



**HAL**  
open science

# Reconstruction tridimensionnelle à partir d'une caméra en mouvement : de l'influence de la precision

Pascal Brand

► **To cite this version:**

Pascal Brand. Reconstruction tridimensionnelle à partir d'une caméra en mouvement : de l'influence de la precision. Interface homme-machine [cs.HC]. Université Claude Bernard - Lyon I, 1995. Français. NNT: . tel-00005031

**HAL Id: tel-00005031**

**<https://theses.hal.science/tel-00005031>**

Submitted on 24 Feb 2004

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

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

# THÈSE

présentée par

Pascal Brand

pour obtenir le grade de Docteur  
de l'Université Claude Bernard – Lyon I  
(Arrêté ministériel du 30 Mars 1992)

Spécialité : **informatique**

## Reconstruction tridimensionnelle à partir d'une caméra en mouvement : *de l'influence de la précision*

Date de soutenance : 13 Octobre 1995

Composition du jury:

Président: M. Philippe Cinquin

Rapporteurs: M. Michel Dhome  
M. Gérard Giraudon

Examineurs: M. Serge Castan  
M. Roger Mohr

Thèse préparée au sein du laboratoire LIFIA – IMAG – INRIA Rhône-Alpes  
sous la direction de Roger Mohr



# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Correction des distorsions</b>	<b>7</b>
1.1 Algorithme . . . . .	10
1.2 Localisation des points d'intersections . . . . .	11
1.2.1 Modélisation à l'aide de fonctions gaussiennes . . . . .	13
1.2.2 Modélisation à l'aide de fonctions B-splines . . . . .	14
1.3 Expérimentations . . . . .	15
1.3.1 Conditions d'expérimentations . . . . .	17
1.3.2 Critères de qualité de la modélisation de croix . . . . .	17
1.3.3 Stabilité de la correction . . . . .	18
1.3.4 Facteurs influençant les distorsions . . . . .	21
1.3.5 Qualité de la correction des distorsions . . . . .	24
1.4 Conclusions . . . . .	24
<b>2 Extraction de positions de points</b>	<b>29</b>
2.1 Extraction de points modélisables . . . . .	30
2.1.1 Algorithme . . . . .	30
2.1.2 Extraction de coins dans l'image . . . . .	31
2.1.3 Extraction de cibles circulaires . . . . .	35
2.2 Validation de la localisation des coins dans une image . . . . .	41
2.2.1 Alignement . . . . .	43
2.2.2 Précision de la reconstruction tridimensionnelle . . . . .	44
2.2.3 Précision du calcul de la géométrie épipolaire . . . . .	45
2.2.4 Stabilité des birapports . . . . .	47
2.2.5 Conclusions . . . . .	49
2.3 Validation de l'extraction des cibles . . . . .	49
2.3.1 Qualité de la géométrie épipolaire . . . . .	51
2.3.2 Précision des reconstructions tridimensionnelles . . . . .	53

2.3.3	Birapports . . . . .	53
2.3.4	Conclusions . . . . .	55
2.4	Mise en correspondance de points entre 2 images . . . . .	55
2.4.1	Déformation affine . . . . .	56
2.4.2	Contrainte sur la déformation affine . . . . .	58
2.4.3	Initialisation des paramètres . . . . .	60
2.4.4	Expérimentations . . . . .	60
2.4.5	Résumé des expérimentations . . . . .	70
2.5	Conclusions . . . . .	71
<b>3</b>	<b>Reconstruction tridimensionnelle</b>	<b>73</b>
3.1	Reconstruction de points . . . . .	74
3.1.1	Mises en correspondance de points . . . . .	76
3.1.2	Reconstruction projective de points . . . . .	77
3.1.3	Passage d'une reconstruction projective à une reconstruction euclidienne . . . . .	78
3.1.4	Reconstruction euclidienne avec un étalonnage des caméras . . . . .	80
3.1.5	En l'absence des points de la base . . . . .	80
3.2	Reconstruction de surface . . . . .	80
3.2.1	Création d'une surface à partir de deux vues . . . . .	82
3.2.2	Fusion de deux surfaces . . . . .	83
3.3	Expérimentations . . . . .	86
3.3.1	Évaluation quantitative de la reconstruction de points . . . . .	87
3.3.2	Évaluation qualitative de la reconstruction de surface . . . . .	92
3.4	Conclusions . . . . .	94
	<b>Conclusion</b>	<b>103</b>
	<b>Annexes</b>	<b>109</b>
<b>A</b>	<b>Ajustement robuste de modèle</b>	<b>109</b>
A.1	Moindres carrés médians . . . . .	110
A.2	M-estimateurs . . . . .	111
<b>B</b>	<b>Résultats de mises en correspondance précises</b>	<b>115</b>
	<b>Bibliographie</b>	<b>123</b>

# Introduction

Cette thèse a pour cadre la vision par ordinateur. Ce domaine tente, par le biais de machines, de reproduire quelques fonctionnalités de la vision naturelle. Parmi les thèmes abordés, il y a notamment la reconnaissance de formes, qui consiste à identifier un objet à partir de son image, et la reconstruction tridimensionnelle d'une scène, qui consiste à localiser les emplacements des différents objets dans la scène.

Pour chaque fonction de la vision par ordinateur, la qualité et la précision de la réponse dépend de l'application qui intéresse l'utilisateur. Ainsi, en reconnaissance des formes, il existe deux types de réponses suivant si on ne recherche que la classe d'un objet, ou si l'on veut l'identifier précisément. La précision de la reconstruction tridimensionnelle d'une scène que l'on veut obtenir dépend également de l'application. Par exemple, si l'on veut seulement se déplacer d'un point d'une pièce vers un autre, il suffit de connaître approximativement la position des objets, et de passer suffisamment loin de chaque objet pour éviter les collisions. Au contraire, lors de la construction d'un bâtiment, il faut pouvoir fournir des mesures exactes, comme celles de la planarité du terrain sur lequel on construit ou le fait que les murs bâtis sont bien verticaux.

Dans ce dernier domaine, les performances de l'homme sont qualitatives (ressemblance de deux caractéristiques,...), les cotés quantitatifs requérant l'usage d'instrument de mesure comme des théodolites. Par exemple, qui ne s'est jamais cogné contre une porte, non pas parce que l'on n'avait pas remarqué qu'elle était basse, mais car elle ne semblait pas *si* basse. Dans ce cas précis, l'évaluation de la hauteur de la porte a été très grossière, à quelques dizaines de centimètres près.

Les systèmes de vision par ordinateur qui ont pour but de réaliser des mesures précises d'objets trouvent de nombreuses applications industrielles. Par exemple, dans [Bey91], les mesures obtenues sont utilisées pour l'évaluation des déformations que subit l'habitacle d'un véhicule lors d'un accident. La précision des mesures est dans ce cas primordiale. Ces systèmes utilisent généralement au moins deux vues d'une scène. Un nombre plus important de vues permet d'accroître la fiabilité des résultats par la fusion de toutes les données recueillies.

Néanmoins, sans aucune information a priori sur les caractéristiques de la scène ou sur la position des points de vues, il est très difficile d'obtenir des mesures précises. Avec seulement deux caméras et sans faire aucune hypothèse, il est impossible de reconstruire une scène dans un repère euclidien. Seule une reconstruction projective peut être réalisée.

Ainsi, il est préférable d'avoir d'autres informations, en plus des images, pour réaliser

des mesures précises. Ces informations dépendent de l'application réalisée. Par exemple, pour la reconstruction d'un bâtiment, on peut facilement préciser qu'un mur (nous dirons plan par la suite) et parallèle à un autre, ou encore que le plancher est confondu avec le plan ( $Oxy$ ) de notre repère. D'autres informations plus riches existent, comme la position tridimensionnelle d'un point dans la scène lorsque l'on dispose d'une mire de calibrage.

L'exploitation au mieux de ces informations est délicate, et dépend de deux critères : la qualité du système d'acquisition des images et la qualité du traitement d'images. La qualité des images n'est désormais plus en cause avec les récents développements du matériel vidéo. Ainsi, les cartes d'acquisitions d'images entièrement numériques font leur apparition. Cette haute technologie permet une très bonne qualité d'images, supprimant les effets de *line-jitter* (décalage d'un pixel d'une ligne sur deux de l'image). De grands progrès ont aussi été réalisés dans la fabrication de caméras depuis l'apparition de la technologie CCD. Les caméras de résolution  $512 \times 512$  pixels sont courantes. Grâce aux progrès technologiques, des caméras ayant une haute-résolution (jusqu'à  $5120 \times 5120$  pixels) font leur apparition. Des appareils photos numériques sont également disponibles. Les photos ne sont plus stockées sur une pellicule, mais sur un disque-dur. Ils procurent ainsi une grande souplesse d'utilisation à tous les systèmes de vision par ordinateur. À court terme, ces techniques devraient devenir courantes, même si elles restent à ce jour relativement onéreuses.

Un des points faibles du processus de la prise d'images réside dans la qualité des objectifs utilisés. En effet, ils focalisent les rayons lumineux sur une cellule CCD. Or, il est généralement fait l'hypothèse que le système caméra-objectif suit un modèle sténopé (dans ce modèle, la caméra est représentée par un plan de projection et par un centre de projection ; voir la figure 0.1). Ce modèle est largement utilisé car il est très simple, et il a de bonnes propriétés permettant de simplifier de nombreux problèmes. Ce modèle est exact avec l'utilisation de lentilles parfaites. Mais il est très difficile d'obtenir de telles lentilles, surtout pour les lentilles de focales courtes. Afin de réduire les erreurs dues à une approximation de la prise d'images suivant un modèle sténopé, il est quelques fois nécessaire de corriger la position des points dans l'image afin de rendre ce modèle l'image la plus fidèle possible à ce modèle. La première contribution de cette thèse est le développement d'une méthode de correction des distorsions optiques du système d'acquisition d'images.

Cette méthode, proche de celle introduite par Peuchot [Peu94], fait l'objet du chapitre 1. À partir de l'observation de lignes droites dans la scène, la correction des distorsions consiste à redresser ces lignes dans l'image. Ces lignes proviennent d'une mire plane qui comporte des points alignés. La précision de la localisation de ces points dans les images fixe la qualité de la correction des distorsions qui est réalisée. Cette tâche est ainsi particulièrement détaillée. De nombreux tests valident cette méthode, tant au niveau de la stabilité des corrections de distorsions qu'au niveau du bien-fondé de leur correction.

Il est à noter que les distorsions ne sont pas caractérisées en distorsions radiales ou tangentielles, et leur calcul ne passe pas par un calibrage de la caméra [Tsa86].

La localisation d'entités dans les images reste néanmoins la principale source d'erreur dans l'exploitation quantitatives des images. Voici un exemple qui illustre leur importance : dans des conditions standards de prise d'images, réalisées avec un objectif de longueur focale 12.5mm, une erreur de positionnement d'un point de 1 pixel dans l'image se concrétise par une erreur de 1mm à une distance de 1m. Cette erreur est loin d'être négligeable dans des applications délicates, telles que les mesures industrielles.

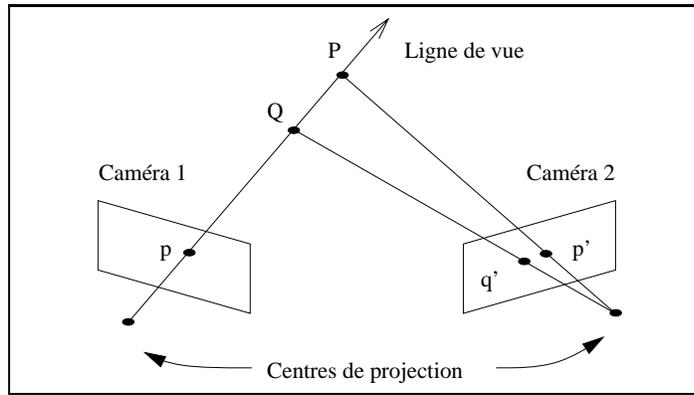
La seconde contribution de cette thèse, traitée dans le chapitre 2, réside dans la proposition de méthodes permettant la localisation de points avec une précision allant de 1/20ème à 1/10ème de la taille d'un pixel, dépendant du type de points extraits. Deux méthodes sont développées, chacune s'appliquant à un type de point particulier. La première section traite de l'extraction de points dont le signal dans l'image est modélisable. Le principe général de la méthode mise en œuvre consiste dans la recherche de la transformation affine qui permet d'ajuster au mieux l'image au modèle, rejoignant les travaux de Deriche [DB93]. Deux sous-sections appliquent ce procédé à deux entités particulières : la localisation de coins et du centre de cibles circulaires. Quatre tests utilisant différents outils et propriétés de la vision par ordinateur valident les précisions atteintes : la conservation des alignements de points dans les images, la qualité de la géométrie épipolaire et de la reconstruction que l'on peut obtenir avec de tels points, et la stabilité de birapports.

En ce qui concerne les points dont le signal dans l'image n'est pas modélisable, la détection est réalisée en mettant en correspondance un même point entre deux images. Ce problème est un des problèmes difficiles de la vision par ordinateur, même s'il est à la base de nombreux algorithmes utilisant plusieurs images. Ce problème est illustré sur la figure 0.1. Un point  $\mathbf{p}$  dans la première image correspond à la projection d'un point  $\mathbf{P}$  ou  $\mathbf{Q}$  de la scène appartenant à la ligne de vue définie par  $\mathbf{p}$ . Le point de la scène correspondant à  $\mathbf{p}$  se projette dans la seconde image en un point  $\mathbf{p}'$  ou  $\mathbf{q}'$ . Le problème qui nous concerne est celui de la recherche du correspondant de  $\mathbf{p}$  dans la seconde image, et ce le plus précisément possible, connaissant seulement les deux images.

