sont marqués, initialement les HOG ont été conçus pour la détection de piétons dans les images. En outre, il est possible de rendre les HOG invariants aux rotations et à l'échelle [158].

Les spatiogrammes Les histogrammes couleurs sont remarquables pour leur invariance aux transformations projectives. Cependant, la perte d'information spatiale peut conduire, soit à une localisation de moindre qualité soit à une fausse détection si des objets de couleur similaire se trouvent dans la scène. Les spatiogrammes sont une bonne alternative permettant de conserver l'information concernant la localisation spatiale des pixels. L'article [28] met d'ailleurs en évidence le gain offert par l'usage d'une telle méthode.



(a) (b)

(c)

FIGURE 3.4 – (a) Image initiale (b) Histogramme 2D concaténé (c) Histogramme 3D 8×8×8 où la taille des sphères est proportionnelle aux nombres de pixels dans la cellule

3.1.4/ LES MÉTHODES DE SUIVI

La littérature comporte de nombreuses méthodes de suivi que nous pouvons diviser en deux principaux groupes : les méthodes déterministes et les méthodes stochastiques. Les méthodes de suivi appartenant au premier groupe comme par exemple (KLT) Kanade-Lucas-Tomasi ou la méthode *mean-shift*, cherchent de manière itérative à maximiser une mesure de similarité locale entre un modèle de la cible et une région dans l'image courante. Les méthodes stochastiques utilisent quant à elles une représentation d'état de l'objet en mouvement afin de modéliser sa dynamique. Parmi ces techniques nous pouvons notamment citer le filtre de Kalman et les filtres particulaires.

3.1.4.1/ LE SUIVI *mean-shift*

L'approche de suivi basée *mean-shift* a été initialement proposée par Comaniciu *et al.* dans [43]. Cette méthode déterministe repose sur la maximisation itérative de la similarité entre l'apparence du modèle et la cible. Au lieu d'effectuer une recherche exhaustive de la cible dans toute l'image, c'est la position précédente de la cible qui est utilisée comme point d'initialisation. Dans la méthode originale Comaniciu *et al.* propose une représentation de la cible par son histogramme couleur pondéré par un noyau. Le calcul

de la distribution colorimétrique représentative de l'apparence d'un objet s'exprime de la manière suivante :

$$q_u = C \sum_{i=1}^{n} k(\| x_i \|^2)\delta[b(x_i) - u],\ u = 1 \cdots m, \tag{3.1}$$

où, m le nombre de classes dans l'histogramme, δ est la fonction de Kronecker, b(x_i) est la classe de l'histogramme correspondant au pixel x_i, $\| x_i \|$ est la distance euclidienne entre le pixel d'intérêt x_i et le centre de la fenêtre, n est le nombre de pixels dans la fenêtre de recherche, k est un noyau d'Epanechnikov et C est une constante de normalisation.

La mesure de similarité entre deux histogrammes est définie par le coefficient de Bhattacharyya :

$$\beta(q, p) = \sum_{u=1}^{m} \sqrt{q_u . p_u}, \tag{3.2}$$

où q et p sont respectivement les histogrammes couleur du modèle et de la cible.

Le vecteur *mean-shift* permettant la convergence d'une localisation initiale y_0 sur l'image précédente vers la nouvelle localisation de l'objet y_1 sur l'image courante se caractérise par :

$$y_1 = \frac{\sum_{i=1}^{n} x_i w_i k(\| (y_0 - x_i)/h \|^2}{\sum_{i=1}^{n} w_i k(\| (y_0 - x_i)/h \|^2}, \tag{3.3}$$

ici h est le nombre de pixels dans la fenêtre de recherche et w_i est le poids associé à chaque pixel candidat. Ce poids est proportionnel à la similarité entre le modèle de référence et la couleur du pixel x_i, il est calculé de la manière suivante :

$$w_i = w(x_i) = \delta[b(x_i) - u] \sqrt{\frac{q_u}{p_u(y_0)}} \tag{3.4}$$

Itérativement, la localisation de la cible se déplace de y_0 vers y_1, le processus peut être stoppé suivant différents critères d'arrêt. Typiquement si la longueur du vecteur *mean-shift* caractérisant la translation entre y_0 et y_1 est inférieur à un certain seuil alors un maximum local est atteint. En pratique on considère que moins de dix itérations sont suffisantes pour obtenir une localisation de la cible dans la nouvelle image.

Le suivi basé *mean-shift* cumule de nombreux avantages, de par la représentation du modèle en histogramme couleur l'algorithme est robuste au flou de mouvement, aux déformations de l'objet et aux occultations partielles. C'est aussi un algorithme pensé pour fonctionner en temps réel puisqu'un nombre réduit d'itérations suffit à assurer une bonne convergence vers la localisation de la cible. Dans la version initiale de l'algorithme on

notera toutefois que celui-ci n'est pas adapté pour les cas d'occultation totale de la cible. Ce problème peut cependant être résolu par l'intégration d'un filtre de Kalman dans le processus de suivi [99].

3.1.4.2/ LE SUIVI VISUEL AVEC FILTRE PARTICULAIRE

Les approches de suivi visuel stochastiques aussi appelées probabilistes reposent sur un mécanisme de prédiction de la nouvelle position de la cible dans une séquence d'images à l'aide d'information provenant des précédents états de la cible.

Le filtrage particulaire est une méthode d'estimation de modèle qui généralise le filtre de Kalman pour les modèles non-linéaires et non-gaussiens. Le principe est d'approximer une fonction de densité de probabilité par un ensemble d'échantillons pondérés $\{x_t^i, w_t^i\}_{i=1\cdots,N}$ (particules), chacun de ces échantillons x_t^i étant l'observation d'un état hypothétique de l'objet avec un poids proportionnel à sont importance w_t^i.
De manière générale, on souhaite estimer l'état de la cible à un instant t à l'aide des observations du système dans le temps, $\mathbf{Z}_t = \{z_0, z_1, \cdots, z_t\}$, c'est-à-dire que l'on souhaite estimer la distribution postérieure $p(x_t \mid \mathbf{Z}_t)$ [154].
La représentation d'état générique permettant de modéliser un système dynamique peut être définie par les équations suivantes :

$$\begin{cases} x_t = f(x_{t-1}) + v_{t-1} \\ z_t = h(x_t) + n_t \end{cases} \tag{3.5}$$

où f et h sont respectivement la fonction de transition et la fonction de mesure, tandis que v_{t-1} et n_t sont les bruits du système et de mesure. Le filtre particulaire au même titre que les autres méthodes bayésiennes séquentielles se décomposent en deux étapes, une prédiction et une correction. L'étape de prédiction consiste à estimer la probabilité *a priori* :

$$p(x_t \mid \mathbf{Z}_{t-1}) = \int p(x_t \mid x_{t-1}) p(x_{t-1} \mid \mathbf{Z}_{t-1}) dx_{t-1}. \tag{3.6}$$

La correction utilise l'observation à l'instant t pour mettre à jour la probabilité *a posteriori* :

$$p(x_t \mid \mathbf{Z}_t) = \frac{p(z_t \mid x_t) p(x_t \mid \mathbf{Z}_{t-1})}{p(z_t \mid \mathbf{Z}_{t-1})}. \tag{3.7}$$

Nous présentons ici une approche standard de suivi visuel avec le filtrage particulaire couleur.
Dans un premier temps un modèle couleur de la cible q est créé à l'aide de l'équation (3.1) (par sélection de la cible sur la première image de la séquence par exemple).
Dans le cas d'une image perspective avec une fenêtre de suivi rectangulaire, chaque

particule correspond à un candidat au modèle, représentée par le vecteur d'état suivant :

$$\mathbf{S} = \{x, y, \dot{x}, \dot{y}, l, h\}, \tag{3.8}$$

où x et y sont les coordonnées du centre de la fenêtre, tandis que $\dot{x}$ et $\dot{y}$ correspondent à la vélocité sur chaque axe, et l et h définissent les dimensions en largeur et en hauteur de la fenêtre.
Il s'agit alors de diffuser ces particules, on parle aussi de prédiction puisque cette étape permet l'estimation de la densité à priori. En pratique cette diffusion, pour chaque particule, est effectuée à l'aide du modèle dynamique suivant :

$$\mathbf{S}_t = \mathbf{A}.\mathbf{S}_{t-1} + \mathbf{W}_{t-1}, \tag{3.9}$$

avec $\mathbf{A}$ la composante déterministe du modèle (ici, nous modélisons la dynamique de l'objet en mouvement par un modèle auto-régressif d'ordre 1) et $\mathbf{W}$ un bruit blanc gaussien. Le bruit additif $\mathbf{W}$ permet une modification de tous les paramètres d'état de chaque particule de manière aléatoire assurant ainsi leur variété et leur diffusion.
On calcule ensuite le poids attribué à chaque particule en fonction de sa similarité avec le modèle de la manière suivante :

$$w^i = e^{(\beta(q,p^i)-1)}, \tag{3.10}$$

où w^i est le poids donné à la $i^{\text{ème}}$ particule ayant pour histogramme couleur p^i. Ces poids permettent un ré-échantillonnage des particules. Cela signifie que les particules ayant un poids important (donc une similitude forte avec le modèle) sont dupliquées tandis que les particules trop dissemblables sont éliminées. De nombreuses approches existent pour effectuer cette étape de ré-échantillonnage des particules [10]. Dans nos algorithmes c'est la méthode de ré-échantillonnage systématique qui est utilisée. Finalement l'état estimé de la cible $\mathbf{E}_t$ est considérée comme étant l'état moyen de notre ensemble de particules :

$$\mathbf{E}_t = \sum w^i \mathbf{S}_t^i. \tag{3.11}$$

Pour chaque nouvelle image les étapes de calcul de poids et de ré-échantillonnage de particules sont effectuées assurant ainsi le suivi de la cible dans une séquence vidéo.

3.2/ SUIVI VISUEL AVEC DES CAMÉRAS OMNIDIRECTIONNELLES

Le suivi d'objet avec des capteurs omnidirectionnels ouvre la voie à de nombreuses applications, principalement dans le domaine de la surveillance et de la navigation robotique. Néanmoins, l'attention portée jusqu'à présent à ce sujet reste relativement limitée et très peu d'articles abordent cette thématique.
Parmi les différentes approches permettant le suivi d'objet avec des capteurs omnidirec-

tionnels, celle consistant à "rectifier" l'image est certainement la plus répandue. Cette méthode nécessite l'utilisation d'un PVU afin d'obtenir une image perspective à partir de celle déformée. Le procédé permettant cette manipulation de l'image est décrit dans [16]. Suite à cette étape, la distorsion de l'image est réduite ce qui permet l'utilisation d'une méthode de suivi traditionnelle. Dans [36], un filtre particulaire est ainsi utilisé sur une séquence d'image désanamorphosée (rectification de la déformation d'une image) tandis que dans [14] l'algorithme de suivi KLT est utilisé.
Le suivi de cible basé sur des connaissances *a priori* de l'objet d'intérêt est également une méthode efficace souvent utilisée pour la navigation robotique, l'asservissement visuel et le suivi 3D. Elle nécessite des connaissances sur la géométrie de l'objet à suivre, tel que les plans ou les droites qui le composent. Le suivi de plan est essentiellement basé sur des méthodes de recalages homographiques. Par exemple, Mei *et al.* [123] utilisent cette technique pour le suivi de modèle planaire dans des images catadioptriques. Cette approche nécessite cependant des conditions optimales pour être pleinement efficace, un déplacement court entre chaque image de la séquence est par exemple requis. Le suivi de droites fait intervenir une autre stratégie. Tout d'abord une détection de droites adaptée aux particularités des capteurs catadioptriques est appliquée sur l'image [19, 126], puis le suivi est effectué à l'aide d'un algorithme classique de suivi de droites [121].
La soustraction d'arrière plan est également une solution simple et efficace lorsque la caméra omnidirectionnelle est statique car il n'est pas nécessaire de tenir compte de la distorsion de l'image. Boult *et al.* [32] donnent un bon aperçu de l'utilisation de cette méthode pour la vidéo surveillance. L'adaptation du voisinage à la géométrie des capteurs afin de prendre en compte les distorsions permet de plus d'obtenir des meilleurs résultats [54].
Une autre approche consiste à adapter les méthodes conventionnelles de suivi sur la sphère unité. Cette adaptation permet une amélioration des résultats de suivi car le voisinage sera adapté aux particularités géométriques et à la non-uniformité de la résolution du capteur. Le concept principal est développé dans [55] avec l'utilisation de distance géodésique calculée sur la sphère. Cependant, seuls [27] et [166] utilisent cette représentation pour le suivi avec un filtre particulaire. Nous présentons dans ce chapitre un travail basé sur cette méthode, avec l'adaptation d'un filtre particulaire et d'un suivi par *mean-shift*. Nous proposons également différentes améliorations, comme l'adaptation du noyau et l'utilisation d'un histogramme multi-parties permettant une meilleure localisation spatiale ainsi qu'une plus grande robustesse aux variations d'échelle et aux fonds "encombrés". De plus, nous présentons une évaluation quantitative des algorithmes basée sur plusieurs type de mesures.

3.3/ L'ADAPTATION DU VOISINAGE POUR LES IMAGES OMNIDIRECTIONNELLES

Dans une image perspective, une région d'intérêt est généralement définie par deux paramètres (largeur et hauteur) dans laquelle le voisinage est uniforme. Nous obtenons alors un rectangle centré sur le pixel d'intérêt. Bien entendu, cet échantillonnage classique n'est pas adapté aux images omnidirectionnelles car il considère la résolution uniforme sur l'ensemble de l'image et ne tient pas compte de la distorsion due à l'utilisation du miroir ou d'une optique grand angle.
Comme nous l'avons vu dans le chapitre 2, nous pouvons modéliser les images acquises à partir d'une caméra respectant le PVU par un équivalent sphérique. La projection sur la sphère permet de créer un voisinage adapté à la distorsion et à la non-uniformité de la résolution de l'image. Nous définissons ce voisinage par une distance géodésique calculée sur la sphère.
Les coordonnées sphériques sont définies de la manière suivante :

$$\mathbf{X} = (\cos(\phi)\sin(\theta), \sin(\phi)\cos(\theta), \cos(\phi)), \tag{3.12}$$

où X est un point sur la sphère S^2, ϕ est la latitude (comprise entre 0 et π) et θ la longitude (de 0 à 2π). Par conséquent, la localisation d'un point peut être définie à l'aide de 2 paramètres (θ, ϕ) comme nous pouvons le constater sur la figure 3.5. Nous déterminons le

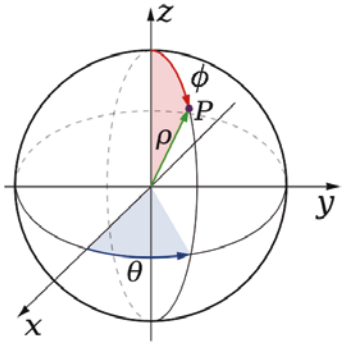
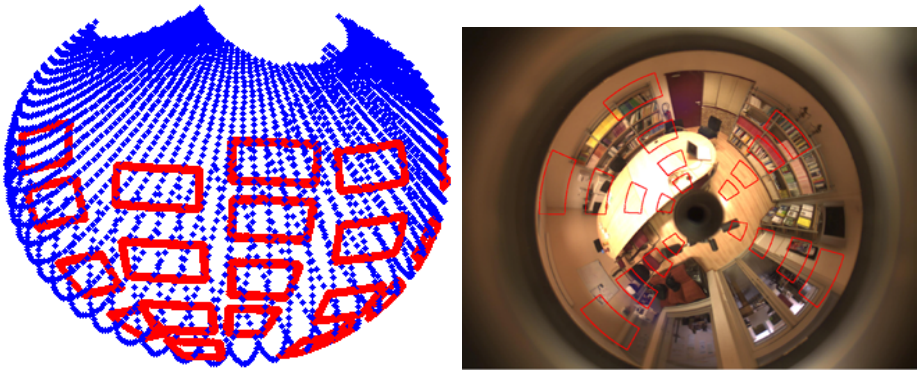


FIGURE 3.5 – Système de coordonnées sphériques

voisinage à partir d'un point central sur la sphère $\mathbf{X}_{sph}(\theta,\phi)$ autour duquel nous définissons une plage de variation de θ et ϕ, respectivement $\delta\theta$ et $\delta\phi$:

$$N_S(\mathbf{X}_{Sph}) = \{\mathbf{X}_{Sph} = (\theta', \phi')\epsilon S^2, \left|\theta' - \theta\right| \leq \delta\theta \text{ et } \left|\phi' - \phi\right| \leq \delta\phi\}.$$

De cette manière nous obtenons les coordonnées sur la sphère que nous re-projetons ensuite sur le plan image afin d'obtenir les coordonnées des pixels concernés.
Comme nous pouvons le voir sur la figure 3.6, des fenêtres créées avec des angles simi-



(a) (b)

FIGURE 3.6 – Voisinage avec des valeurs fixes $\delta\theta$=±0.2 et $\delta\phi$=±0.1 (a) voisinage sur la sphère (b) voisinage projeté sur le plan image

laires sur la sphère auront des dimensions variables en fonction de leur position spatiale sur le plan image.

3.4/ REPRÉSENTATION PAR DES HISTOGRAMMES COULEUR

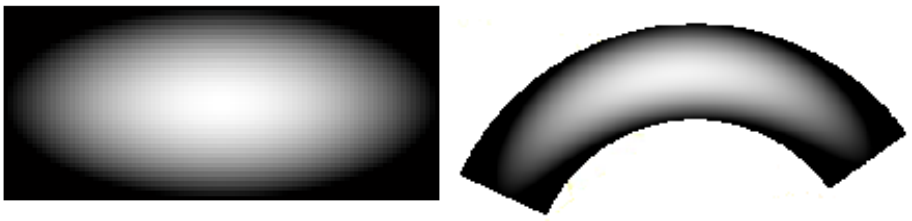
Les méthodes de suivi visuel développées ici sont basées sur les caractéristiques couleur de l'objet à suivre. Les histogrammes sont une méthode rapide et efficace pour représenter l'apparence d'un objet. Le calcul de la distribution colorimétrique peut se faire avec l'équation (3.1).

3.4.1/ ESPACE COULEUR

La modélisation des caractéristiques couleur d'un objet à l'aide d'un histogramme est une méthode rapide mais peu robuste aux changements d'illumination très fréquents notamment avec les systèmes catadioptriques. Le choix d'un espace couleur moins sensible est donc déterminant pour permettre un suivi performant. Partant de ce constat, notre choix c'est porté sur l'espace couleur "transformed RGB", défini dans [173], qui est considéré comme robuste aux changements d'illuminant et d'intensité lumineuse. Cette transformation repose sur la normalisation de chaque canal couleur de manière indépendante :

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \frac{R-\mu_R}{\sigma_R} \\ \frac{G-\mu_G}{\sigma_G} \\ \frac{B-\mu_B}{\sigma_B} \end{pmatrix}, \tag{3.13}$$

σ et μ étant respectivement l'écart type et la moyenne du canal considéré.
La représentation d'un objet basée sur sa couleur ne considère pas l'organisation spatiale



(a) (b)

FIGURE 3.7 – Profil d'un noyau d'Epanechnikov sur l'image, (a) pour une caméra perspective (b) pour une caméra catadioptrique

des pixels de la cible, il est possible d'améliorer cette approche par l'ajout d'un noyau. Le noyau permet d'attribuer un poids à chaque pixel en fonction de sa position dans la région d'intérêt.

3.4.2/ LE NOYAU

Afin de garantir une meilleure localisation de la cible, l'utilisation d'un noyau est fortement recommandée car il permet de prendre en compte des informations spatiales en plus de la distribution couleur d'un objet, une augmentation significative des performances est apportée. Son utilisation est d'ailleurs requise pour certains algorithmes de suivi tel que celui basé sur le *mean-shift*. Cependant, cet aspect n'a pas encore été abordé dans le cas d'images sphériques. Nous utilisons un noyau d'Epanechnikov pour obtenir une estimation non-paramétrique de la distribution couleur. Ce profil initialement conçu pour les images perspectives doit être adapté afin de pouvoir être défini correctement sur la sphère unité (voir figure 3.7).
Le profil d'Epanechnikov est défini par

$$k(x) = \begin{cases} \frac{1}{2}C_d^{-1}(n+2)(1-x) & \text{si } x \leq 1 \\ 0 & \text{sinon} \end{cases}, \tag{3.14}$$

où C_d est le volume de la sphère unité et n le nombre de dimension spatiale. Ici, nous souhaitons calculer ce profil directement sur la sphère unité, nous calculons la distance d'un point avec le centre de la fenêtre par une distance géodésique. Si l'on considère deux points sur la sphère ($\mathbf{X}_{sph}$ and $\mathbf{Y}_{sph}$), la distance géodésique (définie dans [55]) entre ces deux points peut être calculée par

$$\forall \mathbf{X}_{sph}, \mathbf{Y}_{sph} \in S^2, \quad d(\mathbf{X}_{sph}, \mathbf{Y}_{sph}) = \arccos(\mathbf{X}_{sph}.\mathbf{Y}_{sph}). \tag{3.15}$$

3.4.3/ REPRÉSENTATION MULTI-PARTIES

Dans une séquence vidéo omnidirectionnelle, l'échelle de la cible peut varier très rapidement. Il est par conséquent important de pouvoir gérer ce type d'inconvénient en adaptant au mieux la taille de la région d'intérêt. La représentation par histogramme multi-parties permet une meilleure localisation spatiale de la cible, ainsi qu'une meilleure gestion des changements d'échelle et une plus grande robustesse. Dans [119], les auteurs démontrent l'efficacité de cette méthode appliquée aux images perspectives. Au lieu de calculer un seul histogramme, cette approche consiste à en calculer 7 selon l'ajustement proposé dans la figure 3.8. L'histogramme numéro 1 (fig.3.8(a)) correspond

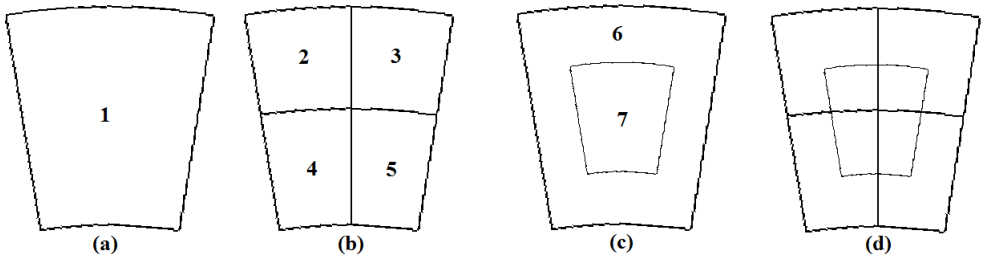


FIGURE 3.8 – Représentation multi-parties (a)région d'intérêt complète (b) division en 4 parties (c) division sensible aux changements d'échelle (d) représentation finale

à l'histogramme de la région d'intérêt entière. Dans un deuxième temps, la fenêtre est scindée en quatre parties, et un histogramme est calculé dans chacune d'entre elles. Cette organisation permet d'obtenir une information sur la rotation de la cible. La dernière étape est la division de la fenêtre en deux parties (intérieure et extérieure) qui permet une meilleure adaptation à l'échelle.

Le calcul de la similarité entre deux histogrammes multi-parties q et p est donné par :

$$\rho(q,p) = \frac{\sum_{i=1}^{N} \beta[q^i, p^i]}{N}, \tag{3.16}$$

où, N est le nombre total d'histogrammes (7 histogrammes dans notre cas) tandis que q^i et p^i sont respectivement, les histogrammes de la $i^{\text{ème}}$ sous-partition de p et q.

3.5/ ALGORITHMES DE SUIVI ADAPTÉS

3.5.1/ FILTRE PARTICULAIRE ADAPTÉ

Le filtre particulaire traditionnel utilise les coordonnées cartésiennes (x, y) afin de diffuser les particules directement sur le plan image. Dans l'approche proposée, les particules sont diffusées sur la sphère unité afin d'obtenir un comportement plus adapté aux par-

ticularités des capteurs omnidirectionnels. Nous pouvons alors définir le vecteur d'état par

$$\mathbf{S} = \{\theta, \phi, \dot{\theta}, \dot{\phi}, \delta\theta, \delta\phi\}, \tag{3.17}$$

où, les variables θ et ϕ sont les coordonnées du centre de la particule, $\delta\theta$ et $\delta\phi$ sont les dimensions de la fenêtre sur la sphère.

3.5.2/ SUIVI *Mean-Shift* ADAPTÉ

L'adaptation de l'algorithme *Mean-shift*(MS) [43] à la sphère unité nécessite une utilisation adéquate des coordonnées sphériques, notamment la gestion de la transition entre 0 et 2π. Le suivi utilisant *Mean-Shift* offre des résultats acceptables mais reste assujetti à de nombreuses contraintes comme l'occultation de la cible ou un changement trop important de son apparence. L'ajout d'un filtre de Kalman à l'algorithme MS est une bonne solution pour corriger ces problèmes. Lorsque le coefficient de similarité entre le modèle et la cible est faible, on peut considérer la cible comme étant perdue. Dans ce cas, le filtre de Kalman peut prendre le relais afin d'estimer la nouvelle position de l'objet en fonction de ses états précédents. Concrètement, après chaque itération de l'algorithme MS, nous calculons l'état prédit de l'objet en fonction de son état précédent. Après cette phase de prédiction, nous comparons le coefficient de similarité obtenu par la méthode MS avec un seuil (Δs). Si ρ est supérieur à Δs, nous considérons la mesure (calculée avec MS) comme suffisamment fiable, dans ce cas l'estimation offerte par le filtre de Kalman est mise à jour afin de corriger la prédiction (ce qui a également pour effet de lisser le suivi). Sinon la mise à jour n'est pas effectuée, ce qui signifie que la nouvelle position de l'objet est uniquement déterminée par la prédiction du filtre de Kalman. Le fonctionnement de la méthode est décrit par la figure 3.9.

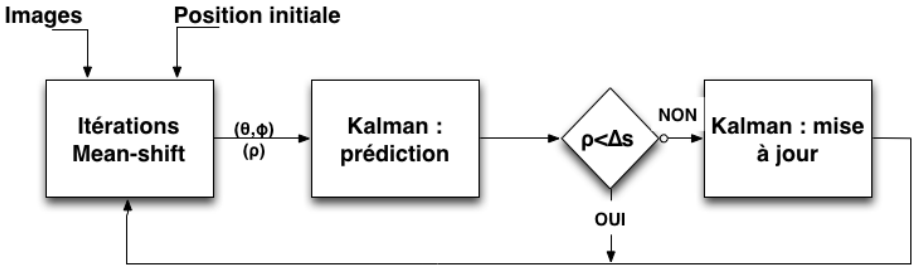
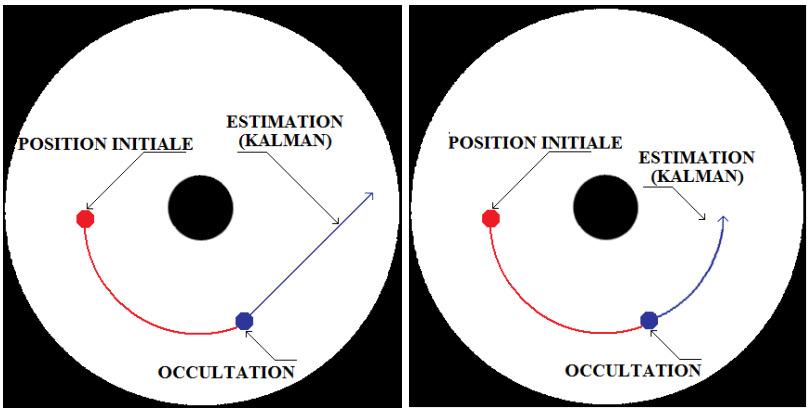


FIGURE 3.9 – Diagramme de fonctionnement Mean-Shift + Kalman

Cependant, la trajectoire d'un objet sur une image catadioptrique sera différente de celle d'un objet sur une image perspective comme l'illustre la figure 3.10.

Ce qui signifie que l'estimation fournie par le filtre de Kalman devra être faite sur la sphère unité afin de respecter la géométrie du capteur. Le vecteur d'état sera alors exprimé de



(a) (b)

FIGURE 3.10 – Filtre de Kalman (a) Non-adapté (b) adapté

la façon suivante :

$$\mathbf{S} = \{\theta, \phi, \dot{\theta}, \dot{\phi}\} \tag{3.18}$$

Dans ce chapitre les algorithmes basés sur la méthode MS incorporent un filtre de Kalman.

3.6/ EXPÉRIENCES ET RÉSULTATS

Dans cette section, nous mettons en œuvre des expériences permettant de quantifier l'efficacité de nos méthodes de suivi. Au regard de l'absence de base de données concernant le suivi d'objet avec des capteurs catadioptriques, les vidéos utilisées dans ce travail ont été directement acquises par nos soins dans différentes conditions afin d'offrir un large panel de situations : scène intérieure ou extérieure, objet ou humain en déplacement, avec ou sans occultation, caméra mobile ou statique... (voir tableau 3.1)
Afin d'apporter un bon aperçu des améliorations offertes par l'approche proposée, nous comparons les méthodes traditionnelles et adaptées. La *toolbox* utilisée pour la calibration est la "OCamCalib" de Davide Scaramuzza [150]. La résolution des images est de 640x480 pixels et la longueur des séquences oscille entre 500 et 800 images.

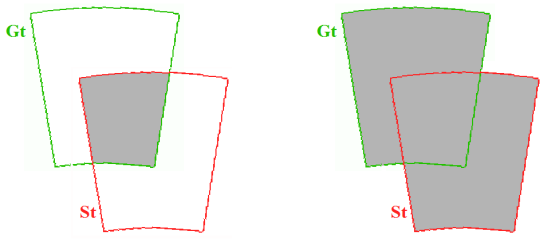
3.6.1/ ÉVALUATION DES PERFORMANCES DES ALGORITHMES DE SUIVI

Il existe dans la littérature de nombreuses méthodes statistiques permettant l'évaluation des algorithmes de suivi. Les tests réalisés ici sont basés sur la création manuelle d'une vérité de terrain. Trois critères représentatifs de la précision des différentes méthodes sont utilisés : la superposition spatiale, la superposition temporelle et la distance entre les centres.

ESTIMATION DES CRITÈRES DE SUPERPOSITION

La superposition spatiale correspond au pourcentage de surface commune A(St,Gt) entre les deux boîtes englobantes (la fenêtre issue de l'algorithme de suivi St et la fenêtre obtenue par notre vérité de terrain Gt), illustré figure 3.11. Ces deux régions sont comparées de cette manière :

$$A(St, Gt) = \frac{\text{Surface}(Gt \cap St)}{\text{Surface}(Gt \cup St)} \tag{3.19}$$



(a) (b)

FIGURE 3.11 – (a)Surface(Gt∩St) (b)Surface(Gt∪St)

En ce qui concerne la superposition temporelle, un seuil est fixé arbitrairement afin de déterminer le pourcentage d'image où le suivi est considéré comme "correct" dans la séquence. Concrètement, on peut par exemple fixer un seuil de superposition spatiale minimum T_{Ov} à 20%, ce qui signifie que le suivi est correct si la superposition entre la vérité de terrain et le résultat de l'algorithme de suivi est supérieure à 20% :

$$TO(St, Gt) = \begin{cases} 1 \text{ si } A(St, Gt) > T_{Ov} \\ 0 \text{ si } A(St, Gt) < T_{Ov} \end{cases} \tag{3.20}$$

La superposition temporelle est le pourcentage d'images avec une superposition correcte (définie par le seuil T_{Ov}) :

$$\tau = \left(\frac{100}{N}\right) \sum_{t=1}^{N} TO(St, Gt), \tag{3.21}$$

où N est le nombres d'images dans la séquence.

DISTANCE ENTRE LES CENTRES

La distance entre le centre de la fenêtre de la vérité terrain et le centre de la fenêtre obtenu par la méthode de suivi peut également être une information représentative de la précision de l'algorithme. Cette distance est calculée de manière euclidienne sur le plan image. Ce critère est certainement moins significatif que ceux présentés précédemment mais offre tout de même une information complémentaire dans l'évaluation proposée.

3.6.2/ RÉSULTATS

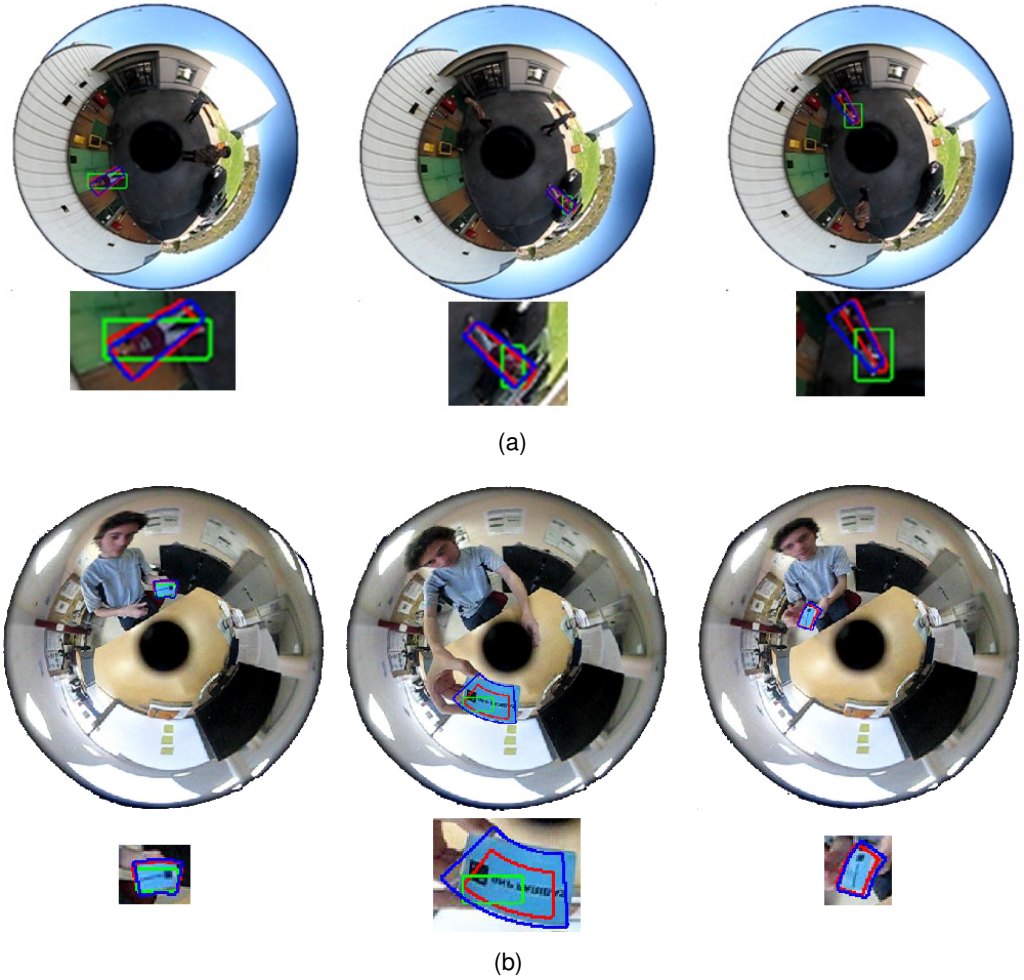
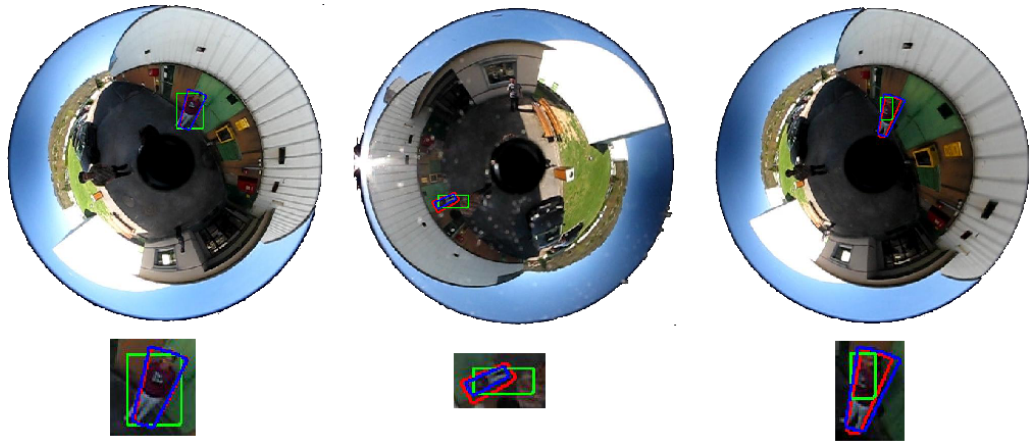


FIGURE 3.12 – Aperçu des résultats de suivi visuel avec un filtre particulaire conventionnel (fenêtre verte), un filtre particulaire adapté (fenêtre rouge) et la vérité terrain (fenêtre bleue) (a) Séquence 1 (b) Séquence 2

L'ensemble des résultats obtenus (tableaux 3.2 et 3.3) nous permet de confirmer l'hypothèse de départ. En effet, l'adaptation des méthodes de suivi visuel dans les images catadioptriques apporte une précision supérieure aux méthodes traditionnelles. Notons tout de même que le filtre particulaire reste la méthode la plus robuste. Dans la première séquence, une personne marche autour d'une caméra fixe avec un éclairage naturel. Malgré l'apparente simplicité de cette vidéo nous remarquons de forts écarts de résultats entre les méthodes adaptées et les méthodes conventionnelles. Ces résultats attestent qu'une fenêtre de sélection rectangulaire n'est pas en mesure de sélectionner correctement des objets sur des images catadioptriques (voir figure 3.12(a)). La seconde séquence nous permet quant à elle d'étudier la robustesse de nos méthodes à d'important changements d'échelle (figure 3.12(b)). Des changements assez importants dans l'apparence de l'objet



(a)

(b)

FIGURE 3.13 – Aperçu des résultats de suivi visuel avec un filtre particulaire conventionnel (fenêtre verte), un filtre particulaire adapté (fenêtre rouge) et la vérité terrain (fenêtre bleue) (a) Séquence 3 (b) Séquence 4

au cours du temps (surface réflective) sont également notables. Dans cette séquence, les méthodes de suivi conventionnelles ne sont pas capables de suivre correctement l'objet durant la totalité de la séquence. Avec le filtre particulaire classique, la cible est d'ailleurs rapidement perdue (voir superposition temporelle dans le tableau 3.2). Dans ce cas notre estimation sur la superposition spatiale n'est effectuée que sur la partie de la séquence correctement suivie. A l'inverse, le filtre particulaire adapté permet de suivre correctement la cible dans la quasi-totalité de la séquence. Nous pouvons attribuer ces bonnes performances à l'utilisation d'un histogramme multi-parties (pour les méthodes adaptées) qui permet de mieux gérer les problèmes de changements d'échelle.