Les méthodes classiques de mise en correspondance de points utilisent des fonctions de corrélation appliquées sur des fenêtres de corrélation de mêmes tailles et de formes identiques. La méthode développée dans le chapitre 2 utilise également une fonction de corrélation, mais la fenêtre de corrélation dans l'image 2 est obtenue par une transformation affine de celle de l'image 1 afin de gagner en précision. Différentes expérimentations réalisées sur des scènes planes valident cette méthodologie à la fin de ce chapitre.

La précision de la localisation de points dans les images peut s'appliquer avec succès dans de nombreuses applications. Par exemple, en reconnaissance des formes, une bonne précision de localisation de points permet une meilleure discrimination des objets les uns par rapport aux autres.

L'application qui nous intéresse dans ce mémoire est la reconstruction tridimensionnelle d'une scène dans un repère euclidien. La précision de la reconstruction obtenue dépend particulièrement de la méthode de la reconstruction utilisée ainsi que de la précision de localisation des points dans les images et des connaissances tridimensionnelles de certains points pour obtenir une reconstruction euclidienne. La qualité de la reconstruction tridimensionnelle suivant la configuration choisie est étudiée au chapitre 3 après



**Figure 0.1** : Une paire d'images ou le problème de la mise en correspondance de points

avoir détaillé les méthodes de reconstruction utilisées.

Une des applications de la reconstruction tridimensionnelle d'une scène concerne la prise d'un objet par une main articulée. Une telle tâche nécessite la reconstruction précise de points de la scène, mais également quelques informations complémentaires comme les normales à la surface. Ainsi, une reconstruction de points n'est pas suffisante pour une telle application, mais il faut pouvoir réaliser une reconstruction de surface. Cette tâche, abordée au chapitre 3, est la dernière contribution de cette thèse. La méthode employée utilise une reconstruction Delaunay des points dans les images qui permet, par projection sur les points de la scène de la topologie obtenue, la reconstruction d'une surface triangulaire. La principale difficulté réside dans la fusion de deux surfaces obtenues à partir de vues différentes. Des expérimentations qualitatives validant cette reconstruction de surface terminent ce chapitre.

Enfin, la conclusion générale résume les différents résultats acquis au cours de ce mémoire, et présente quelques perspectives de recherche qui s'offrent à la suite de ce travail.

## Notations

Ce mémoire comprend quelques parties qui font appel au calcul matriciel. Les notations utilisées dans ce texte sont les suivantes :

Les matrices et les vecteurs sont désignés par des caractères gras (par exemple  $\mathbf{X}$ ). Les scalaires sont quant à eux écrits en italiques (par exemple  $x$ ).

Le transposé d'un vecteur ou d'une matrice  $\mathbf{A}$  est dénoté par  $\mathbf{A}^\top$ .

Le vecteur nul de taille  $n$  est dénoté par  $\mathbf{0}_n$ , et la matrice identité de taille  $n \times n$  par  $\mathbf{I}_n$ .

La matrice du produit vectoriel<sup>1</sup> définie par un vecteur  $\mathbf{t} = (x, y, z)^\top$  est désignée par la notation

$$[\mathbf{t}]_\times = \begin{pmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{pmatrix}$$

Cette matrice antisymétrique possède plusieurs propriétés. Quelques soient les deux vecteurs  $\mathbf{t}$  et  $\mathbf{t}'$  de taille 3 (l'opérateur  $\wedge$  désignant le produit vectoriel) :

$$\begin{aligned} \mathbf{t} \wedge \mathbf{t}' &= [\mathbf{t}]_\times \cdot \mathbf{t}' \\ [\mathbf{t}]_\times \cdot \mathbf{t} &= \mathbf{0}_3 \\ \mathbf{t}'^\top \cdot [\mathbf{t}]_\times \cdot \mathbf{t}' &= 0 \end{aligned}$$

Une matrice perspective pourra être dénotée par

$$\mathbf{P} = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \end{pmatrix} = \begin{pmatrix} \mathbf{p}_1^\top \\ \mathbf{p}_2^\top \\ \mathbf{p}_3^\top \end{pmatrix} = (\mathbf{M} \mid \mathbf{t})$$

Le vecteur  $\mathbf{p}_i$  représente la  $i$ ème ligne de  $\mathbf{P}$  :  $\mathbf{p}_i^\top = (p_{i,1} \ p_{i,2} \ p_{i,3} \ p_{i,4})$

La matrice  $\mathbf{M}$  représente la première sous-matrice de taille  $3 \times 3$  de  $\mathbf{P}$ , et le vecteur  $\mathbf{t}$  est égal à  $\mathbf{t} = (p_{1,4} \ p_{2,4} \ p_{3,4})^\top$

La *distance algébrique* entre un point  $(u \ v)$  et une droite d'équation  $ax + by + c = 0$  est définie par la valeur de

$$\frac{au + bv + c}{\sqrt{a^2 + b^2}}$$

---

<sup>1</sup>en anglais, *skew matrix*



# Chapitre 1

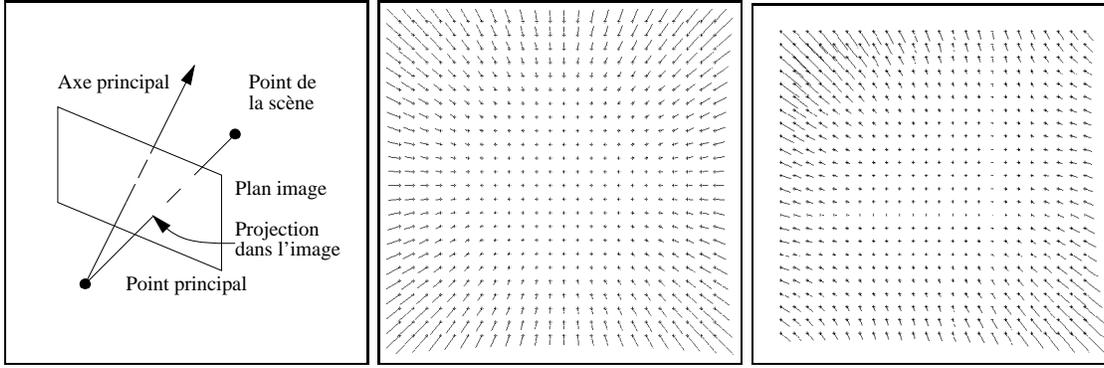
## Correction des distorsions

L'analyse d'image nécessite la compréhension du phénomène physique qui les crée. L'extraction d'informations à partir d'images passent souvent par une modélisation mathématique la plus fidèle possible de ce phénomène physique. Le modèle de formation des images le plus souvent utilisé est le modèle sténopé, schématisé sur la figure 1.1(a). Ce modèle est constitué d'un point principal et d'un plan de projection appelé plan image. La ligne perpendiculaire au plan image et passant par le point principal est appelée l'axe principal. Chaque point de la scène est projeté par rapport au point principal dans l'image sur le plan de projection. Ce modèle perspectif s'exprime algébriquement par une matrice  $3 \times 4$  correspondant à une application de l'espace projectif  $\mathbb{P}^3$  dans  $\mathbb{P}^2$ . Cette modélisation mathématique du processus de formation d'images est linéaire lorsque les coordonnées homogènes sont utilisées. De cette propriété dérive des conséquences géométriques telles que la géométrie épipolaire reliant deux images.

Un système d'acquisition d'images suit ce modèle lorsque les lentilles utilisées et leur montage sur l'objectif sont parfaites. Dans le cas contraire, des distorsions optiques peuvent apparaître. Il existe deux types principaux de distorsion. Les distorsions radiales sont dues au fait que les lentilles ne sont pas parfaitement mince, et qu'elles présentent des défauts de courbure. Leur effet est de déplacer les points images symétriquement au centre de l'image. Les distorsions tangentielles apparaissent lorsque l'assemblage des lentilles sur l'objectif n'est pas parfait (lentilles non alignées, non positionnées perpendiculairement à l'axe principal, ...). Elles sont souvent accompagnées par des distorsions radiales. En plus de ces distorsions, des aberrations peuvent se glisser sur les bords des lentilles [Sla80]. Les figures 1.1(b) et (c) montrent les effets des distorsions radiales et tangentielles. Chaque segment montre la position initiale d'un point dans une image conforme au modèle sténopé (extrémité représentée une croix) ainsi que sa position lorsque des distorsions optiques sont appliquées.

Très vite, le modèle sténopé a été complété afin de mieux représenter le phénomène physique de la formation des images. Les photogrammètres se sont penchés sur ce problème en ajoutant aux paramètres internes de la caméra des paramètres prenant en compte les distorsions radiales et tangentielles [Bro66, Fai75, Bey92a].

Par ailleurs, la communauté scientifique de la vision par ordinateur, notamment avec



(a) Le modèle sténopé

(b) Distorsions radiales

(c) Distorsions tangentielles

**Figure 1.1** : Différents modèles pour l'acquisition d'image

Tsai, a pris en compte les distorsions radiales dans le calibrage des caméras [Tsa86, CR89]. La correction apportée à un point  $(x, y)$ , exprimé dans un repère dont l'origine est confondue avec le centre des distorsions radiales, est réalisée en traduisant ce point par un vecteur  $(\Delta x_r, \Delta y_r)$  défini par :

$$\begin{aligned}\Delta x_r &= x(k_1 r^2 + k_2 r^4 + \dots) \\ \Delta y_r &= y(k_1 r^2 + k_2 r^4 + \dots)\end{aligned}$$

où  $r^2 = x^2 + y^2$  [Sla80]. Le nombre de paramètres  $k_i$  estimés dépend de la précision que l'on veut obtenir sur le calcul de la distorsion et de la qualité de l'objectif utilisé. En pratique, seuls  $k_1$  et  $k_2$  peuvent être calculés de façon significative.

Le calcul précis du facteur d'échelle horizontal et de la position du centre de l'image a fait l'objet de soins particuliers lorsqu'il est calculé pendant un calibrage utilisant une modélisation des distorsions radiales [LT88, Pen91, NSNI92].

Ce modèle prenant en compte seulement les distorsions radiales a été complété en lui ajoutant des paramètres modélisant les distorsions tangentielles [LS90]. La correction des distorsions tangentielles pour un point  $(x, y)$  se modélise par une translation  $(\Delta x_t, \Delta y_t)$  définie par

$$\begin{aligned}\Delta x_t &= [p_1(r^2 + 2x^2) + 2p_2xy] \cdot [1 + p_3r^2 + p_4r^4 + \dots] \\ \Delta y_t &= [p_2(r^2 + 2y^2) + 2p_1xy] \cdot [1 + p_3r^2 + p_4r^4 + \dots]\end{aligned}$$

L'addition de ces 2 types de distorsions (radiales et tangentielles) est parfaitement modélisée par l'ajout d'un prisme mince devant une caméra conforme à un modèle sténopé [Bro66, WCH90, WCH92].

D'autres modélisations de distorsions existent, par exemple par interpolation polynomiale de degré supérieur à 2 [GT90, RK82], par des surfaces de Bézier [Gos89] ou par transformation affine locale [BP91]. Ces différentes interpolations ne se préoccupent pas de l'origine des distorsions.

Pour se détacher de la contrainte du modèle sténopé, mais sans étalonnage utilisant des modèles de distorsions radiales ou tangentielles, le calibrage peut se réaliser par la méthode des biplans [MBK81, WM91]. Son principe est de reconstruire la ligne de vue associée à chaque pixel. La reconstruction des lignes de vue nécessite l'observation de 2 plans de calibrage. Ceci autorise des modèles pour lesquels les lignes de vues ne sont plus nécessairement concourantes. Cette méthode originale a été rendue plus souple en séparant le calcul des paramètres internes et celui des paramètres externes [CG93]. Elle a aussi été généralisée et rendue plus fiable en calculant chaque ligne de vue à partir d'un plus grand nombre de plans [CLSC92].

La modélisation de plus en plus précise des distorsions est intéressante d'un point de vue théorique. Mais il semble qu'une telle complexité ne soit nécessaire que lors de l'utilisation d'objectifs bas de gamme, ou possédant une focale très courte. Pour la plupart des objectifs, la modélisation des distorsions radiales est suffisamment précise [Bey92a], notamment pour les objectifs de grande qualité comme ceux fabriqués par la société Angénieux [RDDL94].

D'autres auteurs pensent que la correction des distorsions est inutile [HS90]. En effet, cela complique à la fois le calibrage et le traitement des images. Par exemple, la projection d'un point de la scène ne peut plus s'exprimer par une relation linéaire, les lignes épipolaires deviennent courbes. Nous verrons que le calcul des distorsions n'est pas toujours nécessaire. Les corrections de distorsions ne font intervenir que des quantités faibles, inférieures le plus souvent à l'unité exprimée en pixel ; il faut donc d'une part que l'application requiert une telle précision, et que d'autre part les algorithmes utilisés détectent les éléments dans les images avec une précision correspondante, ce qui n'est souvent pas le cas actuellement.

La méthode développée au sein de cette partie fut introduite par B. Peuchot [Peu92]. Elle se distingue des méthodes habituelles par le fait qu'aucun calibrage de la caméra n'est effectué. Le but ici est de trouver une transformation point à point qui remplace les points de l'image dans des positions compatibles avec le modèle sténopé, c'est à dire vérifiant le fait que les droites de la scène se projettent en droites dans l'image.

La suite de ce chapitre se présente ainsi : le principe de la méthode sera tout d'abord examiné plus en détail. Pour corriger les distorsions, une mire plane contenant des droites est observée. Cet algorithme consiste essentiellement à redresser les lignes dans l'image.

Comme la précision des corrections dépend essentiellement de la précision d'extraction des points d'intersections des lignes, la localisation de ces points fera l'objet d'une section. Deux algorithmes seront étudiés. Le premier algorithme modélise les intensités lumineuses autour d'une intersection par une fonction composée de deux fonctions gaussiennes. La seconde méthode décompose une intersection en recherchant la position des deux droites qui la composent. Chaque droite est extraite en détectant quelques uns de ces points à l'aide de fonctions B-splines.

Ensuite, les corrections obtenues seront testées et validées. Plusieurs séries de tests montreront la stabilité des corrections en fonction du temps que dure la prise d'images, de l'ouverture de l'iris et de la mise au point. Une dernière expérimentation montrera la

qualité des corrections apportées à l'image. Enfin, nous concluons sur la qualité de la méthode

## 1.1 Algorithme

La méthode développée au sein de cette partie vise à corriger les distorsions sans modélisation de celles-ci. Les corrections sont calculées à partir de l'image d'une mire plane (figure 1.2(a)). Cette mire est composée de droites. Les points d'intersection des lignes (appelés aussi *croix*) forment des alignements horizontaux et verticaux dans la scène. Comme la prise d'image ne suit pas exactement le modèle sténopé, les projetés de ces points ne sont pas exactement alignés dans l'image. La figure 1.2(b) montre les croix de la mire reliées entre elles, mettant en évidence les distorsions optiques de cet objectif par rapport au modèle sténopé.

Or le modèle sténopé est une transformation projective, et toute transformation projective en géométrie sur  $\mathbb{R}$  est caractérisé par le fait que les droites de la scène se projettent en droites dans l'image [SK52]. La correction des déformations de la prise d'image est donc exprimée par une transformation point à point de l'image initiale qui vise à redresser les droites de sorte à être conforme avec le modèle sténopé. Les droites qui sont redressées sont celles formées par les croix de la mire (figure 1.2(b)). La localisation précise de ces points est primordiale pour obtenir de bonnes corrections. Elle fera l'objet de soins particuliers, et deux méthodes seront développées à cette fin dans la section suivante.

Une fois la position des croix détectée, une grille formée de ces points est construite (figure 1.2(b)). Soit la transformation  $T$  point à point de la grille qui place le point en haut à gauche de la grille en la coordonnée  $(0, 0)$ , son voisin immédiat à droite en  $(0, 1)$ , son voisin en dessous en  $(1, 0)$ ,  $\dots$ . Cette transformation permet de repositionner les points de la grille qui se trouvent dans l'image dans un repère conforme au modèle sténopé. En effet, la grille initiale a été transformée en une grille ne contenant plus que des droites horizontales et verticales. Ce repère est appelé le repère virtuel. Cette transformation est illustrée par la figure 1.3(a).

La transformation  $T$  définie pour les points de la mire est alors généralisée pour tous les points de l'image. Pour un point quelconque de l'image, nous définissons son transformé par  $T$  comme une interpolation bilinéaire sur ces 4 voisins dans la grille. D'autres interpolations sont possibles, comme par exemple une interpolation gaussienne ou par B-splines, agissant sur un nombre de points de la grille plus important [Peu94]. En fait, une simple interpolation bilinéaire suffit car l'influence des distorsions est négligeable localement, c'est à dire dans une case de la grille.

Les points exprimés dans le repère virtuel sont alors repositionnés de façon à être proches de leur position initiale, en préservant toutefois le fait qu'ils sont conformes avec un modèle sténopé (cf figure 1.3(b)). Ainsi, une transformation projective  $M$  de  $\mathbb{P}^2$  dans  $\mathbb{P}^2$  permettant de transformer les points du repère virtuel vers le repère pixel est définie. Cette transformation peut être choisie de sorte à minimiser la distance entre la position initiale de chaque point dans l'image et sa position transformée par les 2 transformations successives. Nous préférons la définir en choisissant 4 points non colinéaires 3 à

---



---

Acquisition de la mire
Extraction des positions des croix de la mire
Création de la grille reliant les croix extraites précédemment
Définition de la transformation $T$ pour les points de la grille afin d'obtenir une image compatible avec le modèle sténopé
Généralisation de $T$ pour tous les points de l'image
Définition de la transformation projective $M$ pour replacer les points dans le repère pixel
La nouvelle position d'un point $\mathbf{p}$ est $M \circ T(\mathbf{p})$

---



---

**Algorithme 1.1 :** Algorithme de la correction des distorsions

3 de l'image qui seront considérés comme invariants, c'est à dire ne subissant aucune distorsion. En effet, ce choix des 4 points a l'avantage de permettre la comparaison des distorsions d'objectifs différents.

L'algorithme 1.1 récapitule cette méthode de correction des distorsions.

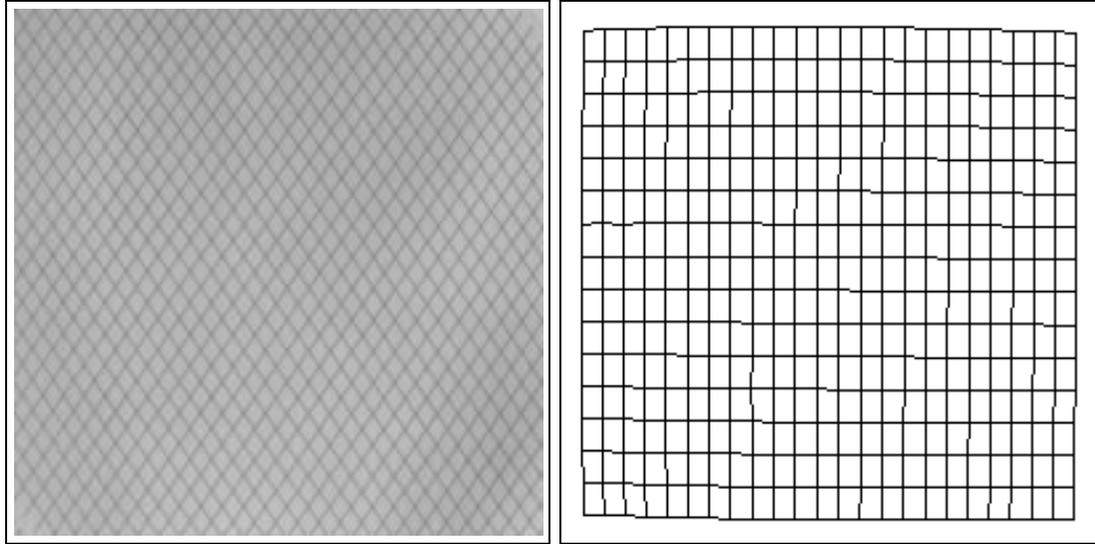
## 1.2 Localisation des points d'intersections

La qualité de la correction des distorsions dépend essentiellement de la précision de la localisation des points d'intersections de la mire. C'est pourquoi cette section est dédiée entièrement à cette tâche délicate.

Deux méthodes ont été étudiées pour localiser précisément la position des points d'intersection de la mire. Elles procèdent par la recherche de minima locaux de fonctions, et nécessitent donc une approximation initiale de la position des points. Cette approximation est obtenue en prenant les minima locaux des intensités lumineuses dans une fenêtre de taille  $15 \times 15$  de l'image. La taille de ce filtre dépend de la taille de la mire dans l'image. Les positions initiales trouvées sont localisées avec une précision de l'ordre de 2 pixels.

Les deux méthodes présentées ci-dessous modélisent mathématiquement le signal de l'image sur une petite fenêtre contenant un point d'intersection de 2 lignes. La première méthode modélise les intensités lumineuses de l'image autour d'une croix par une fonction paramétrique composée de deux fonctions gaussiennes. La seconde méthode aborde le problème de façon plus locale en réalisant l'extraction des deux lignes qui forment cette croix. Cette méthode s'appuie sur une interpolation des intensités lumineuses par des fonctions B-splines bicubiques.

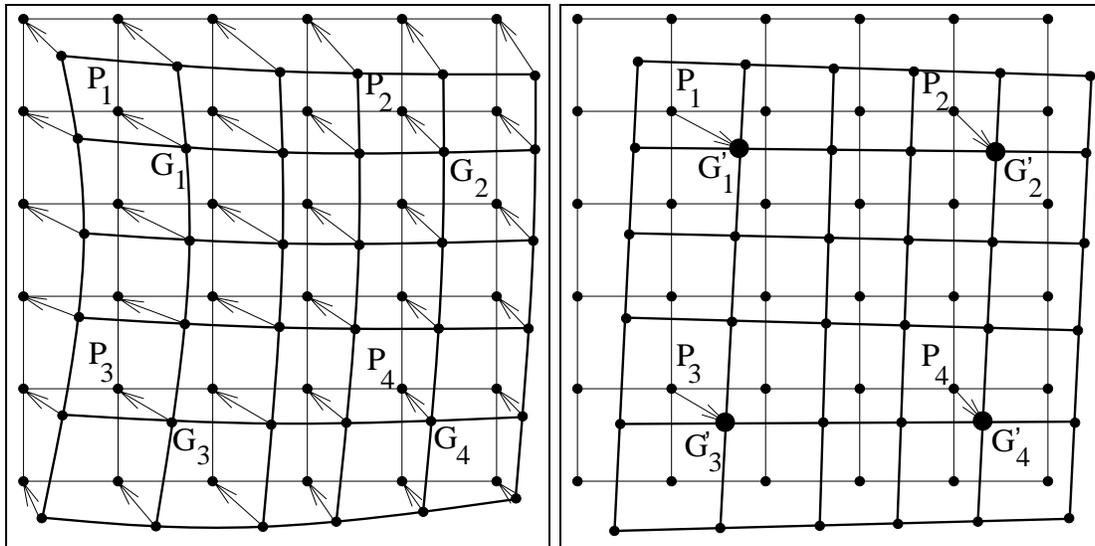
Les distorsions seront considérées comme négligeables localement. Ainsi, sur la petite partie de l'image examinée pour l'extraction d'une croix (typiquement  $21 \times 21$  pixels), les lignes formant cette croix seront supposées comme étant des droites.



(a) Image de la mire utilisée pour corriger les distorsions optiques

(b) Grille obtenue en reliant des points de la mire de la figure (a)

**Figure 1.2 :** Image de la mire utilisée pour la correction des distorsions, prise avec un objectif “CANON TV Zoom Lens J6x12”, de longueur focale 12.5 mm



(a) Transformation  $T$  des points de la grille du repère pixel vers le repère virtuel

(b) Transformation projective  $M$  des points du repère virtuel dans le repère pixel

**Figure 1.3 :** Les 2 transformations successives visant à corriger les distorsions

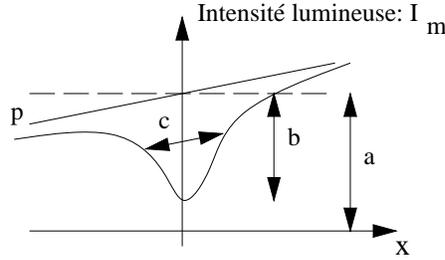


Figure 1.4 : Coupe selon l'axe des abscisses du modèle  $I_m$  d'une croix

### 1.2.1 Modélisation à l'aide de fonctions gaussiennes

Cette méthode a été mise au point par B. Peuchot [Peu94]. Elle consiste à modéliser par une fonction paramétrique les intensités lumineuses de l'image autour d'un point d'intersection. La recherche de la position de la croix se résume à la recherche des paramètres qui permettent d'ajuster au mieux le modèle avec l'image. La fonction utilisée est la suivante :

$$I_m(x, y) = a + b.e^{-c^2(x-x_0)^2} + f.e^{-g^2(y-y_0)^2} + p.(x - x_0) + q.(y - y_0)$$

La figure 1.4 montre une coupe monodimensionnelle de cette fonction. Le rôle de chacun des paramètres est le suivant :

- $(x_0, y_0)$  est la position du point recherché.
- $a$  est l'intensité lumineuse du fond de l'image.
- $b$  et  $f$  représentent les amplitudes de chaque branche gaussienne du modèle. Elles caractérisent la différence entre l'intensité lumineuse du fond de l'image et celle d'une ligne de la mire.
- $c$  et  $g$  représentent la largeur dans l'image des lignes de la mire. Ces 2 variables modélisent la netteté de la prise d'image. Une bonne netteté de l'image est caractérisée par des valeurs importantes de ces variables.
- $p$  et  $q$  sont deux paramètres qui gèrent l'atténuation du contraste le long d'une ligne de la mire.

La recherche des différents paramètres est réalisée par la minimisation des différences au carré entre l'intensité lumineuse des pixels  $I(x, y)$  dans l'image et celle obtenue par le modèle  $I_m(x, y)$  :

$$\chi^2 = \min_{\{\text{paramètres}\}} \left[ \sum_{(x,y) \in S} (I(x, y) - I_m(x, y))^2 \right] \quad (1.1)$$

La minimisation de  $\chi^2$  s'opère sur une partie  $S$  de l'image qui contient la croix. La taille de cette fenêtre est typiquement  $21 \times 21$  pixels. L'optimisation de ce système est

relativement importante. 11 paramètres sont recherchés (les 9 cités ci-dessus, auxquels s'ajoutent l'orientation de la grille et le rapport des facteurs d'échelles sur les axes  $x$  et  $y$ ). La non-linéarité de ce système est de plus importante car la fonctionnelle contient deux fonctions exponentielles. La recherche des paramètres est réalisée par l'algorithme de Levenberg-Marquardt [PFTW88]. Néanmoins, cet algorithme n'assure pas une convergence globale du système. La solution peut donc très bien être un minimum local de l'équation 1.1.