Le suivi MS adapté donne quant à lui une meilleure précision en ce qui concerne la localisation du centre de la cible. La séquence numéro 3 fait intervenir de nombreux phénomènes typiques des capteurs catadioptriques, c'est-à-dire un fort changement de luminosité (éblouissement), une occultation et des changements d'échelle rapides avec une caméra mobile (figure 3.13(a)). Ici encore, les méthodes adaptées fournissent des

résultats très convaincants en tous points comparativement aux méthodes conventionnelles. Finalement, la séquence 4 confronte les algorithmes de suivi à des occultations répétées sur un objet de taille réduite. Cette fois encore, nous observons de meilleurs résultats avec notre approche en particulier en ce qui concerne le MS adapté. Notons que l'utilisation d'un noyau est un élément déterminant qui influe directement sur la qualité du suivi. L'utilisation d'un noyau est d'ailleurs essentielle à l'algorithme MS, les performances atteintes par l'adaptation de cette méthode tendent à démontrer une parfaite adéquation de ce noyau sur la sphère. Pour mettre en exergue le gain de précision offert par l'emploi d'un noyau, une série de tests a été effectuée (voir tableau 3.4), ces expériences confrontent le filtre particulaire avec et sans noyau. Pour montrer l'impact du noyau, la représentation multi-histogrammes n'a pas été utilisée (ce qui explique les scores moins importants et qui donne également un aperçu des améliorations apportées par l'utilisation de plusieurs histogrammes). Dans tous les cas de figures, un gain en terme de performance est apporté. De manière quantitative aussi bien que qualitative, notre méthode d'adaptation permet une meilleure gestion des changements d'échelles (toutes les séquences), des occultations (séquences 3 et 4) et des forts changements d'illumination (séquence 2).

3.7/ CONCLUSION

Dans ce chapitre, nous avons proposé une approche performante permettant l'adaptation des méthodes de suivi visuel basée sur une représentation sphérique de l'image et une modélisation de la cible avec un histogramme couleur (pondéré par un noyau adapté assurant une qualité de suivi élevée). Les résultats expérimentaux fondés sur des critères d'estimation tangible (dans plusieurs séquences et dans différentes conditions) ont permis de montrer une amélioration significative de la précision du suivi. Notre méthode d'adaptation permet de manière générale d'améliorer la robustesse aux occultations, aux changements d'échelle et aux changements d'illumination.

Tableau 3.1 – Particularités des séquences utilisées

	Localisation	Type de caméra	Changement d'illumination	Occultation	Cible	Images/sec	Nombre d'images
Séquence 1	Extérieure	Non centrale	Non	Non	Personne	30	780
Séquence 2	Intérieure	Non centrale	surface réfléchissante	Non	Objet	30	649
Séquence 3	Extérieure	Non centrale	Éblouissement	Partiel	Personne	30	602
Séquence 4	Intérieure	Centrale	Non	Total	Objet	15	481

Tableau 3.2 – Résultats moyens obtenus avec un filtre particulaire conventionnel et avec la méthode adaptée

Séquence	Méthode	superposition spatiale	distance des centres (p)	superposition temporelle
Sequence 1	Méthode conventionnelle	45%	5.7	100%
	Notre méthode	71%	2.6	100%
Sequence 2	Méthode conventionnelle	30%	5.5	44%
	Notre méthode	60%	3.8	99.7%
Sequence 3	Méthode conventionnelle	33%	7.7	88%
	Notre méthode	65%	4.8	99.5%
Sequence 4	Méthode conventionnelle	45%	6	96.8%
	Notre méthode	66%	4.1	97.3%

Tableau 3.3 – Résultats moyens obtenus avec l'algorithme *Mean-Shift* conventionnel et avec la méthode adaptée

Séquence	Méthode	superposition spatiale	distance des centres (p)	superposition temporelle
Sequence 1	Méthode conventionnelle	32.4%	6.38	63.21%
	Notre méthode	64.5%	6.2	96.35%
Sequence 2	Méthode conventionnelle	42.5%	16.2	89%
	Notre méthode	47.79%	8.3	90%
Sequence 3	Méthode conventionnelle	40.55%	7.8	96%
	Notre méthode	61.8%	7.5	97.84%
Sequence 4	Méthode conventionnelle	34.72%	10.52	52.81%
	Notre méthode	67.67%	5.02	97.71%

Tableau 3.4 – Résultats moyens obtenus avec un filtre particulaire adapté, avec et sans noyau

Séquence	Noyau	superposition spatiale	distance des centres (p)	superposition temporelle
Sequence 1	Sans	62,47%	4,93	98,7%
	Avec	65,67%	4,01	98,7%
Sequence 2	Sans	33,28%	4,46	80,7%
	Avec	40,26%	4,46	90,6%
Sequence 3	Sans	54,37%	5,35	96%
	Avec	59,06%	5,04	99%

4

AUTO-CALIBRAGE DE CAMÉRA PTZ

Ce chapitre a pour but d'étudier la caméra mécanisée utilisée dans notre système, nous nous intéressons tout particulièrement à l'auto-calibrage de ce type de caméra (voir figure 4.1).

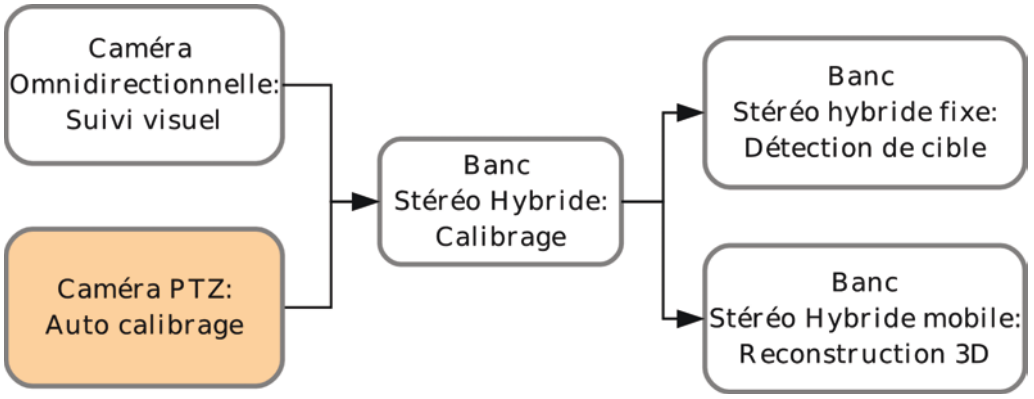


FIGURE 4.1 – Problématique globale

Les caméras PTZ (Pan-Tilt-Zoom) sont dites actives car elles peuvent être mécaniquement orientées dans de multiples directions suivant une rotation sur l'axe Y et une inclinaison sur l'axe X comme l'illustre la figure 4.2. Ces caméras sont également équipées d'un zoom leur permettant d'obtenir des images de hautes résolutions sur une zone particulière. La possibilité de couvrir une large zone et à la fois de fournir des images de grandes résolution sur des régions d'intérêt font des caméras PTZ un outil particulièrement adapté pour la vidéo surveillance [111], mais aussi pour la navigation robotique, la création de panorama, la vidéo conférence,... etc. [156, 167]. Cependant, cette polyvalence offerte par le mouvement et le zoom de la caméra peut aussi représenter un désavantage puisque toute modification du niveau de zoom entraînera un changement de la géométrie interne de la caméra tandis que la rotation affectera sa pose.

Toutes les applications mentionnées précédemment nécessitent une bonne estimation des paramètres internes de la caméra, que ce soit pour son contrôle ou encore pour la création d'images panoramiques. Le calibrage de caméra PTZ est donc une étape cru-

FIGURE 4.2 – Exemple de caméra PTZ

ciale puisqu'elle permet l'obtention de l'ensemble de ces paramètres représentatifs de la projection de l'espace tri-dimensionnel de la scène sur le plan image tel que discuté au chapitre 2. De nombreuses méthodes permettent d'étalonner des caméras hors ligne à l'aide de mires [179]. Ces méthodes sont très efficaces lorsqu'il s'agit de caméras à paramètres intrinsèques fixes. Cependant, dans le cas d'une caméra à focale variable les approches de calibrage classiques sont inadaptées pour déterminer ces paramètres de manière concluante.
C'est dans cette optique que Sturm a proposé dans [161] une méthode consistant à pré-calibrer la caméra pour différents niveaux de zoom afin d'établir un modèle d'interdépendance entre les paramètres internes de la caméra. Cette approche fonctionne en pratique mais cette étape de pré-calibrage peut être particulièrement fastidieuse. L'auto-calibrage, quant à lui permet l'estimation des paramètres de la caméra simplement à l'aide de points mis en correspondance entre images dans une scène inconnue. C'est par conséquent une approche beaucoup plus souple qu'un calibrage nécessitant une mire. Les fondements de l'auto-calibrage de caméras mobiles à paramètres fixes ont été proposés par Faugeras *et al.* il y a maintenant plus de vingt ans dans [65]. Le cas des caméras à rotation pure requiert cependant une formulation différente [3].
Dans ce chapitre nous présentons une méthode permettant d'améliorer la robustesse et la précision de l'auto-calibrage de caméra PTZ. Cette approche est basée sur plusieurs contraintes formulées sous forme d'Inégalité Matricielle Linéaire (LMI) appliquées sur certains paramètres internes de la caméra. Ces contraintes reposent sur des connaissances *a priori* concernant le rapport d'aspect des pixels (*Pixel Aspect Ratio* - PAR) et sur la position du point central. Ces contraintes sont ajustables et nécessitent seulement une connaissance limitée sur la caméra. Des expérimentations menées avec des données réelles et synthétiques montrent une amélioration significative en terme de précision et de stabilité grâce à l'emploi des contraintes susmentionnées.

Ce chapitre est organisé de la manière suivante. La Section 4.1 présente les outils théoriques nécessaires à la compréhension du problème. Dans la Section 4.2 une bibliographie concernant le calibrage de caméra stationnaire est proposée. Nous présentons ensuite notre méthode d'auto-calibrage dans la Section 4.3. La Section 4.4 est dédiée à la présentation des résultats obtenus avec des données réelles et simulées. Finalement,

la Section 4.5 conclura ce chapitre.

4.1/ THÉORIE

4.1.1/ HOMOGRAPHIE À L'INFINI

Comme définie dans la Section 2.4, une homographie infinie $\mathbf{H}_\infty$ est la transformation induite entre le plan à l'infini π_∞ et un plan réel. Pour un plan à l'infini $d = \infty$, on obtient donc une homographie inter-image :

$$\mathbf{H}_\infty = \lim_{d\to\infty} \mathbf{H} = \mathbf{K_2 R_{12} K_1^{-1}}, \tag{4.1}$$

où $\mathbf{K_1}$ et $\mathbf{K_2}$ sont respectivement les paramètres intrinsèques des caméras 1 et 2, tandis que $\mathbf{R_{12}}$ est la matrice exprimant la rotation entre ces deux caméras. On remarquera que cette homographie n'est dépendante que de la rotation et des paramètres intrinsèques existants entre les vues. Ce type d'homographie peut se calculer à l'aide des points de fuite (qui sont par définition sur le plan à l'infini), mais il est également possible de l'estimer dans le cas particulier d'une caméra stationnaire. En effet, une caméra de ce type se démarque par le fait qu'elle n'effectue que des rotations pures (ce qui implique qu'aucune translation n'est possible $t_x = t_y = t_z = 0$). Dans ces conditions, le terme de l'homographie inter-image $\mathbf{t n}^T/d$ sera égal à zéro, ce qui conduit à la relation mathématique démontrée dans l'équation (4.1). Cela signifie également que pour une caméra PTZ (ou Pan-Tilt) $\mathbf{H}_\infty$ peut être calculée à partir de n'importe quelle correspondance de points entre deux vues successives (voir figure 4.3). La relation inter-image peut donc être exprimée par $\mathbf{p_1} = \mathbf{H}_\infty \mathbf{p_2}$.

4.1.2/ CONIQUE ABSOLUE

La conique absolue $\mathbf{\Omega}_\infty$ est une conique particulière, située sur le plan à l'infini, uniquement constituée de points imaginaires. La projection de cette conique sur le plan image est appelée l'image de la conique absolue (ICA), notée ω. L'ICA est invariante à la pose de la caméra, elle est donc exclusivement liée aux paramètres internes de la caméra.

$$\omega = \mathbf{K}^{-T}\mathbf{K}^{-1} \tag{4.2}$$

$$\omega = \begin{bmatrix} 1/(\lambda f)^2 & s & -u_0/(\lambda f)^2 \\ s & 1/f^2 & -v_0/f^2 \\ -u_0/(\lambda f)^2 & -v_0/f^2 & 1 + u_0^2/(\lambda f)^2 + v_0^2/f^2 \end{bmatrix} = \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{bmatrix} \tag{4.3}$$

Cette conique est donc particulièrement intéressante car il est simple d'en extraire les

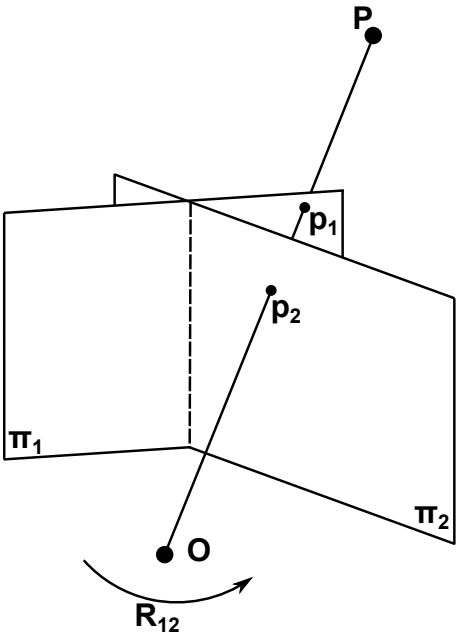


FIGURE 4.3 – Homographie inter-image

paramètres intrinsèques par une décomposition de Cholesky [170].
Dans cette thèse, une autre notion dérivée de l'ICA est également employée. Il s'agit de l'image duale de la conique absolue (ω^*) qui n'est autre que l'inverse de l'image de la conique absolue :

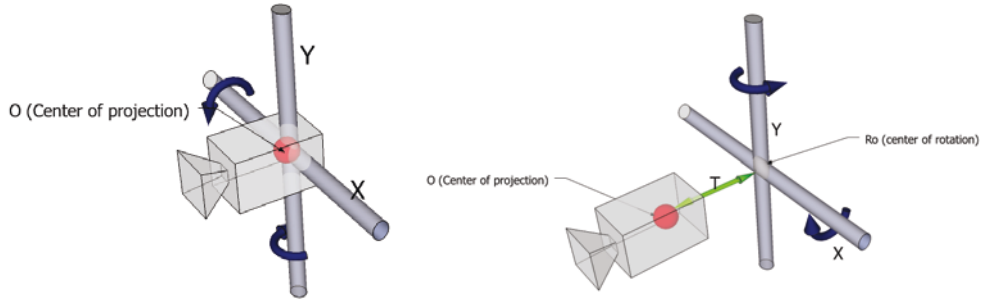
$$\omega^* = \omega^{-1} = \mathbf{K}\mathbf{K}^T \tag{4.4}$$

$$\omega^* = \begin{bmatrix} f^2 + s^2 + u_0^2 & s.u_0.v_0 & u_0 \\ s.u_0.v_0 & \lambda f^2 + v_0^2 & v_0 \\ u_0 & v_0 & 1 \end{bmatrix} = \begin{bmatrix} \omega_{11}^* & \omega_{12}^* & \omega_{13}^* \\ \omega_{21}^* & \omega_{22}^* & \omega_{23}^* \\ \omega_{31}^* & \omega_{32}^* & \omega_{33}^* \end{bmatrix} \tag{4.5}$$

4.1.3/ LES CAMÉRAS STATIONNAIRES

Dans notre étude, nous admettons l'hypothèse que le centre optique et le centre de rotation sont confondus quels que soient les mouvements de la caméra. Cette supposition s'avère cependant fausse en pratique car le mécanisme permettant la rotation ne permet pas d'obtenir ce cas idéal. Une translation est toujours présente et peut même représenter une forte dérive comme on peut le voir sur la figure 4.4. Cette approximation est donc fortement dépendante de la qualité de la caméra utilisée, mais également du rapport existant entre cette translation et la distance des éléments de la scène. Cependant, si la translation est très faible et que l'on considère un ensemble de points 3D éloignés, alors l'influence de la translation sur la formation de l'image sera moindre.

Cette hypothèse de départ nous permet d'établir que la projection d'un point dans le



(a) (b)

FIGURE 4.4 – Emplacement du centre optique et du centre de rotation (a) cas idéal (b) cas réel

monde $\mathbf{P}$ sur deux images successives (obtenue par rotation) peut s'exprimer par $\mathbf{p}_1 = \mathbf{K}_1\mathbf{R}_1\mathbf{P}$ et $\mathbf{p}_2 = \mathbf{K}_2\mathbf{R}_2\mathbf{P}$. On obtient alors la relation suivante :

$$\mathbf{p_2} = \mathbf{K_2R_2R_1^{-1}K_1^{-1}p_1}. \tag{4.6}$$

On peut donc en déduire que :

$$\mathbf{H_{12}} = \mathbf{K_2R_2R_1^{-1}K_1^{-1}} = \mathbf{K_2R_{12}K_1^{-1}}. \tag{4.7}$$

On retrouve bien la formulation ($\mathbf{H_{12}} = \mathbf{H}_\infty$) de l'homographie à l'infini déjà développée dans la section 4.1.1. Il est ensuite possible de réécrire cette équation sous la forme suivante :

$$\mathbf{R_{12}} = \mathbf{K_2}^{-1}\mathbf{H_{12}K_1}. \tag{4.8}$$

Sachant qu'une matrice de rotation satisfait certaines propriétés telles que $\mathbf{R_{12}} = \mathbf{R_{12}}^{-T}$ et $\mathbf{R_{12}}^{-1} = \mathbf{R_{12}}^{T}$. Nous obtenons donc deux formulations dépendantes uniquement de l'homographie à l'infini et de l'ICA (ou de sa duale) :

$$\mathbf{K}_2\mathbf{K}_2^T = \mathbf{H_{12}}(\mathbf{K}_1\mathbf{K}_1^T)\mathbf{H_{12}}^T \tag{4.9}$$

$$\omega^{2*} = \mathbf{H_{12}}\omega^{1*}\mathbf{H_{12}}^T \tag{4.10}$$

$$\mathbf{K_1}_2^{-T}\mathbf{K_1}_2^{-1} = \mathbf{H_{12}}^{-T}(\mathbf{K}_1^{-T}\mathbf{K}_1^{-1})\mathbf{H_{12}}^{-1} \tag{4.11}$$

$$\omega^2 = \mathbf{H_{12}}^{-T}\omega^1\mathbf{H_{12}}^{-1} \tag{4.12}$$

Chacune des formulations (équations (4.10) et (4.12)) permet d'imposer un certain nombre de contraintes.

Le problème d'échelle dans des relations homogènes peut être résolu en normalisant $\mathbf{H}$ de manière à obtenir $det(\mathbf{H}) = 1$:

$$\mathbf{H} = \frac{\mathbf{H}}{det(\mathbf{H})^{\frac{1}{3}}} \tag{4.13}$$

4.1.4/ INÉGALITÉ MATRICIELLE LINÉAIRE

Dans ce chapitre nous proposons de résoudre le problème d'auto-calibrage d'une caméra de type PTZ de manière convexe à l'aide d'inégalités matricielles linéaires.
Une inégalité matricielle linéaire (*Linear matrix Inequality* - LMI) est une expression de la forme :

$$\mathbf{F}(\mathbf{X}) = \mathbf{F}_0 + \sum_{i=1}^{n} x_i \mathbf{F}_i \succ 0, \tag{4.14}$$

où $\mathbf{X} = (x_1, \cdots, x_n)$ est un vecteur de scalaires réels appelés variables de décision. $\mathbf{F}_0 \cdots \mathbf{F}_n$ sont des matrices réelles symétriques tandis que l'expression $\mathbf{F}_i \succ 0$ signifie que la matrice $\mathbf{F}_i$ est semi-définie positive (toutes les valeurs propres de $\mathbf{F}_i$ sont positives). Dans ce travail, nous utilisons des contraintes formulées sous forme de LMI afin de résoudre un problème d'optimisation convexe de la forme :

$$\min_X \mathbf{c}^T \mathbf{X} \quad \text{s.t.} \quad \mathbf{F}(\mathbf{X}) \succ 0 \tag{4.15}$$

où $\mathbf{c}$ est un vecteur modélisant le problème.
Il s'agit d'une généralisation du problème de programmation linéaire au cone des matrices définies positives. On parle ici d'optimisation convexe puisque le problème à résoudre $\mathbf{c}^T\mathbf{X}$ est linéaire tandis que la contrainte est imposée par la condition de positivité de $\mathbf{F}(\mathbf{X})$ qui est elle même convexe par définition. L'avantage des problèmes convexes est de ne posséder qu'un seul minimum, c'est-à-dire le minimum global.
Des détails plus poussés concernant l'optimisation à l'aide d'inégalité matricielle linéaire sont disponibles dans [33].

Les LMIs peuvent donc servir à résoudre des problèmes linéaires sous contraintes. Ces contraintes peuvent parfois être non-linéaires si il est possible de les reformuler à l'aide du complément de Schur sous forme de LMIs. C'est à l'aide de cet outil mathématique que certaines des contraintes d'optimisation présentées dans cette thèse ont été conçues.
Une matrice $\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{bmatrix}$ est définie positive, si et seulement si le complément de Schur suivant est satisfait :

$$\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T > 0. \tag{4.16}$$

4.2/ TRAVAUX ANTÉRIEURS

La popularité des caméras stationnaires motorisées dans les systèmes de vidéo-surveillance a attiré l'attention de la communauté sur l'aspect de l'auto-calibrage depuis plus de vingt ans. Par exemple, dans [59] et [23], les auteurs proposent différentes solutions pour auto-calibrer des caméras Pan-Tilt (PT). Ces méthodes sont cependant parti-

culièrement sensibles au bruit et ne peuvent en pratique qu'être utilisées pour initialiser un processus d'optimisation itératif. Dans [82], Hartley propose d'utiliser l'homographie entre deux images consécutives de façon à calculer linéairement les inconnues de l'image de la conique absolue duale (ICAD) dans le cas où les paramètres internes de la caméra restent inchangés sur toutes les images. Une fois l'ICAD résolue, les paramètres de la caméras $\mathbf{K}$ peuvent être obtenus à l'aide d'une décomposition de Cholesky. Notons qu'il est également possible de résoudre le problème d'auto-étalonnage linéairement avec des informations additionnelles concernant la rotation [159] de la caméra. De plus, Ji *et al.* [98] ont développé une stratégie permettant d'annuler l'effet de la translation résiduelle décrite dans la figure 4.4. Cependant, toutes les méthodes décrites précédemment ne prennent en compte que le cas où les paramètres intrinsèques de la caméra restent constants.

Avec une caméra PTZ, les paramètres intrinsèques peuvent être amenés à varier à chaque nouvelle image. Dans [3], Agapito *et al.* fournissent une reformulation du problème avec laquelle il est possible de prendre en compte les changements des paramètres internes de la caméra. Pour ce faire, les auteurs ont exprimé le problème à l'aide de l'image de la conique absolue plutôt qu'avec sa duale (comme cela était le cas avec la méthode proposée par Hartley pour les caméras à paramètres intrinsèques fixes). Cette nouvelle formulation permet d'imposer un ensemble de contraintes linéaires sur différents paramètres de la caméra (ces contraintes sont résumées dans le tableau 4.1). En pratique, la contrainte d'orthogonalité des pixels, *i.e.* $s = 0$, donne de mauvais résultats lorsqu'elle est utilisée seule. En général une contrainte additionnelle sur le rapport d'aspect des pixels $\lambda = 1$ est employée et permet alors de déterminer les paramètres intrinsèques de la caméra avec seulement trois images. Le désavantage majeur de cette approche est sa sensibilité au bruit menant à une estimation biaisée de l'ICA. Si l'ICA calculée n'est pas strictement définie positive la décomposition de Cholesky nécessaire à l'extraction de $\mathbf{K}$ échouera. Ce problème se manifeste souvent en présence de bruit comme le souligne les expériences menées dans [115]. Lorsque les paramètres intrinsèques sont extraits avec succès, la solution optimale peut être obtenue par raffinement avec l'utilisation d'un ajustement de faisceaux [156] ou en minimisant une fonction de coût dérivée de (4.12) [3].

Plus récemment, Agrawal *et. al* ont introduit l'optimisation basée sur des systèmes de LMI (Linear Matrix Inequality) dans la communauté de la vision par ordinateur. Les LMIs étant initialement utilisées dans le domaine de l'automatisme afin d'analyser ou

Condition	Contrainte	Type	Nb d'images
$s = 0$	$\omega_{12} = 0$	linéaire	5
$u_0 = v_0 = 0$	$\omega_{12} = \omega_{33} = 0$	linéaire	3
r=f_x/f_y	$\omega_{11} = r^2\omega_{22}$	linéaire	2
λ constant	$\omega^i_{11}/\omega^i_{22} = \omega^j_{11}/\omega^j_{22}$	quadratique	2

Tableau 4.1 – Contraintes possible sur ω

d'asservir des systèmes (l'exemple basique étant la stabilité de Lyapunov). Dans [7], une méthode d'auto-calibrage d'une caméra à paramètres internes fixes basée sur l'utilisation de sphères est proposée. Le même auteur publie par la suite une approche d'auto-calibrage plus générale, utilisant également la programmation semi-définie [6]. L'utilisation de LMI est ensuite appliquée au problème de l'auto-calibrage de caméra rotative par Li dans [115], où il reprend l'équation (4.12) qu'il reformule sous forme d'inégalité matricielle.
L'utilisation de la programmation semi-définie dans le contexte du calibrage de caméra présente des avantages non-négligeables. Elle permet en effet de prendre en compte la spécificité des matrices définies positive (ce qui est le cas pour l'ICA ω) de manière à éviter les problèmes de décomposition en matrice intrinsèque évoqués précédemment. De plus, les LMIs permettent toujours de converger vers une solution (résolution de problème convexe) au contraire des méthodes de résolution non-linéaires.

Ce mode de résolution nécessite cependant une reformulation du problème, et dans [115] l'auteur propose l'approche suivante :
Si l'on considère 2 homographies ($\mathbf{H}_{12}$, $\mathbf{H}_{13}$) résultant de 3 prises vues (I_1,I_2,I_3), on aura alors les équations suivantes :

$$\omega = \mathbf{H}_{12}^{-T}\omega\mathbf{H}_{12}^{-1} \tag{4.17}$$
$$\omega = \mathbf{H}_{13}^{-T}\omega\mathbf{H}_{13}^{-1} \tag{4.18}$$

On peut reformuler cet ensemble d'équations comme un problème à minimiser :

$$\min_{\omega}\{\|\omega - \mathbf{H}_{12}^{-T}\omega\mathbf{H}_{12}^{-1}\|_2 + \|\omega - \mathbf{H}_{13}^{-T}\omega\mathbf{H}_{13}^{-1}\|_2\} \tag{4.19}$$
$$\text{avec } \omega \succ 0 \tag{4.20}$$

Ici $\|\cdot\|_2$ correspond à la norme $L2$. Afin de résoudre ce problème sous forme d'inégalité matricielle, il est nécessaire d'y introduire des bornes $t_1 \geqslant 0$ et $t_2 \geqslant 0$ de la manière suivante :

$$\min_{\omega}(t1 + t2) \tag{4.21}$$
$$\text{tel que} \quad \|\omega - \mathbf{H}_{12}^{-T}\omega\mathbf{H}_{12}^{-1}\|_2 < t1 \tag{4.22}$$
$$\|\omega - \mathbf{H}_{13}^{-T}\omega\mathbf{H}_{13}^{-1}\|_2 < t2 \tag{4.23}$$
$$\omega \succ 0 \tag{4.24}$$

Il est à présent possible de réécrire ces équations comme 3 conditions d'un système d'inégalité matricielle linéaire :
Condition 1 :

$$C1 = \begin{bmatrix} t_1\mathbf{I} & \omega - \mathbf{H}_{12}^{-T}\omega\mathbf{H}_{12}^{-1} \\ (\omega - \mathbf{H}_{12}^{-T}\omega\mathbf{H}_{12}^{-1})^T & t_1\mathbf{I} \end{bmatrix} \succ 0 \tag{4.25}$$

Cette condition correspond à la contrainte de l'équation (4.22).

Condition 2 :

$$C2 = \begin{bmatrix} t_2\mathbf{I} & \omega - \mathbf{H}_{13}^{-T}\omega\mathbf{H}_{13}^{-1} \\ (\omega - \mathbf{H}_{13}^{-T}\omega\mathbf{H}_{13}^{-1})^T & t_2\mathbf{I} \end{bmatrix} \succ 0 \tag{4.26}$$

Cette condition correspond à la contrainte de l'équation (4.23).

Condition 3 :

$$C3 = \omega \succ 0 \tag{4.27}$$

Cette condition correspond à la contrainte de l'équation (4.24).

Avec $\mathbf{I}$ une matrice identité 3×3, cette formulation permet d'imposer la norme spectrale déjà utilisée dans [6]. Cette norme est la norme subordonnée de la norme euclidienne, *i.e* : $\| A \|_{sp} = \max_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|}$ où $\| \cdot \|$ est la norme euclidienne.
La méthode décrite précédemment nécessite au minimum 3 images afin d'obtenir les paramètres de la caméra, cependant un plus grand nombre d'homographies peut être utilisé.

4.3/ L'APPROCHE PROPOSÉE

La méthode présentée ici est une extension de l'approche d'auto-calibrage basée sur l'utilisation de LMI de Li *et al.* [115]. Dans leur papier, les auteurs ont reformulé les méthodes d'auto-calibrage proposées dans [82] et [3] sous forme d'optimisation LMI permettant par conséquent de forcer la "positivité" de l'ICA ω. Grâce à cette nouvelle contrainte, la stabilité de l'algorithme est fortement améliorée sans pour autant apporter un gain significatif en terme de précision. La méthode présentée précédemment [115] permet de tirer avantage des LMIs mais n'utilise pas le plein potentiel de la programmation semi-définie. Il est en effet possible d'imposer de nouvelles contraintes afin de rendre plus robuste l'estimation des paramètres constitutifs de l'image de la conique absolue (ou de sa duale). De plus, la majorité des approches d'auto-calibrage impose une contrainte "dure" uniquement sur la première ICA (sur la première image), on peut par conséquent parler de contraintes "souples" sur les autres coniques. Dans notre méthode, nous utilisons les LMIs de manières à forcer les mêmes contraintes sur l'ensemble des images. Il en résulte une meilleure description du problème. Nous présentons ici les nouvelles conditions que nous imposons dans notre système de LMIs.

Afin de faciliter l'accès aux éléments de la conique une série de vecteurs est déclarée dans un premiers temps : $\mathbf{L}_1 = [1,0,0], \mathbf{L}_2 = [0,1,0], \mathbf{L}_3 = [0,0,1], \mathbf{R}_1 = \mathbf{L}_1^T, \mathbf{R}_2 = \mathbf{L}_2^T, \mathbf{R}_3 = \mathbf{L}_3^T$. A présent si l'on veut par exemple accéder à l'élément ω_{23} on aura : $\omega_{23} = \mathbf{L}_2\omega\mathbf{R}_3$.
Dans toutes les contraintes présentées, i correspond au numéro de l'image.

Condition 1 : Orthogonalité des pixels $-\varepsilon <$s$< \varepsilon$

A l'instar de nombreuses autres approches nous imposons $s = 0$. Ce paramètre ne représente pas une hypothèse forte puisque celle-ci reste très souvent vérifiée pour les caméras actuelles, cette contrainte peut s'exprimer de la manière suivante :

$$C_1^i = \begin{bmatrix} \varepsilon & \omega_{12}^i \\ \omega_{12}^i & \varepsilon \end{bmatrix} > 0.$$

Puisque nous passons par l'ICA, il est aisé de forcer le paramètre s à une valeur proche de zéros en fixant un ε très petit (ou en le minimisant).

Conditions 2 et 3 : rapport d'aspect des pixels (PAR) $\delta_1 < \lambda < \delta_2$

Lorsqu'une caméra tourne autour d'un seul de ses axes, l'information relative à la distance focale sur l'autre axe est perdue. Nous pouvons alors parler de mouvement dégénéré. Pour pallier ce problème, nous proposons une contrainte directement incorporée dans notre système d'inégalité matricielle pour forcer le paramètre λ.
Dans le cas d'une caméra PT, nous offrons la possibilité de fixer le PAR entre deux bornes ajustables, par exemple $0.75 < \lambda < 1.25$, ce qui est particulièrement raisonnable, et permet d'améliorer la robustesse de l'estimation de l'ICA (notamment en cas de mouvements dégénérés).
Pour une caméra PTZ, notre approche permet de forcer λ à une valeur choisie. Cette contrainte permet en conséquence de gérer tous les mouvements de rotation, même dans les cas dégénérés. Si aucun *a priori* n'est connu, il est possible de considérer un rapport d'aspect de pixel unitaire ($\lambda = 1$) comme c'est le cas dans la plupart des papiers traitants de l'auto-calibrage de caméras à paramètres variables. Cependant, puisque λ est constant quels que soient les autres paramètres, il est possible de calibrer la caméra préalablement à n'importe quel niveau de zoom pour en déterminer sa valeur réelle et l'imposer dans notre système de résolution.
Ces deux bornes peuvent être imposées séparément :

borne inférieure	borne supérieure
$\frac{f_y}{f_x} < \delta_2$	$\frac{f_y}{f_x} > \delta_1$
$C_2^i = \omega_{11}^i - \omega_{22}^i \delta_2^2 > 0$	$C_3^i = \omega_{22}^i \delta_1^2 - \omega_{11}^i > 0$

Conditions 4 et 5 : Le point principal proche du centre de l'image

Comme Agapito l'a démontré dans [3], si le point principal n'est pas contraint il est très sensible au bruit. Il est généralement admis [175] que le point principal (u_0, v_0) est proche du centre de l'image (x_c, y_c). Cela reste vrai même dans le cas d'une caméra

munie d'une focale variable.
Dans [85], Hartley *et al.* ont démontré que la relaxation du point principal mène à une meilleure estimation de la distance focal. La détermination de ce point central peut être limitée dans une zone proche du centre de l'image (voir figure 4.5) :

$$(u_0 - x_c)^2 < d^2 \rightarrow d^2 - (u_0 - x_c)^2 > 0 \tag{4.28}$$
$$(v_0 - y_c)^2 < d^2 \rightarrow d^2 - (v_0 - y_c)^2 > 0 \tag{4.29}$$

où d est la distance maximale (en pixels) du point principal par rapport au centre de l'image. Les équations précédentes peuvent être reformulées à partir des éléments disponibles dans ω :

$$d^2\omega_{11} - (\omega_{13} - x_c\omega_{11}) > 0 \rightarrow \frac{d^2}{f_x^2} - \frac{(u_0 - x_c)^2}{f_x^2} > 0 \tag{4.30}$$
$$d^2\omega_{22} - (\omega_{23} - y_c\omega_{22}) > 0 \rightarrow \frac{d^2}{f_y^2} - \frac{(v_0 - y_c)^2}{f_y^2} > 0. \tag{4.31}$$

On a donc deux termes supplémentaires dans notre système :

$$C_4^i = d^2\omega_{11}^i - (\omega_{13}^i - x_c\omega_{11}^i) > 0$$
$$C_5^i = d^2\omega_{22}^i - (\omega_{23}^i - y_c\omega_{22}^i) > 0$$

Cette contrainte doit être utilisée avec précaution. Si les bornes fixées sont trop restrictives et si le point principal est en faite hors de la zone définie par ces bornes, alors l'estimation des autre paramètres s'en trouvera affectée. Prendre des bornes considérant le point principal à $\pm 10\%$ du centre de l'image est une contrainte raisonnable pour la plupart des caméras.