Un résultat qualitatif de la localisation d'une croix par cette méthode est présenté en figure 1.5. La figure (a) montre le graphe des intensités lumineuses  $I(x, y)$  dans l'image sur une partie de l'image contenant une croix, et la figure (b) indique les valeurs du modèle  $I_m(x, y)$  qui minimisent  $\chi^2$ .

### 1.2.2 Modélisation à l'aide de fonctions B-splines

Contrairement à la méthode précédente qui modélise globalement l'image autour d'un point à extraire, cette méthode reste plus locale à chaque élément formant ce point. Ici, la détection de la position de la croix se résume à l'extraction des deux droites  $D_1$  et  $D_2$  qui la génèrent. Le processus de la détection d'une droite composant la croix consiste principalement dans la détection de points appartenant à cette droite.

Les orientations approximatives  $D_1^{(a)}$  et  $D_2^{(a)}$  des 2 droites  $D_1$  et  $D_2$  peuvent être estimées en utilisant les minima locaux des intensités lumineuses, connaissant une approximation de la position de la croix.

Le point d'intersection entre une droite  $D'_1$  parallèle à  $D_1^{(a)}$ , et la droite  $D_2$  est le point d'intensité lumineuse la plus faible le long de  $D'_1$ . Afin de détecter ce point, la droite  $D'_1$  est discrétisée par l'algorithme de Bresenham [FvDFH91]. Les intensités lumineuses le long de cette droite sont alors interpolées par une fonction B-spline cubique [WC90, PFTW88] :

$$s(x) = \sum_{j=-3}^{n-1} \alpha_j B_j(x)$$

où

- $n$  est le nombre de points qui sont interpolés. Dans la pratique,  $n \approx 30$ .
- $B_j(x)$  est un polynôme de degrés 3
- $s(x)$  est l'interpolation par une B-spline cubique des points discrets de  $D'_1$ .  $s(x)$  correspond à l'intensité lumineuse d'un point se trouvant à la position  $x$  sur la droite  $D'_1$ . Cette fonction est de classe  $\mathcal{C}^2$ .

Le point recherché est alors celui qui minimise  $s(x)$  :

$$p = \underset{x}{\operatorname{argmin}} s(x)$$

La figure 1.6(a) montre un exemple d'interpolation par une fonction B-spline des intensités lumineuses le long d'une coupe parallèle à  $D_1^{(a)}$ .

---



---

```

Approximation de la position de la croix
Approximation de son orientation
Faire pour chaque droite formant la croix
    Détection de plusieurs points appartenant à cette droite
        par coupes successives sur les intensités lumineuses
        par des B-splines
    Calcul de l'équation de la droite passant le mieux, au sens
        des moindres carrés, par les points détectés précédemment
La position du centre de la croix est le point d'intersection
des 2 droites extraites

```

---



---

**Algorithme 1.2:** Algorithme de l'extraction de la position du centre d'une croix, par la méthode utilisant des B-splines

Cet algorithme permet la localisation d'un point de  $D_2$ . En déplaçant  $D'_1$ , plusieurs points  $(x_i, y_i)$  de  $D_2$  sont extraits. Afin que  $D'_1$  n'intersecte que  $D_2$  et non  $D_1$ , la droite  $D'_1$  est toujours choisie à plus de 5 pixels de  $D_1^{(a)}$

L'ajustement sur les points  $(x_i, y_i)$  de la meilleure droite, au sens des moindres carrés, fournit l'équation de  $D_2$ . La minimisation effectuée est celle de

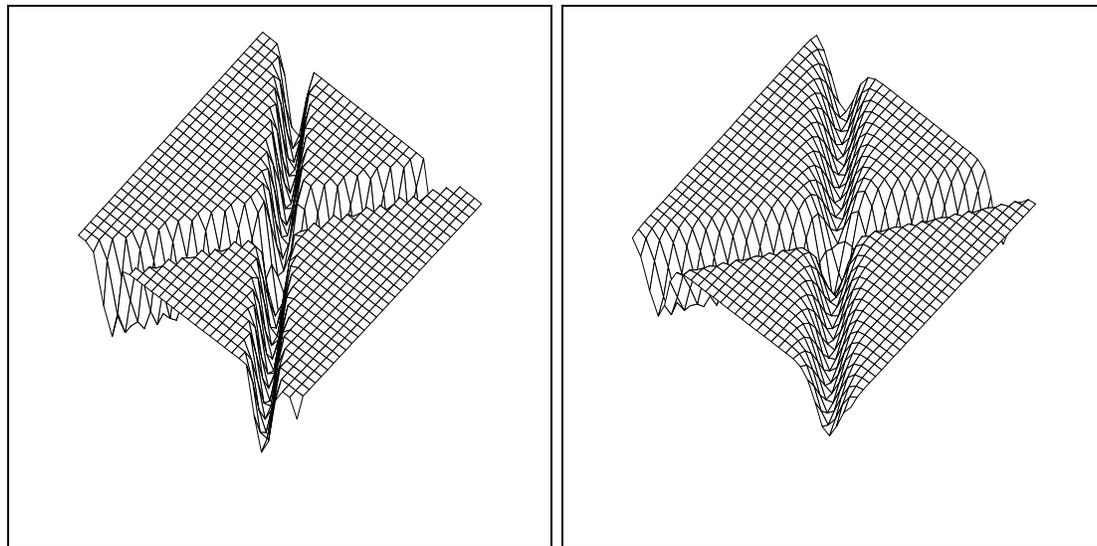
$$\sum_i (ax_i + by_i + c)^2$$

sous la contrainte  $a^2 + b^2 = 1$ . Cette contrainte permet de minimiser la somme du carré des différences entre chaque point et la ligne ajustée. Cette optimisation contrainte a été résolue par la méthode des lagrangiens [Fau93].

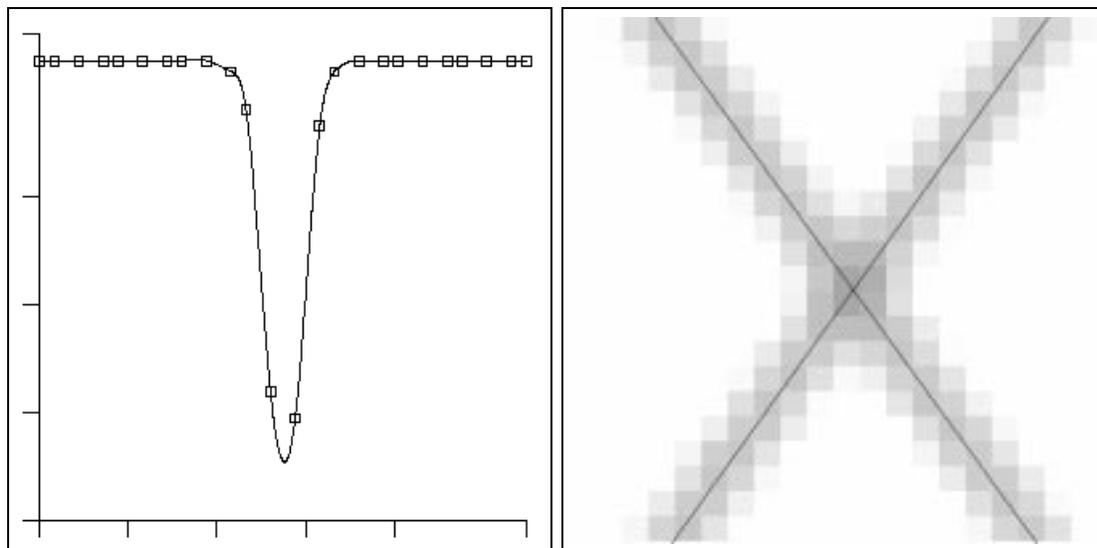
L'algorithme 1.2 présente un résumé de cette méthode. La figure 1.6(b) montre un zoom sur l'extraction d'une croix par l'extraction des deux droites qui la composent.

### 1.3 Expérimentations

Afin d'évaluer la qualité des corrections effectuées, plusieurs phénomènes ont été étudiés. Après avoir précisé les conditions matérielles (carte d'acquisition d'images, mire utilisée, ...) dans lesquelles ont eu lieu les différents tests, la précision d'extraction des croix de la mire sera évaluée. Ensuite, la stabilité de la prise d'image sera étudiée, tout comme la stabilité de la correction des distorsions en fonction de divers paramètres tels que la position de la mire dans l'image, le temps de la prise d'images, l'ouverture de l'iris et le réglage de la mise au point. Enfin, la qualité des corrections sera analysée.

(a) Niveaux de gris dans l'image  $I(x, y)$ (b) Niveaux de gris dans le modèle  $I_m(x, y)$ 

**Figure 1.5** : Visualisation du graphe des intensités lumineuses d'une croix, dans l'image et dans le modèle utilisant des fonctions gaussiennes



(a) intensités lumineuses le long d'une coupe perpendiculaire à une ligne de la mire et son interpolation par une fonction B-spline

(b) Zoom sur une partie de l'image de la mire contenant une croix. Les 2 lignes extraites par la méthode des B-splines sont affichées

**Figure 1.6** : Résultat de l'extraction d'une croix par la méthode des B-splines

### 1.3.1 Conditions d'expérimentations

Les tests de qualité des corrections des distorsions ont été effectués avec une caméra PULNIX TM-6EX. L'objectif utilisé est un objectif KINOPTIK, de longueur focale 12.5mm. La carte d'acquisition est une carte "Imaging Technology Incorporated Series 150". Cette carte fournit des images digitalisées de taille  $512 \times 512$  pixels sur 8 bits (256 niveaux de gris).

La mire utilisée a été réalisée par un procédé de phototraçage laser (Disk), ayant une résolution de 2000 lignes par pouce. Le support utilisé est un film polyester transparent clair, de  $18/100^{\text{ème}}$  de millimètre d'épaisseur (Dupond de Nemours).

Lors de ces expériences, la caméra se trouve à peu près à la verticale de la mire. Dans une telle position, tous les points d'intersection des lignes de la mire sont nets dans l'image, ce qui favorise leur bonne extraction, et la grille extraite recouvre toute l'image.

La mire a été placée de façon à avoir ses lignes obliques dans l'image. Cette position permet d'éviter un biais systématique dans la détection des croix. Ce biais pourrait apparaître si les lignes de la mire se projetaient dans l'image suivant les directions horizontales et verticales.

La figure 1.7 montre une image de distorsions obtenue dans de telles conditions d'expérience. La correction a été réalisée pour des points répartis de façon homogène dans l'image (tous les 40 pixels). Chaque ligne partant d'un point montre la direction de la correction à apporter à ce point. La longueur de ces lignes est 40 fois supérieure au module du vecteur de correction à appliquer aux croix respectives. Dans cet exemple, le plus grand déplacement à un module égal à 0.52 pixels.

Les 4 points entourés sont les points fixés comme invariants, c'est à dire qu'ils sont considérés comme ne subissant aucune distorsion. Ce sont ces 4 points qui définissent la transformation projective  $M$  qui reprojette les points exprimés dans le repère virtuel vers un repère plus proche du repère pixel (cf section 1.1).

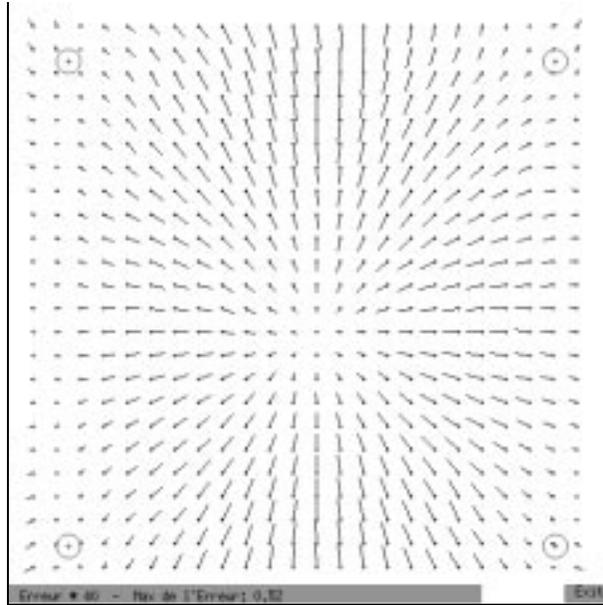
### 1.3.2 Critères de qualité de la modélisation de croix

La différence au carré des intensités lumineuse entre l'image et le modèle est un bon critère de qualité de la modélisation globale d'une croix par une double gaussienne. Ce critère est défini par

$$ssn = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n \varepsilon_i^2}$$

où

- $n$  est le nombre de pixels de la surface  $S$  sur laquelle la minimisation est effectuée. Typiquement,  $n \approx 21^2$ .
- $\varepsilon_i^2$  avec  $\varepsilon_i = I_m(x_i, y_i) - I(x_i, y_i)$  pour les  $n$  points  $(x_i, y_i)$  de  $S$  (cf équation 1.1).



**Figure 1.7 :** Affichage des distorsions de l'objectif KINOPTIK, de longueur focale 12.5mm. Les points entourés sont les points fixés comme invariants. Chaque croix représente un pixel, et chaque segment montre la direction de la correction à apporter à ce point, multipliée par un facteur 40

Expérimentalement, la valeur de  $ssn$  est proche de 2.5. Cette valeur est faible comparée aux 256 niveaux de gris possibles de l'image. Néanmoins, aucun support théorique n'est disponible pour bien apprécier cette valeur.