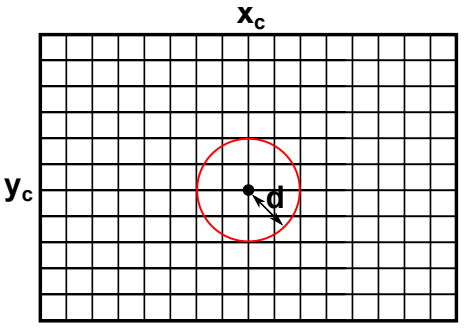


FIGURE 4.5 – Exemple où le cercle rouge limite la zone de recherche du point principal dans l'image

Condition 6 : L'ICA est une matrice définie positive $C_6^i = \omega^i > 0$

Les conditions de 1 à 6 doivent être forcées pour chacune des ICAs, c'est à dire pour chaque nouvelle image. Formant ainsi un système de LMI où tous les ω sont affectés par ces contraintes dures.

Condition 7 : $\omega^{i-1} = \mathbf{H}^{-T}\omega^{i}\mathbf{H}^{-1}$ Finalement, il est possible de minimiser l'erreur définie dans l'équation (4.12) (pour chaque homographie) en utilisant la norme spectrale :

$$C_7^i = \begin{bmatrix} t_i\mathbf{I} & \omega^{i-1} - \mathbf{H}^{-T}\omega^{i}\mathbf{H}^{-1} \\ (\omega^{i-1} - \mathbf{H}^{-T}\omega^{i}\mathbf{H}^{-1})^T & t_i\mathbf{I} \end{bmatrix} \succ 0$$

avec $\mathbf{I}$ la matrice identité 3×3 et t_i un scalaire à minimiser. Cette dernière condition permet de lier toutes les coniques entre elles, elles sont de cette manière toutes inter-dépendantes. L'unique différence lorsqu'il est question de calibrer une caméra PT est que $\omega^i = \omega^1$, c'est donc la même ICA que l'on recherche pour toutes les vues.

L'optimisation complète du problème peut être résumée de la manière suivante :

$$\min_{\omega, t_1 \cdots t_n} \sum_{i=0}^{n} t_i \tag{4.32}$$

$$\text{tel que} \begin{bmatrix} C_1^1 & 0 & \cdots & 0 \\ 0 & C_2^1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_7^n \end{bmatrix} \succ 0 \tag{4.33}$$

L'approche de résolution de ce problème est basée sur une méthode des points intérieurs décrite dans [134].

4.4/ EXPÉRIMENTATIONS

4.4.1/ ÉVALUATION AVEC DES DONNÉES SYNTHÉTIQUES

Afin d'évaluer les performances de l'algorithme proposé, une série de tests a été effectuée à l'aide d'un nuage de 5000 points 3D générés de manière aléatoire dans un cube de dimensions 1000x1000x1000. Nous avons restreint nos caméras virtuelles à un champ de vue limité de manière à avoir un nombre de points visibles de l'ordre d'une centaine par image pour calculer les homographies. La caméra synthétique est localisée au centre du nuage de points où elle effectue des rotations aléatoires comprises entre $-30°$ et $+30°$. Les points 3D sont alors projetés sur le plan image de 640x480 pixels. L'homographie inter-image est calculée linéairement sans raffinement. Afin de déterminer la robustesse de différents algorithmes, un bruit aléatoire est ajouté à la position des points sur l'image.

Différents niveaux de bruit sont utilisés et 1000 exécutions sont effectuées pour chacun d'eux. Tous les résultats présentés utilisent un ensemble de trois homographies pour les caméras PTZ et de deux homographies pour les caméras PT.

4.4.1.1/ CAMÉRA À PARAMÈTRES FIXES

Dans le cas d'une caméra PT l'évaluation est réalisée avec des paramètres intrinsèques choisis arbitrairement : $f = 900$, $\lambda = 0.8889$, $u_0 = 325$ et $v_0 = 240$. Dans ce cas trois algorithmes sont comparés : La méthode décrite par Hartley dans [82], l'approche basée LMI de Li *et al.* [115] et notre approche.
Les résultats visibles dans la figure 4.6 sont obtenus avec 3 rotations, ce qui est le minimum pour les autres méthodes de résolution (tandis que la nôtre n'en nécessite que deux). Ici notre algorithme est configuré de cette manière : $0.75 < \lambda < 1.25$, le point principal est forcé à être dans l'image et $s = 0$.

Dans cette série de tests, notre stratégie offre des résultats significativement meilleurs que les autres méthodes, pour plusieurs raisons. Tout d'abord, nous imposons des contraintes dures sur chaque conique, de plus le point principal est restreint à l'intérieur de l'image et le PAR est fixé entre deux bornes. Tous ces éléments conduisent à une meilleure robustesse du système sans même posséder de fortes connaissances *a priori* sur la caméra.

4.4.1.2/ CAMÉRA À PARAMÈTRES VARIABLES

Dans les tests suivants la caméra effectue un zoom entre chaque nouvelle image, la distance focale ainsi que le point principal vont donc varier. Dans le cas d'une caméra PTZ nous comparons notre méthode avec les techniques décrites dans [3] et [115]. Par convenance nous fixons le PAR à un : $\lambda = 1$. Notre configuration considère simplement un rapport d'aspect pixellique unitaire, une orthogonalité parfaite des pixels et un point principal dans l'image. Sans l'utilisation de contrainte a-priori, nous obtenons de meilleurs résultats que les méthodes existantes (voir figure 4.7). La forte inter-dépendance imposée entre toutes les images de la conique absolue justifie cette amélioration. En outre, prendre des bornes plus restrictives peut conduire à des résultats plus robustes encore. Notons que le nombre minimum d'images requis pour l'auto-calibrage est le même que pour les méthodes [3, 115] dans le cas d'une caméra à paramètres variables.

4.4.1.3/ INFLUENCE DE LA CONTRAINTE SUR LE PAR

Dans le cas d'une caméra Pan-Tilt, il est possible de restreindre le paramètre λ dans un intervalle défini, le test présenté figure 4.8 montre une amélioration en terme de robustesse lorsque la fourchette est proche de la valeur réelle de λ.

4.4.1.4/ INFLUENCE DE LA CONTRAINTE SUR LE POINT CENTRAL

Pour une caméra PTZ, il est préférable de garder un λ fixe pour assurer une meilleure stabilité, dans ce cas seule la restriction concernant la position du point principal sera importante dans l'estimation des paramètres internes. La seule contrainte existante concernant le point principal apparaissant dans la littérature est celle permettant de le fixer à une position connue. Nous suggérons ici une approche plus flexible en limitant sa position dans un intervalle. Les résultats présentés dans la figure 4.9 montrent l'influence des contraintes appliquées à la position du point principal sur l'estimation de la distance focale.
Même dans le cas où le PAR et le point principal sont connus, leurs véritables valeurs dévierons en présence de bruit. Cela signifie qu'imposer des contraintes trop strictes peut amener à une accumulation de l'erreur affectant par conséquent le calcul des paramètres inconnus. Au contraire, laisser le point principal totalement libre (non-contraint) peut pousser le système à converger vers un mauvais minimum. Forcer la position du point principal dans un intervalle est donc un très bon compromis pour éviter une mauvaise convergence mais aussi pour équilibrer l'erreur induite par le bruit. Les courbes visibles dans la figure 4.9 valident ces hypothèses, en présence de bruit le point principal limité par des bornes admet une meilleure précision qu'un point principal plus libre ou que celui fixé à sa valeur réelle.

4.4.2/ TESTS AVEC DES DONNÉES RÉELLES

4.4.2.1/ CAMÉRA PT

Les tests qui suivent ont été effectués à l'aide d'une simple webcam *Logitech Quickcam Sphere AF Web camera - pan / tilt*. Cette caméra fournie une image très bruitée de taille 640×480 pixels. Un calibrage initial à l'aide de la toolbox de Bouguet [31] fournie une estimation des paramètres intrinsèques que nous utiliserons comme vérité terrain.
Nous avons auto-calibré la caméra successivement avec deux et quatre images (voir les résultats table 4.2). Dans ce test, nous ne considérons aucun *a priori* sur la caméra : donc le PAR est restreint de la manière suivante $0.75 < \lambda < 1.25$ et la recherche du point principal se fait dans toute l'image. La comparaison entre notre méthode et les approches conventionnelles montre clairement une forte amélioration dans l'estimation de presque tous les paramètres de la caméra.

4.4.2.2/ CAMÉRA PTZ

Évaluer les performances d'auto-étalonnage pour ce type particulier de caméra est difficile puisqu'aucune vérité terrain n'est disponible. C'est la raison pour laquelle nous avons décidé de projeter les images sur une sphère à l'aide du modèle sphérique unifié pour

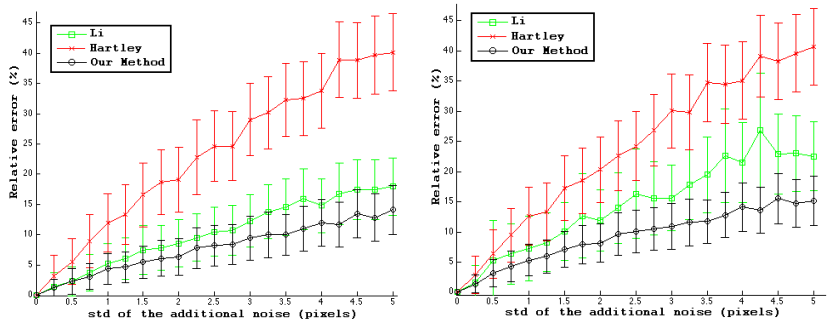
		fx		U_0		V_0		λ	
		pixels	Error (%)	pixels	Error (%)	Pixels	Error (%)		Error (%)
GT : Bouguet		904.6	0	327.6	0	277.5	0	1.0007	0
Hartley	4 images	959	5.6726	365	11.4164	265.2	4.4324	0.9265	7.4148
Our method	2 images	919.3	1.6250	325.7	0.5800	243.48	12.2595	1.0511	5.0365
	4 images	913.4	0.9728	321.22	1.9475	232.66	16.1586	1.0103	0.9593

Tableau 4.2 – Résultats obtenus avec des images réelles (acquises avec une caméra PT) avec différentes méthodes

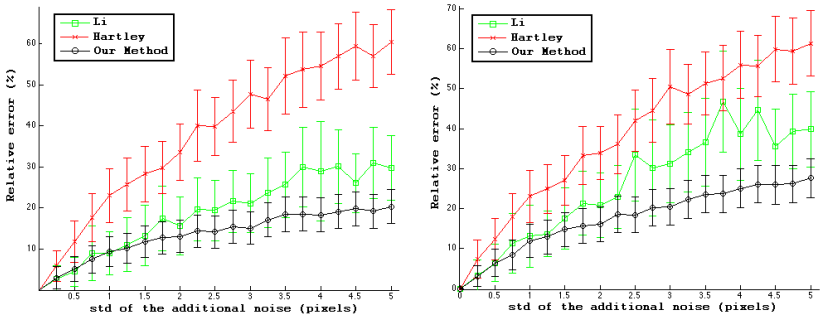
comparer la qualité des panoramas obtenus. En effet, cette modélisation permet de représenter les images acquises avec une caméra admettant le PVU sur une sphère, cette projection nécessite une bonne estimation des paramètres internes et externes. La qualité de l'image sphérique résultante est directement associée à la précision de l'autocalibrage de la caméra PTZ. Les panoramas obtenus sont visibles dans les figures 4.11 et 4.12, ceux calculés à l'aide de notre approche sont beaucoup plus précis et contiennent moins d'artefacts.

4.5/ CONCLUSION

Dans ce chapitre nous avons présenté une nouvelle approche utilisant les LMIs où plusieurs contraintes ajustables sont appliquées sur chaque conique. Nous avons également montré que les LMIs sont un outil remarquablement efficace et fiable pour les problèmes d'optimisation convexe avec des connaissances *a priori*. La programmation semi-définie n'est pas une stratégie courante dans le domaine de la vision par ordinateur alors qu'elle peut être particulièrement utile. Les résultats expérimentaux (particulièrement ceux concernant le point principal) ont montré une bien meilleure robustesse et permettent une estimation précise de tous les paramètres d'une caméra PTZ.

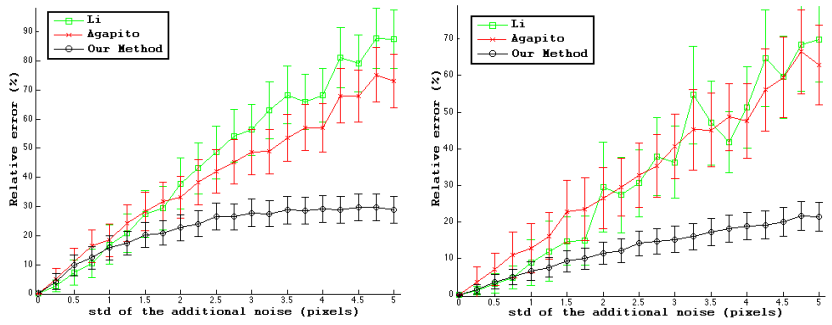


(a) f_x (b) f_y

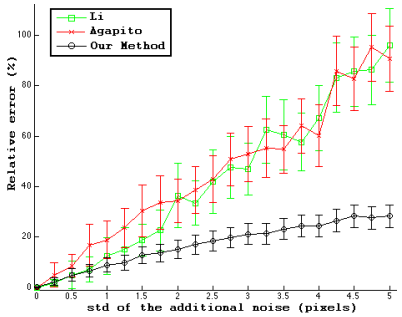


(c) u_0 (d) v_0

FIGURE 4.6 – Moyenne et écart type de l'erreur obtenue pour l'estimation des paramètres intrinsèques sur des données synthétiques avec des paramètres intrinsèques fixes



(a) f (b) u_0



(c) v_0

FIGURE 4.7 – Moyenne et écart type de l'erreur obtenue sur des données synthétiques avec des paramètres intrinsèques variables et $\lambda = 1$

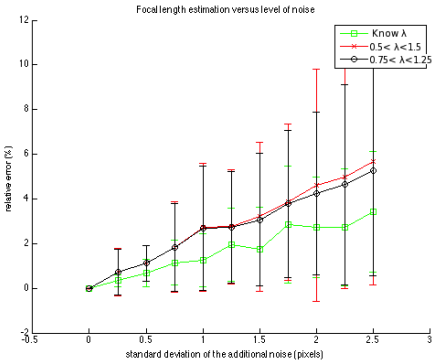


FIGURE 4.8 – Influence sur l'estimation de la distance focale f des contraintes imposées sur le PAR

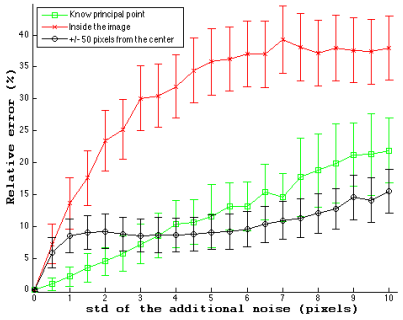


FIGURE 4.9 – Influence sur l'estimation de la distance focale f des contraintes imposées sur le point principal

(a)

(b)

(c)

(d)

FIGURE 4.10 – Images obtenues avec notre caméra PT

(a)

(b)

FIGURE 4.11 – Mosaïque sphérique multi-résolution de 5 images obtenues avec (a) Notre méthode (b) [115]. La colonne de gauche étant la vue sphérique complète et la colonne de droite un aperçu détaillé d'une région d'intérêt

(a)

(b)

FIGURE 4.12 – Mosaïque sphérique multi-résolution de 7 images obtenues avec (a) Notre méthode (b) [115]. La colonne de gauche étant la vue sphérique complète et la colonne de droite un aperçu détaillé d'une région d'intérêt

5

CALIBRAGE D'UN SYSTÈME DE STÉRÉO-VISION HYBRIDE

Dans ce chapitre nous développons le concept de système de vision hybride, et tout particulièrement du cas qui nous concerne c'est-à-dire l'usage conjoint d'une caméra omnidirectionnelle (voir chapitre 3) et d'une caméra PTZ (voir chapitre 4). Nous nous focalisons particulièrement sur le calibrage de ce type de capteur multi-caméras. Cette étape de calibrage est essentielle à toute procédure de reconstruction 3D de l'environnement mais est également nécessaire pour la plupart des approches de navigation robotique basée vision (voir figure 5.1). Si le calibrage d'un système de stéréo-vision

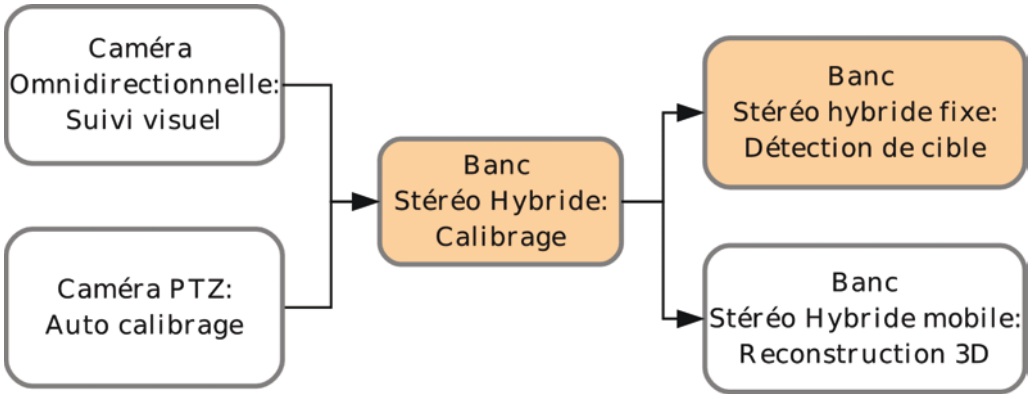


FIGURE 5.1 – Problématique globale

constitué de caméras perspectives est déjà bien connu, calibrer un système composé de caméras de différentes natures est plus spécifique et nécessite l'utilisation d'outils adaptés. Nous verrons au travers de ce chapitre que l'utilisation du modèle sphérique nous permet un calibrage simple et précis pour tout banc de caméras respectant le PVU. Notre calibrage sera évalué de manière quantitative et comparé avec une autre méthode de l'état de l'art. De plus, nous montrerons différentes applications rendues possible par notre étape de calibrage ; telles que la rectification d'images dans un système hétérogène et un système de surveillance reposant sur l'utilisation de la géométrie du capteur.

Ce chapitre est organisé de la manière suivante, dans les Sections 5.1 et 5.2 nous définissons ce que sont les systèmes de vision homogènes et hétérogènes. Dans la Section 5.3 nous proposons un état de l'art concernant le calibrage de systèmes de vision hybrides. Les Sections 5.4 et 5.5 décrivent la méthode de calibrage proposée respectivement pour l'estimation des paramètres intrinsèques et extrinsèques. Enfin, dans la Section 5.7, nous présentons une méthode de détection de cible pour la vidéo surveillance avec notre système de vision hybride. Finalement, la Section 5.8 conclura ce chapitre.

5.1/ LES SYSTÈMES DE VISION HOMOGÈNES

Les systèmes dit homogènes sont composés d'un réseau de caméras de même nature. La vision humaine peut d'ailleurs être assimilée à ce type de système [122]. Le fait d'employer des caméras possédants la même géométrie (voir les mêmes paramètres intrinsèques) facilite grandement l'étude de tels systèmes.

Le système homogène le plus fréquent dans le domaine de la vision par ordinateur, est le couple de deux caméras perspectives. Ce dispositif est couramment employé afin d'extraire des informations 3D de la scène. L'usage de deux caméras perspectives similaires permet un calibrage aisé ainsi qu'une mise en correspondance stéréoscopique efficace et cela pour un coût relativement faible. Dans [34], Bradley *et al.* décrivent une méthode de reconstruction 3D par mise en correspondance de points utilisant un système de vision binoculaire. Les possibilités offertes par les systèmes stéréoscopiques "conventionnels" sont vastes et concernent l'asservissement visuel [78], la vidéo surveillance [13, 93, 102, 181], la navigation robotique [62] la biométrie [172]...
Les bancs de caméras perspectives ont récemment connu un regain de popularité notamment avec les tournages des films en relief mais aussi avec l'apparition de solution "clé en main" comme celle offerte par le *Bumblebee* manufacturé par *Point Grey* (figure 5.2). Les plateformes stéréoscopiques utilisant des caméras d'un type plus spécifique

FIGURE 5.2 – Le système de stéréo-vision *Bumblebee*.

ont également été étudiées, bien que leur emploi soient moins courants ; ils fournissent généralement une plus grande flexibilité d'utilisation. C'est par exemple le cas pour les bancs de vision équipés de plusieurs caméras omnidirectionnelles. Cette configuration

fournit un champ de vision très étendu tout en conservant les avantages offerts par la stéréo-vision. On retrouve ce type de dispositif appliqué à la navigation robotique [68], où une vision grand angle de la scène assure l'existence de points d'intérêts sur les images. Dans le domaine de la vidéo surveillance, il n'est pas rare d'utiliser des réseaux de caméra PTZ [58]. Un assemblage de ce type de caméra, offre par exemple la possibilité d'utiliser différents modes de fonctionnement : coopératif ou indépendant. Le mode coopératif permet par exemple un suivi de la cible à l'aide des deux caméras et l'extraction d'informations 3D, comme la distance d'une cible, alors que l'utilisation des caméras de manière indépendante permet une surveillance autonome d'une zone et la recherche de cibles. Ce processus est décrit dans [182]. Ce système est cependant beaucoup plus complexe à gérer qu'un ensemble de caméras statiques car de nombreux paramètres, tels que les rotations et les paramètres intrinsèques des caméras, sont à prendre en compte.
D'autres systèmes de vision homogènes plus marginaux comme ceux faisant intervenir des caméras sensibles à des longueurs d'ondes hors du visible peuvent constituer un avantage majeur, comme c'est le cas dans [108] où des caméras infrarouges permettent une meilleure détection de piétons qu'un système de stéréo-vision classique.

5.2/ LES SYSTÈMES DE VISION HÉTÉROGÈNES

Le terme de stéréo-vision hybride signifie que les caméras employées au sein du banc sont de natures ou de modalités différentes [61]. Cette association permet l'obtention d'informations supplémentaires comme par exemple, une extension du champ de vision, l'ajout d'informations 3D ou l'étude d'une gamme de longueur d'onde plus importante.
Par exemple, dans [108] les auteurs proposent d'extraire des informations provenant à la fois d'un système binoculaire conventionnel mais aussi de deux caméras infrarouges afin de permettre une détection des piétons plus précise.

Les capteurs RGB-D de type "*Kinect*" (figure 5.3) employant une approche de reconstruction par projection de motif infrarouge peuvent également s'apparenter à des systèmes de vision hybride dans le sens où une caméra RGB est utilisée pour texturer la reconstruction tandis que la caméra infrarouge permet l'analyse d'un motif projeté permettant ainsi de reconstruire la scène. Ces deux capteurs sont très complémentaires. Par exemple, dans [91] la mise en correspondance des images couleurs permet de faciliter le recalage des nuages de points 3D.

FIGURE 5.3 – La *Kinect 2*.

	Caméra omnidirectionnelle	Caméra PTZ
Avantages	-Large champ de vision	-Grande mobilité -possibilité de zoomer sur une zone d'intérêt -haute résolution
Inconvénients	-Résolution non uniforme -Impossible d'avoir une observation précise d'une cible particulière -distorsion géométrique	-Champ de vision restreint

Tableau 5.1 – Comparaison caméra omnidirectionnelle et PTZ

Il existe beaucoup d'autres dispositifs originaux, on peut citer [90], où un capteur dédié à la vidéo surveillance est constitué d'une caméra haute résolution couplée avec deux caméras de faible résolution afin de simplifier la phase de détection d'événements.

Dans cette thèse, nous nous concentrerons sur le cas faisant intervenir des caméras possédant des propriétés géométriques différentes ; plus explicitement il s'agira de l'association d'une caméra perspective et omnidirectionnelle. Ce type de système a déjà été étudié auparavant notamment pour des applications dans le domaine de la vidéo-surveillance. En effet, l'association d'une caméra omnidirectionnelle et d'un réseau de caméras perspectives permet d'avoir une vision globale de la scène mais également un aperçu de la cible depuis plusieurs points de vue. Ce dispositif permet de fournir des informations plus précises concernant l'apparence ou la localisation dans l'espace d'un objet d'intérêt. Le calibrage d'un système de ce type est abordé dans [40] utilisant une caméra catadioptrique et un réseau de caméras perspectives.
On retrouve également cette configuration pour la navigation robotique, comme dans [2] où l'association d'un capteur catadioptrique et d'une caméra perspective est utilisée afin de permettre la détection d'obstacles et la navigation du robot. Dans un cadre différent, les auteurs de [63] proposent une méthode permettant de faire collaborer deux robots mobiles équipés de caméras ayant des géométries différentes.
Eynard *et al.* utilisent également ce type d'assemblage afin d'estimer l'attitude et l'altitude d'un drone [61] en tirant partie des deux caméras à l'aide d'une méthode de recalage dense par "plane sweeping".

En substituant la caméra perspective statique par une caméra active (PTZ), nous obtenons un système beaucoup plus flexible. Ces deux types de caméras possèdent des caractéristiques différentes (*cf.* tableau 5.1). Le système ainsi décrit offre à la fois une vision globale de la scène mais également la possibilité d'obtenir une image de haute précision sur des zones d'intérêts.

Clairement, l'ajout de la caméra PTZ permet de compenser les défauts inhérents à la caméra omnidirectionnelle en fournissant une plus grande flexibilité dans l'observation de la cible.
De nombreuses approches faisant intervenir un tel montage existent déjà, elles sont cependant presque exclusivement dédiées à la vidéo surveillance, comme dans [138, 152,

176, 50, 12]. Dans les sections suivantes, nous fournirons plus de détails sur ce type de méthodes et sur les limitations dont elles souffrent. Quelques application dédiées à la robotique existent tout de même, c'est par exemple le cas dans [135] où un robot "footballeur" peut naviguer et détecter une balle à l'aide d'une caméra de type catadioptrique tandis qu'une caméra perspective est utilisée afin de fournir une vue à la première personne du jeu. De la même manière, Adorni *et al.* [2] proposent une autre approche pour l'évitement d'obstacles en fusionnant des informations provenant d'une caméra PT et d'une caméra omnidirectionnelle. D'autres innovations destinées à la domotique voient aussi le jour comme dans [26] où un robot équipé des caméras susmentionnées peut automatiquement prendre des photos de visages lors d'évènements festifs.
Cependant, l'utilisation de ces systèmes pour la vidéo surveillance, le suivi visuel où la reconstruction 3D par stéréoscopie hybride n'est pas triviale et nécessite dans un premier temps un calibrage du système.

5.3/ LES MÉTHODES DE CALIBRAGE POUR LES SYSTÈMES DE VISION HYBRIDE

Le calibrage de système hétérogène dans la littérature peut explicitement se diviser en deux groupes.
D'une part, il existe les stratégies assujetties à la scène, elles sont presque exclusivement destinées à la vidéo surveillance avec un système de vision fixe. Dans [11, 116], l'approche proposée consiste à créer une table de correspondance représentative de la relation entre la commande de la caméra dynamique et les coordonnées image de la caméra omnidirectionnelle, ce qui suppose donc un environnement fixe. Toute modification de l'environnement nécessite donc un nouveau calibrage du système. De plus, c'est un processus long à mettre en œuvre dans le cas où des balises doivent être disposées dans la scène afin de permettre la création de cette table de correspondance. Certaines méthodes ne nécessitent pas la mise en place d'un tel dispositif en optant pour une mise en correspondance automatique des images [39, 11]. On trouve également une grande variété d'articles prenant avantage de différentes spécificités de la scène ou de l'installation du dispositif. Par exemple dans [152], la hauteur des caméras par rapport au sol est connue tandis que dans [30] c'est la planéité du sol qui est utilisée. Notons également l'existence de systèmes où les axes optiques des caméras sont alignés afin de simplifier le contrôle de la caméra rotative [1]. Dans [138] une approche plus élaborée permet de confondre le centre de rotation de la caméra PTZ avec le centre optique de la caméra grand angle à l'aide d'un dispositif optique particulier.
L'autre type de calibrage est un calibrage géométrique, c'est une approche plus flexible dans le sens où elle reste valide quelle que soit la scène observée. C'est le même type de calibrage standard qui permet la reconstruction 3D à l'aide d'un banc de vision stéréo classique. En d'autres termes, ce calibrage permet de retrouver les paramètres intrin-

sèques et extrinsèques liant les deux caméras. Il existe de nombreuses références et outils afin de calibrer un banc de stéréo vision classique, cependant peu de travaux se sont penchés sur le cas très particulier des systèmes de vision hybride. On notera tout de même l'existence des travaux de Caron et Eynard [38] traitant du calibrage d'un réseau de caméra hybride à PVU (perspective, *fisheye* et catadioptrique). Cette approche se base sur l'utilisation d'Asservissement Visuel Virtuel (AVV), un logiciel cross-plateforme -dénommée HysCas [1]- est également disponible.
A notre connaissance, très peu de travaux dans la littérature s'intéressent cependant au cas particulier d'une caméra mécanisée et d'une caméra grand angle. Dans cette thèse nous proposons une approche d'étalonnage où un calibrage géométrique est effectué pour une commande particulière de la caméra motorisée, puis incrémenté à chaque nouvelle rotation à l'aide de la commande angulaire.

5.4/ CALIBRAGE INTRINSÈQUE DES CAMÉRAS

Dans les chapitres 2 et 4, nous avons rappelé les différents paramètres responsables de la formation d'une image : les paramètres intrinsèques. Dans cette section nous verrons comment calibrer une caméra hors-ligne, c'est-à-dire comment estimer les paramètres du modèle choisi, ce modèle peut être le modèle sténopé, sphérique... Généralement ces paramètres sont calculés à l'aide d'un objet de calibrage connu, il peut s'agir de sphères, de cubes, ou plus souvent d'un simple motif planaire (échiquier, cercles concentriques ...). Une ou plusieurs images de cet objet permettent de remonter à la géométrie interne de la caméra.

5.4.1/ CALIBRAGE DE CAMÉRA PERSPECTIVE

Dans le cas d'une caméra perspective, c'est le modèle sténopé qui est généralement choisi. C'est donc la matrice intrinsèque $\mathbf{K}$ (définie dans le chapitre 2) et les coefficients de distorsion qui doivent être déterminés. Pour ce type de caméra, le sujet a déjà été très étudié puisque les premières approches de calibrage géométrique remontent aux années 70. Trois approches majeures ont marqué l'histoire du calibrage géométrique des caméras, il s'agit de celles développées par Tsai [171], Heikkila et Silven [88] et Zhang [179].

Le calibrage proposé par Tsai nécessite la connaissance *a priori* de certain paramètres fournis par le fabricant afin de réduire le nombre de paramètres à estimer. Cette approche permet un calibrage précis et complet de la caméra avec une ou plusieurs images.
Par la suite, Heikkila et Silven proposent une technique plus souple. Dans un premier

1. http ://hyscas.com/blog/

temps, un calcul initial des paramètres est opérée linéairement. Ces paramètres initiaux sont ensuite raffinés non-linéairement avec l'algorithme d'optimisation Levenberg-Marquardt.
La méthode de Zhang [179] reste aujourd'hui la référence la plus utilisé lorsqu'il est question de calibrer une caméra perspective. C'est cette méthode que nous décrirons brièvement ici, et c'est celle que nous utiliserons par la suite afin de calibrer notre caméra perspective via la boite à outils Matlab proposée par Bouguet [31].

La méthode de Zhang, étape par étape Cette méthode de calibrage nécessite l'utilisation d'un damier de dimensions connues. Il est essentiel que ce motif soit aussi plat que possible puisque cette approche s'appuie sur la planéité de la mire. Une série de clichés avec la caméra que l'on souhaite calibrer est alors requise. Avec la méthode présentée ici, au moins 3 images sont nécessaires afin d'estimer tous les paramètres intrinsèques de la caméra. Sur chacune des images une homographie entre la mire et son image est calculée $\mathbf{p} \sim \mathbf{HP}$ où $\mathbf{p}$ est un point image, $\mathbf{P}$ Le point correspondant sur la mire et $\mathbf{H}$ l'homographie décrivant la transformation projective entre ces deux points (voir figure 5.4).

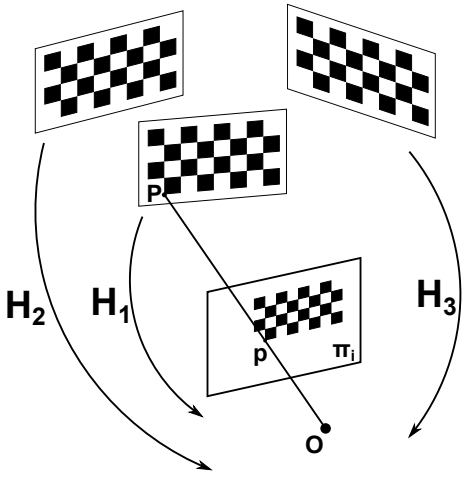


FIGURE 5.4 – Calibrage avec une mire planaire

On sait également que les points 3D sur la mire possèdent une coordonnée sur l'axe Z nulle $\mathbf{P} = [X\ Y\ 0\ 1]^T$ relativement au repère de la mire, la matrice de projection peut donc être reformulée ainsi :

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \mathbf{M} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = \mathbf{K}[\mathbf{r_1}\ \mathbf{r_2}\ \mathbf{t}] \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = \mathbf{H} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}, \tag{5.1}$$

avec $\mathbf{r}_1$ et $\mathbf{r}_2$ les deux premières colonnes de la matrice de rotation $\mathbf{R}$. Cette relation

permet d'imposer deux contraintes sur l'image de la conique absolue :

$$\mathbf{h}_1\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_1 = \mathbf{h}_2\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2, \tag{5.2}$$

$$\mathbf{h}_1\omega\mathbf{h}_1 = \mathbf{h}_2\omega\mathbf{h}_2, \tag{5.3}$$

ainsi que :

$$\mathbf{h}_1\omega\mathbf{h}_2 = 0. \tag{5.4}$$

Les deux équations (5.3) et (5.4) peuvent ainsi se réécrire sous forme d'un système linéaire $\mathbf{Ac} = 0$, où $\mathbf{c}$ est un vecteur contenant des entrées de ω nécessaires à l'extraction des paramètres intrinsèques, $\mathbf{c} = [\omega_{11}\ \omega_{12}\ \omega_{22}\ \omega_{13}\ \omega_{23}\ \omega_{33}]$. Les paramètres intrinsèques peuvent alors être extraits individuellement de la manière suivante :

$$v_0 = (\omega_{12}\omega_{13} - \omega_{11}\omega_{23})/(\omega_{11}\omega_{22} - \omega_{12}^2)^2, \tag{5.5}$$

$$\lambda = \omega_{33} - [\omega_{13}^2 + v_0(\omega_{12}\omega_{13} - \omega_{11}\omega_{23})]/\omega_{11}, \tag{5.6}$$

$$f_x = \sqrt{\lambda/\omega_{11}}, \tag{5.7}$$

$$f_y = \sqrt{\lambda\omega_{11}/\omega_{11}\omega_{22} - \omega_{12}^2}, \tag{5.8}$$

$$s = -\omega_{12}f_x^2 f_y/\lambda, \tag{5.9}$$

$$u_0 = sv_0/f_x - \omega_{13}f_x^2/\lambda. \tag{5.10}$$

Le calcul de la rotation et de la translation est ensuite triviale. Par ailleurs, l'ensemble de ces paramètres est raffiné à l'aide d'un ajustement de faisceaux réalisé avec un algorithme de type Levenberg-Marquardt minimisant la fonction de coût suivante :

$$\min_{\mathbf{K},\mathbf{R},\mathbf{t}} \sum_{i=1}^{N}\sum_{j=1}^{m} \| \mathbf{p}_{ij} - \widehat{\mathbf{p}}(\mathbf{K}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{P}_j) \|^2, \tag{5.11}$$

avec $\mathbf{p}_{ij}$ le $j^{\text{ème}}$ point détecté sur la $i^{\text{ème}}$ image et $\widehat{\mathbf{p}}$ le point re-projeté sur le plan image avec les paramètres estimés.