Lors de la recherche de la position d'une croix par sa décomposition en deux lignes, la qualité de l'extraction de chaque ligne peut être un bon critère. La moyenne des distances entre chaque point extrait d'une ligne et la droite ajustée est calculée par

$$m = \frac{1}{n} \sum_{i=1}^n |ax_i + by_i + c|$$

sachant que  $a^2 + b^2 = 1$ . La valeur expérimentale de  $m$  est  $m \approx 0.06$  pixel, pour une droite ajustée en utilisant environ 30 points. Cette valeur est faible, ce qui fait penser que l'extraction des droites composant une croix se déroule bien, et que l'extraction du point d'intersection est ainsi précis.

### 1.3.3 Stabilité de la correction

Cette série de tests a pour but de mettre en évidence la stabilité du calcul des distorsions. Plusieurs phénomènes peuvent en effet perturber les résultats obtenus. Tout d'abord, il faut s'assurer que les corrections de distorsions ne dépendent pas de la *position de la mire* dans l'image.

De plus, le *système d'acquisition d'image* n'est pas stable. La caméra utilisée est de type CCD<sup>1</sup>. Elle convertit chaque photon en un signal par le biais d'ionisation de cellules [BB82]. Cette ionisation introduit un bruit lors de la génération de l'image qui suit une loi de poisson [BC92]. Cette loi possède une variance plus ou moins forte suivant l'intensité lumineuse des pixels. De plus, ce bruit est non corrélé entre pixels voisins. Néanmoins, la variance du bruit intrinsèque à la cellule CCD peut être estimée [HK92].

La carte d'acquisition qui digitalise l'image fournie par la caméra contribue aussi à rajouter du bruit dans l'image. Par exemple, la désynchronisation de l'horloge de la caméra avec celle du digitaliseur peut atteindre 5% [LT88]. Le phénomène de line-jitter (translation jusqu'à 1 pixel d'une ligne de l'image) peut aussi se produire lors du transfert de l'image de la caméra vers la carte d'acquisition [Bey92b].

De plus, le matériel contient de nombreux composants électroniques très sensibles à la température ambiante. Ainsi, une variation de 8 à 10 degrés de la température ambiante peut doubler le bruit thermique de la cellule CCD [Bey91]. L'instabilité de ces composants est surtout sensible à la mise en route du système d'acquisition. Il est ainsi préférable d'attendre 3 heures avant toute prise d'image. En effet, la position dans l'image d'une cible peut varier de 0.1 pixel entre deux prises d'images faites immédiatement après la mise en route du système d'acquisition [Bey92b].

Enfin, *l'éclairage par néons* fournit une lumière instable qui peut influencer sur l'image acquise.

### Conditions d'expérimentations

Afin mettre en évidence la stabilité du calcul des distorsions, 7 images de la mire ont été prises. Les conditions des prises d'images sont schématisées sur la figure 1.8. Les images sont numérotées  $I_{1b}, I_1 \dots I_5, I_{5b}$ . Les images  $I_{1b}$  et  $I_1$  ont été prises à des moments très proches dans le temps, avec deux positions différentes de la mire. Les images  $I_1 \dots I_5$  ont été prises toutes les demi-heures, en laissant la position de la mire identique. Les mires  $I_5$  et  $I_{5b}$  ont été photographiées presque simultanément, pour des positions différentes de la mire.

Ainsi, différentes corrections de distorsions sont évaluées pour le même objectif, dans des conditions légèrement différentes. Les 4 points invariants considérés comme ne subissant aucune distorsion par rapport au modèle sténopé classique sont identiques pour toutes les corrections. Ils sont placés à chaque coin de l'image, décalés de 50 pixels.

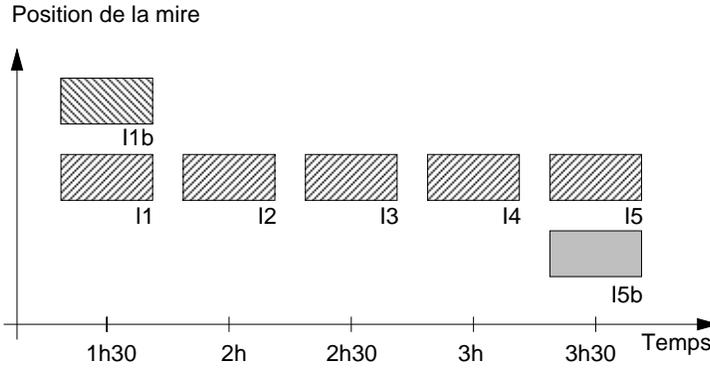
Dans les résultats des tests qui vont suivre,  $I_j(x, y)$  représentera l'intensité lumineuse du pixel  $(x, y)$  dans l'image  $j$ . Deux types de résultats sont analysés : tout d'abord, la stabilité de la prise d'image proprement dite, puis la stabilité des corrections de distorsions calculées avec les différentes images de la mire utilisées.

### Stabilité des prises d'images

Dans un premier temps, le bruit dû à la prise d'images a été estimé. La différence d'intensité lumineuse de chaque pixel entre les images 1 et 5 a été calculée. Ces images

---

<sup>1</sup>Charge Coupled Device



**Figure 1.8 :** Conditions expérimentales afin de mettre en évidence la bonne stabilité du processus de correction des distorsions

étant prises dans des conditions rigoureusement identiques (position de la mire par rapport à la caméra, éclairage, ...), ces différences mettent en évidence le bruit induit par l'acquisition d'une image. Les statistiques calculées sont :

$$\begin{aligned} \max &= \max_{x,y} |I_i(x,y) - I_j(x,y)| \\ \text{moy} &= \frac{1}{n} \sum_{x,y} (I_i(x,y) - I_j(x,y)) \\ \text{rms} &= \left( \frac{1}{n} \sum_{x,y} (I_i(x,y) - I_j(x,y))^2 \right)^{1/2} \end{aligned}$$

où  $n$  est le nombre de pixel dans une image. Les initiales rms signifient *Root Mean Square*. Le tableau 1.1 montre les résultats obtenus pour chaque couple d'images. Les valeurs écrites en gras sont les erreurs les plus importantes. Les différences d'intensités lumineuses peuvent être relativement importantes, surtout en ce qui concerne son maximum qui peut atteindre 58 niveaux de gris, soit plus de 20% d'erreur comparé aux 256 niveaux de gris possible de l'image.

### Stabilité des corrections

Les corrections fournies par les diverses images de la mire ont été comparées. Le tableau 1.2 contient les résultats de stabilité obtenus en utilisant la méthode de modélisation gaussienne de la croix. Le tableau 1.3 contient ceux obtenus par une extraction des croix utilisant les fonctions B-splines. Les statistiques calculées ont été :

$$\begin{aligned} \max &= \max_{(x,y)} \|C^i(x,y) - C^j(x,y)\| \\ \text{rms} &= \left( \frac{1}{n} \sum_{(x,y)} \|C^i(x,y) - C^j(x,y)\|^2 \right)^{1/2} \end{aligned}$$

où  $C^i(x,y)$  est le vecteur de la correction apportée au point de coordonnées  $(x,y)$ , lorsque la mire de l'image  $i$  est utilisée pour corriger l'image. Les 7 images ( $i, j = 1b, 1 \dots 5, 5b$ )

ont été utilisées pour calculer les distorsions de la caméra.

La correction des distorsions semble plus stable lorsque l'extraction des croix est réalisée par la méthode utilisant des B-splines. En effet, les corrections diffèrent au maximum de 0.068 à 0.116 pixels si la modélisation des croix par deux gaussiennes est utilisée. Ces chiffres tombent entre 0.026 et 0.080 pixels en utilisant des B-splines. De plus, les valeurs statistiques obtenues avec la méthode des B-splines sont toujours inférieures à celles obtenues par la double-gaussienne.

Ces expériences montrent que l'instabilité de cette méthode de correction des distorsions est de l'ordre de  $2/100^{\text{ème}}$  de pixels lorsque les croix sont modélisées à l'aide de fonctions B-splines, et au  $3/100^{\text{ème}}$  de pixels lorsque les croix sont modélisées par une fonction composée de deux fonctions gaussiennes.

### 1.3.4 Facteurs influençant les distorsions

Les expérimentations réalisées dans la sous-section précédente ont montré que la stabilité des corrections est de l'ordre de  $2/100^{\text{ème}}$  de pixel lorsque la configuration de l'acquisition de l'image reste la même. La série de tests de cette sous-section va souligner l'influence sur les distorsions optiques de deux paramètres de la prise d'image : l'ouverture de l'iris de l'objectif et la mise au point.

#### Ouverture de l'iris

L'influence de l'ouverture de l'iris sur l'image est de plusieurs natures. Son rôle immédiat est la modification de la luminosité des images. Mais l'ouverture de l'iris est aussi directement liée à la profondeur de champ. Plus l'iris est ouvert, moins la profondeur de champ est importante et plus il est difficile d'obtenir une image nette [Hor86]. Dans un tel cas, la lumière créant l'image traverse la lentille de l'objectif sur des zones relativement éloignées de son centre. Ces parties de la lentille sont les moins parfaites, pouvant engendrer des distorsions optiques par rapport au modèle sténopé [Sla80]. Les distorsions peuvent donc être modifiées lorsque l'ouverture de l'iris est changée.

Afin d'évaluer les distorsions optiques en fonction de l'ouverture de l'iris, 6 images  $I_i$  ont été prises avec un objectif SCHNEIDER de longueur focale 17mm. Seule l'ouverture de l'iris de l'objectif a changé entre deux prises de vues. Les valeurs de l'ouverture d'iris sont 2.0 (pour  $I_1$  ; l'iris est très ouvert), 2.8, 4.0, 5.6, 8.0 et 11.0 (pour  $I_6$  ; l'iris est pratiquement fermé). Les corrections de distorsions obtenues avec les images  $I_1$  et  $I_6$  sont schématisées sur la figure 1.9. Comme précédemment, des statistiques sur les différences de corrections sont calculées pour chaque couple  $(I_i, I_j)$  d'images :

$$\begin{aligned} \max &= \max_{(x,y)} \|C^i(x,y) - C^j(x,y)\| \\ \text{rms} &= \left( \frac{1}{n} \sum_{(x,y)} \|C^i(x,y) - C^j(x,y)\|^2 \right)^{1/2} \end{aligned}$$

Le tableau 1.4 contient les valeurs expérimentales de ces statistiques. L'instabilité des

	$I_1 \leftrightarrow I_2$	$I_1 \leftrightarrow I_3$	$I_1 \leftrightarrow I_4$	$I_1 \leftrightarrow I_5$	$I_4 \leftrightarrow I_5$
max	22	28	32	45	<b>58</b>
moy	0.31	3.44	<b>3.67</b>	0.39	-3.28
rms	2.30	6.58	7.16	8.93	<b>12.13</b>
	$I_2 \leftrightarrow I_3$	$I_2 \leftrightarrow I_4$	$I_2 \leftrightarrow I_5$	$I_3 \leftrightarrow I_4$	$I_3 \leftrightarrow I_5$
max	34	35	43	20	57
moy	3.12	3.36	0.08	0.28	-3.05
rms	6.39	7.20	7.56	2.04	11.05

**Table 1.1** : Statistiques sur les différences, en intensité lumineuse, d'images prises dans des conditions identiques

	$I_{1b} \leftrightarrow I_1$	$I_{1b} \leftrightarrow I_2$	$I_{1b} \leftrightarrow I_3$	$I_{1b} \leftrightarrow I_4$	$I_{1b} \leftrightarrow I_5$	$I_{1b} \leftrightarrow I_{5b}$	$I_4 \leftrightarrow I_5$
$\max^{(g)}$	<b>0.116</b>	0.082	0.090	0.075	0.080	0.104	0.068
$\text{rms}^{(g)}$	0.029	0.030	0.027	0.029	0.027	0.038	0.029
		$I_1 \leftrightarrow I_2$	$I_1 \leftrightarrow I_3$	$I_1 \leftrightarrow I_4$	$I_1 \leftrightarrow I_5$	$I_1 \leftrightarrow I_{5b}$	$I_4 \leftrightarrow I_{5b}$
$\max^{(g)}$		0.109	0.101	0.113	0.082	0.108	0.096
$\text{rms}^{(g)}$		0.030	0.024	0.027	0.023	0.032	<b>0.039</b>
			$I_2 \leftrightarrow I_3$	$I_2 \leftrightarrow I_4$	$I_2 \leftrightarrow I_5$	$I_2 \leftrightarrow I_{5b}$	$I_5 \leftrightarrow I_{5b}$
$\max^{(g)}$			0.090	0.106	0.095	0.114	0.083
$\text{rms}^{(g)}$			0.029	0.028	0.031	0.032	0.035
				$I_3 \leftrightarrow I_4$	$I_3 \leftrightarrow I_5$	$I_3 \leftrightarrow I_{5b}$	
$\max^{(g)}$				0.089	0.086	0.087	
$\text{rms}^{(g)}$				0.026	0.027	0.034	

**Table 1.2** : Stabilité des corrections de distorsions lorsque la modélisation d'une croix par une double-gaussienne est utilisée