5.4.2/ CALIBRAGE DE CAMÉRA OMNIDIRECTIONNELLE

Dans le cas d'une caméra omnidirectionnelle à point de vue unique le modèle utilisé est le plus souvent le modèle sphérique unifié. Cela signifie qu'un paramètre de plus est à estimer, à savoir l le paramètre représentatif de la distorsion radiale de l'image. Les méthodes de calibrage employées sont par conséquent quelque peu différentes de celles présentées précédemment. Les approches les plus souvent utilisées dans la communauté de la vision par ordinateur sont [150] et [127]. Ces deux méthodes ont été rendues particulièrement populaires grâce aux boites à outils associées.
La première est proposée par Scaramuzza *et al.* [150], le modèle de projection est ici quelque peu différent du modèle sphérique unifié de Barreto [18]. Les auteurs modé-

lisent la projection des points 3D sur le plan image à l'aide d'une décomposition en série de Taylor, les points ainsi projetés ne le sont pas sur la sphère unité mais directement sur la surface du miroir ou de la lentille avant d'être projetés sur l'image.
La méthode proposé par Mei [127] repose sur le modèle sphérique de Barreto [18]. Afin de calibrer la caméra, l'auteur introduit une série d'hypothèses initiales : les coefficients de distorsion sont nuls, le rapport d'aspect de pixel est unitaire, le point principal est au centre de l'image et l est considéré dans un premier temps égal à 1. Dans ces conditions, il ne reste donc plus que la distance focale f à estimer. Il est possible d'obtenir une approximation de f linéairement à l'aide d'au moins 4 points colinéaires (sur une ligne non radiale de l'image). Une fois cette distance calculée, une détection sub-pixellique des coins de la mire est effectuée suivi d'une optimisation non-linéaire de l'ensemble des paramètres du modèle.
Même si les modèles de projection choisis par les méthodes présentées précédemment semblent très différents, Puig *et al.* [142] ont démontré leur équivalence.

5.5/ ESTIMATION DES PARAMÈTRES EXTRINSÈQUES

Partant du postulat qu'il est possible d'obtenir les paramètres intrinsèques avec les méthodes décrites dans les sections précédentes, on peut utiliser le modèle sphérique pour chacune des caméras. L'avantage de cette représentation est de conserver les propriétés de la géométrie projective. Pour la caméra PTZ (pour un niveau de zoom donné) l'utilisation de la *toolbox* développée par Bouguet [31] nous a permis le calcul de ses paramètres intrinsèques tandis que la caméra *fisheye* a été étalonnée avec la méthode Mei [127]. Ici on utilise la géométrie épipolaire afin d'initialiser la rotation et la translation entre nos caméras pour une rotation donnée, et donc d'estimer la position relative des caméras pour une commande mécanique de rotation $\mathbf{R}^p_{ptz}(0,0) = \mathbf{I}$. Notons $\mathbf{K}_p$ et $\mathbf{K}_f$ respectivement les matrices intrinsèques de la caméra perspective et omnidirectionnelle. La rotation et la translation entre nos deux caméras (pour une commande mécanique donnée) seront notées $\mathbf{R}^o_p$ et $\mathbf{t}^o_p$ (figure 5.5).

Afin de calculer la matrice essentielle liant les deux caméras, nous utilisons m points sur un plan dans le monde, visibles simultanément par les deux caméras dans un ensemble de n prises de vues. Soient $\mathbf{P}^i_{oj}$ et $\mathbf{P}^i_{pj}$ les positions du $i^{\text{ème}}$ point de correspondance (sur la sphère) visible dans la $j^{\text{ème}}$ image *fisheye* et PTZ. L'algorithme des huit points décrit dans le chapitre 2 nous permet de facilement calculer cette matrice $\mathbf{E}^o_p$ respectant :

$$(\mathbf{P}^i_{pj})^T \mathbf{E}^o_p \mathbf{P}^i_{oj} = 0 \quad \forall i, j \tag{5.12}$$

Notons que les points ne sont pas coplanaires si plusieurs prises de vues de la mire sont effectuées, ce qui permet l'usage de l'algorithme mentionné précédemment. Si l'on

souhaite calibrer notre système avec une seule vue, il est possible d'utiliser l'algorithme des 5 points permettant de faire face à ce cas dégénéré. Il aurait également été possible de nous servir des homographies existantes entre l'échiquier et les images afin de déterminer les paramètres extrinsèques. En pratique le calcul de la matrice essentielle nous fournit une très bonne initialisation. A partir de cette matrice essentielle, on extrait $\mathbf{R}_p^o$ et $\mathbf{t}_p^o$ à l'aide de la contrainte de chiralité. Cette initialisation nous permet d'alimenter un ajustement de faisceaux afin de déterminer plus précisément ces paramètres. Le critère à minimiser consiste en la somme de l'erreur de reprojection au carré sur les sphères :

$$\{\mathbf{R}_p^{o*}, \mathbf{t}_p^{o*}\} = \underset{\mathbf{R}_p^o, \mathbf{t}_p^o}{argmin} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\| \mathbf{P}_{oj}^i - \widehat{\mathbf{P}_{oj}^i}(\mathbf{R}_p^o, \mathbf{t}_p^o) \|^2 + \| \mathbf{P}_{pj}^i - \widehat{\mathbf{P}_{pj}^i}(\mathbf{R}_p^o, \mathbf{t}_p^o) \|^2 \right], \tag{5.13}$$

avec $\widehat{\mathbf{P}_{oj}^i}$ et $\widehat{\mathbf{P}_{pj}^i}$ la reprojection des points 3D respectivement sur les sphères S_o et S_p.

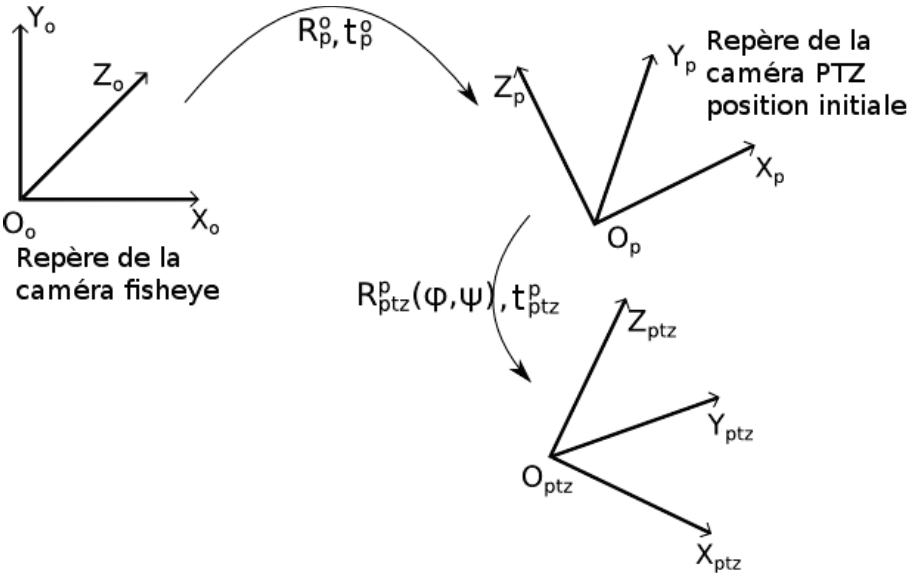


FIGURE 5.5 – Modélisation des différentes rotations et translations du système

5.6/ RÉSULTATS

Cette section s'intéresse au calibrage géométrique d'un système de vision hybride. Le schéma 5.6 détaille les étapes nécessaires à l'obtention de l'ensemble des paramètres de notre système. Nous passons ici ces différentes étapes en revue et nous proposons une étude comparative avec l'outil de calibrage développé par Caron et Eynard dans [38]. Pour l'évaluation de notre méthode de calibrage, les caméras utilisées sont deux caméras IDS μEye de résolution 1280×1024p dont une est équipée d'une lentille *fisheye* permettant un champ de vision de 180°, l'autre est dotée d'une optique plus conventionnelle n'induisant pas de fortes distorsions dans l'image. Ces caméras sont fixées sur banc rigide, comme le montre la figure 5.7.

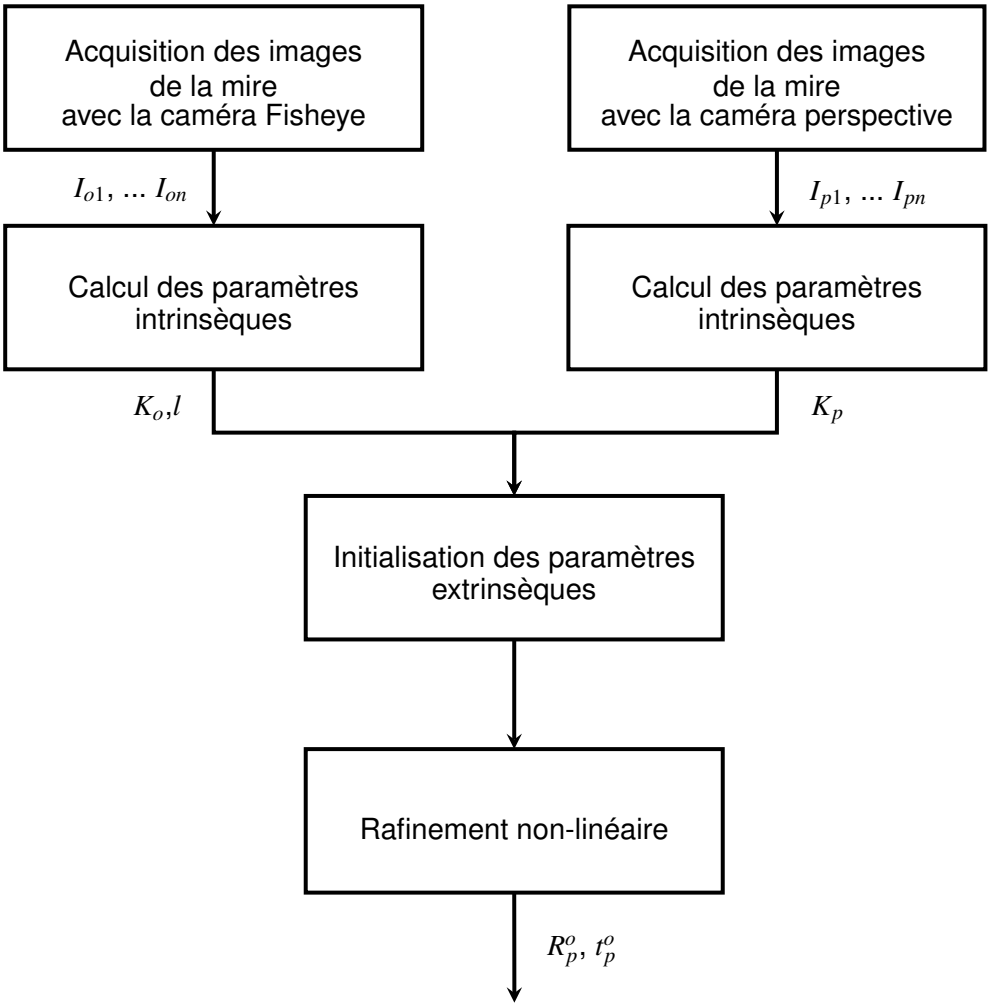


FIGURE 5.6 – Étapes du calibrage de notre système de stéréo-vision

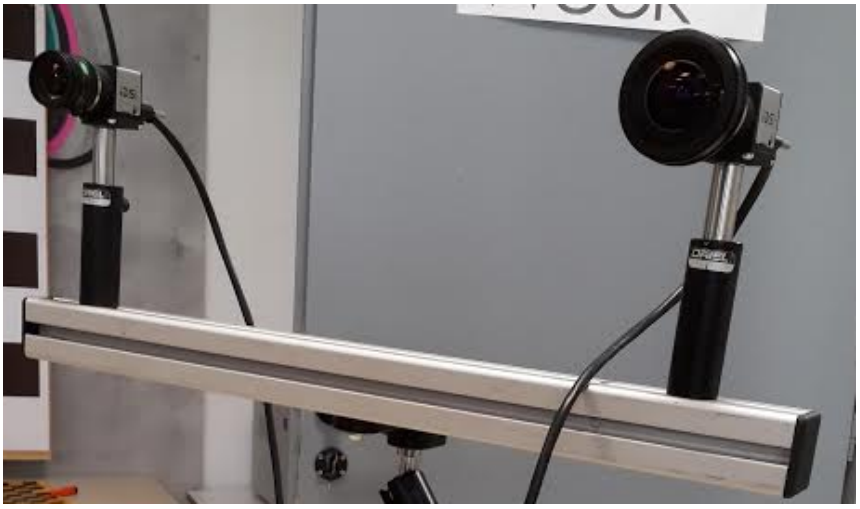


FIGURE 5.7 – Banc stéréo hybride utilisé pour évaluer notre approche de calibrage

5.6.1/ ACQUISITION DES IMAGES

Le calibrage de systèmes de vision hybride nécessite l'adaptation de la procédure classique de calibrage, par exemple une mire aux dimensions particulières doit être choisie. Pour la plupart des bancs de vision homogène une mire de taille A4 est bien souvent suffisante. Ici, il est important de prendre en compte la très forte dissimilarité entre les images, l'échiquier a donc été imprimé à un format supérieur afin d'être pleinement visible sur les deux images. L'acquisition d'une série de 11 images par caméra a été effectuée (voir figure 5.8). Sur chaque image de la mire un ensemble de 42 coins peuvent être détectés soit un total de 462 points de correspondance.

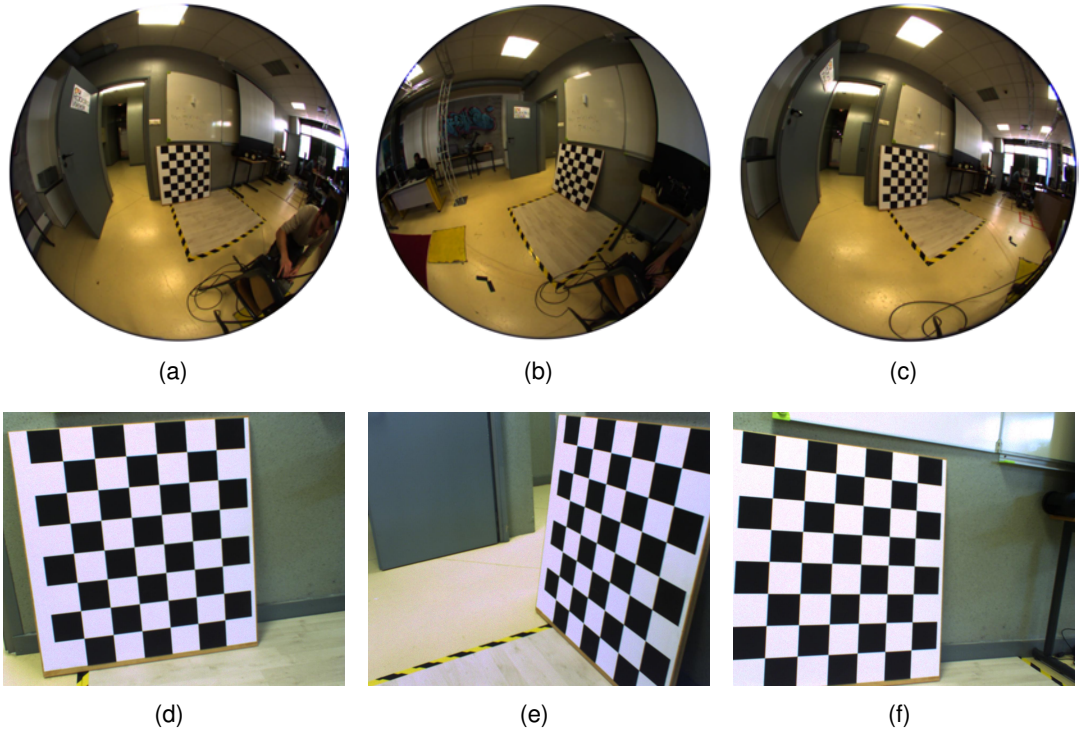


FIGURE 5.8 – Extrait d'images utilisées pour la phase de calibrage (a-c) Les images acquises avec la caméra *fisheye* (d-f) Images perspectives correspondantes

5.6.2/ CALIBRAGE INTRINSÈQUE

Le calibrage intrinsèque et la détection des coins $\mathbf{p}^i_{oj}$ et $\mathbf{p}^i_{pj}$ -respectivement sur les images *fisheye* et perspectives- ont été effectuées à l'aide des boites à outils [31] et [127]. La détection de ces points est très performante et assure une précision sous pixellique. Les paramètres intrinsèques ainsi obtenus sont présentés dans le tableau 5.2.

	f	λ	s	u_0	v_0	l
Caméra *Fisheye*	750,79p	1,002	0	610,38p	480,55p	1,7619
Caméra Perspective	1586,3p	1,0021	0	614,0p	477,6p	0

Tableau 5.2 – Paramètres intrinsèques

5.6.3/ CALCUL DES PARAMÈTRES EXTRINSÈQUES

A l'aide des paramètres intrinsèques de nos caméras nous pouvons re-projeter les points détectés sur leur sphère respective :

$$\mathbf{p}^i_{oj} \rightarrow \mathbf{P}^i_{oj}, \tag{5.14}$$

$$\mathbf{p}^i_{pj} \rightarrow \mathbf{P}^i_{pj}. \tag{5.15}$$

Les paramètres extrinsèques sont initialisés à l'aide de l'algorithme des huit points pour être ensuite optimisés par l'ajustement de faisceaux décrit par l'équation (5.13). Les résultats ainsi obtenus sont résumés dans le tableau 5.3.

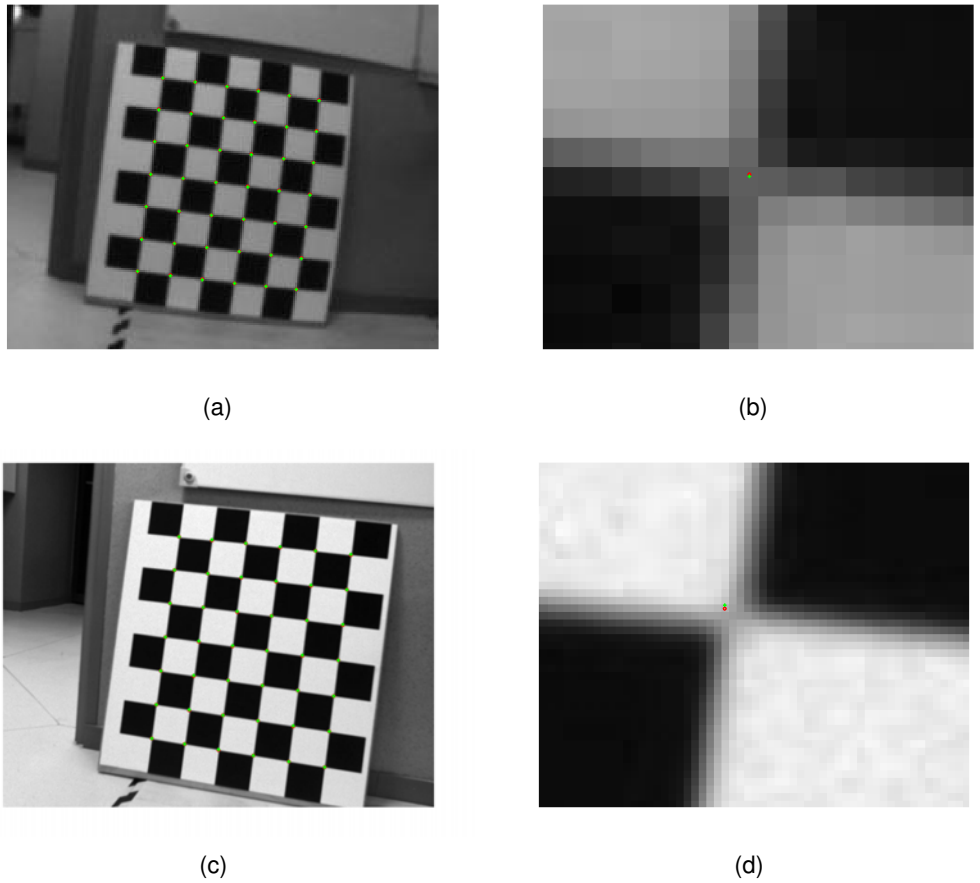
Les résultats de notre méthode sont très proches de ceux calculés avec l'asservissement visuel virtuel (AVV) [38]. Notons que contrairement à cette méthode [38], nous fournissons les erreurs de re-projection de chaque caméra indépendamment, c'est un choix délibéré permettant de fournir une mesure d'erreur plus représentative. L'échelle de la mire est très différente sur chacune des images rendant les mesures d'erreur non équivalentes. Par exemple une erreur de re-projection d'un pixel sur la caméra *fisheye* est beaucoup plus dommageable que la même erreur avec la caméra perspective.
Avec une évaluation similaire à [38], nous obtenons une erreur de re-projection moyenne $\mu_u = 0,25$ pixels et un écart type $\sigma_u = 0,259$ pixels, pour l'ensemble des points sur les images *fisheye* et perspectives. La méthode basée AVV nous donne quant à elle $\mu_u = 0,415$ pixels et $\sigma_u = 0,492$ pixels. Les résultats obtenus avec notre approche sont nettement supérieurs à ceux déterminés par le logiciel Hyscas. Cette différence peut s'expliquer par plusieurs facteurs notamment une détection des coins plus précises.
La figure 5.9 contient un exemple de la re-projection des points dans les images. Une reconstruction 3D des mires est également proposée dans la figure 5.10 où la couleur des points correspond à leur appartenance à une image.

5.6.4/ LIGNES/CONIQUES ÉPIPOLAIRES

Afin d'évaluer la précision de notre calibrage qualitativement nous proposons de visualiser les lignes (images perspectives) et coniques (images *fisheye*) épipolaires obtenues à l'aide des paramètres calculés précédemment. Ce test est effectué dans un autre environnement que celui utilisé pour le calibrage.
Dans la figure 5.11, plusieurs points choisis sur l'image *fisheye* sont utilisés afin de tracer leurs lignes épipolaires correspondantes sur l'image perspective. Malgré la très forte dif-

	Notre approche		HysCas	
	Caméra *fisheye*	Caméra perspective	Caméra *fisheye*	Caméra perspective
f	750.79p	1586.3p	715.96p	1587.23p
u_0	610.38p	614.0p	613.96p	608.5p
v_0	480.55p	477.6p	479.65p	467.07p
λ	1.002	1.0021	1.007	1.005
s	0	0	0	0
l	1.76	0	1.58	0
θ_x	\	-0.49°	\	-0.75°
θ_y	\	6.36°	\	7.44°
θ_z	\	-0.18°	\	-0.24°
t_x	\	-0.977	\	-0.989
t_y	\	0.0066	\	0.0166
t_z	\	0.213	\	0.1435
μ_u	0.076p	0.44p	\	\
σ_u	0.041p	0.23p	\	\

Tableau 5.3 – Comparaison des résultats obtenus avec l'approche présentée dans [38], μ_u et σ_u étant l'erreur moyenne et l'écart type de reprojection ; tandis que θ_w représente la rotation angulaire autour d'un axe w.



(a) (b)

(c) (d)

FIGURE 5.9 – Reprojection des points sur les images, les cercles rouges représentent les points détectés et les croix vertes les points reprojetés (a) Image *fisheye* de la mire (b) détail de l'image *fisheye* (c) Image Perspective de la mire (d) détail de l'image perspective

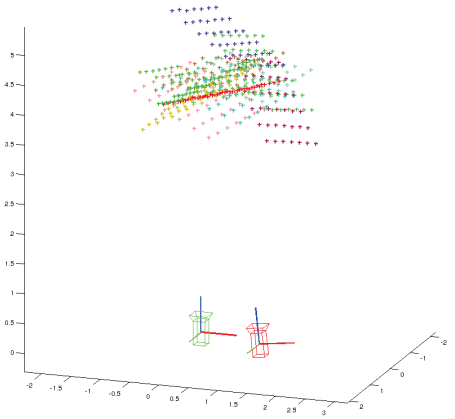
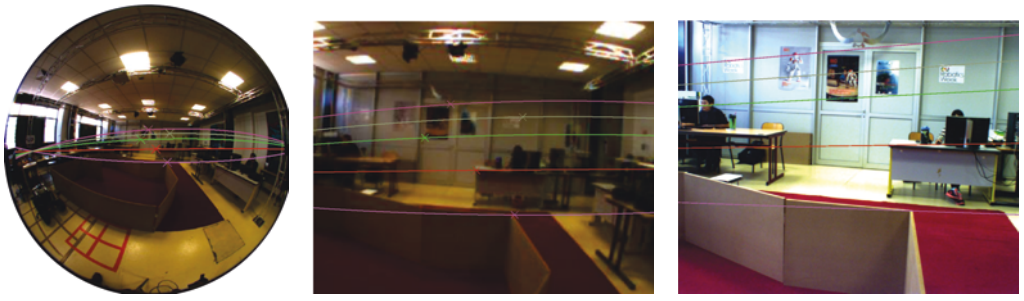


FIGURE 5.10 – Reconstruction 3D des différentes positions de la mire

férence de résolution, il est notable que les lignes épipolaires passent bien par les points sélectionnés. Le même constat est fait dans le cas où les points sont choisis sur l'image perspective (voir figure 5.12). Cette expérience est complémentaire de l'erreur de reprojection présentée dans la section précédente et permet un aperçu visuel de la justesse de notre calibrage.



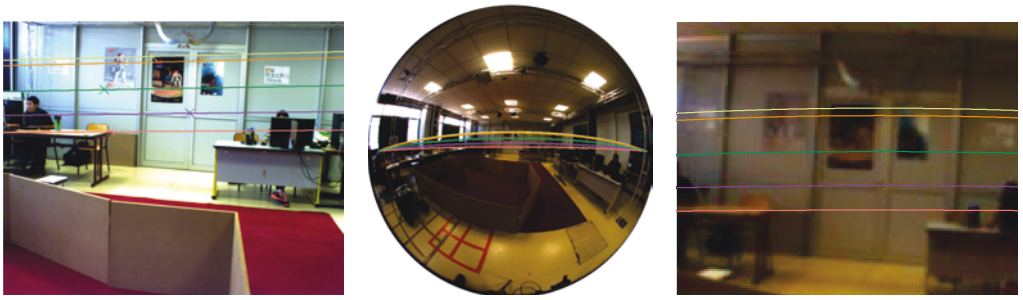
(a) (b) (c)

FIGURE 5.11 – (a) Image *fisheye* où les points d'intérêt indiqués par des croix sont utilisés pour calculer les lignes épipolaires sur l'image perspective (b) détail de l'image *fisheye* (c)Image perspective avec les lignes épipolaires

5.6.5/ RECTIFICATION D'IMAGE HYBRIDE

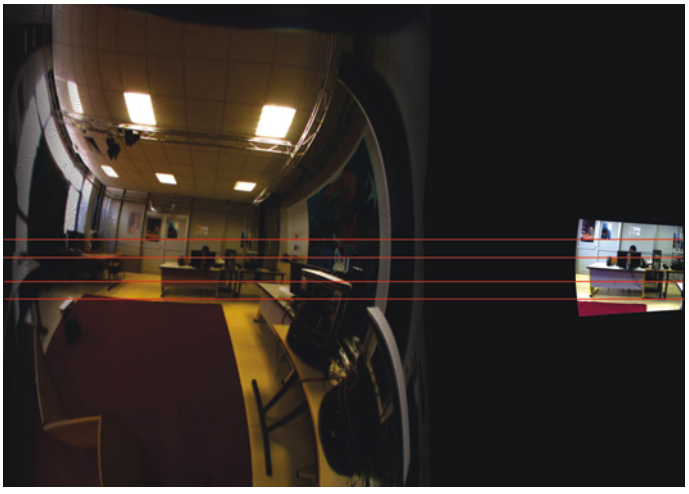
La rectification stéréo consiste à rendre les lignes épipolaires parallèles à l'axe horizontal des images, ce qui revient à déplacer les épipoles à l'infini. Au même titre qu'il est possible de rectifier des images perspectives, une rectification sphérique des images est possible [74].

La figure 5.13 est un exemple de rectification obtenu par calibrage de notre système

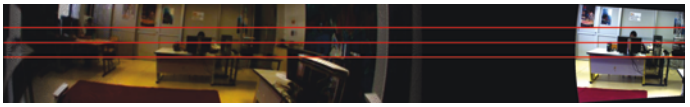


(a) (b) (c)

FIGURE 5.12 – (a) Image perspective où les points d'intérêt indiqués par des croix sont utilisés pour calculer les coniques épipolaires sur l'image *fisheye* (b) Image *fisheye* avec les coniques épipolaires (c) détail de l'image *fisheye*



(a)



(b)

FIGURE 5.13 – Rectification d'images avec un système de vision hybride, (a)Rectification complète, (b)Une autre rectification où seule la région en commun est conservée

de vision, nous remarquons que les lignes rouges (lignes horizontales) passent par les mêmes points sur les images *fisheye* et les image perspectives rectifiées.

5.7/ CONTRÔLE D'UNE CAMÉRA PTZ DANS UN SYSTÈME DE STÉRÉO VISION HYBRIDE

Dans cette section, nous proposons une nouvelle approche permettant d'orienter la caméra mécanisée sur une cible visible depuis l'image omnidirectionnelle de manière à obtenir une image de bonne définition de l'objet d'intérêt à partir de la caméra PTZ. Connaissant le calibrage complet du système, la méthode présentée ici exploite la géométrie épipolaire existante entre les deux caméras afin de réduire la zone de recherche de la région d'intérêt.
Les méthodes dédiées à la vidéo surveillance présentées dans la section 5.3, utilisent de fortes contraintes liées à l'environnement dans lequel le système est installé. Il en résulte que ces approches nécessitent un re-calibrage à chaque changement d'environnement, par exemple si le système de surveillance est déplacé d'une pièce à une autre, tandis que l'approche décrite ici permet le contrôle du système dans une scène totalement inconnue.

5.7.1/ MODÉLISATION DE NOTRE SYSTÈME DE STÉRÉO VISION HYBRIDE

Les figures 5.14 et 5.5 résument l'ensemble des interactions entre les caméras, toutes les notations utilisées dans notre système sont résumées dans le tableau 5.4. Par convenance, nous prendrons le repère de la caméra omnidirectionnelle $(\vec{X}_o, \vec{Y}_o, \vec{Z}_o)$ situé en $\mathbf{O}_o$ comme référence pour l'ensemble du système. La position et l'orientation de la caméra PTZ dans son propre référentiel $(\vec{X}_{ptz}, \vec{Y}_{ptz}, \vec{Z}_{ptz})$ devront alors s'exprimer en fonction de la référence globale. Notons qu'une translation $\mathbf{t}^p_{ptz}$ existe en pratique, elle correspond à la translation résiduelle (déjà développée dans le chapitre 4) inhérente à l'utilisation de mécanismes servant à pivoter la caméra sur ses axes.

Un point $\mathbf{P}_{ptz}$ dans la référence de la caméra PTZ pourra être exprimé dans la référence système de la manière suivante :

$$\mathbf{P}_o = \mathbf{R}^{oT}_p \mathbf{R}^p_{ptz}(\varphi, \psi)^T (\mathbf{P}_{ptz} - \mathbf{t}^p_{ptz}) - \mathbf{t}^o_p, \tag{5.16}$$

où $\mathbf{P}_o$ est le point dans le repère de la caméra omnidirectionnelle. Dans notre cas on considère la translation $\mathbf{t}^p_{ptz}$ comme négligeable on obtient donc la relation suivante :

$$\mathbf{P}_o = \mathbf{R}^{oT}_p \mathbf{R}^p_{ptz}(\varphi, \psi)^T \mathbf{P}_{ptz} - \mathbf{t}^o_p. \tag{5.17}$$

5.7.2/ MÉTHODOLOGIE

Nous souhaitons ici trouver les consignes angulaires de la caméra mécanisée (φ, ψ) permettant de visualiser une cible localisée sur la caméra *fisheye* ayant un centroïde $\mathbf{P}^c_o$. La

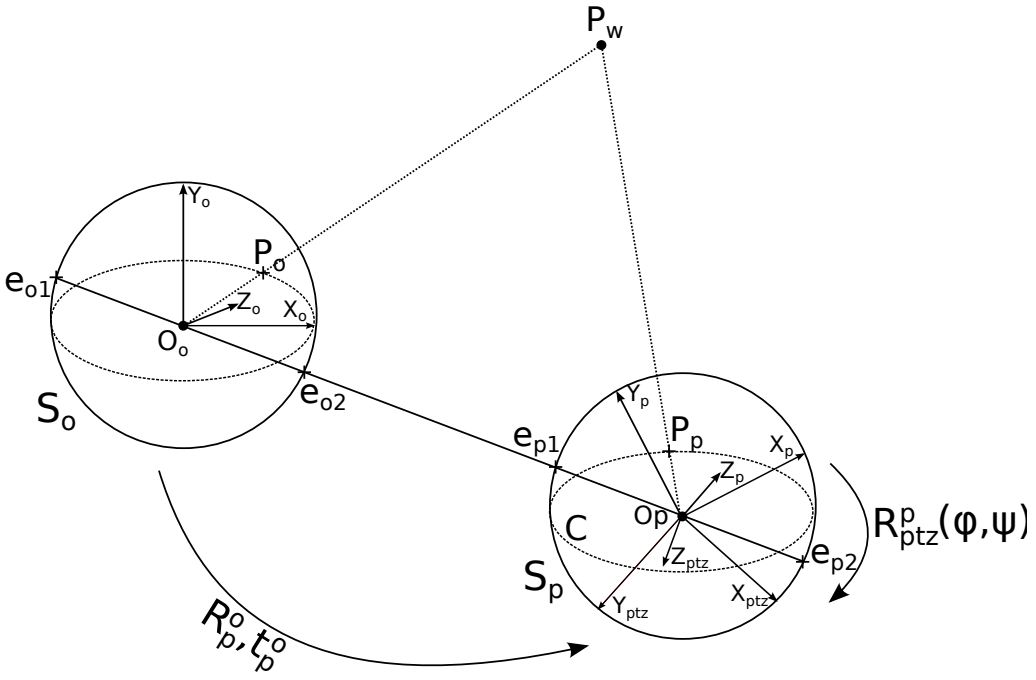


FIGURE 5.14 – modélisation complète du système

méthode proposée peut se décomposer en deux parties. La première consiste à scruter un grand cercle épipolaire à l'aide des rotations de la caméra PTZ. Cette étape est suivi par la détection de la région d'intérêt parmi les images acquises le long de $\mathbf{C}$.

5.7.2.1/ COMMANDE DE LA CAMÉRA LE LONG DU CERCLE ÉPIPOLAIRE

Dans le cas d'un capteur fixe et si aucune information concernant la scène n'est disponible, il est impossible de commander directement la caméra PTZ dans la direction désirée (ambiguïté sur la distance entre la cible et la caméra *fisheye*). Cependant la géométrie épipolaire permet de réduire la zone de recherche à un ensemble de points $\mathbf{P}_p^c \, \epsilon \, \mathbf{S}_p$ définissant le grand cercle $\mathbf{C}$ et satisfaisant donc $\mathbf{P}_p^{c\,T} \mathbf{t}^o_{p_{[\times]}} \mathbf{R}^o_p \mathbf{P}^c_o = 0$. La méthode proposée ici consiste à orienter la caméra PTZ afin de scruter le cercle épipolaire $\mathbf{C}$ pour y repérer l'objet d'intérêt. Obtenir les consignes angulaires en *pan* et *tilt* est simple, elles correspondent en effet aux coordonnées sphériques des points se trouvant sur $\mathbf{C}$:

$$\forall \mathbf{P}_p^c(X,Y,Z) \, \epsilon \, \mathbf{S}_p,$$

$$\begin{cases} \varphi = arccos(Z/\sqrt{(X^2+Y^2+Z^2)}) \\ \psi = arctan(Y/X) \end{cases}$$

Cette série de commandes peut être considérablement réduite en éliminant les angles non atteignables par le mécanisme de la caméra PTZ.
Connaître la matrice de rotation inter-caméras à chaque instant est également un élément essentiel pour la majorité des applications de stéréo-vision, par exemple cette information

$\mathbf{S}_o$	Modèle sphérique de la caméra omnidirectionnelle
$\mathbf{S}_p$	Modèle sphérique de la caméra perspective
$\mathbf{O}_o$	Centre de $\mathbf{S}_o$, repère monde
$\mathbf{O}_p$	Centre de $\mathbf{S}_p$, $\mathbf{O}_p$=$\mathbf{t}_p^o$
$\mathbf{P}_o$	Point cible ϵ $\mathbf{S}_o$
$\mathbf{P}_w$	Position 3D de la cible dans le monde
$\mathbf{E}_o^p$	Matrice essentielle $\mathbf{E} = \mathbf{t}_{[\mathrm{X}]}\mathbf{R}$
π_e	Plan épipolaire définit par $\mathbf{E}_o^p\mathbf{P}_o$
$\mathbf{C}$	Grand cercle épipolaire ϵ $\mathbf{S}_p$
$\mathbf{P}_p$	Point recherché ϵ $\mathbf{C}$
ψ	Commande angulaire permettant d'orienter la caméra en *Tilt*
φ	Commande angulaire permettant d'orienter la caméra en *Pan*
$\mathbf{K}_p$	Paramètres intrinsèques de la caméra PTZ
$\mathbf{K}_o$	Paramètres intrinsèques de la caméra omnidirectionnelle
$\mathbf{t}_p^o$	Translation entre les caméras
$\mathbf{R}_p^o$	Rotation entre les caméras pour $\psi = \varphi = 0$
$\mathbf{R}_{ptz}^p(\varphi, \psi)$	Rotation de la caméra PTZ dans son repère, dépendant des commandes moteurs
$(\vec{X}_o, \vec{Y}_o, \vec{Z}_o)$	Repère de la caméra *fisheye* et du système
$(\vec{X}_{ptz}, \vec{Y}_{ptz}, \vec{Z}_{ptz})$	Repère de la caméra PTZ
$(\vec{X}_p, \vec{Y}_p, \vec{Z}_p)$	Repère intermédiaire décrivant la rotation inter-caméra pour $\mathbf{R}_{ptz}^p = \mathbf{I}$

Tableau 5.4 – Notations relatives au système

est particulièrement utile pour tout procédé de reconstruction 3D ou de mise en correspondances. Cette matrice de rotation est directement liée aux consignes envoyées à la caméra. L'ensemble des rotations vérifiant les équations suivantes permettent d'aligner le centre de la caméra PTZ sur le cercle $\mathbf{C}$:

$$\mathbf{N}.\mathbf{R}_p^o\mathbf{R}_{ptz}^p(\varphi, \psi)\mathbf{Z}_{ptz} = 0, \tag{5.18}$$

$$\mathbf{N}.\mathbf{R}_p^o\mathbf{R}_\varphi\mathbf{R}_{\psi-\frac{\pi}{2}}\mathbf{Z}_{ptz} = 0, \tag{5.19}$$

avec $\mathbf{N} = \mathbf{t}_{p_{[\times]}}^o\mathbf{R}_p^o\mathbf{P}_o^c$ la normale du plan épipolaire, et où $\mathbf{R}_\varphi$ et $\mathbf{R}_\psi$ sont respectivement les matrices de rotations en *pan* et *tilt*. Tandis que $\mathbf{Z}_{ptz} = [0\ 0\ 1]^T$ correspond à l'axe optique de la caméra PTZ à aligner sur le grand cercle épipolaire.
Notons que notre approche est totalement fonctionnelle pour tout niveau de zoom, puisqu'il est possible de pré-calibrer [161] ou d'auto-calibrer [144] la caméra à focale variable.

5.7.2.2/ DÉTECTION DE LA RÉGION D'INTÉRÊT

Cette étape consiste à reconnaître un zone sélectionnée depuis l'image *fisheye* parmi une série d'images perspectives. La mise en correspondance d'image est l'une des tâches les plus complexes dans le contexte d'un système de vision hybride. Les raisons avancées sont nombreuses, il peut s'agir d'une forte dissimilarité dans la résolution, l'échelle, l'orientation ou la réponse colorimétrique des images. C'est une problématique

déjà abordée par le passé, dans [49] ce sont les descripteurs SIFT qui sont adaptés à la géométrie sphérique tandis que dans [110] c'est l'approche géométrique qui est adaptée. A ce problème particulier nous proposons une approche capable de faire face aux différentes difficultés mentionnées.

Lorsque l'on scrute la ligne épipolaire avec la caméra PTZ nous capturons un ensemble d'images (une image pour chaque position (φ, ψ)). Le rôle de la détection est alors de localiser la cible dans les images acquises durant cette étape de recherche, et de trouver la commande permettant d'orienter la caméra dans cette direction.

Tout d'abord une imagette de la cible est sélectionnée manuellement sur l'image *fisheye*. Une détection de points caractéristiques est ensuite effectuée à la fois sur ce *patch* mais également sur l'ensemble des images perspectives. Considérons $\mathbf{p}_o$ et $\mathbf{p}_p$ comme les points détectés par le détecteur de Harris respectivement sur l'image omnidirectionnelle et sur les images perspectives. Ces points sont reprojetés sur leur sphère respectives en $\mathbf{P}_o$ et $\mathbf{P}_p$. A ce stade tous les points sont possiblement en correspondance entre eux. La plupart de ces correspondances peuvent être rejetées à l'aide de la contrainte épipolaire $\mid \mathbf{P}_p^T \mathbf{E}_p^o \mathbf{P}_o \mid < \varepsilon$, ou ε est un seuil choisi arbitrairement. Cette décimation aura pour effet de ne conserver que les points de correspondance potentiels, c'est-à-dire les points présent dans l'enveloppe épipolaire.

Nous ne faisons donc pas appel à des descripteurs photométriques mais simplement à la géométrie du capteur stéréo calibré pour déterminer les correspondances possibles. Cette approche permet une très grande robustesse aux changements d'illumination, de rotation ou d'échelle, spécifique à ce type de système. Avec notre configuration de caméras les tests menés avec des descripteurs SURF et MSER ne se sont d'ailleurs pas montrés concluants. Une approche purement géométrique telle que celle présentée ne permet cependant pas une élimination performante de tous les *outliers*, il restera en somme une très grande quantité de correspondances incorrectes.

Si l'on considère la région d'intérêt comme étant localement planaire il est cependant possible de calculer une homographie permettant sa détection. Une estimation robuste de cette transformation à l'aide d'un algorithme de type RANSAC [66] permet en outre une élimination efficace des *outliers*. Une homographie est généralement calculée à l'aide d'un ensemble de 4 points minimum. Dans le cas présent nous possédons plusieurs informations sur les éléments constituants l'homographie inter-image $\mathbf{H}_p^o$, à savoir la rotation et la translation entre les caméras de notre banc de vision. Ces paramètres extrinsèques peuvent en conséquence servir à réduire le nombre de points nécessaires au calcul de la matrice $\mathbf{H}_p^o$. Une réduction du nombre de points minimum à la résolution du modèle permet de décroître exponentiellement la complexité calculatoire de l'algorithme RANSAC. Une simple pré-rotation des points $\mathbf{P}_p^i$ sur la sphère $\mathbf{S}_p$ permet une réécriture de l'homographie sous la forme suivante :

$$\mathbf{H} \sim \mathbf{I} - \frac{\mathbf{t}_p^o \mathbf{n}^T}{d}. \tag{5.20}$$

Connaissant $\mathbf{t}_p^o$, le nombre de degrés de liberté de $\mathbf{H}_p^o$ se trouve réduit à 3. Ces trois degrés de liberté étant les entrées du vecteur $\mathbf{N}_d = \frac{\mathbf{n}^T}{d}$. $\mathbf{H}_p^o$ peut alors s'écrire ainsi :

$$\mathbf{H}_p^o = \begin{bmatrix} 1 - \frac{n_x}{d} t_x & -\frac{n_y}{d} tx & -\frac{n_z}{d} t_x \\ -\frac{n_x}{d} t_y & 1 - \frac{n_y}{d} t_y & -\frac{n_z}{d} t_y \\ -\frac{n_x}{d} t_z & -\frac{n_y}{d} t_z & 1 - \frac{n_z}{d} t_z \end{bmatrix}, \tag{5.21}$$

avec $\mathbf{t}_p^o = [t_x \; t_y \; t_z]$ et $\mathbf{N}_d = [n_x \; n_y \; n_z]^T / d$.
Pour déterminer $\mathbf{N}_d$ on peut résoudre $\mathbf{P}_p \times \mathbf{P}_o \mathbf{H}_p^o = 0$. Chaque point de correspondance $P_o(x_o, y_o, z_o)$ et $\mathbf{P}_p(x_p, y_p, z_p)$ fournit 3 équations (de la forme $\mathbf{A}\mathbf{N}_d = \mathbf{b}$) toutes linéairement dépendantes à l'équation suivante :

$$[t_y x_o z_p - t_z x_o y_p \quad t_y y_o z_p - t_z y_o y_p \quad t_y z_o z_p - t_z y_p z_o] \mathbf{N}_d = y_p z_o - y_o z_p. \tag{5.22}$$

Trois points de correspondances sont donc suffisants pour résoudre $\mathbf{N}_d$. Dans notre algorithme RANSAC, trois points de correspondance sont sélectionnés aléatoirement parmi les points $\mathbf{P}_o$ et $\mathbf{P}_p$ afin de calculer une matrice d'homographie, le nombre d'*inliers* (N_I) est calculé de la manière suivante :

$$k_i = \begin{cases} 1 & \text{si } \|\mathbf{P}_o^i - \mathbf{H}_p^o \mathbf{P}_p^i\|^2 + \|\mathbf{P}_p^i - \mathbf{H}_p^{o\,-1} \mathbf{P}_o^i\|^2 < \tau \\ 0 & \text{sinon} \end{cases},$$

$$N_I = \sum_{i=1}^{n} k_i,$$

où τ est un seuil arbitraire et n le nombre total de points potentiellement en correspondance. Après une première évaluation du nombre d'*inliers* avec le modèle calculé, le processus est réitéré jusqu'à l'obtention d'une homographie admettant un nombre suffisant d'*inliers*.
Finalement, la commande permettant d'orienter la caméra est calculée à l'aide de la transformation des coordonnées du point central de la cible $\mathbf{P}_o^c$:

$$\mathbf{P}_p^c = \mathbf{H}_p^o \mathbf{P}_o^c. \tag{5.23}$$

Toute les étapes nécessaires à la détection de la région d'intérêt sont résumées dans l'algorithme 1 :

1 Sélection d'un *patch* sur l'image *fisheye*;
2 Détection de points caractéristiques sur la *fisheye* et sur l'ensemble des images PTZ;
3 Projection sur les sphères équivalentes;
4 Filtrage par la contrainte épipolaire;
5 Estimation de l'homographie (RANSAC);

Algorithme 1 : Algorithme de détection d'une région d'intérêt

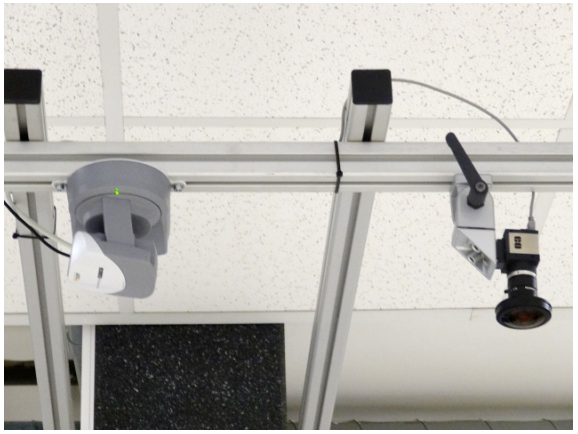


FIGURE 5.15 – Banc stéréo hybride

5.7.3/ RÉSULTATS

Cette section est dédiée à l'évaluation de la méthode proposée. Pour ce faire, nous présentons ici une analyse quantitative obtenue dans un environnement photo-réaliste généré avec un logiciel de tracé de rayon. Une série d'expériences qualitatives en conditions réelles est également proposée afin de témoigner de la robustesse de notre approche.
Afin de tester notre algorithme, nous utilisons le logiciel PovRay (Persistance of Vision Ray Tracer) pour synthétiser les vues *fisheye* et perspectives dans un environnement totalement contrôlé. La programmation PovRay permet en effet de spécifier les paramètres intrinsèques mais aussi la position et l'orientation des caméras dans la scène. Ce procédé d'évaluation permet de travailler dans une scène au rendu quasi-réaliste où toutes les informations 3D sont connues. Dans la série de tests proposée les deux caméras possèdent une résolution spatiale de 640×480 pixels.
Concernant les expériences en conditions réelles, les résultats expérimentaux ont été obtenus en utilisant une caméra fixe d'une résolution de 640×480p dotée d'une optique *fisheye* permettant d'obtenir un champ de vision de $180°$. La caméra PTZ employée est une AXIS 2130R permettant une rotation panoramique sur $338°$ et un angle compris entre 0 et $90°$ en *tilt*. Elle est également pourvue d'un zoom optique $16\times$. Cette structure est suspendue au plafond d'une salle (voir figure 5.15). La figure 5.16 correspond à la représentation sphérique du système dans sa position de référence pour laquelle il a été calibré, c'est-à-dire que nous considérerons les commandes de rotation de la caméra ψ et φ égales à zéro. On distingue également sur cette figure le grand cercle épipolaire (en rouge) sur la sphère relative à la caméra PTZ, ce cercle correspond au point indiqué par un point rouge sur la caméra *fisheye*.

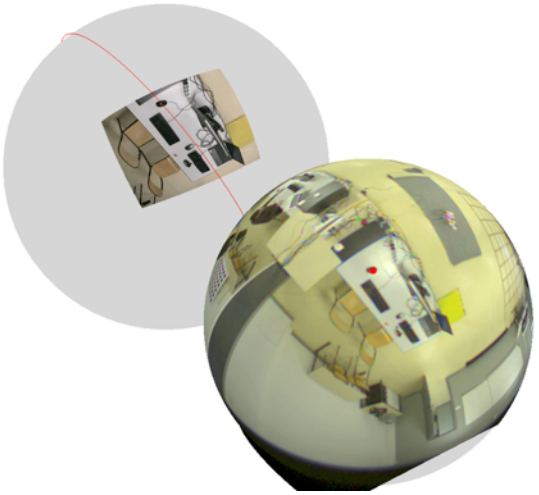


FIGURE 5.16 – Représentation sphérique du système après calibrage

5.7.3.1/ SCAN DU GRAND CERCLE ÉPIPOLAIRE

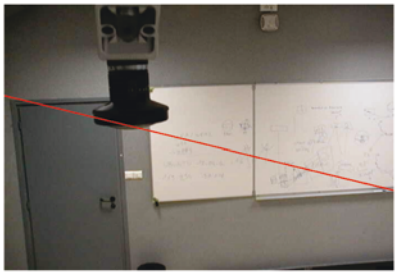
Les simulations effectuées dans PovRay concernant le suivi du cercle épipolaire sont toujours très précises puisque que les paramètres intrinsèques et extrinsèques sont connus et que le point focal de la caméra rotative reste toujours à la même position quelle que soit la rotation de la caméra. La figure 5.17, présente un exemple représentatif d'un des essais réalisés à partir de notre système, dans cette séquence d'images réelles nous avons délibérément choisi un objet lointain et assez difficilement distinguable sur l'image *fisheye*. Dans la séquence d'images proposées on note bien la présence de l'objet d'intérêt dans la figure 5.17(d). De plus, la ligne épipolaire (illustrée en rouge) est bien localisée sur l'objet cible. Il est ici difficile de quantifier l'erreur provenant de l'omission de $\mathbf{t}^{p}_{ptz}$, elle semble cependant négligeable au vu des résultats obtenus qui sont visuellement très acceptables.

5.7.3.2/ DÉTECTION DE L'OBJET

Dans nos expérimentations nous utilisons le détecteur de Harris directement sur les images. Les images omnidirectionnelles ne sont par conséquent pas rectifiées afin de corriger les distorsions, ce qui constitue un gain de temps notable. Pour la détection de tous les objets qui suivent les même seuils τ=0.01 et ε=0.1 ont été conservés.

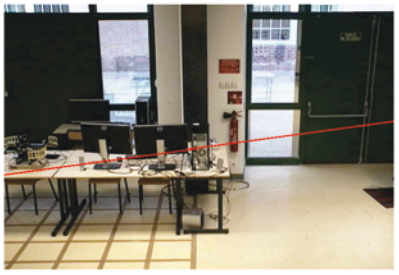
Tests avec des images synthétiques : Dans nos expériences nous mettons en évidence la robustesse de notre approche dans différents scénarios. La métrique choisie pour juger de la qualité de la localisation des objets est la distance angulaire entre le centre de la région d'intérêt désirée et l'orientation effective de la caméra PTZ.

(a)

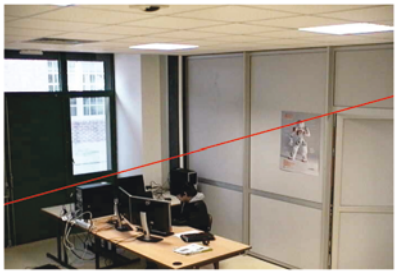


(b)

(c)



(d)



(e)

FIGURE 5.17 – (a) Image *fisheye* où l'extincteur rouge a été désigné comme cible (boite englobante rouge à droite de l'image), (b)(c)(d)(e) images obtenues par rotation de la caméra le long du cercle épipolaire

Tout d'abord, nous proposons un test de robustesse au bruit. L'ajout de bruit sur l'intensité des pixels a notamment pour effet de biaiser la détection de points et par conséquent d'introduire un plus grand nombre d'*outliers*. Dans les tests présentés par la figure 5.18,

variance du bruit	0	5.1	10.2	15.3	20.4	25.5	30.6	35.7
erreur angulaire (°)	6.89	7.01	3.96	5.13	5.66	5.04	4.62	6.07

Tableau 5.5 – Erreur angulaire avec ajout de bruit gaussien

différents niveaux de bruit blanc gaussien ont été ajoutés à la fois sur l'image *fisheye* et sur les images perspectives. Afin de rendre les résultats plus lisibles la région d'intérêt détectée est affichée sur les images non-bruitées. Le tableau 5.5 référence quant à lui la précision angulaire obtenue avec notre méthode pour différents niveaux de bruit. Ce tableau met en évidence une grande stabilité de notre algorithme, puisque les résultats obtenus sont relativement constants même en présence d'un bruit additif important. Il est notable que même en absence de bruit notre approche ne permet pas une détection parfaite de la région d'intérêt puisque celle ci ne satisfait pas l'hypothèse de planéité. Malgré cela la région désirée est toujours très proche du centre de l'image quel que soit le niveau de bruit. Ces résultats suggèrent une très bonne robustesse à ce type de bruit, ceci est essentiellement lié au fait que la mise en correspondance proposée n'est aucunement basé sur l'intensité des pixels. Pour l'application demandée, cette approche est donc un très bon compromis entre précision et robustesse.

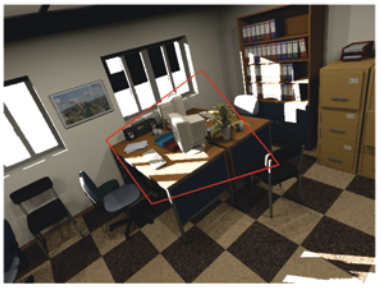
Dans la figure 5.19, nous avons testé notre méthode sur différentes surfaces afin de déterminer quelle influence pouvait avoir le calcul d'une homographie sur une surface non-planaire dans la détection de notre objet. Dans la figure 5.19(a) il s'agit d'une région parfaitement planaire, (b) est légèrement non planaire tandis que dans la dernière image c'est une région complètement non-planaire qui a été choisie. Dans le premier cas de figure c'est une erreur angulaire 0.58° qui a été calculée, 0.6° pour le second et finalement 7.34° pour le dernier. Cette expérience met en évidence une relation directe entre la précision de notre algorithme et la géométrie de l'objet recherché. Cependant, cela démontre également que même dans le cas d'une cible absolument non-planaire nous sommes capable d'orienter la caméra dans la direction adéquate, permettant ainsi d'obtenir l'objet d'intérêt dans le champs de vue de la caméra PTZ.

La figure 5.20 montre les résultats obtenus pour la localisation du même objet avec différentes distances focales (donc différents niveaux de zoom). Les erreurs angulaires obtenues sont comprises entre 2.5 et 4° ce qui confirme que l'approche développée fonctionne quel que soit le niveau de zoom.

Tests avec des images réelles : L'image *fisheye* "maître" (figure 5.21(a)) utilisée pour toutes nos expériences reste inchangée tandis que les images prises avec la caméra PTZ sont sujettes à de nombreux changements dans l'environnement où dans d'illumination de la scène.
Dans un premier temps, une surface parfaitement plane et très texturée a été choisie sur l'image *fisheye* (voir la boite englobante rouge sur la figure 5.21(a)), ces éléments consti-

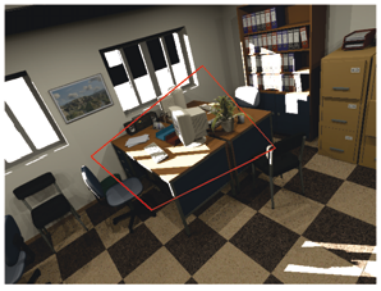
(a)

(b)



(c)

(d)



(e)

FIGURE 5.18 – Résultats de détection obtenus avec ajout de bruit , (a) master image, (b),(c) résultats avec un bruit additif de variance 0.0, 15, 25 et 38.25 pixels respectivement

tuent donc la meilleure configuration pour notre approche. La zone détectée parmi les images PTZ acquises le long du cercle épipolaire est visible sur la figure 5.21(b). Dans ce cas de figure et conformément à nos attentes la détection est très précise.

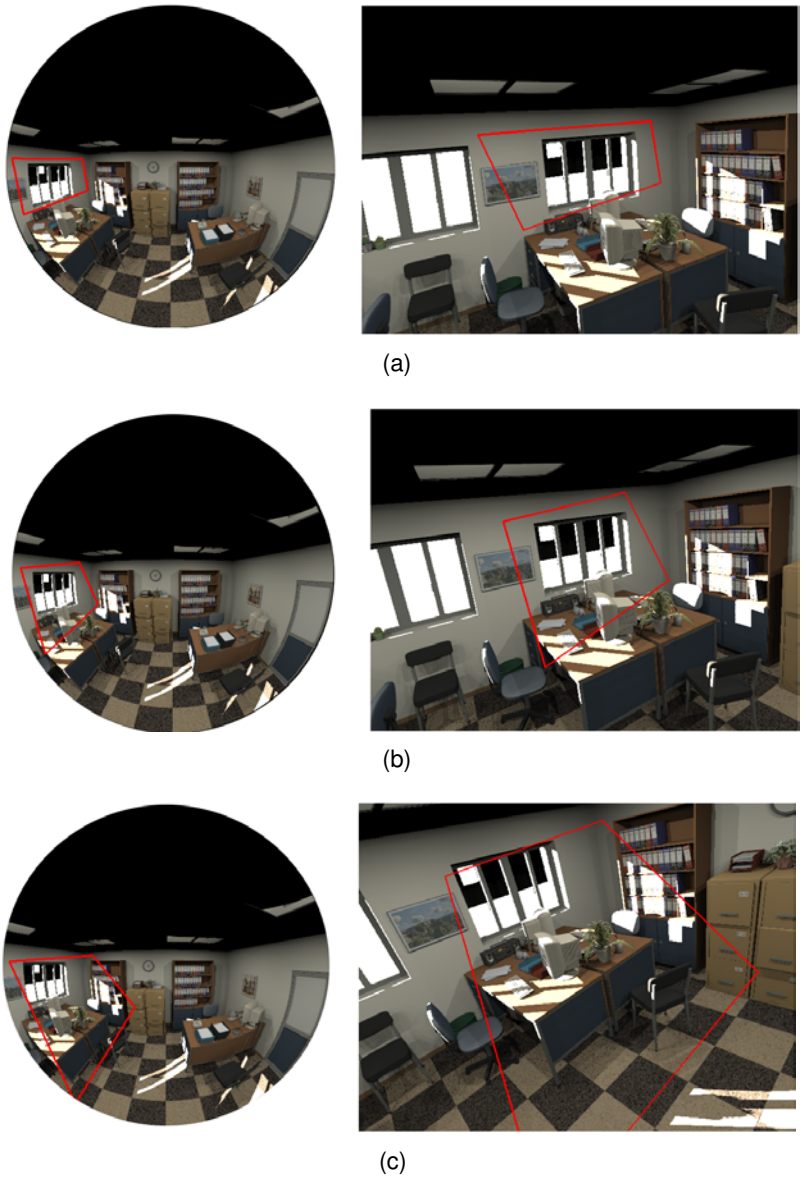


FIGURE 5.19 – Test avec différents type de surfaces (a) planaire (b) quasi-planaire (c) non-planaire

La figure 5.21(e) correspond au résultat d'un test dans des conditions moins favorables. Bien qu'il s'agisse ici encore d'une surface planaire, celle ci contient très peu de textures et les images perspectives couvrant cette zone souffre d'une occultation partielle. Même dans ces circonstances la détection reste très précise.
Nous avons également effectué des tests sur des régions non-planaires, comme dans les figures 5.21(c)(d). Malgré cela leur détection reste assez précise considérant l'application visée.
La figure 5.21(f) présente le résultats d'une détection de cible dans des conditions particulièrement désavantageuses pour notre approche. En effet la région sélectionnée ne respecte par la contrainte de planéité et contient une occultation importante. Même dans

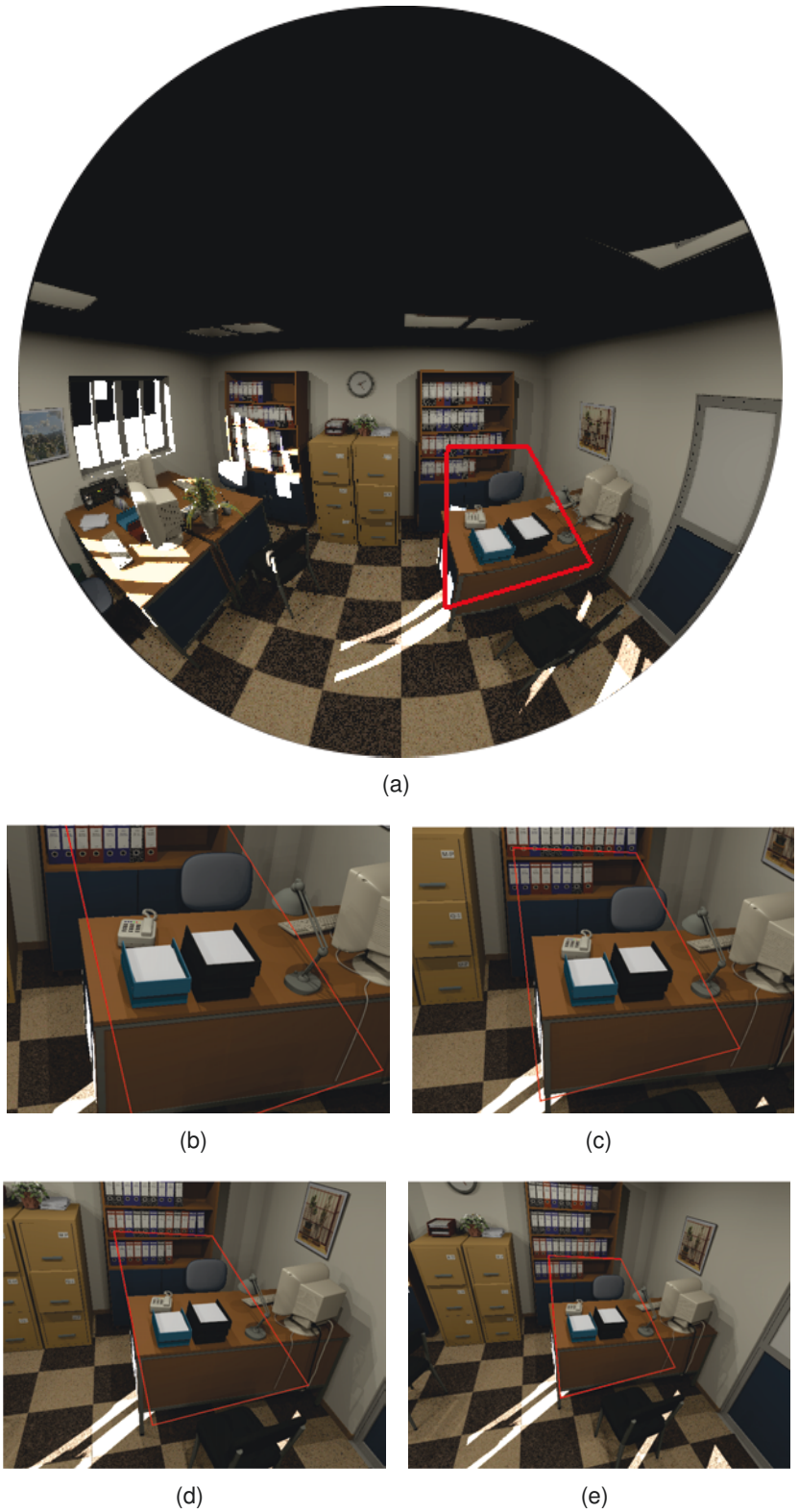


FIGURE 5.20 – Détection de la région d'intérêt avec différents niveaux de zoom, (a) Image *fisheye* avec la zone sélectionnée ; (a-e) résultats pour un champ de vue horizontal de 40,50,65 et 80°

ces conditions notre algorithme permet d'obtenir des résultats très convainquants. Tandis que sur l'ensemble des autre images un zoom optique de notre caméra PTZ est de 1×, la figure 5.21(d) a été acquise avec un zoom de 5×. L'usage de différent niveau de zoom n'affecte que peu les performances de la méthode proposée. Finalement la dernière expérience présenté dans la figure 5.21(h) prouve la robustesse dans le cas de fort changement d'illumination.

5.8/ CONCLUSION

Dans ce chapitre nous avons traité de deux points importants. D'une part nous proposons une méthode de calibrage adaptée à la géométrie particulière de nos caméras, cette approche est particulièrement simple à reproduire et une comparaison avec une approche de l'état de l'art a permis d'évaluer son efficacité.

Nous avons également présenté une approche flexible et efficace pour contrôler une caméra PTZ dans un système de vision hybride. Nous avons ici prouvé qu'il est possible de localiser une cible avec la caméra mécanisée seulement à l'aide d'information provenant d'une caméra omnidirectionnelle à PVU.
Contrairement aux méthodes existantes dans la littérature notre approche fonctionne dans un environnement inconnu sans autre a-priori que le calibrage du banc stéréo. De plus la détection de la cible elle même est purement géométrique tandis que la plupart des approches de mise en correspondance de points utilisent des descripteurs photométriques. Ce qui nous permet de faire face aux problèmes liés à l'utilisation de caméra de types différents tels que, les fortes distorsions, les changements d'illuminations, les différences d'échelle. Cette approche a été testée de manière aussi bien qualitative que quantitative au travers d'une série de tests synthétiques et réels. Ces résultats montrent que sans utilisation de connaissances sur la scène, notre approche permet la localisation d'une cible précise avec ce système de vision hybride.

(a) (b)

(c) (d)

(e) (f)

(g) (h)

FIGURE 5.21 – Détection d'objets divers (a) Image *fisheye* avec les cibles sélectionnées, (b) (c) (d) (e) (f) (g) et (h) cibles détectées sur les images PTZ

6

NAVIGATION ROBOTIQUE AVEC UN SYSTÈME DE STÉRÉO-VISION HYBRIDE OMNIDIRECTIONNELLE/PERSPECTIVE

Ce chapitre est dédié à la navigation robotique à l'aide d'un capteur de vision hybride (décrit par la figure 6.1) constitué d'une caméra omnidirectionnelle et d'une caméra perspective fixe. Un capteur de ce type est particulièrement intéressant pour des applications en odométrie visuelle pour la robotique. L'utilisation d'une caméra omnidirectionnelle offre l'avantage de visualiser l'ensemble de la scène. De plus, il a été démontré dans [75] que l'utilisation de capteur sphérique permet de résoudre des ambiguités lors de déplacements de faibles amplitudes. D'autre part, l'ajout d'une caméra perspective peut servir à retrouver le facteur d'échelle du déplacement et a obtenir une vision plus détaillée de la scène face au robot comme c'est le cas dans [135].
Il sera ici question d'estimer le déplacement à l'échelle d'un robot mobile. Dans un premier temps, nous parlerons des différents types de robots mobiles mais également des capteurs leur permettant de naviguer et tout particulièrement de l'utilisation de caméras pour la robotique. Nous discuterons ensuite des méthodes permettant la reconstruction 3D de la scène et l'estimation de la pose des caméras avec notre dispositif. Finalement, une solution basée sur une approche de type Structure-From-Motion (SFM) sans recouvrement est proposée.

6.1/ LES ROBOTS MOBILES

On appelle robot mobile tout robot pouvant effectuer un déplacement dans son environnement contrairement aux robots manipulateurs fixes majoritairement utilisés pour des tâches industrielles. Les robots mobiles peuvent être classés en plusieurs catégories en fonction de leur mode de locomotion, on retrouvera entre autre les robots aériens, terrestre (à roues, à chenilles, marcheur, etc), sous-marin (voir figure 6.2)... A l'inverse des

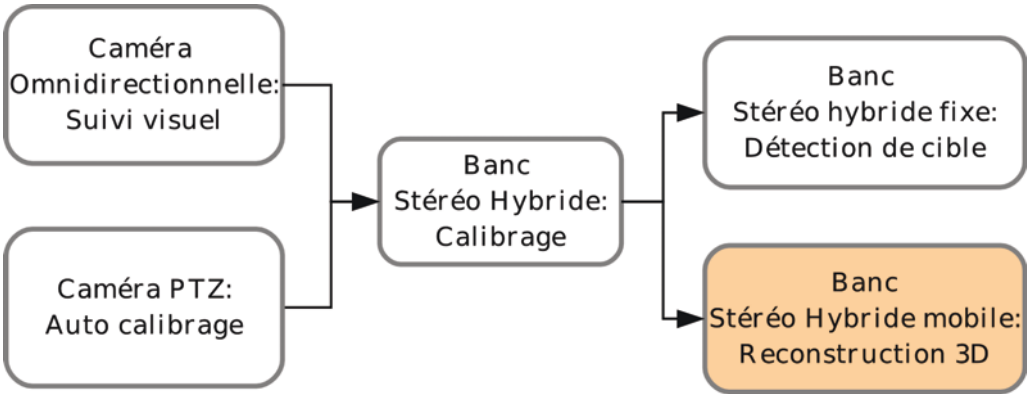
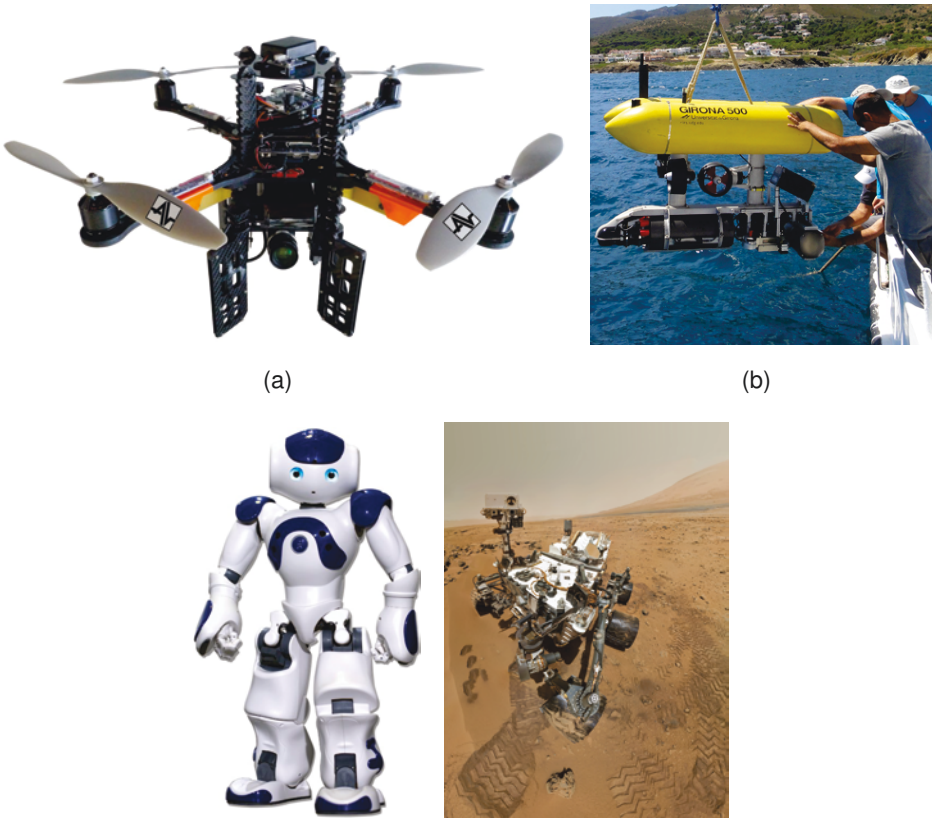


FIGURE 6.1 – Problématique globale

robots manipulateurs qui fonctionnent généralement dans un environnement connu, la difficulté majeure à laquelle sont confrontés les robots mobiles est la navigation dans un espace inconnu. Il est donc essentiel d'équiper les robots d'un certain nombre de capteurs leurs permettant de percevoir leur environnement et d'estimer leurs déplacements dans celui-ci.



(a) (b)

(c) (d)

FIGURE 6.2 – Différents types de robots mobiles ; (a) le Pelican de Astec (b) un robot submersible utilisé à l'université de Gérone (c) le robot humanoide NAO de Aldebaran (d) le rover Curiosity durant sa mission sur Mars

6.1.1/ LES CAPTEURS

Les capteurs embarqués sur les robots mobiles peuvent être classés en deux catégories, les capteurs proprioceptifs et extéroceptifs. La première catégorie de capteur permet d'estimer l'état du robot comme par exemple son changement d'orientation où de position entre deux instants. Les capteurs extéroceptifs permettent quant à eux de fournir des informations sur l'environnement dans lequel évolue le robot, il peut s'agir de la température ambiante, la reconstruction 3D de l'espace environnant, une mesure de distance à un obstacle ...