	$I_{1b} \leftrightarrow I_1$	$I_{1b} \leftrightarrow I_2$	$I_{1b} \leftrightarrow I_3$	$I_{1b} \leftrightarrow I_4$	$I_{1b} \leftrightarrow I_5$	$I_{1b} \leftrightarrow I_{5b}$	$I_4 \leftrightarrow I_5$
$\max^{(b)}$	0.047	0.051	0.049	0.051	0.055	<b>0.080</b>	0.059
$\text{rms}^{(b)}$	0.020	0.018	0.017	0.016	0.023	<b>0.028</b>	0.021
		$I_1 \leftrightarrow I_2$	$I_1 \leftrightarrow I_3$	$I_1 \leftrightarrow I_4$	$I_1 \leftrightarrow I_5$	$I_1 \leftrightarrow I_{5b}$	$I_4 \leftrightarrow I_{5b}$
$\max^{(b)}$		0.051	0.043	0.040	0.051	0.059	0.063
$\text{rms}^{(b)}$		0.017	0.015	0.016	0.019	0.023	0.027
			$I_2 \leftrightarrow I_3$	$I_2 \leftrightarrow I_4$	$I_2 \leftrightarrow I_5$	$I_2 \leftrightarrow I_{5b}$	$I_5 \leftrightarrow I_{5b}$
$\max^{(b)}$			0.043	0.048	0.042	0.058	0.054
$\text{rms}^{(b)}$			0.017	0.020	0.017	0.018	0.022
				$I_3 \leftrightarrow I_4$	$I_3 \leftrightarrow I_5$	$I_3 \leftrightarrow I_{5b}$	
$\max^{(b)}$				0.026	0.060	0.068	
$\text{rms}^{(b)}$				0.012	0.020	0.027	

**Table 1.3** : Stabilité des corrections de distorsions lorsque la modélisation d'une croix est réalisée en utilisant des fonctions B-splines

distorsions lorsque l'ouverture de l'iris est modifiée peut atteindre 0.1 pixel, ce qui est important face aux 0.02 pixel d'instabilité intrinsèque à la méthode. Cette instabilité est d'autant plus forte que l'iris est ouvert. En revanche, lorsque l'iris est pratiquement fermé, une petite modification de son ouverture n'a que très peu d'influence sur les distorsions obtenues. Ces conclusions ne sont bien entendu valables que pour l'objectif testé car elles dépendent de la qualité de la lentille qui peut être non-homogène sur toute sa surface.

### Mise au point

Les effets sur la prise d'images d'un changement de la mise au point sont de plusieurs natures. En premier lieu, il modifie les paramètres internes du couple caméra-objectif. Ce changement peut être néanmoins approximé par une transformation affine ayant trois paramètres indépendants [EVF93]. Cette transformation peut être déterminée en suivant au moins deux points entre deux images.

Les distorsions dépendent aussi du réglage de la netteté effectué. En effet, une modification de la mise au point translate la lentille. Cette translation est effectuée par rotation sur la plupart des objectifs. Les distorsions sont ainsi modifiées car la lentille ne peut être parfaitement symétrique, et ne peut être positionnée parfaitement perpendiculairement à l'axe optique. Une modélisation des changements de distorsions en fonction d'un changement de mise au point est complexe, et dépend beaucoup de l'objectif utilisé. Nous nous contenterons dans cette section d'observer ces modifications pour un objectif SCHNEIDER 17mm

Cette expérience a été réalisée en prenant 7 images de la mire. Dans chaque image, seule la mise au point a été modifiée. L'ouverture de l'iris est 15, c'est à dire qu'il est pratiquement fermé afin d'augmenter la profondeur de champ. Ainsi, la qualité de l'image ne dépendra pas trop de la mise au point. Les distances du réglage sont 0.35 mètres pour  $I_1$ , 0.4, 0.5, 0.6, 0.8, 2.0 et l'infini (pour  $I_7$ ). Comme pour le test précédent, nous avons calculé

$$\text{rms} = \left( \frac{1}{n} \sum_{(x,y)} \|C^i(x,y) - C^j(x,y)\|^2 \right)^{1/2}$$

Les valeurs rms pour les différents couples d'images sont contenues dans le tableau 1.5. Ces chiffres montrent que les distorsions sont instables lorsque le réglage de la netteté est modifié. Par contre, les instabilités sont homogènes autour de  $\text{rms} \approx 0.05$  lorsque le réglage est fortement modifié, et ceci quelque soit les réglages choisis. L'homogénéité obtenue est due au fait que la lentille subit une rotation, mais que la partie de la lentille utilisée reste toujours la même.

Si l'on s'autorise en plus une modification de l'ouverture de l'iris, il faut donc rajouter une variation de 0.10 pixel, soit une stabilité globale de l'ordre de  $\sqrt{0.05^2 + 0.10^2} = 0.11$  point image si l'on s'autorise toutes les configurations de l'objectif. Il faut noter que ces chiffres ne sont significatifs que pour cet objectif SCHNEIDER, les distorsions dépendant principalement des défauts de la lentille. Aucune règle générale ne peut être déduite de ces tests.

### 1.3.5 Qualité de la correction des distorsions

La stabilité des corrections de distorsions est de l'ordre de  $2/100^{\text{ème}}$  de pixel lorsque la configuration de la prise d'image est fixe. Cette sous-section étudie la précision proprement dite des corrections. Pour cela, la qualité des alignements, dans l'image, des points de la mire avant et après correction est étudiée. Les points corrigés ne sont pas ceux de la mire qui servent à calculer les corrections afin de ne pas introduire de biais dans ces tests.

Une droite est ajustée au sens des moindres carrés sur les points théoriquement alignés de la mire  $I_{1b}$  (cf les tests de stabilité de la section 1.3.3). L'ajustement se fait avant et après correction des distorsions. Les corrections des distorsions optiques ont été réalisées en utilisant les images  $I_1$ ,  $I_5$  et  $I_{5b}$  de la mire. La qualité des ajustements de droites est exprimée dans les graphes des figures 1.10(a) et 1.11(a) (recherche des croix par une double-gaussienne) et dans les figures 1.10(b) et 1.11(b) (croix recherchées en utilisant des B\_splines). Chaque figure a en abscisse l'indice de la ligne ou de la colonne de la grille qui est ajustée (ligne formée des  $n$  points de coordonnées  $(x_i, y_i)$ ), et en ordonnée la valeur rms des distances des points à la droite ajustée d'équations  $ax + by + c = 0$ , sous la contrainte  $a^2 + b^2 = 1$  :

$$\text{rms} = \left( \frac{1}{n} \sum_{i=1}^n (ax_i + by_i + c)^2 \right)^{\frac{1}{2}}$$

Chaque figure montre quatre courbes. La courbe composée de points matérialisés par des carrés montre la valeur rms avant la correction des distorsions. Les trois autres courbes montrent la valeur rms après correction des distorsions suivant l'image de la mire utilisée.

Ces graphes montrent que les corrections sur les alignements sont bien réelles. Le facteur gagné avec cet objectif est un facteur 10 pour les points qui se trouvent sur le bord de l'image. En effet, les non-alignements de points dans l'image atteignent  $\text{rms} \approx 0.20$  pixel, alors qu'après correction, la valeur de rms tombe à  $\text{rms} \approx 0.02$ .

La qualité des alignements obtenu après correction des distorsions est équivalente à la stabilité de cette méthode. De plus, les corrections sont homogènes. En effet, la qualité de l'alignement obtenu est équivalente quelque soit la mire utilisée et quelque soit la position dans l'image de la droite ajustée. Elle ne dépend donc pas de l'amplitude des distorsions.

Le chapitre suivant traitera de l'extraction précise de points dans les images. Les tests d'évaluation de la précision d'extraction atteinte concerneront notamment la qualité de la reconstruction tridimensionnelle de points, ainsi que la qualité de la géométrie épipolaire obtenue avec ces points. Ces tests mettront en évidence la nécessité de corriger les distorsions optiques de certains objectifs (section 2.3).

## 1.4 Conclusions

Une méthode de correction de distorsions a été présentée et expérimentée. Cette méthode ne repose que sur la seule hypothèse que l'image observée est une déformation

continue d'une projection perspective parfaite, hypothèse très générale. Elle s'appuie sur l'observation d'une mire plane qui contient des lignes s'intersectant. Ces lignes s'intersectent en des points qui forment des alignements. Les distorsions sont corrigées en définissant une transformation point à point qui exprime les points de la mire dans un repère appelé repère virtuel. Le repère virtuel est choisi de façon à être compatible avec le modèle sténopé, c'est à dire que les droites de la scène sont projetées suivant des droites dans l'image. Une transformation projective est ensuite définie afin de reprojeter le repère virtuel dans un repère proche du repère pixel.

La précision des corrections obtenues dépend essentiellement de la précision d'extraction des positions des croix de la mire. Deux algorithmes ont été implantés dans ce but. La première méthode utilise une modélisation globale de l'image d'une croix par une fonction qui a la forme d'une double-gaussienne. La seconde méthode décompose une croix en deux lignes, et recherche à l'aide de B-splines des points de chacune de ces deux droites. Cette seconde méthode s'est avérée de part les tests effectués plus stable et plus simple à mettre en œuvre, cette modélisation étant linéaire.

Pour les objectifs que nous avons testés, les expériences montrent que les corrections de distorsions peuvent être reportées pour des ouvertures d'iris et des mises au point différentes, tant que la précision requise n'est pas supérieure à 0.10 pixels. Ceci ne signifie pas que les paramètres internes des caméras peuvent être reportés pour différentes ouvertures ou mises au point, mais seulement qu'après ces corrections, les projections dans l'image peuvent être considérées comme perspectives avec la précision annoncée, permettant ainsi l'usage de méthodes projectives.

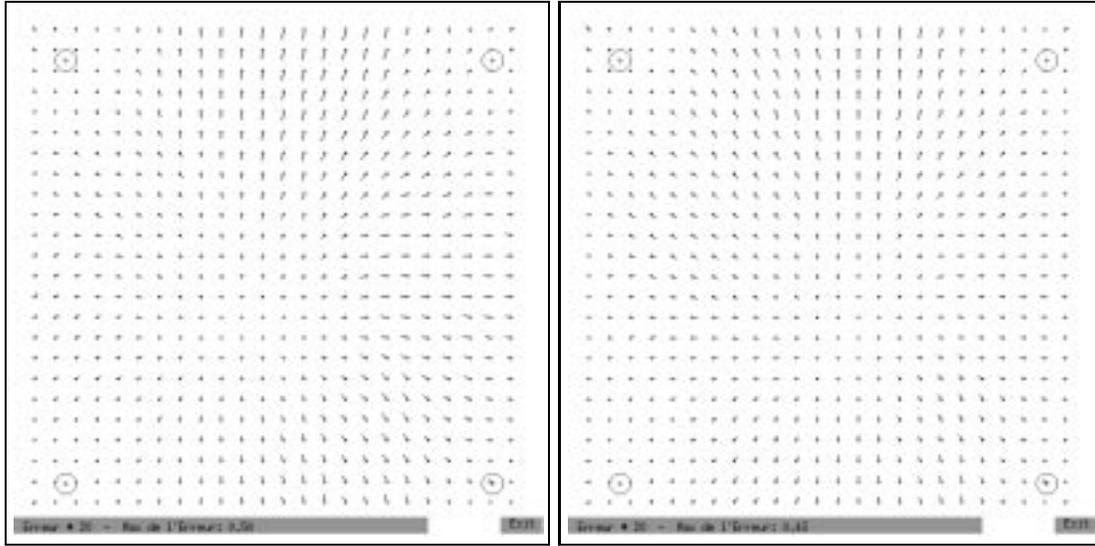
La stabilité des distorsions optiques de l'objectif SCHNEIDER, lorsque l'ouverture de l'iris et la mise au point sont modifiées, a été étudiée. Elle est de l'ordre de 0.10 pixel. Différentes expérimentations ont aussi montré que la stabilité au cours du temps et en fonction de la position de la mire (mais sans modifier le système de prise d'images) est de l'ordre de 0.02 pixel. Ces conclusions ont été confirmées par des tests effectués par la société ITMI APTOR<sup>2</sup> [BdPS94]. Leurs expérimentations ont eu pour but de quantifier la qualité de 3 objectifs de longueur focale courte : TEGEA KINOPTIK 5.7mm, COMPUTAR M618 6mm et un objectif XENOPLAN SCHNEIDER 6.5mm. Les distorsions optiques de l'objectif XENOPLAN sont très importantes, atteignant 20 pixels sur les bords de l'image. Celles de l'objectif TEGEA sont nettement plus faible (5 pixels sur les bord de l'image). Ces tests permettent de conclure que la stabilité des corrections restent la même quelque soit l'objectif utilisé, et est de l'ordre de 0.05 pixel.

Le fait que les distorsions ne soient pas classées suivant des composantes radiales ou tangentielles s'est avérée particulièrement adapté pour l'objectif TEGEA KINOPTIK. En effet, cet objectif a la particularité de disposer d'un dispositif atténuant les distorsions, qui ne suivent plus alors les schémas classiques. De plus, cette méthode ne nécessite aucun calibrage des paramètres internes et externes de la caméra. Seule une prise de vue d'une mire plane suffit, alors que généralement, il est nécessaire d'observer un volume contenant des points dont la position tridimensionnelle est connue.

---

<sup>2</sup>ITMI APTOR - 61, chemin du vieux chêne - BP 177 - 38244 Meylan Cedex

La correction des déformations géométriques de l'image permet de supposer dans la suite de la thèse que les données extraites d'images ont été obtenues avec une image conforme au modèle sténopé.