6.1.1.1/ MESURE DE POSITION

La mesure de position d'un robot mobile terrestre la plus courante est basée sur l'odométrie, cette approche consiste à déterminer le déplacement du robot à l'aide de la mesure de la rotation des roues. On peut ainsi déterminer la position courante du robot par rapport à une position initiale. Les odomètres sont donc des capteurs proprioceptifs. Les odomètres permettant de mesurer la vitesse de rotation des roues sont bien souvent très peu coûteux et compacts. La précision d'un tel dispositif est cependant très dépendante de la surface sur laquelle l'engin progresse, les sols meubles sont par exemple particulièrement problématique en raison du glissement des roues entraînant une dérive de la mesure. En règle générale, l'odométrie permet une mesure assez fidèle sur une courte distance mais reste peu efficace sur de longs trajets.
Lorsqu'il s'agit de déterminer la position du robot dans un repère fixe donné, il est courant d'avoir recours au système GPS (Global Positioning System). Le principe du GPS repose sur un réseau de satellites équipés d'horloges atomiques envoyant des signaux synchronisés. Connaissant la position des satellites et l'instant de réception de ces signaux, un récepteur peut -par triangulation- être localisé si la réception d'au moins quatre satellites est assurée (le quatrième signal permettant d'améliorer la robustesse de la localisation). La précision d'une localisation standard par GPS est de l'ordre de 25m verticalement et 15m horizontalement, ce qui est relativement imprécis si l'on souhaite faire naviguer un robot de manière autonome. De plus, ce système est inutilisable en intérieur où les signaux provenant des satellites ne peuvent pas être reçus. Ce système a initialement été développé par le département de la défense des États-Unis et reste sous leur contrôle.
Une localisation plus précise peut être effectuée à l'aide du système DGPS (Differencial Global Positioning System). Ce dispositif permet de réduire l'erreur de localisation d'un GPS classique à l'aide d'un réseau de stations au sol. La mise en pratique d'un tel système est donc particulièrement compliquée et coûteuse.

6.1.1.2/ MESURE DE L'ORIENTATION

Les capteurs proprioceptifs tels que les magnétomètres, gyroscopes, gyromètres et accéléromètres permettent d'estimer l'orientation du robot. D'autres capteurs extéroceptifs comme les compas ou les boussoles électroniques peuvent déterminer l'orientation par rapport au nord magnétique. Généralement ces capteurs sont embarqués dans une centrale inertielle où l'ensemble des données provenant de ces différents appareils sont fusionnées afin d'obtenir une estimation précise de l'orientation du robot. Les centrales inertielles au même titre que l'odométrie sont particulièrement sensibles à l'accumulation d'erreur sur une longue période de mesure.

6.1.1.3/ MESURE DE LA SCÈNE

Certains capteurs permettent une estimation ponctuelle de distance avec l'environnement comme les capteurs infrarouge (inefficace au delà d'un mètre) et les capteurs ultrason. Ils sont surtout utilisés pour éviter les collisions avec les éléments présents dans la scène mais ne peuvent généralement pas être utilisés pour déterminer la position du robot dans l'espace. D'autres outils plus polyvalents sont conçus pour effectuer une reconstruction 3D de l'environnement, les plus populaires étant les télémètres à balayage laser permettant la mesure d'un profil 3D de la scène. Ils sont relativement abordables et admettent une précision de l'ordre du centimètre. Des dispositifs plus complexes peuvent aussi être mis en oeuvre afin de reconstruire plus densément l'environnement comme les caméras à temps de vol et les Lidars (light detection and ranging).
Ces différents outils sont très polyvalents car ils permettent à la fois une reconstruction 3D mais aussi un calcul de la pose du robot sur lequel ils sont montés. Cette approche de cartographie et localisation simultanées est connu sous le nom de SLAM (Simultaneous Localization and Mapping). La mise en commun des reconstructions peut se faire par fusion avec les informations provenant d'autres capteurs équipant le robot. Beaucoup d'approches proposent également un calcul de pose basé sur le recalage de reconstructions 3D consécutives [69, 131].

6.2/ LA VISION PAR ORDINATEUR POUR LA NAVIGATION ROBOTIQUE

L'utilisation de caméra pour la robotique mobile est un moyen privilégié et très efficace pour extraire un maximum d'information de l'environnement et estimer les déplacements du robot. En plus de fournir des informations concernant la couleur et la texture des objets composant la scène, il est possible de simultanément calculer la structure tridimensionnelle de l'environnement et la pose de la caméra en déplacement. Cette stratégie est connu sous le nom de VSLAM (Visual Simultaneous Localization and Mapping)

[52]. Rappelons que le VSLAM et la SFM sont des méthodes très similaires, le terme de VSLAM est surtout retenu pour les applications robotiques en temps réel où des approches probabilistes sont utilisées, tandis que l'appellation SFM est plus générique. La quasi totalité des méthodes de reconstruction 3D à base d'images consistent d'ailleurs en une résolution conjointe à la fois de la structure et du mouvement, les deux étant interdépendants. Il n'est pas rare de faciliter la reconstruction 3D à l'aide de paramètres mesurés par d'autres capteurs présent sur le robot [100].
L'utilisation des capteurs RGB-D pour la navigation robotique est également de plus en plus répandue [60]. L'usage de ce type de capteur permet une localisation rapide et efficace du robot dans son environnement mais n'est utilisable qu'en intérieur puisque la visualisation du motif infra-rouge est très sensible à la lumière du jour.
D'autres approches plus spécifiques permettent également l'estimation des paramètres d'état d'un robot mobile à l'aide de différentes contraintes ou *a priori* sur la scène. Dans [57] l'assiette d'un drone est calculée à partir de la ligne d'horizon visible avec une caméra omnidirectionnelle. De la même manière, l'attitude d'un satellite peut être estimée à l'aide du suivi des étoiles [15]. Plusieurs méthodes plus adaptées au milieu urbain proposent de calculer la rotation de la caméra avec les points de fuites obtenus par la détection des lignes parallèles détectées dans les images [25].
D'autres travaux suggèrent l'utilisation de balises visuelles afin de faciliter la localisation du robot [9], ces méthodes sont cependant plus restrictives puisque des marqueurs visuels doivent dans un premier temps être installés dans l'environnement.
Des contraintes d'ordre géométrique peuvent également être utilisées, notamment l'utilisation de surfaces planaires [125] pour déterminer la pose de la caméra.
Les méthodes présentées sont cependant inefficaces dans le cas qui nous concerne, c'est-à-dire l'association d'une caméra perspective et d'une caméra omnidirectionnelle. Nous verrons dans la section suivante que peu de méthodes existent spécifiquement pour cette configuration.

6.3/ RECONSTRUCTION 3D ET LOCALISATION AVEC UN BANC DE CAMÉRA HYBRIDE

Les méthodes de "structure-from-motion" classiques consistent en l'estimation de la structure et du mouvement d'une seule caméra en déplacement à partir d'une séquence d'images [106]. Quand cette méthode est utilisée pour le calcul de pose d'un robot les images sont ordonnées temporellement, elles sont donc traitées les unes après les autres à chaque nouvelle acquisition, on parle alors de SFM séquentiel. La reconstruction 3D ainsi que les poses calculées par ce procédé seront à un facteur d'échelle près, cela pose un problème lorsqu'il est question -comme c'est souvent le cas- de connaître le déplacement du robot à l'échelle réelle. L'utilisation de plusieurs caméras calibrées permet de résoudre ce problème, le cas le plus basique étant un système de stéréo-vision. Cette

configuration nécessite cependant un calibrage préalable afin de déterminer la rotation et la translation entre les deux caméras, de plus une synchronisation est nécessaire afin de permettre une prise d'image au même instant. Contrairement au cas mono-caméra présenté précédemment ; en plus des correspondances de mouvements s'ajoutent les correspondances stéréo, cette mise en correspondance permet à chaque instant une reconstruction de l'environnement à l'échelle réelle (voir figure 6.3). La plupart des approches peuvent se décomposer en deux étapes, la première étant la mise en correspondance à un instant t des deux images du capteur de vision stéréoscopique permettant le calcul d'un premier nuage de points, ces primitives sont suivies dans les images suivantes à l'instant $t + 1$. Un calcul de pose peut alors être effectué par minimisation de l'erreur de reprojection des points 3D sur les images acquises à $t + 1$ [71].

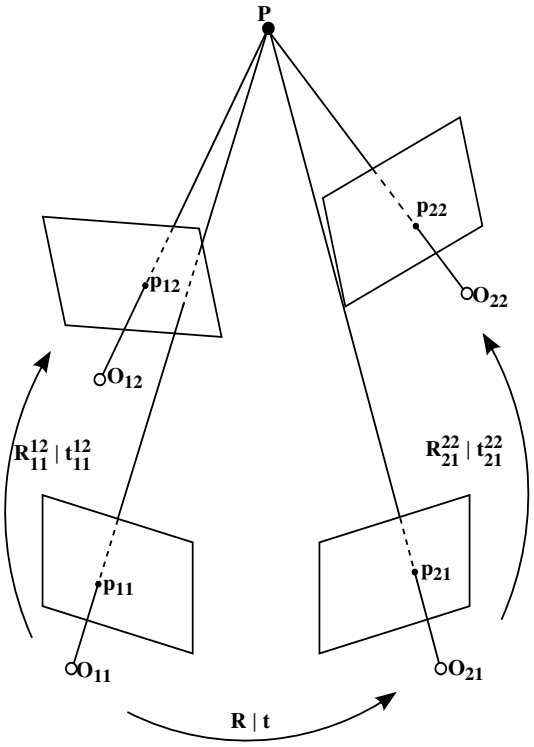


FIGURE 6.3 – Capteur de stéréo-vision en mouvement

D'autres techniques sont possibles, [105] propose une estimation du mouvement à l'aide de deux tenseurs trifocaux "entrelacés" afin d'estimer les six degrés de libertés constituant la pose du banc de caméras. Une approche plus sophistiquée présentée dans [44] décrit l'utilisation d'un tenseur quadrifocal permettant un calcul du mouvement par recalage dense.

Ces approches sont très efficaces comme le montre le comparatif sur la base de données KITTI [70], cependant elles ne concernent que les bancs de stéréo-vision homogènes constitués de caméras perspectives. Très peu de travaux se sont pour l'instant concentrés sur l'odométrie visuelle avec un capteur de stéréo-vision hybride. Dans [61],

le déplacement d'un drone est estimé à l'aide d'un banc de vision de ce type mais reste limité au cas des robots aériens puisque l'hypothèse permettant de calculer le déplacement de l'appareil considère un sol parfaitement planaire. A notre connaissance aucun travail ne s'est pour l'instant intéressé à la reconstruction 3D à l'aide d'un banc de vision hybride équipé à la fois d'une caméra omnidirectionnelle et d'une caméra perspective. La principale difficulté étant la mise en correspondance stéréo entre deux images de natures très différentes. Plusieurs travaux existent sur ce problème particulier, la plupart des approches actuelles reposent sur la recherche de correspondances entre les images à l'aide de descripteurs adaptés (comme Harris [55] où SIFT [49]) à la géométrie des caméras associées à un modèle géométrique approprié afin d'éliminer les points aberrants. Ces méthodes permettent une reconstruction 3D de l'environnement ainsi qu'une localisation des caméras, cependant elles ne concernent pas un banc calibré comme c'est le cas dans notre étude. Ce calibrage nous fournit des informations qui peuvent faciliter la mise en correspondance entre les images comme cela a déjà été souligné dans le chapitre 5. La rectification des images -à l'aide de ces paramètres de calibrage- peut également être une solution viable pour faciliter la mise en correspondance entre des images de natures différentes. Cependant, malgré une étape de rectification épipolaire la mise en correspondance entre deux images de résolutions très différentes ne permet pas une précision satisfaisante pour une estimation correcte du mouvement. La mise en correspondance inter-caméra est donc une étape particulièrement complexe et nécessite l'utilisation de plusieurs outils sophistiqués, de plus la précision d'un tel processus peut être très variable en fonction de la dissimilarité entre les deux images. Cette différence est accentuée à mesure que la distance focale de la caméra perspective augmente, ce qui peut rendre les approches décrites précédemment inefficaces.
En revanche l'appariement de points entre caméra de même nature est une étape relativement basique que ce soit pour les images perspectives où pour les images de type omnidirectionnelles puisque les descripteurs usuels y sont très efficaces. La méthode proposée ici prend avantage de cela en s'affranchissant de la mise en correspondance stéréoscopique à l'aide d'une méthode de SFM sans recouvrement.

6.4/ ESTIMATION DE LA STRUCTURE ET DU MOUVEMENT STÉRÉOSCOPIQUE SANS RECOUVREMENT

La figure 6.4 est une description grossière d'un dispositif de vision multi-caméra sans recouvrement où la matrice $\mathbf{M} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix}$ correspond à la transformation rigide entre les deux caméras fixées sur un support commun, tandis que $\mathbf{M}_{L1}^{L2}$ et $\mathbf{M}_{R1}^{R2}$ représentent respectivement le mouvement de la caméra de gauche et de la caméra de droite d'une position 1 à une position 2.

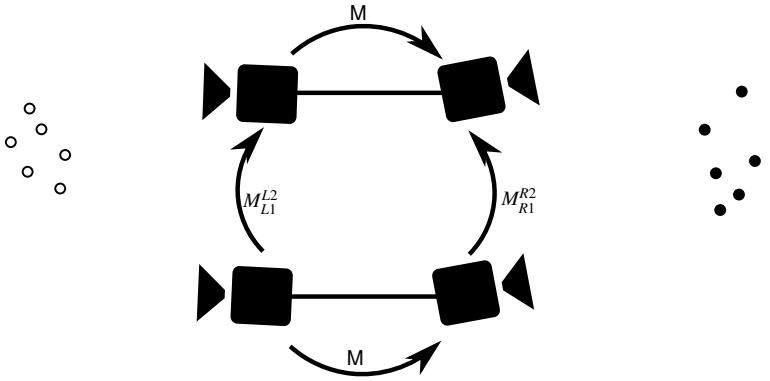


FIGURE 6.4 – stéréo vision sans recouvrement

Si l'on exclut l'utilisation du modèle de projection généralisé, deux approches se distinguent parmi les méthodes de SFM avec une telle configuration.
La première méthode développée dans [103] et [101], consiste dans un premier temps à considérer les caméras comme étant totalement indépendantes afin d'appliquer une approche de SFM monoculaire sur chacune d'elles, permettant ainsi de calculer $\mathbf{M}_{L1}^{Ln}(\lambda_L) = \begin{pmatrix} \mathbf{R}_{L1}^{Ln} & \mathbf{t}_{L1}^{Ln}\lambda_L \\ 0 & 1 \end{pmatrix}$ et $\mathbf{M}_{R1}^{Rn}(\lambda_R) = \begin{pmatrix} \mathbf{R}_{R1}^{Rn} & \mathbf{t}_{R1}^{Rn}\lambda_R \\ 0 & 1 \end{pmatrix}$ pour le déplacement n. Ces mouvements évalués séparément sont estimés aux échelles λ_L et λ_R près. Puisque les caméras sont fixées sur un support rigide alors $\mathbf{R}$ et $\mathbf{t}$ sont constants et connus à chaque instant, il est donc possible d'imposer cette contrainte entre les caméras afin d'estimer le déplacement à l'échelle réelle. Considérant un déplacement unique comme présenté sur la figure 6.4, la relation entre toutes les transformations rigides du système peut s'exprimer ainsi :

$$\mathbf{M}(\mathbf{M}_{L1}^{L2}(\lambda_L))^{-1}\mathbf{M}^{-1}\mathbf{M}_{R1}^{R2}(\lambda_R) = \mathbf{I}. \tag{6.1}$$

Cette relation peut être développée sous la forme suivante :

$$[-\mathbf{R}_{R1}^{R2}\mathbf{t}_{R1}^{R2} \quad \mathbf{R}(\mathbf{R}_{L1}^{L2})^T\mathbf{t}_{L1}^{L2}]\begin{bmatrix}\lambda_L \\ \lambda_R\end{bmatrix} = [\mathbf{t} - (\mathbf{R}_{L1}^{L2})^T\mathbf{t}] \tag{6.2}$$

Il est donc possible de résoudre linéairement et conjointement les deux facteurs d'échelle λ_L et λ_R. Un seul mouvement n'est cependant pas suffisant pour une estimation robuste des facteurs d'échelle, il est en effet préférable de choisir un ensemble de déplacements afin d'effectuer ce calcul. L'équation (6.2) peut être généralisée pour un ensemble de n déplacements :

$$\begin{bmatrix} -\mathbf{R}_{R1}^{R2}\mathbf{t}_{R1}^{R2} & \mathbf{R}(\mathbf{R}_{L1}^{L2})^T\mathbf{t}_{L1}^{L2} \\ & \vdots \\ -\mathbf{R}_{R1}^{Rn}\mathbf{t}_{R1}^{Rn} & \mathbf{R}(\mathbf{R}_{L1}^{Ln})^T\mathbf{t}_{L1}^{Ln} \end{bmatrix}\begin{bmatrix}\lambda_L \\ \lambda_R\end{bmatrix} = \begin{bmatrix} \mathbf{t} - (\mathbf{R}_{L1}^{L2})^T\mathbf{t} \\ \vdots \\ \mathbf{t} - (\mathbf{R}_{L1}^{Ln})^T\mathbf{t} \end{bmatrix} \tag{6.3}$$

Pour cette approche le déplacement de chaque caméra constituant le banc doit être estimé individuellement, l'approche minimale pour le calcul de poses de caméras calibrées

nécessite un minimum de 5 points (voir chapitre 3). Les méthodes [103] et [101] nécessitent donc au minimum 5 points de correspondance par déplacement et par caméra, ce qui représente un total de 10 points de correspondance pour résoudre les 6 degrés de liberté d'un déplacement du banc de vision.
Le travail de Clipp *et al.* [42] permet une résolution du problème à l'échelle réelle avec seulement 6 points de correspondance au total (5 sur une caméra et 1 sur la seconde). Cette méthode propose d'estimer d'abord le déplacement d'une des deux caméras (prenons ici la caméra de gauche L) $\mathbf{M}_{L1}^{L2}(\lambda_L)$ à l'échelle λ_L près à l'aide de l'algorithme des cinq points. Il est donc possible de connaître le déplacement de la seconde caméra $\mathbf{M}_{R1}^{R2}(\lambda_L)$, d'après la relation :

$$\mathbf{M}_{R1}^{R2}(\lambda_L) = \mathbf{M}\mathbf{M}_{L1}^{L2}(\lambda_L)\mathbf{M}^{-1}. \tag{6.4}$$

La matrice essentielle liant la première et la seconde vues de la caméra de droite $\mathbf{E}_{R1}^{R2}$ est donc connue à l'échelle λ_L près, cette inconnue peut alors être déterminée à l'aide d'un seul point. L'utilisation d'un algorithme RANSAC permet d'estimer ce paramètre de manière robuste et d'éliminer les mauvaises correspondances entre les vues.
Ces approches offrent une solution relativement simple au problème de SFM sans recouvrement, on notera cependant une forte sensibilité au cas dégénérés tels que les translations pures et les déplacements autour d'un même axe de rotation.

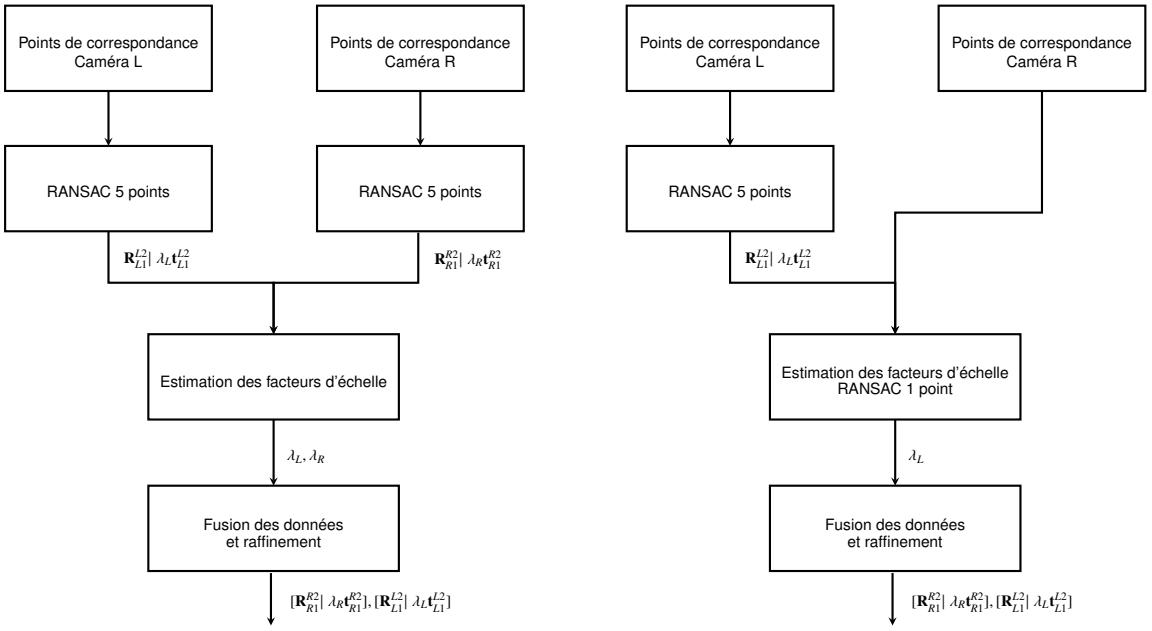


FIGURE 6.5 – Algorithmes de SFM sans recouvrement, à gauche méthode de[103], à droite méthode de [42]

6.5/ MÉTHODOLOGIE

Nous proposons ici une adaptation de l'approche proposée par Clipp *et al.* [42] à notre cas c'est-à-dire avec un système de vision hybride en mouvement comme décrit sur la figure 6.6. La méthode d'origine étant très sensible aux mouvements dégénérés, nous proposons ici une nouvelle formulation à l'aide de tenseurs trifocaux. Nous verrons aussi que la substitution des matrices essentielles utilisées dans la méthode originale par des tenseurs permet une plus grande robustesse au bruit.

6.5.1/ NOTRE SYSTÈME EN MOUVEMENT

Dans cette section, nous analysons les relations existantes entre les vues de notre système. Nous considérons ici deux déplacements consécutifs à un temps $t+1$ et $t+2$. Notre système est considéré comme fixe et calibré cela signifie que la transformation inter-caméra entre la caméra *fisheye* o et la caméra perspective p exprimée par la matrice $\mathbf{M}_o^p = \begin{pmatrix} \mathbf{R}_o^p & \mathbf{t}_o^p \\ 0 & 1 \end{pmatrix}$ est fixe et connue à chaque instant. Les déplacements de la caméra omnidirectionnelle et de la caméra perspective sont donc directement liés par les relations suivantes :

$$\mathbf{M}_o^p \mathbf{M}_{o1}^{o2} (\mathbf{M}_o^p)^{-1} = \mathbf{M}_{p1}^{p2} \tag{6.5}$$

$$\mathbf{M}_o^p \mathbf{M}_{o1}^{o3} (\mathbf{M}_o^p)^{-1} = \mathbf{M}_{p1}^{p3} \tag{6.6}$$

Ce lien rigide entre les deux caméras réduit le nombre de degrés de liberté de notre système à 6 dans le cas d'un seul mouvement, ce qui est le cas traité dans [42], et à 11 dans le cas où deux déplacements successifs sont effectués par le banc de vision, c'est le cas que nous traiterons ici.
Malgré l'existence d'un champ recouvrant entre les deux caméras, notre approche ne considère aucune correspondance stéréo, cela signifie que les points détectés sur une caméra sont considérés uniquement visibles par celle-ci.
La seule condition requise pour le fonctionnement de notre approche est la détection de six triplets de points de correspondance temporelle sur une des caméras de notre système et un triplet de points sur l'autre.

Notre méthode se base essentiellement sur le fait qu'il est possible d'estimer le déplacement d'une des caméras à l'aide du calcul d'un tenseur trifocal avec un ensemble minimum de six triplets de points de correspondance temporelle [87]. Il existe également une solution minimale pour le cas calibré proposée par Nister et Schaffalitzky dans [136], elle reste cependant relativement difficile à mettre en œuvre. Il est possible d'extraire les matrices de projection d'un tenseur, à l'aide des différentes méthodes décrites dans [87]. Au même titre que la matrice essentielle, cette estimation est effectuée à

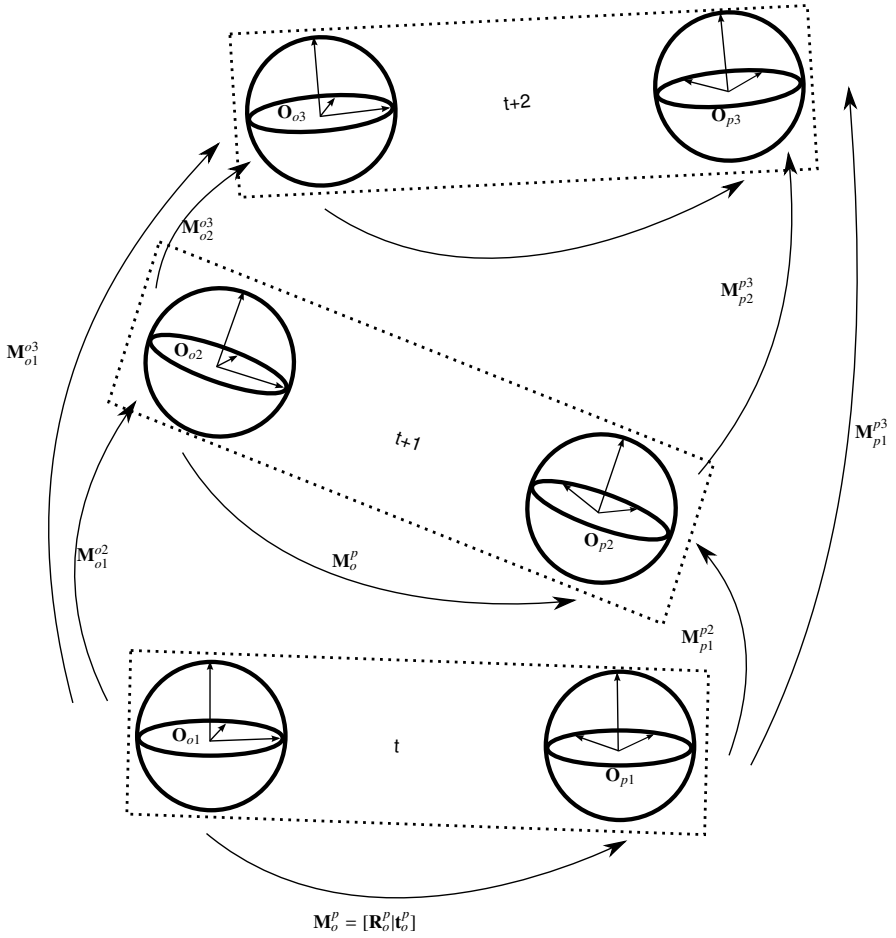


FIGURE 6.6 – Notre système en déplacement

l'échelle près, ne laissant ainsi qu'un seul degré de liberté à estimer. Celui-ci peut être résolu à l'aide d'un triplet de correspondance unique sur l'autre caméra constituant le banc de vision. Au total cela représente une solution minimale nécessitant seulement 7 triplets de points afin de résoudre à la fois le déplacement mais aussi son échelle. Notons que cette approche peut facilement être étendue à un système faisant intervenir un plus grand nombre de caméras. De plus, elle est compatible avec toutes les caméras à point de vue unique grâce à l'utilisation du modèle de projection sphérique unifié.

6.5.1.1/ ESTIMATION DES MOUVEMENTS DE LA CAMÉRA OMNIDIRECTIONNELLE

Dans le chapitre 3 nous avons vu le formalisme du tenseur trifocal, sa construction à partir des matrices de projection et les transferts de lignes et de points possibles entre trois vues. Nous étudierons dans cette section le calcul d'un tenseur trifocal à partir d'une caméra omnidirectionnelle calibrée ainsi que l'extraction des matrices de projections le composant. Dans nos explications nous omettons volontairement les

étapes de normalisation et de dénormalisation des points (inutiles dans le cas calibré qui nous concerne). Les explications concernant le calcul du tenseur sont directement inspirées du livre de Hartley et Zisserman [87].

Calcul du tenseur trifocal Le transfert point-point-point, entre trois vues omnidirectionnelles $o1$, $o2$ et $o3$, présenté dans le chapitre 3, peut s'écrire sous la forme :

$$\mathbf{P_{o1}}^{i}\mathbf{P_{o2}}^{j}\mathbf{P_{o3}}^{k}\varepsilon_{ipw}\varepsilon_{jqs}\varepsilon_{krt}\mathbf{T}_{o}^{i,q,r} = 0^{st} \tag{6.7}$$

avec $\mathbf{P}_{o1}$, $\mathbf{P}_{o2}$ et $\mathbf{P}_{o3}$ respectivement les points d'un triplet de correspondance sur la première, seconde et troisième vue et $\mathbf{T}_o$ le tenseur trifocal existant entre les trois images. La résolution du tenseur $\mathbf{T}_o$ nécessite la reformulation de la relation 6.7 sous la forme d'une équation matricielle $\mathbf{At_o} = 0$ avec $\mathbf{t_o}$ un vecteur contenant les entrées du tenseur. La relation 6.7 contient 9 équations dont seulement 4 sont linéairement indépendantes. Cela signifie qu'il n'est pas nécessaire d'utiliser la totalité des équations et que seulement deux valeurs pour les indices s et t (par exemple 1 et 2) sont suffisantes pour extraire les quatre équations désirées. Pour un choix quelconque de s et t on peut réécrire la relation d'incidence précédente sous la forme :

$$\mathbf{P_{o1}}^{k}(\mathbf{P_{o2}}^{i}\mathbf{P_{o3}}^{m}\mathbf{T}_{o}^{j,l,k} - \mathbf{P_{o2}}^{j}\mathbf{P_{o3}}^{m}\mathbf{T}_{o}^{i,l,k} - \mathbf{P_{o2}}^{i}\mathbf{P_{o3}}^{l}\mathbf{T}_{o}^{j,m,k} + \mathbf{P_{o2}}^{j}\mathbf{P_{o3}}^{l}\mathbf{T}_{o}^{i,l,k}) = \mathbf{0}^{ijlm} \tag{6.8}$$

avec $i, j \neq s$ et $l, m \neq t$. La sélection des équations linéairement indépendantes peut alors se faire en forçant $j = m = 3$ (avec $s = 1$ et $t = 2$) et en choisissant différentes valeurs pour les indices i et l. Avec par exemple i,l=1,2 on obtient les quatre équations nécessaires à la résolution du tenseur.

Un tenseur trifocal contient 27 entrées (matrice de taille 3x3x3), puisque la résolution du tenseur se fait à un facteur d'échelle près, 26 paramètres sont à résoudre (pour 12 degrés de liberté dans le cas calibré). Chaque point fournissant 4 équations, il faut un minimum de 7 points pour résoudre le tenseur. Cette résolution purement linéaire du tenseur trifocal est la plus simple mais d'autres algorithmes tels que celui présenté dans [169] permettent une résolution du tenseur à l'aide de 6 triplets de points seulement. Cette méthode fournit cependant 3 solutions possibles dont une d'entre elles est géométriquement correcte. Pour le cas calibré, il existe aussi des solutions minimales où un ensemble de seulement 4 triplets de points de correspondance est nécessaire [136], ce type d'approche est cependant moins robuste [136] que les autres méthodes courantes.

Extraction des matrices de projection Une fois le tenseur calculé il est possible d'en extraire les matrices de projection, dans le cas d'une caméra calibrée ce sont donc les poses des caméras que l'on obtient. Si l'on considère la première caméra comme réfé-

rentiel ($\mathbf{M}_{o1}^{o1} = [\mathbf{I} \mid \mathbf{0}]$), il est possible d'extraire les épipoles des deux autres vues. Avec la notation matricielle du tenseur suivante $\mathbf{T}_o = [\mathbf{T}_{o1}, \mathbf{T}_{o2}, \mathbf{T}_{o3}]$. Les épipoles des deux vues $\mathbf{e}_2$ et $\mathbf{e}_3$ ayant pour matrice de projection $\mathbf{M}_{o1}^{o2}$ et $\mathbf{M}_{o1}^{o3}$ peuvent être calculés à l'aide des relations suivantes :

$$\mathbf{e}_2^T[\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] = \mathbf{0} \qquad \text{et} \qquad \mathbf{e}_3^T[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = \mathbf{0} \tag{6.9}$$

où $\mathbf{u}_i$ et $\mathbf{v}_i$ correspondent respectivement aux vecteurs singuliers à gauche et à droite des matrices $\mathbf{T}_{oi}$. Les épipoles sont ensuite normalisés afin de calculer les matrices de projection de la manière suivante :

$$\mathbf{M}_{o1}^{o2} = [[\mathbf{T}_{o1}, \mathbf{T}_{o2}, \mathbf{T}_{o3}]\mathbf{e}_3 \mid \mathbf{e}_2] \qquad \text{et} \qquad \mathbf{M}_{o1}^{o3} = [(\mathbf{e}_3\mathbf{e}_3^T - \mathbf{I})[\mathbf{T}_{o1}, \mathbf{T}_{o2}, \mathbf{T}_{o3}]\mathbf{e}_2 \mid \mathbf{e}_3]. \tag{6.10}$$

Ces matrices de projection sont calculées à un facteur d'échelle près : $\mathbf{M}_{o1}^{o1} = [\mathbf{I} \mid \mathbf{0}]$, $\mathbf{M}_{o1}^{o2}(\lambda) = [\mathbf{R}_{o1}^{o2} \mid \mathbf{t}_{o1}^{o2}\lambda]$ et $\mathbf{M}_{o1}^{o3}(\lambda) = [\mathbf{R}_{o1}^{o3} \mid \mathbf{t}_{o1}^{o3}\lambda]$.

6.5.1.2/ ESTIMATION DU FACTEUR D'ÉCHELLE

L'estimation des deux premiers déplacements de la caméra *fisheye* rendue possible par le calcul d'un tenseur trifocal nous fournit $\mathbf{M}_{o1}^{o2}(\lambda)$ et $\mathbf{M}_{o1}^{o3}(\lambda)$ où λ correspond à un facteur d'échelle.
Dans cette section nous décrivons une approche permettant de retrouver le facteur d'échelle λ à l'aide d'un seul triplet de points sur la caméra perspective.
Le tenseur trifocal $\mathbf{T}_p^{123}$ liant les vues de la caméra perspective $p1$, $p2$ et $p3$ peut s'exprimer de la manière suivante :

$$\mathbf{T}_p^{123} = [\mathbf{T}_{p1},\ \mathbf{T}_{p2},\ \mathbf{T}_{p3}] \tag{6.11}$$

$$\mathbf{T}_{p1} = \mathbf{M}_{p1}^{p2}[({}^2\mathbf{M}_{p1}^{p1})^T.{}^3\mathbf{M}_{p1}^{p1} - {}^3\mathbf{M}_{p1}^{p1}.{}^2\mathbf{M}_{p1}^{p1}]_{[\times]}(\mathbf{M}_{p1}^{p3})^T \tag{6.12}$$

$$\mathbf{T}_{p2} = \mathbf{M}_{p1}^{p2}[({}^3\mathbf{M}_{p1}^{p1})^T.{}^1\mathbf{M}_{p1}^{p1} - {}^1\mathbf{M}_{p1}^{p1}.{}^3\mathbf{M}_{p1}^{p1}]_{[\times]}(\mathbf{M}_{p1}^{p3})^T \tag{6.13}$$

$$\mathbf{T}_{p3} = \mathbf{M}_{p1}^{p2}[({}^1\mathbf{M}_{p1}^{p1})^T.{}^2\mathbf{M}_{p1}^{p1} - {}^2\mathbf{M}_{p1}^{p1}.{}^1\mathbf{M}_{p1}^{p1}]_{[\times]}(\mathbf{M}_{p1}^{p3})^T \tag{6.14}$$

où ${}^i\mathbf{M}$ correspond à la $i^{\text{ème}}$ lignes de la matrice $\mathbf{M}$. Les relations (6.5) permettent d'aboutir aux équations suivantes qui nous servirons à réécrire le tenseur de la caméra perspective avec les matrices de projections de la caméra omnidirectionnelle :

$$\mathbf{M}_{p1}^{p1} = \mathbf{M}_o^p, \tag{6.15}$$

$$\mathbf{M}_{p1}^{p2} = \mathbf{M}_o^p\mathbf{M}_{o1}^{o2}(\lambda), \tag{6.16}$$

$$\mathbf{M}_{p1}^{p3} = \mathbf{M}_o^p\mathbf{M}_{o1}^{o3}(\lambda). \tag{6.17}$$

Le tenseur peut donc se réécrire ainsi :

$$\mathbf{T}_p^{123} = [\mathbf{T}_{p1},\ \mathbf{T}_{p2},\ \mathbf{T}_{p3}] \tag{6.18}$$

$$\mathbf{T}_{p1} = \mathbf{M}_o^p \mathbf{M}_{o1}^{o2}(\lambda)[({}^2\mathbf{M}_o^p)^T.{}^3\mathbf{M}_o^p - {}^3\mathbf{M}_o^p.{}^2\mathbf{M}_o^p]_{[\times]}\mathbf{M}_o^p\mathbf{M}_{o1}^{o3}(\lambda), \tag{6.19}$$

$$\mathbf{T}_{p2} = \mathbf{M}_o^p \mathbf{M}_{o1}^{o2}(\lambda)[({}^3\mathbf{M}_o^p)^T.{}^1\mathbf{M}_o^p - {}^1\mathbf{M}_o^p.{}^3\mathbf{M}_o^p]_{[\times]}\mathbf{M}_o^p\mathbf{M}_{o1}^{o3}(\lambda), \tag{6.20}$$

$$\mathbf{T}_{p3} = \mathbf{M}_o^p \mathbf{M}_{o1}^{o2}(\lambda)[({}^1\mathbf{M}_o^p)^T.{}^2\mathbf{M}_o^p - {}^2\mathbf{M}_o^p.{}^1\mathbf{M}_o^p]_{[\times]}\mathbf{M}_o^p\mathbf{M}_{o1}^{o3}(\lambda). \tag{6.21}$$

On connaît maintenant l'ensemble des entrées de ce tenseur à l'exception du facteur d'échelle λ.
La fonction de transfert point-point-point ($\mathbf{P}_{p1}-\mathbf{P}_{p2}-\mathbf{P}_{p3}$) validant ce tenseur s'écrit conventionnellement sous la forme suivante :

$$\mathbf{P}_{p2[\times]}(\sum_{i=1}^{3}\mathbf{P}_{p1}^i\mathbf{T}_{pi})\mathbf{P}_{p3[\times]} = \mathbf{0}_{3\times3}. \tag{6.22}$$

L'échelle inconnue est une composante du tenseur $\mathbf{T}_{pi}$, chaque triplet de point fournie 9 équations toutes linéairement dépendantes de la forme 6.22, à partir desquelles on peut calculer le facteur d'échelle λ en les réécrivant sous la forme $a\lambda + b = 0$ à l'aide d'un logiciel de calcul formel. Ces équations permettent de résoudre λ à l'aide d'un seul triplet de points de correspondance.

6.5.2/ LES CONFIGURATIONS DÉGÉNÉRÉES

L'ensemble des méthodes de SFM sans recouvrement souffre de plusieurs configurations dégénérées. Dans [103] et [42] les auteurs font état de cas de figures où l'obtention du facteur d'échelle n'est pas possible. Ces deux mouvements étant les déplacements autour d'un même axe de rotation et les translations pures.
Dans [42], l'auteur propose une détection simple des cas dégénérés. Pour chaque déplacement du capteur, le facteur d'échelle calculé est multiplié par deux et le nombre d'*inliers* est recalculé, si cela n'affecte pas le nombre de correspondances correctes il s'agit alors d'un cas dégénéré.
Avec notre approche ce n'est plus une matrice essentielle mais un tenseur trifocal qui est utilisé, les cas dégénérés restent les mêmes. Cependant, si au moins une des deux poses composant le tenseur est non dégénérée alors le calcul de l'échelle reste possible. Les démonstrations arithmétiques et géométriques de ces cas dégénérés sont disponibles dans [103] et [42].

6.5.3/ L'ALGORITHME

Dans un premier temps, les points détectés sur les trois vues *fisheye* et perspectives sont projetés sur la sphère unité. Ces trois vues sont nécessaires au calcul du tenseur

trifocal sur lequel repose notre méthode.
Ensuite, six triplets de points détectés sur la caméra *fisheye* sont utilisés afin d'estimer le tenseur trifocal entre les trois vues. La méthode employée pour ce calcul est celle décrite dans [87]. On notera toutefois qu'il est possible d'utiliser des approches de détermination du tenseur trifocal (7 points) plus simple aboutissant à des résultats similaires mais nécessitant plus de points. L'estimation robuste de ce tenseur est assuré par l'algorithme RANSAC nous permettant ainsi une élimination efficace des mauvaises correspondances. L'erreur de re-projection sur les sphères est utilisée comme critère de réjection dans RANSAC.
Les poses de la caméra *fisheye* peuvent alors être extraites du tenseur trifocal calculé à l'aide de méthodes telles que celles proposées dans [87]. A cette étape, nous avons calculé les translations à un facteur d'échelle près.

Pour calculer le facteur d'échelle à l'aide d'un seul triplet de points nous utilisons l'équation dérivée du transfert point-point-point (6.22). Afin de permettre un calcul robuste du facteur d'échelle nous utilisons à nouveau l'algorithme RANSAC où un seul point est choisit aléatoirement à chaque itération.

Une fois que l'ensemble des poses de nos caméras sont connues, les solutions obtenues linéairement doivent être optimisées afin de garantir une précision suffisante. Pour ce faire nous proposons un ajustement de faisceaux adapté à la particularité de notre système :

$$\{\mathbf{M}_{o1}^{o2*}, \mathbf{M}_{o1}^{o3*}\} = \underset{\mathbf{M}_{o1}^{o2*}, \mathbf{M}_{o1}^{o3*}}{argmin} (\varepsilon_o + \varepsilon_p) \tag{6.23}$$

avec :

$$\varepsilon_o = \sum_{i=1}^{m} \| \mathbf{P}_{o1}^{i} - \widehat{\mathbf{P}_{o1}^{i}}(\mathbf{I}) \|^2 + \| \mathbf{P}_{o2}^{i} - \widehat{\mathbf{P}_{o2}^{i}}(\mathbf{M}_{o1}^{o2}) \|^2 + \| \mathbf{P}_{o3}^{i} - \widehat{\mathbf{P}_{o3}^{i}}(\mathbf{M}_{o1}^{o3}) \|^2 \tag{6.24}$$

et

$$\varepsilon_p = \sum_{j=1}^{n} \| \mathbf{P}_{p1}^{j} - \widehat{\mathbf{P}_{p1}^{j}}(\mathbf{M}_{o}^{p}) \|^2 + \| \mathbf{P}_{p2}^{j} - \widehat{\mathbf{P}_{p2}^{j}}(\mathbf{M}_{o}^{p}\mathbf{M}_{o1}^{o2}) \|^2 + \| \mathbf{P}_{p3}^{j} - \widehat{\mathbf{P}_{p3}^{j}}(\mathbf{M}_{o}^{p}\mathbf{M}_{o1}^{o3}) \|^2 \tag{6.25}$$

où $\mathbf{P}_{oi}^{j}$ et $\mathbf{P}_{pi}^{j}$ correspondent respectivement au $j^{\text{ème}}$ point de correspondance sur la $i^{\text{ème}}$ vue omnidirectionnelle et perspective. Avec $\widehat{\mathbf{P}_{o1}^{j}}$ et $\widehat{\mathbf{P}_{pi}^{j}}$ leur re-projection sur leur sphère respective.
Cette fonction de coût est adaptée à notre problème car elle force la transformation rigide inter-caméra obtenue par calibrage sans correspondances stéréo.

6.6/ RÉSULTATS

6.6.1/ EXPÉRIMENTATIONS AVEC DES DONNÉES SYNTHÉTIQUES

Afin d'évaluer la qualité et la pertinence de notre approche dédiée à l'estimation de poses pour caméras à champ non recouvrant, nous proposons dans cette section une série de tests synthétiques. Notre méthode utilisant le formalisme des tenseurs trifocaux y est comparée à [42], où l'auteur développe une méthode basée sur l'estimation de la matrice essentielle.
L'environnement synthétique est constitué d'un nuage de points 3D généré aléatoirement à l'intérieur d'un cube de dimensions $500\times500\times500$, les caméras du banc de stéréo-vision sont espacées par une distance de 20 unités. Les points 3D sont donc projetés sur le plan image de deux caméras, la caméra *fisheye* est modélisée à l'aide du modèle sphérique unifié et admet un champs de vue horizontal comme vertical de $180°$. La seconde caméra est modélisée par le modèle sténopé avec un angle de vue restreint à $45°$. Nous simulons ainsi un système très proche des cas réels auxquels nous sommes confrontés.
Le banc de vision est initialement situé au centre du cube, à chaque itération de notre procédure de test, deux nouveaux mouvements de notre banc de vision sont générés aléatoirement. Pour chaque mouvement les trois axes de rotation subissent une rotation comprise aléatoirement entre $\pm6°$ tandis la translation est comprise entre une distance de ±10 unités. Dans les expériences qui suivent seuls les cas non-dégénérés sont pris en compte.
Afin de déterminer la robustesse de notre approche, un bruit blanc gaussien est ajouté sur les points de correspondance entre les images. Cent itérations sont effectuées pour chaque niveau de bruit.
La figure 6.7 correspond aux résultats obtenus avec notre méthode comparée à celle de Clipp *et al.* avec une et deux paires de vues. La métrique utilisée est celle proposée par les auteurs cités précédemment, il s'agit de $\| \mathbf{t}_{est} - \mathbf{t}_{Vraie} \| / \| \mathbf{t}_{Vraie} \|$ avec $\mathbf{t}_{Vraie}$ la vérité terrain et $\mathbf{t}_{est}$ la translation estimée sur la caméra perspective. Cette mesure a l'avantage de permettre d'évaluer à la fois la justesse de l'estimation de l'échelle mais aussi de la direction de la translation. Il est notable que notre approche reste plus robuste à l'ajout d'un bruit gaussien.

6.6.2/ EXPÉRIMENTATIONS AVEC DES IMAGES RÉELLES

Dans cette section nous présentons les résultats expérimentaux obtenus avec notre banc de stéréo-vision hybride composé d'une caméra *fisheye* et d'une caméra perspective, les caméras sont les mêmes que celles utilisées dans le chapitre 5 dans la partie calibrage. Les essais ont été effectués en intérieur afin de permettre une mesure de la pièce qui nous servira de vérité terrain. Le banc de caméra effectue dans cette séquence des mouvements importants et non dégénérés. Le détecteur et descripteur de point utilisé

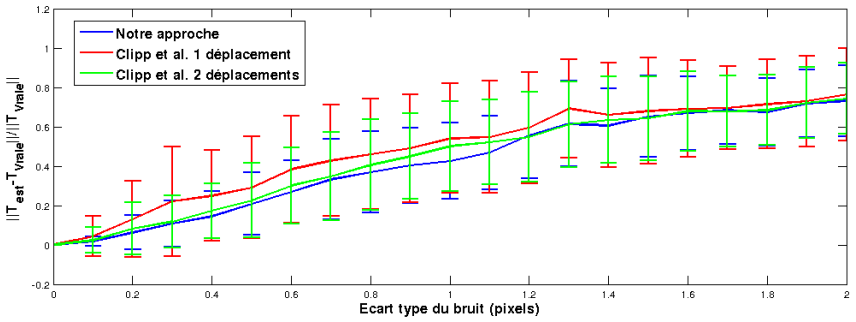


FIGURE 6.7 – Résultats synthétiques avec deux caméras pour 100 tests par niveau de bruit

est SURF pour la mise en correspondance temporelle des images *fisheye* et perspectives.

Nous présentons ici chacune des étapes de notre approche. La figure 6.8 montre la mise en correspondance de trois images *fisheye* à l'aide du tenseur trifocal décrit précédemment, pour un calcul robuste de ce tenseur les points admettant une erreur de reprojection de plus de 1° sur la sphère sont rejetés à l'aide d'un algorithme RANSAC.

La figure 6.9(a) montre la mise en correspondance à l'aide des descripteurs SURF des trois images perspectives acquises simultanément avec les images *fisheye*, on remarque que cette mise en correspondance à l'aide de descripteurs photométriques contient un grand nombre d'*outliers*. La figure 6.9(b) est le résultat obtenu avec notre approche RANSAC permettant de déterminer conjointement la valeur du facteur d'échelle des déplacements et les bonnes correspondances à l'aide d'un seul triplet de points sur les images perspectives.

La figure 6.10 correspond à la reconstruction 3D calculée avec un ensemble de 6 vues. Sur cette figure les points rouges correspondent aux points reconstruits à partir des vues *fisheye*, on a donc une reconstruction couvrant une large zone. La reconstruction avec les vues perspectives est affichée en bleue, on remarque que cette reconstruction ne correspond qu'à une petite portion de la scène visualisée par la caméra *fisheye*.

La mesure de la pièce nous a fournit une largeur de 10 mètres, cette même largeur estimée avec notre banc stéréo est du même ordre de grandeur, on peut en conclure que l'estimation du mouvement à l'échelle réelle est valide.

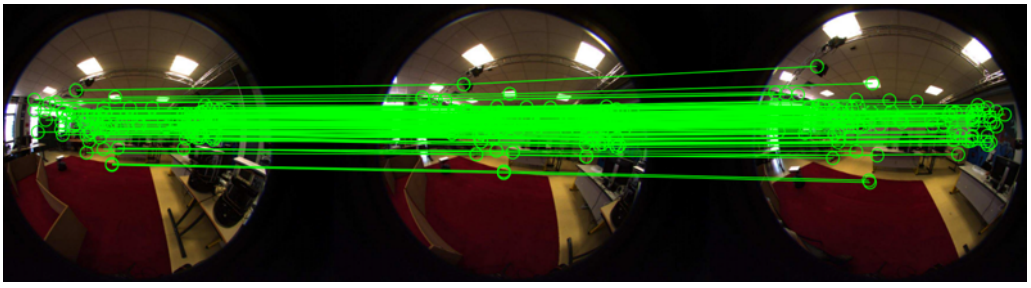
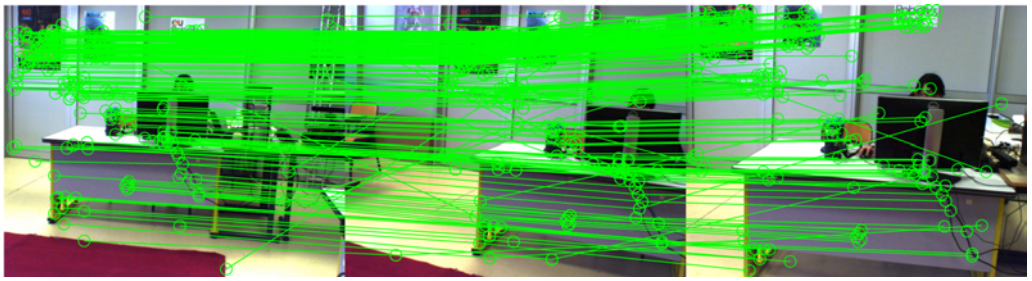


FIGURE 6.8 – Appariement de trois images *fisheye*



(a)

(b)

FIGURE 6.9 – Appariement de trois images perspectives, (a) avant calcul du facteur d'échelle, (b) après le calcul du facteur d'échelle (RANSAC 1 triplet de points)

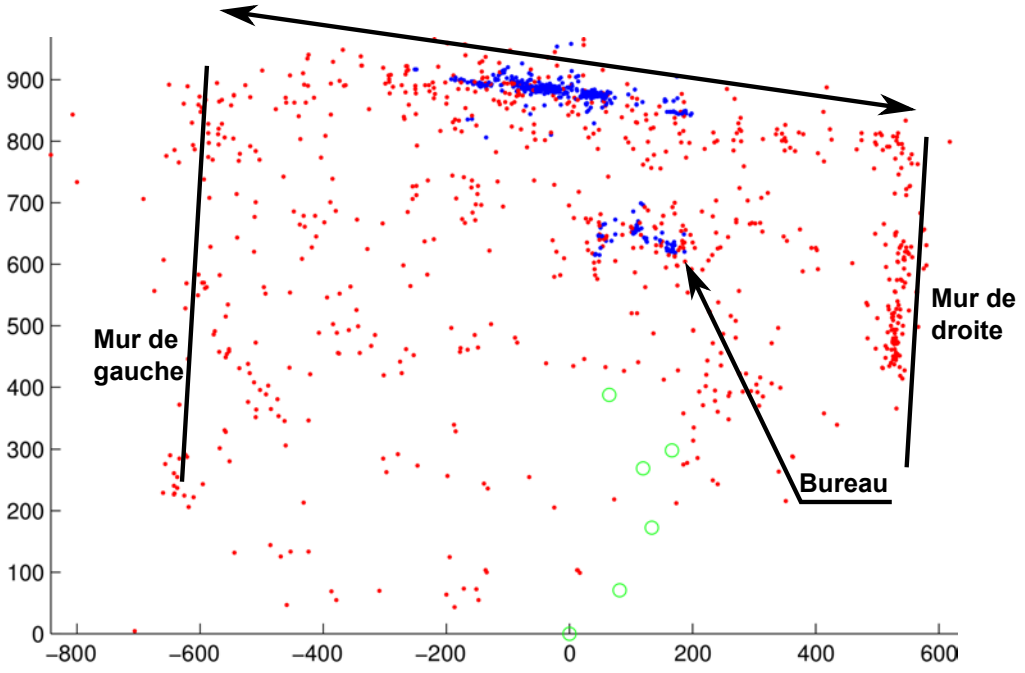
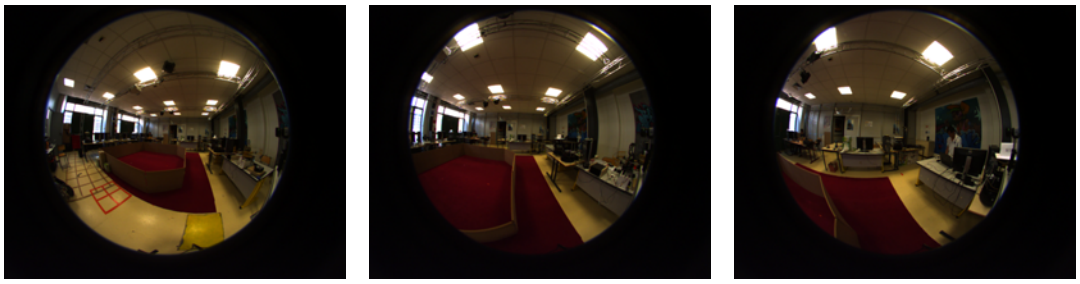


FIGURE 6.10 – Reconstruction 3D de l'environnement (vue du dessus), les points rouges sont les points 3D reconstruits avec la caméra *fisheye* tandis que les points bleus sont reconstruits avec la caméra perspective. Les cercles verts représentent la position de la caméra *fisheye* pour les vues successives.

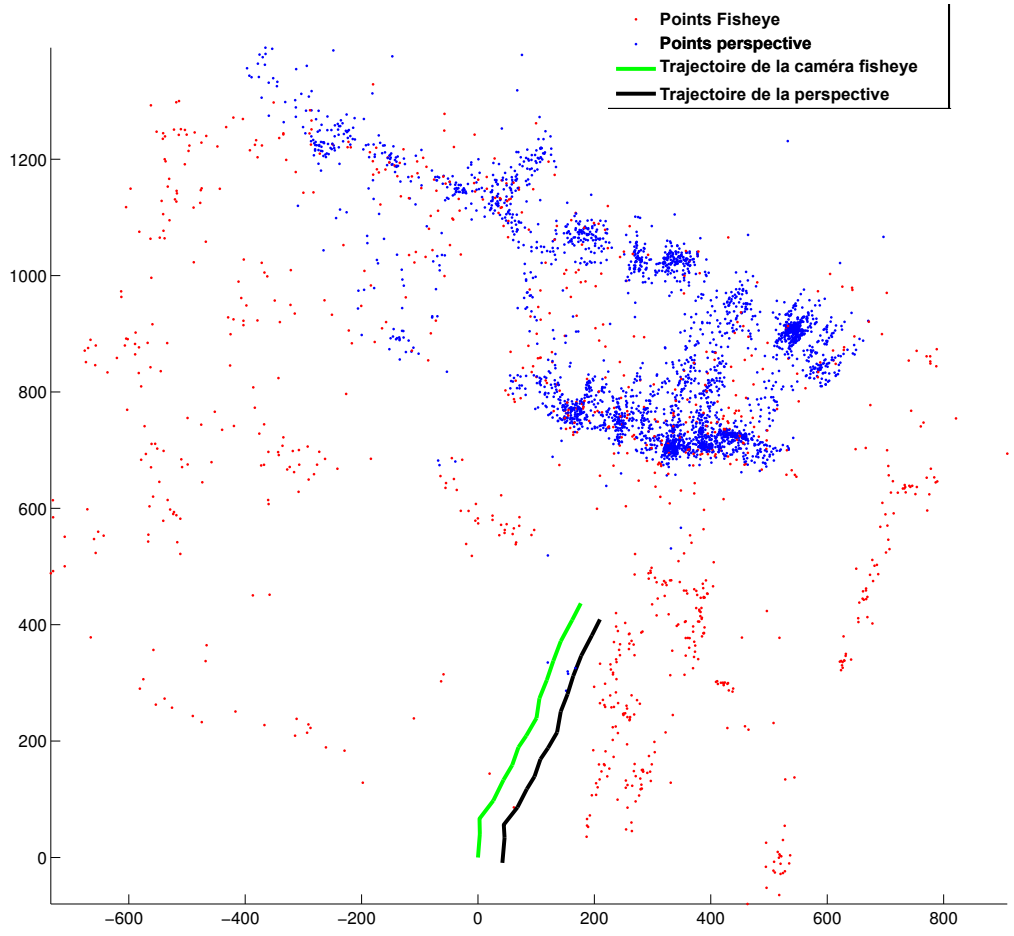
La figure 6.11 propose une autre expérimentation menée cette fois avec 20 paires d'images. Dans la reconstruction obtenue on remarque à nouveau un nuage de points

éparse fournie par la caméra *fisheye* qui compte environs 1000 points 3D mais qui couvre la quasi-totalité de la pièce. D'autre part, on note une très forte densité de points reconstruits à l'aide de la caméra perspective (plus de 3000 points) sur une zone plus restreinte.
Cet exemple souligne bien le grand avantage de ce capteur de stéréo-vision hybride puisqu'un système binoculaire standard -en dépit de leur grande précision- ne permet pas de reconstruire une zone aussi large avec un nombre de mouvements limités. De plus un système de vision stéréo purement omnidirectionnel (par exemple deux caméras *fisheye* ou deux caméras catadioptriques) peuvent permettre une reconstruction complète de la pièce mais pas aussi précise et dense que celle calculée avec notre caméra perspective.



(a) (b) (c)

(d) (e) (f)



(g)

FIGURE 6.11 – Résultats obtenus avec 20 images par caméra, (a-c) échantillon d'images fisheye, (d-f) images perspectives correspondantes, (g) reconstruction 3D avec le système de vision hybride

6.6.3/ EXPÉRIMENTATIONS AVEC LA BASE DE DONNÉES KITTI

En l'absence de vérité terrain fiable il n'est pas possible d'offrir une évaluation pertinente de notre algorithme, c'est pour cette raison que nous proposons ici plusieurs tests effectués à l'aide de la base de données en ligne KITTI [1].
Les données librement partagées en ligne contiennent les informations provenant d'un grand nombre de capteurs tels qu'un banc stéréo de caméra monochrome, un banc stéréo couleur, un vélodyne, un IMU ainsi qu'un GPS. L'ensemble est monté sur un véhicule circulant dans les rues de la ville de Karlsruhe. La base de données KITTI pour l'odométrie visuelle contient 22 séquences couvrant des distances allant de plusieurs centaines de mètres à plusieurs kilomètres. Une vérité terrain très précise (une erreur de localisation inférieure à 10cm) des poses des caméras est fournit.
Pour nos expérimentations nous utiliserons les images en niveau de gris provenant de deux caméras d'une résolution de 1.4 Megapixels de type Point Grey Flea 2. Ces caméras partagent un large champ de vue (voir figure 6.12), dans nos tests nous ne considérerons cependant aucune correspondance inter-caméra comme cela était déjà le cas dans le test avec notre système de vision hybride.

(a)

(b)

FIGURE 6.12 – Deux images (prises au même instant) provenant des deux caméras monochromes synchronisées équipant le véhicule KITTI

Les métriques permettant de quantifier la dérive de notre approche sont les mêmes que celles implémentées dans le kit de développement de KITTI. L'erreur de rotation est cal-

1. http ://www.cvlibs.net/datasets/kitti/

culée de la manière suivante :

$$\varepsilon_R = acos\left(\frac{1}{2}\left(tr(\mathbf{R}_R^{-1}\mathbf{R}_{GT}) - 1\right)\right), \tag{6.26}$$

avec $\mathbf{R}_{GT}$ et $\mathbf{R}_R$ respectivement les rotations provenant de la vérité terrain et de notre algorithme et ε_R l'erreur de rotation.
L'erreur de translation correspond à la distance euclidienne (en mètres) entre la position mesurée et la position réelle (donnée par la vérité terrain) :

$$\varepsilon_t = \sqrt{\sum(\mathbf{t}_R - \mathbf{t}_{GT})^2}. \tag{6.27}$$

avec $\mathbf{t}_{GT}$ et $\mathbf{t}_R$ respectivement les translations provenant de la vérité terrain et de notre algorithme et ε_t l'erreur de translation.

La figure 6.13 montre les résultats obtenus sur une séquence d'une centaine de mètres contenant 160 images. Le tracé rouge correspond au résultat obtenu avec notre approche de SFM sans recouvrement tandis que la courbe bleue est la vérité terrain. On remarque que dans cette séquence assez simple nous obtenons à la fois une assez bonne estimation de l'échelle mais aussi du mouvement de nos caméras comme le souligne les figures 6.13(f) et 6.13(g). On note toutefois une dérive inhérente à notre approche, celle-ci pouvant être corrigée à l'aide d'un ajustement de faisceaux non-utilisé pour ces séquences.

La figure 6.14 contient les résultats avec une autre séquence, où le véhicule parcours environs 165m pour 204 images. Une fois encore les résultats obtenus sont particulièrement satisfaisants, malgré une fois encore une dérive dans l'estimation de la translation et de la rotation au cours du temps.

Ces évaluations prouvent que notre approche permet d'assurer une estimation correcte du mouvement à l'échelle réelle sans nécessiter aucun recouvrement entre les caméras. En comparaison avec les résultats disponibles sur le site de KITTI la méthode développée est globalement moins performante que les approches de stéréo-vision classiques. En revanche, elle est beaucoup plus générique grâce à l'utilisation conjointe du modèle sphérique -compatible avec différents types de caméra à PVU- et d'une méthode d'estimation sans recouvrement efficace avec tout les systèmes stéréo-vision calibrés.

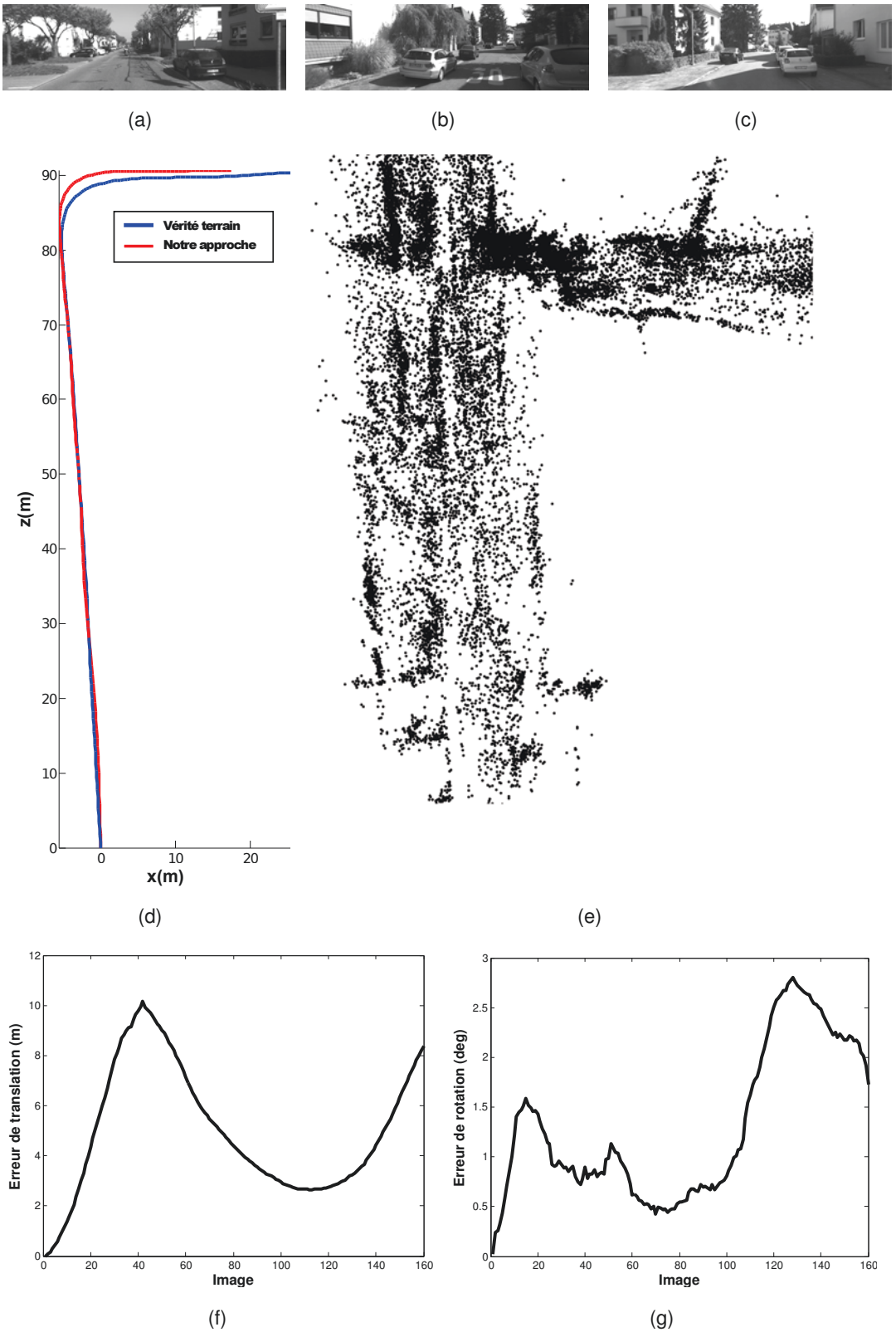


FIGURE 6.13 – Résultats obtenus pour une séquence extraite de la base de données KITTI avec 160 images, (a-c) échantillon d'images composant la séquence, (d) Trajectoire estimée par odométrie visuelle, (e) aperçu de la reconstruction 3D obtenue à partir de la caméra de gauche (vue du dessus) sans ajustement de faisceaux, (f) erreur de translation, (g) erreur de rotation

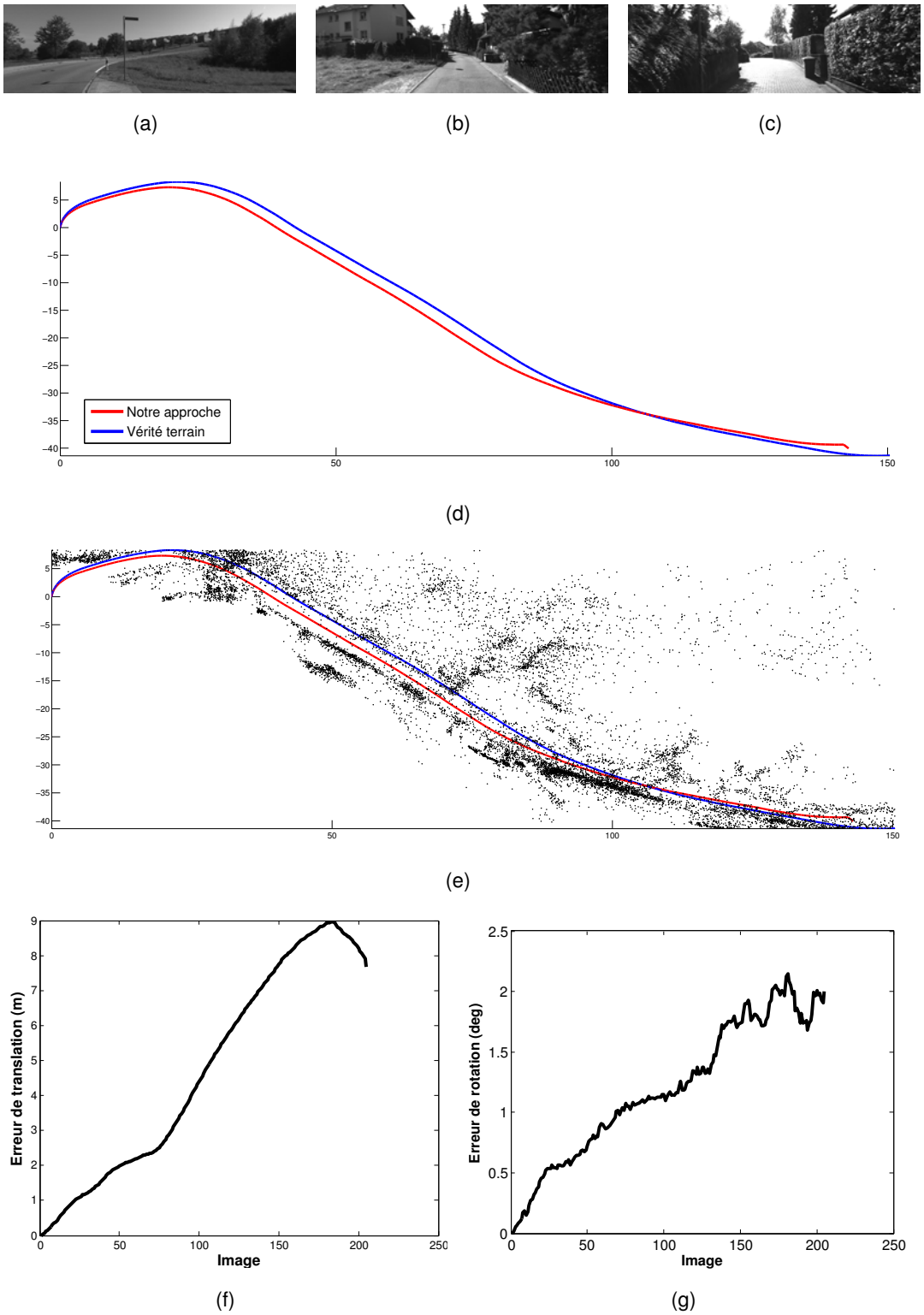


FIGURE 6.14 – Résultats obtenus pour une séquence extraite de la base de données KITTI avec 204 images, (a-c) échantillon d'images composant la séquence, (d) Trajectoire estimée par odométrie visuelle, (e) aperçu de la reconstruction 3D obtenue à partir de la caméra de gauche (vue du dessus) sans ajustement de faisceaux, (f) erreur de translation, (g) erreur de rotation

6.7/ CONCLUSION

Dans ce chapitre nous avons proposé l'adaptation d'une approche de SFM sans recouvrement à notre assemblage particulier de caméra. L'application de ce type de méthode à notre problème permet l'utilisation de tous les points visibles sur la caméra grand angle afin de fournir une reconstruction 3D d'une large zone et à la fois une reconstruction plus détaillée d'une zone particulière à l'aide de la caméra perspective sans utiliser de correspondances inter-caméra. L'autre avantage offert par cette approche est de permettre une localisation et une reconstruction de l'environnement à l'échelle réelle grâce à notre nouvelle formulation basée sur un tenseur trifocal. De plus, cette approche c'est également montrée plus robuste au bruit que la méthode originale sur une série d'essais synthétique.
Il a donc été prouvé au travers de ce chapitre qu'il s'agit d'une approche générique parfaitement adaptée au problème de banc de vision incorporant des caméras de types très différents où la mise en correspondance inter-caméra est difficile.

7

CONCLUSION

Les travaux présentés dans ce manuscrit portent sur l'étude d'un système de vision hybride constitué d'une caméra de type omnidirectionnelle et d'une caméra perspective mécanisée (PTZ). Ils avaient pour objectif l'étude géométrique complète du système et l'apport combiné des deux capteurs. Ils avaient notamment pour but le développement d'un mécanisme de fovéation adapté. Les travaux effectués dans le cadre de cette thèse ont permis de lever plusieurs verrous scientifiques et techniques encore peu étudiés :

Le suivi d'objet sur des images omnidirectionnelles : La distorsion induite par les systèmes de vision panoramique ne permet pas l'utilisation des algorithmes conventionnels de manière performante. La méthode que nous avons développée permet d'appliquer des algorithmes de suivi visuel à des images omnidirectionnelles. Cette approche repose sur une représentation de la cible adaptée à la géométrie des images grâce à l'utilisation du modèle sphérique unifié. Cette méthode est par conséquent compatible avec toutes les caméras à point de vue unique (ou légèrement décentré). De plus, toutes les approches de suivi visuel où la cible peut être représentée à l'aide d'un histogramme peuvent être adaptées sur la sphère grâce à la méthodologie développée dans le chapitre 3.
De nombreuses expérimentations appuyées par des critères d'évaluation pertinents ont permis de démontrer le gain offert par notre adaptation des algorithmes de suivi à la géométrie sphérique.

Auto-calibrage de caméra PTZ : Nos travaux ont permis l'élaboration d'une méthode d'auto-calibrage adaptée à la spécificité de ce type de caméra. La méthode proposée met en œuvre un nouvel ensemble d'inégalités matricielles linéaires (LMI) afin de contraindre l'estimation des paramètres du système entre des bornes déterminées empiriquement mais dont la validité et la consistance sont assurées. En conséquence l'espace de recherche des composantes de l'image de la conique absolue peut être réduit à l'aide de connaissances *a priori*. Une étude synthétique et expérimentale de notre algorithme a

permis de démontrer un gain significatif en terme de robustesse et de précision comparativement aux méthodes d'auto-étalonnage existantes.

Calibrage de système de vision hybride : Après une étude approfondie des deux caméras composants notre système (décrites dans le chapitres 3 pour la caméra omnidirectionnelle et 4 pour la caméra PTZ). Cette thèse s'est concentrée sur la collaboration de ces deux capteurs avec dans un premier temps la mise en place d'une méthode de calibrage adaptée à la géométrie de nos caméras. Les résultats obtenus à l'aide de notre approche ont été comparés avec une méthode de l'état de l'art. Cette comparaison a permis de certifier la qualité de notre approche puisque les paramètres calculés permettent une reprojection des points avec une précision supérieure.