(a)  $I_1$  : ouverture d'iris = 2.0(b)  $I_6$  : ouverture d'iris = 11.0

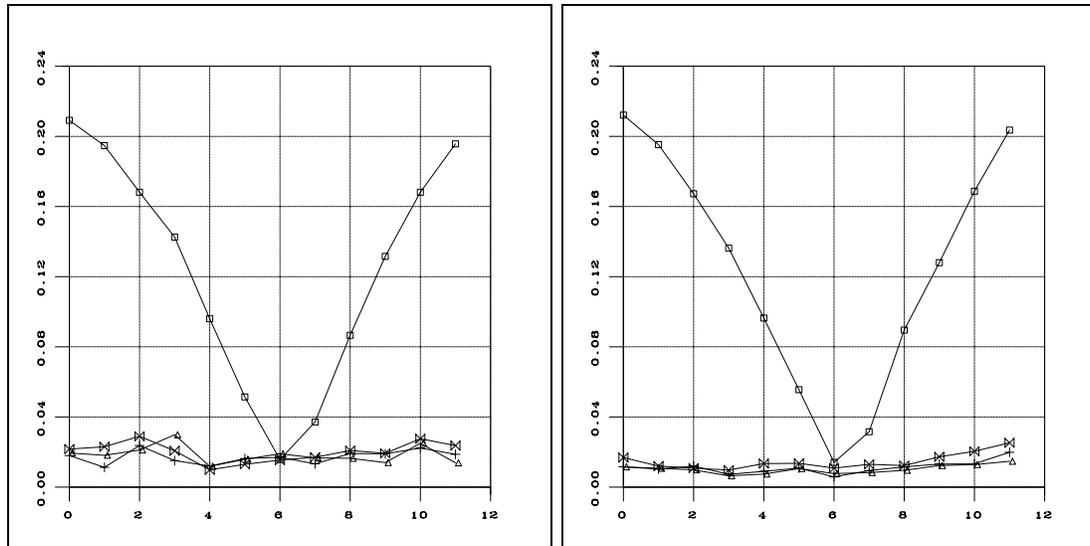
**Figure 1.9** : Corrections de distorsions obtenues avec deux ouvertures d'iris différentes de l'objectif SCHNEIDER, de longueur focale 17mm

	$I_1 \leftrightarrow I_2$	$I_1 \leftrightarrow I_3$	$I_1 \leftrightarrow I_4$	$I_1 \leftrightarrow I_5$	$I_1 \leftrightarrow I_6$	$I_3 \leftrightarrow I_4$	$I_3 \leftrightarrow I_5$	$I_3 \leftrightarrow I_6$
max	0.135	0.158	0.147	<b>0.161</b>	0.155	0.062	0.049	0.079
rms	0.072	0.096	0.086	<b>0.098</b>	0.091	0.027	0.021	0.0302
		$I_2 \leftrightarrow I_3$	$I_2 \leftrightarrow I_4$	$I_2 \leftrightarrow I_5$	$I_2 \leftrightarrow I_6$	$I_4 \leftrightarrow I_5$	$I_4 \leftrightarrow I_6$	
max		0.068	0.069	0.072	0.086	0.058	0.046	
rms		0.033	0.030	0.035	0.034	0.025	0.019	

**Table 1.4** : Influence de l'ouverture de l'iris de la caméra sur les distorsions

	$I_1 \leftrightarrow I_2$	$I_1 \leftrightarrow I_3$	$I_1 \leftrightarrow I_4$	$I_1 \leftrightarrow I_5$	$I_1 \leftrightarrow I_6$	$I_1 \leftrightarrow I_7$	$I_5 \leftrightarrow I_6$
rms	0.036	0.034	0.053	0.040	<b>0.061</b>	0.054	0.036
		$I_2 \leftrightarrow I_3$	$I_2 \leftrightarrow I_4$	$I_2 \leftrightarrow I_5$	$I_2 \leftrightarrow I_6$	$I_2 \leftrightarrow I_7$	$I_5 \leftrightarrow I_7$
rms		0.026	0.034	0.032	0.044	0.044	0.029
			$I_3 \leftrightarrow I_4$	$I_3 \leftrightarrow I_5$	$I_3 \leftrightarrow I_6$	$I_3 \leftrightarrow I_7$	$I_6 \leftrightarrow I_7$
rms			0.036	0.026	0.042	0.038	0.037
				$I_4 \leftrightarrow I_5$	$I_4 \leftrightarrow I_6$	$I_4 \leftrightarrow I_7$	
rms				0.033	0.024	0.039	

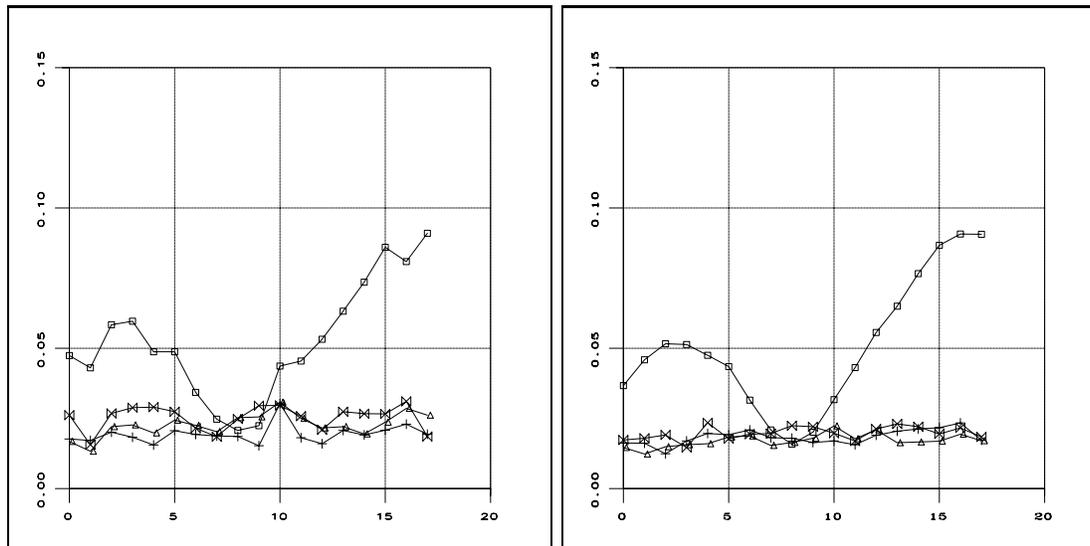
**Table 1.5** : Stabilité des distorsions lorsque la mise au point est modifiée



(a) Extraction des croix réalisée par la double-gaussienne

(b) Extraction des croix obtenue avec des B-splines

**Figure 1.10 :** Qualité des alignements des droites horizontales. Pour chaque ligne de la mire (en abscisse), valeurs rms, comprises entre 0.0 et 0.24 pixel, des distances entre les points et la droite ajustée (en ordonnée), avant et après correction des distorsions



(a) Extraction des croix réalisée par la double-gaussienne

(b) Extraction des croix obtenue avec des B-splines

**Figure 1.11 :** Qualité des alignements de droites verticales. Pour chaque colonne de la mire (en abscisse), valeurs rms, comprises entre 0.0 et 0.15 pixel, des distances entre les points et la droite ajustée (en ordonnée), avant et après correction des distorsions

## Chapitre 2

# Extraction de positions de points

L'extraction d'indices dans les images est une étape indispensable en analyse d'images. Ces indices sont divers, allant du plus simple (le point) aux plus complexes (contours, régions, ...). Ce chapitre s'intéresse à l'extraction de points particuliers dans les images. Ces points extraits peuvent être utilisés dans de nombreux traitements d'images, comme par exemple le calibrage, le calcul de la géométrie épipolaire ou la reconstruction tridimensionnelle de points dans une scène.

L'extraction de points dans les images a déjà beaucoup été étudiée. Des détecteurs de points d'intérêts sont apparus très tôt, comme le détecteur de Moravec [Mor77] dont le but est d'extraire des coins de la scène. D'un point de vue général, un bon extracteur de points en vue de leur suivi dans une séquence d'images possède trois qualités :

- une bonne *détection*, au sens du nombre de points extraits. Il faut être en mesure d'extraire le plus de points possible afin d'acquérir le maximum de données sur la scène.
- une bonne *stabilité* de leur localisation afin de s'assurer que les points extraits le long d'une séquence d'image correspondent bien à la projection d'un même point tridimensionnel de la scène.
- une bonne *localisation* des points trouvés : il faut que les positions détectées soient le plus proche possible des positions exactes.

Ce chapitre s'intéresse particulièrement à la précision de la localisation de points dans les images. L'extraction de deux types de points est réalisée dans ce chapitre. La première section traite des points dont le signal est modélisable. Deux cas particuliers sont considérés : celui des cibles circulaires et celui des coins. Différents tests valident les précisions de localisations obtenues pour ces deux entités. Ces expérimentations sont basées sur quatre type de mesure :

- la conservation des alignements de points dans les images.
- la qualité de la reconstruction tridimensionnelle de ce type de points.
- la qualité de la géométrie épipolaire évaluée avec ces points.

- la stabilité des positions des points extraits à l'aide de birapports.

La dernière section de ce chapitre traite de la mise en correspondance précise entre deux images de points quelconques. Après une mise en correspondance classique utilisant une fonction de corrélation, le point obtenu est repositionné.

## 2.1 Extraction de points modélisables

Cette section traite de l'extraction dans les images de points modélisables. Le principe général de la méthode est exposé dans la première partie. Il consiste essentiellement dans la recherche d'une transformation affine entre l'image et le modèle de l'entité recherchée. Afin de rester suffisamment général, la forme du modèle n'est pas précisée. L'extraction de deux entités particulières est ensuite étudiée : les coins et les cibles. Pour chacune de ces deux entités, le modèle utilisé est explicité, ainsi que le processus de l'approximation préalable de la position des indices dans l'image.

### 2.1.1 Algorithme

Les points modélisables ont l'avantage d'être facilement caractérisables. On peut donc espérer localiser leur position précisément dans les images. R. Deriche [DB93] et K. Rohr [Roh92] ont utilisé des algorithmes faisant explicitement appel à un modèle paramétrique des niveaux de gris pour la détection de droites et de coins dans l'image. Ces deux algorithmes recherchent la valeur des paramètres du modèle afin que celui-ci s'ajuste au mieux, au sens des moindres carrés, avec l'image. Les paramètres permettent ensuite de trouver la position de l'entité recherchée dans l'image.

Notre optique est quelque peu différente. Nous allons recherché la meilleure transformation affine qui permet d'ajuster notre modèle paramétrique à l'image. Le modèle est paramétré par rapport aux caractéristiques de la forme observée, mais non par rapport à son orientation et à sa position dans l'image. Il est donc bien plus contraints que les modèles utilisés par Deriche et Rohr. La prise d'image est en fait modélisée par la transformation affine recherchée.

Notre méthode se positionne ainsi entre la méthode de Deriche où les paramètres du modèle contiennent toute l'information sur les caractéristiques du point à modéliser, et celle de Gruen [Gru85, GS91] où seule la transformation affine est prise en compte, avec un modèle totalement figé.

Cet algorithme revient donc à minimiser la valeur de

$$\chi^2 = \sum_{x_i, y_i} (I(x_i, y_i) - I_m(x'_i, y'_i, \mathbf{p}))^2$$

avec les notations

- $I(x_i, y_i)$  est le niveau de gris, dans l'image, du pixel  $(x_i, y_i)$ .
- $I_m(x'_i, y'_i, \mathbf{p})$  est l'intensité du pixel  $(x'_i, y'_i)$  dans le modèle lorsque les paramètres ont les valeurs de  $\mathbf{p}$ .  $(x'_i, y'_i)$  est obtenu à partir de  $(x_i, y_i)$  en lui appliquant une

transformation affine représentée par la matrice  $2 \times 3$   $\mathbf{A}$  :

$$\begin{pmatrix} x'_i \\ y'_i \end{pmatrix} = \begin{pmatrix} a & b & c \\ a' & b' & c' \end{pmatrix} \cdot \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \mathbf{A} \cdot \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}$$

La minimisation de  $\chi^2$  se fait sur une petite partie de l'image qui contient la primitive à détecter. L'algorithme qui a été choisit pour l'optimisation du  $\chi^2$  est l'algorithme itératif de Levenberg-Marquardt [PFTW88]. L'optimisation fournit la matrice  $\mathbf{A}$  de la transformation affine ainsi que la valeur des paramètres  $\mathbf{p}$  qui minimisent la valeur de  $\chi^2$ .

La position de l'indice recherchée (la position d'un coin ou du centre d'une cible circulaire par exemple) est connue exactement dans le modèle. Sa position dans l'image est obtenue en appliquant la transformation affine inverse de  $\mathbf{A}$  à sa position dans le modèle.

Les deux sous-sections suivantes appliquent cet algorithme à deux types de points particuliers : les coins de l'image et les cibles circulaires.

### 2.1.2 Extraction de coins dans l'image

Les coins de l'image sont définis comme la jonction de plusieurs droites séparant des régions de l'image dont les intensités lumineuses sont plus ou moins homogènes. Dans cette section, seuls les coins formés par deux droites seront considérés. Ces coins sont appelés les coins en "*L*".

Une multitude d'algorithmes a vu le jour afin de détecter cette entité car elle est facilement caractérisable et elle est présente dans de nombreuses scènes. Ces méthodes peuvent se diviser en deux grandes catégories : les méthodes basées sur les contours de l'image, et celles utilisant directement les niveaux de gris.

La première catégorie de détecteurs considère les points où la direction du contour varie fortement comme étant des coins. Ainsi, des détecteurs de coins de l'image utilisent directement les contours. Une simple approximation polygonale des chaînes de contours permet, après filtrage des points qui se trouvent effectivement sur des droites, d'extraire les coins [HJF89].