Contrôle de la caméra PTZ au sein de notre système de vision stéréo-hybride dédiée à la vidéo surveillance : Dans le chapitre 5, nous avons élaboré une stratégie permettant d'orienter la caméra mécanisée sur une cible visible depuis l'image omnidirectionnelle.
Ce travail a permis la mise en place d'une approche originale autorisant le contrôle de la caméra rotative dans un environnement fixe mais inconnu. Cette méthode se base sur l'utilisation du modèle sphérique unifié ainsi que sur la géométrie épipolaire afin de réduire la zone de recherche de la région d'intérêt. De plus, une approche innovante de détection de cible particulièrement adaptée à ce couple de caméras a été développée. Cette détection de cible est basée exclusivement sur des contraintes géométriques (les descripteurs photométriques usuels étant peu adaptés à la distorsion induite par la lentille fisheye) rendant l'appariement de points robuste aux changements d'illumination, aux rotations ou encore à l'échelle.

Navigation robotique avec un système de stéréo-vision hybride omnidirectionnelle/perspective : Dans cette thèse une adaptation d'une méthode de SFM sans recouvrement est proposée permettant la localisation de notre système de vision et la reconstruction 3D de l'environnement. Cette nouvelle approche est une amélioration du travail de Clipp *et al.* [42]. La méthode développée a été validée à l'aide de données synthétiques et d'images réelles.

PERSPECTIVES

Dans ce travail, nos expérimentations ont prouvé l'efficacité de nos différentes approches pour le suivi de cible, l'auto-calibrage de caméra PTZ, la vidéo-surveillance et le SFM avec un système de vision hybride. Ce travail peut cependant être amélioré et complété

de différentes manières.

Amélioration de la méthode de suivi et intégration au sein d'un système de vision hybride La possibilité de suivre plusieurs cibles simultanément ainsi que l'utilisation de caractéristiques autres que la couleur (e.g. texture, forme) sont des pistes intéressantes, qui nécessiteraient une adaptation des attributs choisis à la géométrie sphérique. Le suivi multi-cibles ouvre en effet d'autres perspectives concernant la stratégie à adopter dans le contrôle de la caméra PTZ associée à la caméra omnidirectionnelle, notamment en ce qui concerne les critères permettant de prioriser la fovéation sur une cible particulière. Il existe déjà des travaux allant dans ce sens tel que [39] mais seule la disposition des cibles dans l'image omnidirectionnelle sert de critère afin de faire naviguer la caméra PTZ de cible en cible . Nous pensons que l'utilisation d'attributs de plus haut niveau, qualifiés sémantiquement (par exemple, piétons, véhicules, ...) constituerait une stratégie mieux adaptée à des scénarios en situation réelle.
Pour l'instant nos algorithmes de suivi visuel et de détection de cible reposent sur une sélection manuelle sur l'image omnidirectionnelle. Dans le cas d'un banc de vision fixe, il est tout à fait envisageable d'inclure une étape de détection de cible automatique, par exemple, à l'aide de méthodes de soustraction de fond adaptées aux caméras omnidirectionnelles [54]. Cela permettrait de rendre nos méthodes totalement autonomes.
D'autre part, dans le cas des capteurs catadioptriques l'ajout du miroir accentue les effets d'éblouissement du capteur, ce phénomène provoque souvent l'échec du suivi. L'utilisation de méthodes d'élimination de l'éblouissement dans les images comme [5] est une solution intéressante pour réduire l'impact de cet effet sur nos approches de suivi de cible.
Une autre perspective intéressante est l'élaboration d'une méthode de mise à jour de l'apparence de la cible à l'aide d'un processus d'apprentissage intégré au suivi. Ce qui, dans le cas de caméras à forte distorsion, n'est pas une tâche triviale.
Un autre aspect, qui n'a pas été abordé dans cette thèse est l'utilisation des concepts d'asservissement visuel pour le suivi de la cible à partir des informations conjointes provenant de la caméra PTZ et de la caméra omnidirectionnelle.

Amélioration de l'approche de SFM pour système de vision hybride Dans ce document nous avons cherché à proposer une solution géométrique permettant d'éviter la mise en correspondance stéréo de points d'intérêt entre nos images. Cependant nos caméras partagent un champ de vue commun pas pris en compte par notre approche, cette information est par conséquent non-utilisée. A partir de nos travaux, il est possible d'initialiser une méthode de recalage dense entre les vues, par exemple à l'aide d'un tenseur quadri-focal comme c'est le cas dans [44]. Cette extension offrirait une meilleure précision dans l'estimation du déplacement. Pour effectuer au mieux cette tâche il est

cependant essentiel de choisir des critères de mesure robuste à la forte dissemblance entre les prises de vues omnidirectionnelles et perspectives. L'utilisation de l'information mutuelle [168] est une piste privilégiée pour résoudre ce problème, de manière plus générale, la définition d'outils de traitement des images génériques (ou sphériques) constitue un axes de recherche qui nous semble porteur.

Implémentation sur une plateforme robotique L'intégration des algorithmes et méthodes développés sur une plateforme logicielle et matérielle commune constitue l'étape finale de ce travail de thèse.
Il est a présent possible d'implémenter nos algorithmes sur une plateforme robotique mobile équipée du système d'exploitation ROS (Robot Operating System). Nous avons déjà commencé le déploiement de nos approches sous l'intergiciel ROS. Les contraintes de calcul en temps réel peuvent réalistement être respectées par des calculs déportés ou avec une implémentation sur carte graphique à l'aide d'outil tel que CUDA.

Le scénario réalisable à l'aide des approches développées est la navigation d'un robot mobile terrestre ou aérien où la localisation peut s'effectuer à l'aide de l'algorithme développé dans le chapitre 6. Le suivi de cible depuis la caméra omnidirectionnelle et l'orientation de la caméra PTZ sur la zone d'intérêt, nécessite la combinaison des méthodes décrites dans les chapitres 3 et 5 lorsque le robot est à l'arrêt.
Cependant, pour rendre réalisable un tel scénario il est nécessaire d'améliorer encore la procédure de calibrage de la caméra PTZ. Dans la version actuelle, les commandes mécaniques sont utilisées afin de mettre à jour la rotation entre les caméras. Bien que cette stratégie se soit montrée efficace pour notre détection de cible, la précision offerte par la commande des servo-moteurs de la caméra rotative est une approximation trop forte pour permettre une localisation efficace. La mise en place d'une méthode de pré-calibrage des rotations de cette caméra constitue donc une piste intéressante. Cette pré-calibration peut, par exemple, être effectuée à l'aide de notre approche d'auto-calibrage afin de déterminer un modèle entre la commande des moteurs de la caméra et les rotations effectives du capteur. Il serait dans ce cas également nécessaire de s'intéresser à la répétabilité mécanique des caméras, c'est-à-dire à la fiabilité des servo-moteurs et à leur capacité à retourner à une position définie avec précision.

L'autre avantage de l'intégration de la caméra PTZ pour la navigation robotique est la possibilité d'élaborer des stratégies de positionnement de la caméra rotative afin d'éviter les configuration dégénérées dont souffrent les algorithmes de structure-from-motion sans recouvrement.

Extension à la collaboration multi-agents Dans cette thèse, nous avons considéré le cas d'un système de vision hybride constitué d'une caméra omnidirectionnelle et d'une

caméra perspective montées sur un ensemble fixe. Ce travail peut être étendu à la collaboration multi-robots équipés de caméras de différentes natures (infrarouge, omnidirectionnelle, RGB-D, ...) afin d'effectuer des tâches de surveillance, de reconnaissance et de reconstruction 3D d'un environnement. Dans un tel scénario, la nouvelle inconnue est la position des robots entre eux, l'utilisation d'autres instruments, comme le GPS, peuvent dans ce cas servir à obtenir cette information afin de rendre nos approches compatibles avec ce mode de fonctionnement.

Cette extension soulève des problématiques liées à la physique des capteurs, aux traitements et à la fusion de données multi-modales, etc. Elle ouvre sur un large champ de recherche, nécessitant le développement d'outils photométriques et géométriques génériques et adaptatifs, permettant d'appréhender des scénarios réalistes.

BIBLIOGRAPHIE

[1] H A. Iraqui, Y. Dupuis, R. Boutteau, J. Ertaud, and X. Savatier. Fusion of omnidirectional and ptz cameras for face detection and tracking. In *EST*, pages 18–23, 2010.

[2] G. Adorni, L. Bolognini, S. Cagnoni, and M. Mordonini. Stereo obstacle detection method for a hybrid omni-directional/pin-hole vision system. In *RoboCup 2001 : Robot Soccer World Cup V*, pages 244–250. 2002.

[3] L. Agapito, E. Hayman, and I. Reid. Self-calibration of rotating and zooming cameras. *IJCV*, 45(2) :107–127, 2001.

[4] A. Agarwal and B. Triggs. Tracking articulated motion using a mixture of autoregressive models. In *ECCV*, pages 54–65. Springer, 2004.

[5] A. Agrawal, R. Raskar, S. K Nayar, and Y. Li. Removing photography artifacts using gradient projection and flash-exposure sampling. In *TOG*, volume 24, pages 828–835, 2005.

[6] M. Agrawal. On automatic determination of varying focal lengths using semidefinite programming. In *ICPR*, Singapore, 2004.

[7] M. Agrawal and L. Davis. Camera calibration using spheres : A semi-definite programming approach. *ICCV*, 2, 2003.

[8] A. Alahi, Y. Boursier, L. Jacques, and P. Vandergheynst. Sport players detection and tracking with a mixed network of planar and omnidirectional cameras. In *ICDSC*, pages 1–8, 2009.

[9] H. Aliakbarpour, O. Tahri, and H. Araujo. Visual servoing of mobile robots using non-central catadioptric cameras. *RAS*, 2014.

[10] M. Sanjeev Arulampalam, S. Maskell, and N. Gordon. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50 :174–188, 2002.

[11] J. Badri, C. Tilmant, J. Lavest, Q. Pham, and P. Sayd. Camera-to-camera mapping for hybrid pan-tilt-zoom sensors calibration. In *SCIA*, pages 132–141, Berlin, Heidelberg, 2007. Springer-Verlag.

[12] J. Badri, C. Tilmant, J. Lavest, P. Sayd, and Q. Pham. Automatic calibration of hybrid dynamic vision system for high resolution object tracking. *PR*, pages 978–3, 2008.

[13] S. Bahadori and L. Iocchi. A stereo vision system for 3d reconstruction and semi-automatic surveillance of museum areas. In *AI* IA*, 2003.

[14] H. Bakstein and A. Leonardis. Catadioptric image-based rendering for mobile robot localization. In *ICCV*, pages 1–6, 2007.

[15] D. Baldini, M. Barni, A. Foggi, G. Benelli, and A. Mecocci. A new star-constellation matching algorithm for satellite attitude determination. *ESA journal*, 17 :185–198, 1993.

[16] J. P. Barreto. General central projection systems, modeling, calibration and visual servoing. Technical report, University of Coimbra, 2003.

[17] J. P. Barreto. Lifted fundamental matrices for mixtures of central projection systems. *ROBOMAT*, page 161, 2007.

[18] J. P. Barreto and H. Araujo. Issues on the geometry of central catadioptric image formation. In *CVPR*, pages 422–427, 2001.

[19] J.P. Barreto and H. Araujo. Direct least square fitting of paracatadioptric line images. In *OMNIVIS*, page 78, 2003.

[20] M. Barth and C. Barrows. A fast panoramic imaging system and intelligent imaging technique for mobile robots. In *IROS*, volume 2, pages 626–633, 1996.

[21] Y. Bastanlar, A. Temizel, Y. Yardimci, and P. Sturm. Effective structure-from-motion for hybrid camera systems. *transformation*, 1 :2, 2010.

[22] Y. Bastanlar, A. Temizel, Y. Yardimci, and P. Sturm. Multi-view structure-from-motion for hybrid camera scenarios. *Image and Vision Computing*, 30(8) :557–572, 2012.

[23] A. Basu. Active calibration : Alternative strategy and analysis. In *CVPR*, pages 495–500, 1993.

[24] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features. *CVIU*, 110(3) :346–359, June 2008.

[25] J. C. Bazin, C. Demonceaux, P. Vasseur, and IS. Kweon. Motion estimation by decoupling rotation and translation in catadioptric vision. *CVIU*, 114(2) :254–273, 2010.

[26] J.C. Bazin, S.G. Kim, D.G. Ghoi, J.Y.Lee, and I. Kweon. Mixing collaborative and hybrid vision devices for robotics applications. *journal of Korea Robotics Society*, 2011.

[27] J.C. Bazin, K.J. Yoon, , I.S. Kweon, C. Demonceaux, and P. Vasseur. Particle filter approach adapted to catadioptric images for target tracking application. In *BMVC*, 2009.

[28] S. Birchfield and S. Rangarajan. Spatiograms versus histograms for region-based tracking. In *CVPR*, volume 2, pages 1158–1163, 2005.

[29] C. Bishop. *Pattern recognition and machine learning*. springer New York, 2006.

[30] R. Bodor, R. Morlok, and N. Papanikolopoulos. Dual-camera system for multi-level activity recognition. In *IROS*, pages 643–648, 2004.

[31] J. Y. Bouguet. Camera calibration toolbox for Matlab, 2008.

[32] T.E. Boult, R. Micheals, X. Gao, P. Lewis, C. Power, W. Yin, and A. Erkan. Frame-rate omnidirectional surveillance and tracking of camouflaged and occluded targets. In *International Workshop on Visual Surveillance*, pages 48–55, 1999.

[33] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15. SIAM, Philadelphia, PA, 1994.

[34] D. Bradley, T. Boubekeur, T. Berlin, and W. Heidrich. Accurate multiview reconstruction using robust binocular stereo and surface meshing. In *CVPR*, 2008.

[35] J. Canny. A computational approach to edge detection. *TPAMI*, (6) :679–698, 1986.

[36] Z. Cao and S. Liu. Dynamic omnidirectional vision localization using a beacon tracker based on particle filter. In *Optics East*, pages 538–554, 2008.

[37] D. Capel. *Image mosaicing and super-resolution*. Springer, 2004.

[38] G. Caron and D.Eynard. Multiple camera types simultaneous stereo calibration. In *ICRA*, pages 2933–2938, 2011.

[39] C. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, and M. Abidi. Heterogeneous fusion of omnidirectional and ptz cameras for multiple object tracking. *TCSVT*, 18(8) :1052–1063, 2008.

[40] X. Chen, J. Yang, and A. Waibel. Calibration of a hybrid camera network. In *ICCV*, Washington, DC, USA, 2003.

[41] D. Claus and A. Fitzgibbon. A rational function lens distortion model for general cameras. In *CVPR*, volume 1, pages 213–219, 2005.

[42] B. Clipp, J. Kim, J-M Frahm, M. Pollefeys, and R. Hartley. Robust 6dof motion estimation for non-overlapping, multi-camera systems. In *WACV*, pages 1–8, 2008.

[43] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *TPAMI*, 25 :564–575, 2003.

[44] A. Comport, E. Malis, and P. Rives. Accurate quadrifocal tracking for robust 3d visual odometry. In *ICRA*, pages 40–45, 2007.

[45] J. Coolidge. The rise and fall of projective geometry. *American Mathematical Monthly*, pages 217–228, 1934.

[46] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *ECCV*, pages 484–498. Springer, 1998.

[47] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *CVIU*, 61(1) :38–59, 1995.

[48] J. Courbon, Y. Mezouar, L. Eck, and P. Martinet. A generic fisheye camera model for robotic applications. In *IROS*, pages 1683–1688, 2007.

[49] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire, and J. Thiran. Scale invariant feature transform on the sphere : Theory and applications. *IJCV*, 98(2) :217–241, 2012.

[50] Yuntao Cui, S Samarasckera, Qian Huang, and M Greiffenhagen. Indoor monitoring via the collaboration between a peripheral sensor and a foveal sensor. In *IWVS*, pages 2–9, 1998.

[51] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893, 2005.

[52] A. Davison, I. D Reid, N. D Molton, and O. Stasse. Monoslam : Real-time single camera slam. *TPAMI*, 29(6) :1052–1067, 2007.

[53] C. Demonceaux. Etude de caméras sphériques : du traitement des images aux applications en robotique, 2012. Universite de Bourgogne, These HDR.

[54] C. Demonceaux and P. Vasseur. Markov random fields for catadioptric image processing. *Pattern Recognition Letters*, 27(16) :1957–1967, 2006.

[55] C. Demonceaux and P. Vasseur. Omnidirectional image processing using geodesic metric. In *ICIP*, pages 221–224, 2009.

[56] C. Demonceaux, P. Vasseur, and Y. Fougerolle. Central catadioptric image processing with geodesic metric. *Image and Vision Computing*, 29(12) :840–849, 2011.

[57] C. Demonceaux, P. Vasseur, and C. Pégard. Omnidirectional vision on uav for attitude computation. In *ICRA*, pages 2842–2847, 2006.

[58] C. Ding, B. Song, A. Morye, J. Farrell, and A. Roy-Chowdhury. Collaborative sensing in a distributed ptz camera network. *TIP*, 21(7) :3282–3295, 2012.

[59] F. Du and M. Brady. Self-calibration of the intrinsic parameters of cameras for active vision systems. In *CVPR*, pages 477–482, 1993.

[60] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard. An evaluation of the rgb-d slam system. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1691–1696. IEEE, 2012.

[61] D. Eynard, P. Vasseur, C. Demonceaux, and V. Frémont. Uav altitude estimation by mixed stereoscopic vision. In *IROS*, pages 646–651, 2010.

[62] A. Maryum F. Development of a stereo vision system for outdoor mobile robots, 2006.

[63] C. Soria R. F. Vassallo F. Roberti, J. M. Toibero and R. Carelli. Hybrid collaborative stereo vision system for mobile robots formation. *International Journal of Advanced Robotic Systems*, 2010.

[64] O. Faugeras. *Three-dimensional computer vision : a geometric viewpoint*. MIT press, 1993.

[65] O. Faugeras, Q. T. Luong, and S. Maybank. Camera self-calibration : Theory and experiments. In *ECCV*, pages 321–334, 1992.

[66] M. A. Fischler and R. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) :381–395, 1981.

[67] A. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *CVPR*, volume 1, pages I–125, 2001.

[68] S. Fleck, F. Busch, P. Biber, W. Straßer, and H. Andreasson. Omnidirectional 3d modeling on a mobile robot using graph cuts. In *ICRA*, pages 1748–1754, 2005.

[69] D. Fox, W. Burgard, F. Dellaert, and S. Thrun. Monte carlo localization : Efficient position estimation for mobile robots. *AAAI/IAAI*, 1999 :343–349, 1999.

[70] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics : The kitti dataset. *The International Journal of Robotics Research*, 2013.

[71] A. Geiger, J. Ziegler, and C. Stiller. Stereoscan : Dense 3d reconstruction in real-time. In *IV*, pages 963–968. IEEE, 2011.

[72] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical implications. In *ECCV*, pages 445–461. 2000.

[73] C. Geyer and K. Daniilidis. Structure and motion from uncalibrated catadioptric views. In *CVPR*, volume 1, pages I–279, 2001.

[74] C. Geyer and K. Daniilidis. Conformal rectification of omnidirectional stereo pairs. In *CVPRW*, volume 7, pages 73–73, 2003.

[75] J. Gluckman and S. K Nayar. Ego-motion and omnidirectional cameras. In *ICCV*, pages 999–1005, 1998.

[76] S. Godber, R. S Petty, M. Robinson, and J. Evans. Panoramic line-scan imaging system for teleoperator control. In *IS&T*, pages 247–257, 1994.

[77] J. Gonzalez-Barbosa. *Vision panoramique pour la robotique mobile : stérovision et localisation par indexation d'images*. PhD thesis, Toulouse, INPT, 2004.

[78] Gregory D. Hager, Wen-Chung Chang, and A. S. Morse. Robot hand-eye coordination based on stereo vision. *IEEE Control Systems Magazine*, 15 :30–39, 1995.

[79] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988.

[80] R. Hartley. Euclidean reconstruction from uncalibrated views. In *Applications of invariance in computer vision*, pages 235–256. Springer, 1994.

[81] R. Hartley. In defense of the eight-point algorithm. *TPAMI*, 19(6) :580–593, 1997.

[82] R. Hartley. Self-calibration of stationary cameras. *IJCV*, 22(1) :5–23, 1997.

[83] R. Hartley and R. Gupta. Computing matched-epipolar projections. In *CVPR*, pages 549–555. IEEE, 1993.

[84] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *CVPR*, pages 761–764. IEEE, 1992.

[85] R. Hartley and R. Kaucic. Sensitivity of calibration to principal point position. In *ECCV*, ECCV '02, pages 433–446, London, UK, 2002.

[86] R. Hartley and P. Sturm. Triangulation. *CVIU*, 68(2) :146–157, 1997.

[87] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[88] J. Heikkila and O. Silvén. A four-step camera calibration procedure with implicit image correction. In *CVPR*, pages 1106–1112. IEEE, 1997.

[89] L. Heng, G. Hee Lee, and M. Pollefeys. Self-calibration and visual slam with a multi-camera system on a micro aerial vehicle. *RSS*, 2014.

[90] S. Hengstler, D. Prashanth, S. Fong, and H. Aghajan. Mesheye : a hybrid-resolution smart camera mote for applications in distributed intelligent surveillance. In *IPSN*, pages 360–369, 2007.

[91] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. Rgb-d mapping : Using kinect-style depth cameras for dense 3d modeling of indoor environments. *IJRR*, 31(5) :647–663, 2012.

[92] T. Ho, C. Davis, and S. Milner. Using geometric constraints for fisheye camera calibration. In *OMNIVIS*, 2005.

[93] R. Horaud, D. Knossow, and M. Michaelis. Camera cooperation for achieving visual attention. Research Report RR-5216, INRIA, 2004.

[94] P. Hough. Method and means for recognizing complex patterns, December 18 1962. US Patent 3,069,654.

[95] T. Huang and A. Netravali. Motion and structure from feature correspondences : A review. *Proceedings of the IEEE*, 82(2) :252–268, 1994.

[96] M. Isard and A. Blake. Condensation : conditional density propagation for visual tracking. *IJCV*, 29(1) :5–28, 1998.

[97] A. Jepson, D. Fleet, and T. El-Maraghi. Robust online appearance models for visual tracking. *TPAMI*, 25(10) :1296–1311, 2003.

[98] Q. Ji and S. Dai. Self-calibration of a rotating camera with a translational offset. *Robotics and Automation*, 20(1) :1–14, 2004.

[99] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D) :35–45, 1960.

[100] N. Karlsson, E. Di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. Munich. The vslam algorithm for robust localization and mapping. In *ICRA*, pages 24–29, 2005.

[101] T. Kazik, L. Kneip, J. Nikolic, M. Pollefeys, and R. Siegwart. Real-time 6d stereo visual odometry with non-overlapping fields of view. In *CVPR*, pages 1529–1536, 2012.

[102] S. Khan, O. Javed, Z. Rasheed, and M. Shah. Human tracking in multiple cameras. In *ICCV*, pages 331–336, 2001.

[103] J. Kim and M. Chung. Absolute motion and structure from stereo image sequences without stereo correspondence and analysis of degenerate cases. *PR*, 39(9) :1649–1661, 2006.

[104] J. Kim, M. Hwangbo, and T. Kanade. Spherical approximation for multiple cameras in motion estimation : Its applicability and advantages. *CVIU*, 114(10) :1068–1083, 2010.

[105] B. Kitt, A. Geiger, and H. Lategahn. Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme. In *IV*, pages 486–492. IEEE, 2010.

[106] G. Klein and D. Murray. Parallel tracking and mapping for small ar workspaces. In *ISMAR*, pages 225–234, 2007.

[107] P. Koschorrek, T. Piccini, P. Oberg, M. Felsberg, L. Nielsen, and R. Mester. A multi-sensor traffic scene dataset with omnidirectional video. In *CVPRW*, pages 727–734, 2013.

[108] S. Krotosky and M. Trivedi. On color-, infrared-, and multimodal-stereo approaches to pedestrian detection. *ITS*, 8(4) :619–629, 2007.

[109] Z. Kukelova and T. Pajdla. A minimal solution to the autocalibration of radial distortion. In *CVPR*, pages 1–7. IEEE, 2007.

[110] J. J. Guerrero L. Puig and P. Sturm. Matching of omnidirectional and perspective images using the hybrid fundamental matrix. In *OMNIVIS*, 2008.

[111] M. Lalonde, S. Foucher, L. Gagnon, E. Pronovost, M. Derenne, and A. Janelle. A system to automatically track humans and vehicles with a ptz camera. In *DS*, volume 6575, page 657502, 2007.

[112] G. Lemaître, E. Vargiu, J. Lorenzo Fernández, and F. Miralles. Real-time 2d face detection and features-based tracking in video. In *IADIS*, pages 91–98, 2012.

[113] A. Lévy-Schoen. Le champ d'activité du regard : données expérimentales. *L'annee Psychologique*, 74(1) :43–65, 1974.

[114] H. Li and R. Hartley. Five-point motion estimation made easy. In *ICPR*, volume 1, pages 630–633, 2006.

[115] H. Li and C. Shen. An lmi approach for reliable ptz camera self-calibration. In *AVSS*, Washington, DC, USA, 2006.

[116] H C Liao and Y C Cho. A new calibration method and its application for the cooperation of wide-angle and pan-tilt-zoom cameras. *Information Technology Journal*, 7(8) :1096–1105, 2008.

[117] H C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Readings in Computer Vision : Issues, Problems, Principles, and Paradigms*, pages 61–62, 1987.

[118] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2) :91–110, 2004.

[119] E. Maggio and A. Cavallaro. Multi-part target representation for color tracking. In *ICIP*, pages 729–732, 2005.

[120] E. Maggio and A. Cavallaro. *Video Tracking : Theory and Practice*. Wiley, February 2011.

[121] E. Marchand and F. Chaumette. Fitting 3d models on central catadioptric images. In *ICRA*, pages 52–58, Roma, Italy, 2007.

[122] D. Marr and T. Poggio. A theory of human stereo vision. Technical report, Cambridge, MA, USA, 1977.

[123] C. Mei, S. Benhimane, E. Malis, and P. Rives. Constrained multiple planar template tracking for central catadioptric cameras. In *BMVC*, pages 4–7, 2006.

[124] C. Mei, S. Benhimane, E. Malis, and P. Rives. Homography-based tracking for central catadioptric cameras. In *IROS*, 2006.

[125] C. Mei, S. Benhimane, E. Malis, and P. Rives. Efficient homography-based tracking and 3-d reconstruction for single-viewpoint sensors. *TRO*, 24(6) :1352–1364, 2008.

[126] C. Mei and E. Malis. Fast central catadioptric line extraction, estimation, tracking and structure from motion. In *IROS*, pages 4774–4779, 2006.

[127] C. Mei and P. Rives. Single view point omnidirectional camera calibration from planar grids. In *ICRA*, pages 3945–3950, 2007.

[128] M. Meilland, A. Ian Comport, and P. Rives. Dense visual mapping of large scale environments for real-time localisation. In *IROS*, pages 4242–4248. IEEE, 2011.

[129] E. Michaelsen, W. von Hansen, M. Kirchhof, J. Meidow, and U. Stilla. Estimating the essential matrix : Goodsac versus ransac. In *Symposium on photogrammetric computer vision*. Citeseer, 2006.

[130] B. Micusik and T. Pajdla. Structure from motion with wide circular field of view cameras. *TPAMI*, 28(7) :1135–1149, 2006.

[131] J. Minguez, L. Montesano, and F. Lamiraux. Metric-based iterative closest point scan matching for sensor displacement estimation. *TRO*, 22(5) :1047–1054, 2006.

[132] P. Moëssard. *Le cylindrographe, appareil panoramique.... : Le cylindrographe photographique...* Gauthier-Villars et fils, 1889.

[133] S. K Nayar and S. Baker. Catadioptric image formation. In *Proceedings of the 1997 DARPA Image Understanding Workshop*, pages 1431–1437, 1997.

[134] Y. Nesterov, A. Nemirovskii, and Y. Ye. *Interior-point polynomial algorithms in convex programming*, volume 13. SIAM, 1994.

[135] A. JR Neves, D. A Martins, and A. Pinho. A hybrid vision system for soccer robots using radial search lines. In *ICARSC*, pages 51–55, 2008.

[136] D. Nistér and F. Schaffalitzky. Four points in two or three calibrated views : Theory and practice. *IJCV*, 67(2) :211–231, 2006.

[137] Y. Onoe, N. Yokoya, K. Yamazawa, and H. Takemura. Visual surveillance and monitoring system using an omnidirectional video camera. In *ICPR*, volume 1, pages 588–592, 1998.

[138] U. Park, H. Choi, A. Jain, and S. Lee. Face tracking and recognition at a distance : A coaxial and concentric ptz camera system. *TIFS*, 8(10) :1665–1677, 2013.

[139] R Pless. Using many cameras as one. In *CVPR*, volume 2, pages II–587, 2003.

[140] J. Plucker. On a new geometry of space. *Proceedings of the Royal Society of London*, 14 :53–58, 1865.

[141] J. Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1) :15–19, 1970.

[142] L. Puig, J. Bermúdez, P. Sturm, and J. Guerrero. Calibration of omnidirectional cameras in practice : A comparison of methods. *CVIU*, 116(1) :120–137, 2012.

[143] L. Puig, P. Sturm, and J. Guerrero. Hybrid homographies and fundamental matrices mixing uncalibrated omnidirectional and conventional cameras. *MVA*, 24(4) :721–738, 2013.

[144] F. Rameau, A. Habed, C. Demonceaux, D. Sidibé, and D. Fofi. Self-calibration of a ptz camera using new lmi constraints. In *ACCV*, 2012.

[145] D. Rees. Panoramic television viewing system, April 7 1970. US Patent 3,505,465.

[146] D. A Ross, J. Lim, R. Lin, and M. Yang. Incremental learning for robust visual tracking. *IJCV*, 77(1-3) :125–141, 2008.

[147] J. Salvi. *An approach to coded structured light to obtain three dimensional information*. Universitat de Girona, 1998.

[148] T. Sato, S. Ikeda, and N. Yokoya. Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system. In *ECCV*, pages 326–340. 2004.

[149] T. Sato, T. Pajdla, and N. Yokoya. Epipolar geometry estimation for wide-baseline omnidirectional street view images. In *ICCVW*, pages 56–63, 2011.

[150] D. Scaramuzza. Omnidirectional camera and calibration toolbox for matlab. http://robotics.ethz.ch/~scaramuzza/Davide_Scaramuzza_files/Research/OcamCalib_Tutorial.htm, May 2009.

[151] D. Scaramuzza. 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. *IJCV*, 95(1) :74–85, 2011.

[152] G. Scotti, L. Marcenaro, C. Coelho, F. Selvaggi, and CS. Regazzoni. Dual camera intelligent sensor for high definition 360 degrees surveillance. *IEE Proceedings-Vision, Image and Signal Processing*, 152(2) :250–257, 2005.

[153] J. Shi and C. Tomasi. Good features to track. In *CVPR*, pages 593–600, 1994.

[154] D. Sidibé, D. Fofi, and F. Mériaudeau. Using visual saliency for object tracking with particle filters. *Eusipco 2010*, pages 1776–1780, 2010.

[155] G. Silveira and E. Malis. Unified direct visual tracking of rigid and deformable surfaces under generic illumination changes in grayscale and color images. *IJCV*, 89(1) :84–105, 2010.

[156] S. N. Sinha and M. Pollefeys. Pan–tilt–zoom camera calibration and high-resolution mosaic generation. *CVIU*, 103(3) :170–183, 2006.

[157] I. Sobel and G. Feldman. A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project in*, pages 271–272, 1968.

[158] S. Stefanou and A. Argyros. Efficient scale and rotation invariant object detection based on hogs and evolutionary optimization techniques. In *Advances in Visual Computing*, pages 220–229. 2012.

[159] G. P. Stein. Accurate internal camera calibration using rotation, with analysis of sources of error. In *ICCV*, pages 230–236, 1995.

[160] R. Strand and E. Hayman. Correcting radial distortion by circle fitting. In *BMVC*, 2005.

[161] P. Sturm. Self-calibration of a moving zoom-lens camera by pre-calibration. *Image and Vision Computing*, 15(8) :583–589, 1997.

[162] P. Sturm. Mixing catadioptric and perspective cameras. In *OMNIVIS*, pages 37–44, 2002.

[163] P. Sturm and S. Ramalingam. A generic concept for camera calibration. In *ECCV*, pages 1–13. 2004.

[164] R. Szeliski. Image alignment and stitching : A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1) :1–104, 2006.

[165] O. Tahri, Y. Mezouar, F. Chaumette, and P. Corke. Decoupled image-based visual servoing for cameras obeying the unified projection model. *TRO*, 26(4) :684–697, 2010.

[166] M. Taiana, J. Gaspar, J. Nascimento, R. Bernardino, and Pedro Lima. 3d tracking by catadioptric vision based on particle filters. In *RoboCup International Symposium*, 2007.

[167] R. Tanawongsuwan, A. Stoytchev, and I. A. Essa. Robust tracking of people by a mobile robotic agent. 1999.

[168] P. Thévenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *TIP*, 9(12) :2083–2099, 2000.

[169] P. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15(8) :591–605, 1997.

[170] L. N. Trefethen and D. Bau. *Numerical Linear Algebra*. SIAM, 1997.

[171] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *RA*, 3(4) :323–344, 1987.

[172] N. Uchida, T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi. 3d face recognition using passive stereo vision. In *ICIP*, pages 950–953, 2005.

[173] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *TPAMI*, 32(9) :1582–1596, 2010.

[174] P. Viola, M. J Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *ICCV*, pages 734–741. IEEE, 2003.

[175] R. Willson and S. Shafer. What is the center of the image ? *Journal of the Optical Society of America A*, 11(1) :2946 – 2955, November 1994.

[176] Y. Xu and D. Song. Systems and algorithms for autonomously simultaneous observation of multiple objects using robotic ptz cameras assisted by a wide-angle camera. In *IROS*, pages 3802–3807, 2009.

[177] X. Ying and Z. Hu. Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model. In *ECCV*, pages 442–455. 2004.

[178] J. Yu, A. Eriksson, T. Chin, and D. Suter. An adversarial optimization approach to efficient outlier removal. *JMIV*, 48(3) :451–466, 2014.

[179] Z. Zhang. A flexible new technique for camera calibration. *TPAMI*, 22(11) :1330–1334, 2000.

[180] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial intelligence*, 78(1) :87–119, 1995.

[181] T. Zhao, M. Aggarwal, R. Kumar, and H. S. Sawhney. Real-time wide area multi-camera stereo tracking. In *CVPR (1)*, pages 976–983, 2005.

[182] J. Zhou, D. Wan, and Y. Wu. The chameleon-like vision system. *IEEE Signal Processing Magazine*, 27(5) :91–101, 2010.

[183] S. Zhou, R. Chellappa, and B. Moghaddam. Visual tracking and recognition using appearance-adaptive models in particle filters. *TIP*, 13(11) :1491–1506, 2004.

TABLE DES FIGURES

LISTE DES TABLES

Résumé :

L'objectif principal de ce travail de thèse est l'élaboration d'un système de vision binoculaire mettant en œuvre deux caméras de types différents. Le système étudié est constitué d'une caméra de type omnidirectionnelle associée à une caméra PTZ. Nous appellerons ce couple de caméras un système de vision hybride. L'utilisation de ce type de capteur fournit une vision globale de la scène à l'aide de la caméra omnidirectionnelle tandis que l'usage de la caméra mécanisée permet une fovéation, c'est-à-dire l'acquisition de détails, sur une région d'intérêt détectée depuis l'image panoramique.
Les travaux présentés dans ce manuscrit ont pour objet, à la fois de permettre le suivi d'une cible à l'aide de notre banc de caméras mais également de permettre une reconstruction 3D par stéréoscopie hybride de l'environnement nous permettant d'étudier le déplacement du robot équipé du capteur.

Mots-clés : Vision hybride, Caméra omnidirectionnelle, Fisheye, PTZ, Suivi visuel, Auto-Calibrage, SFM

Abstract:

The primary goal of this thesis is to elaborate a binocular vision system using two different types of camera. The system studied here is composed of one omnidirectional camera coupled with a PTZ camera. This heterogeneous association of cameras having different characteristics is called a hybrid stereo-vision system. The couple composed of these two cameras combines the advantages given by both of them, that is to say a large field of view and an accurate vision of a particular Region of interest with an adjustable level of details using the zoom.
In this thesis, we are presenting multiple contributions in visual tracking using omnidirectional sensors, PTZ camera self calibration, hybrid vision system calibration and structure from motion using a hybrid stereo-vision system.

Keywords: Hybrid vision system, Omnidirectionnal camera, Fisheye, PTZ, Visual Tracking, Self-calibration, SFM

SPIM

■ École doctorale SPIM - Université de Bourgogne/UFR ST BP 47870 F - 21078 Dijon cedex
■ tél. +33 (0)3 80 39 59 10 ■ ed-spim@univ-fcomte.fr ■ www.ed-spim.univ-fcomte.fr

UB UNIVERSITÉ DE BOURGOGNE