Les coins peuvent également être détectés en approximant les contours par des fonctions B-splines [MY87]. Les coins sont alors définis comme les extrémaux locaux de la courbure de la fonction B-spline.

La courbure des contours peut être calculée directement à partir des chaînes de contours, sans les interpoler par des fonctions B-splines. Les contours étant représentés par des chaînes de points discrets, ce calcul de la courbure des contours nécessite une régularisation [SAKS94].

La seconde catégorie de détecteurs de coins utilise directement les niveaux de gris de l'image. Les premiers détecteurs réalisés font usage de masques. Un coin est alors défini

comme étant un point ayant une forte réponse à un masque ([MN90] fournit un état de l'art des extractions de coins utilisant des masques). Le masque peut être réalisé de façon à minimiser le rapport signal/bruit [RSB89].

D'autres détecteurs de coins utilisent les dérivées des intensités lumineuses de l'image. Ainsi, les points dont la direction du gradient varie fortement, et qui sont des maxima locaux du gradient, sont détectés comme étant des coins [SS90, SS94]. D'autres opérateurs utilisant des fonctions du gradient et du laplacien ont aussi été étudiés. [Nag83] en fournit un état de l'art.

Néanmoins, la position du coin ne se trouve pas exactement sur un maximum local du gradient. En remarquant d'une part que le coin se trouve sur un passage par 0 du laplacien (qui ne dépend pas de l'échelle à laquelle il est calculé), et d'autre part que la position obtenue par l'opérateur DET<sup>1</sup> [Bea78] se déplace sur une droite selon l'échelle à laquelle est calculée la valeur de DET, la position du coin peut être recalée en réalisant l'intersection de ces deux entités [DG90] [DG93].

Le recalage de coins est facilité lorsqu'un modèle est disponible. Dans [OG93], le modèle est créé pour chaque coin à partir d'une segmentation préalable de l'image. La détection du coin est alors réalisée par une mesure de corrélation classique entre l'image et ce modèle. Des modèles plus perfectionnés et ayant plus de degrés de libertés peuvent être utilisés (cf section précédente [DB93, Roh92]). Le paragraphe suivant indique comment l'algorithme présenté précédemment (section 2.1.1) est adapté à la détection de coins, notamment en explicitant le modèle utilisé.

## Modélisation

Le modèle des niveaux de gris autour d'un coin est composé d'un angle droit qui sépare 2 régions de niveaux de gris homogènes  $\mathbf{p} = (n_1, n_2)$ . Le coin de cet angle est positionné en  $(0, 0)$ . Les 2 droites composant cet angle sont parallèles aux axes  $(Ox)$  et  $(Oy)$ . Ce modèle est exprimé par la fonction

$$I'(x, y, \mathbf{p}) = \begin{cases} n_1 & \text{si } x \geq 0 \text{ et } y \leq 0 \\ n_2 & \text{sinon} \end{cases}$$

Ce modèle ne prend pas en compte le flou induit dans l'image par une profondeur de champs non parfaite de l'objet. La convolution avec une fonction correspondant à la réponse impulsionnelle du système d'acquisition permet de prendre en compte ce phénomène [BB82, SHB93]. La fonction utilisée est la fonction de densité bidimensionnelle de la loi normale afin de générer un flou gaussien [BB82, Pen87, Pra78]

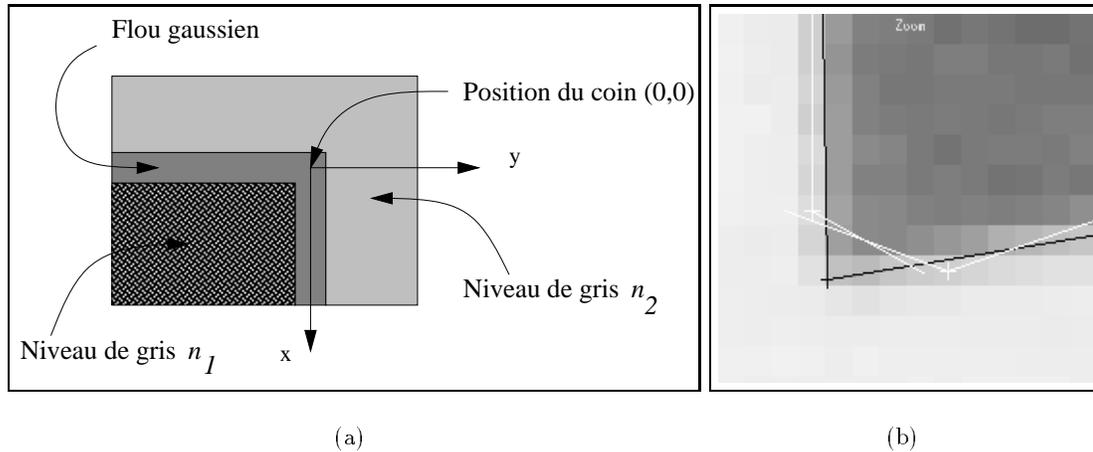
$$g_2(\alpha, \beta) = \frac{1}{2\pi\sigma^2} \exp^{-(\alpha^2+\beta^2)/(2\sigma^2)}$$

Ainsi, les intensités lumineuse de notre modèle sont égales à

$$I_m(x, y, \mathbf{p}) = g_2 * I'(x, y, \mathbf{p}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g_2(\alpha, \beta) I'(x - \alpha, y - \beta, \mathbf{p}) d\alpha d\beta$$

---

<sup>1</sup>DET= $I_{xx}I_{yy} - I_{xy}^2$  où  $I_{xx}$  (resp.  $I_{yy}$ ) est la dérivée seconde de l'image suivant l'axe de  $x$  (resp.  $y$ ), et  $I_{xy}$  la dérivée suivant l'axe  $y$  de la dérivée suivant l'axe  $x$ .



**Figure 2.1 :** (a) Schématisation du modèle d'un coin. (b) Exemple d'extraction d'un coin. Ce zoom sur l'image montre deux positions initiales de la position de ce coin (en blanc) fournies par les jonctions de l'interpolation polygonale des contours (voir texte). Les deux positions finales obtenues à partir de ces approximations sont éloignées de 0.02 pixels, et une seule est affichée (en noir)

Comme les pixels qui sont éloignés de plus de  $3\sigma$  des segments de droites formant l'angle ont leur niveau de gris modifié de façon négligeable ( $I_m(x, y, \mathbf{p}) \approx I'(x, y, \mathbf{p})$  pour ces points), le modèle peut être simplifié en

$$I_m(x, y, \mathbf{p}) = \begin{cases} n_1 & \text{si } x \geq 3\sigma \text{ et } y \leq -3\sigma \\ n_2 & \text{si } x \leq -3\sigma \text{ ou } y \geq 3\sigma \\ g_2 * I'(x, y, \mathbf{p}) & \text{sinon} \end{cases}$$

La figure 2.1(a) schématise le modèle  $I_m(x, y)$ . La figure (b) montre quant à elle un exemple d'extraction d'un coin de l'image pour deux positions initiales du coin et les deux positions (pratiquement confondues) après optimisation de la transformation affine. La figure 2.2(a) montre le graphe des niveaux de gris dans l'image autour d'un coin, et la figure (b) montre le graphe du modèle une fois la transformation affine trouvée.

Une telle paramétrisation de ce problème n'est pas minimale. En effet, 4 paramètres suffisent à décrire un coin (2 pour la translation ainsi que 2 angles) comme le font Deriche [DB93] et Rohr [Roh92], alors que nous en utilisons 6 pour décrire la transformation affine. La paramétrisation affine a cependant l'avantage de rendre le système à minimiser pratiquement linéaire, alors qu'une paramétrisation incluant des angles devient fortement non-linéaire avec l'introduction de fonctions trigonométriques. La paramétrisation affine aide ainsi la convergence de l'algorithme de Levenberg-Marquardt. Il est à noter que le minimum se trouve dans une région plane de la fonction à minimiser due à la surparamétrisation du modèle utilisé.

L'écart-type  $\sigma$  du flou gaussien ne fait pas parti du vecteur de paramètres  $\mathbf{p}$ . L'inclure dans  $\mathbf{p}$  aurait comme conséquence de rendre l'optimisation fortement non-linéaire. Il s'est

avéré que la valeur de  $\sigma$  a peu d'influence sur la précision de la localisation obtenue. Quantitativement, il y a une différence de seulement 0.01 pixel de précision entre les positions des coins obtenues avec  $\sigma = 0.5$ ,  $\sigma = 1.0$  et  $\sigma = 1.5$ . La valeur de  $\sigma = 1.0$  a été utilisée par la suite car elle correspond à un bon compromis entre une image très nette et une image floue.

La taille de la fenêtre où l'optimisation est réalisée est de taille fixée par l'utilisateur. La taille standard utilisée dans la suite est  $21 \times 21$  pixels. Le temps d'exécution moyen, sur une station SUN SPARC 10.51 est de 332ms par coin extrait avec cette taille de fenêtre.

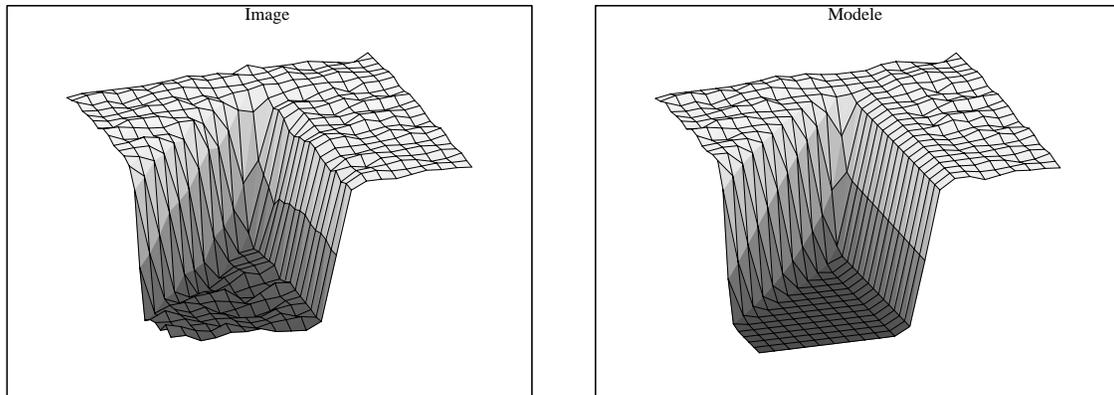
Lorsque la partie convexe du coin correspond à une partie occultée de la scène, les niveaux de gris de cette partie de l'image ne sont pas uniformes. Le modèle utilisé n'est alors plus adapté. Afin de diminuer les effets de ce phénomène dans le processus d'optimisation, la fenêtre de corrélation est érodée au fur et à mesure des itérations de façon à restreindre le nombre de pixel de la partie convexe du coin pris en compte dans la minimisation de la valeur de  $\chi^2$ . L'exemple de la figure 2.2(b) illustre ce procédé. Dans ce graphe, les points du modèle qui n'appartiennent plus à la fenêtre de corrélation se voient attribuer le niveau de gris du pixel correspondant dans l'image.

## Initialisation

L'initialisation des paramètres est obtenue à partir des contours de l'image, obtenus par un filtre de Canny-Deriche. Ils sont approximés par des lignes polygonales [HJF89]. On forme ensuite les jonctions des segments de droites. Une jonction est caractérisée par un ensemble de segments se rejoignant dans une même région de l'image [HSV90]. La direction de chaque segment d'une jonction permet d'initialiser la transformation affine  $\mathbf{A}$ . Les niveaux de gris  $n_1$  et  $n_2$  du modèle sont initialisés en réalisant un moyennage des niveaux de gris des points de l'image proche du coin à détecter.

La précision du positionnement des coins obtenu avec cette méthode est de l'ordre de 2 pixels. Ceci est largement suffisant car l'algorithme fonctionne avec des approximations éloignées jusqu'à 5 pixels. Néanmoins, deux coins fournis par cet algorithme peuvent en fait correspondre à deux approximations différentes du même coin (la figure 2.1(b) en montre un exemple). Les positions obtenues à partir de ces deux approximations seront sensiblement les mêmes (à 0.02 pixel près). Afin de traiter ces ambiguïtés, deux positions très proches (moins de 1 pixel) seront fusionnées en un même point en utilisant une moyenne pondérée par la valeur de  $1/\chi^2$  (en effet, une valeur de  $\chi^2$  faible signifie que le modèle est bien ajusté à l'image).

Une autre méthode d'approximation des points a été testée. Elle consiste à approximer les contours de l'image par des fonctions B-splines [MY87]. Les extrémaux de courbures sont alors des coins potentiels. Cette méthode s'est avérée moins efficace car les contours sont quelques fois morcelés, faisant ainsi perdre quelques coins.



**Figure 2.2 :** Niveau de gris de l'image et du modèle (après transformation affine) d'un coin dont l'angle est inférieur à 90 degrés. Le modèle a été obtenu avec un flou gaussien d'écart-type  $\sigma = 1$  (voir texte)

### 2.1.3 Extraction de cibles circulaires

Dans le domaine de la photogrammétrie, la pose de cibles dans une scène est une méthode couramment utilisée afin de calibrer un système de prise d'image ou de réaliser des mesures de dimensions. De nombreuses applications industrielles mettent en pratique ces procédés, par exemple en aéronautique [Bey95], en automobile pour mesurer les effets des impacts dus aux accidents [Bey91], ou dans l'industrie maritime pour la mesure de la qualité de la coque d'un bateau [MK94].

Plusieurs formes de cibles ont été étudiées [BA90, Hav91] : carrée, circulaire, en forme de diamants, ou même de forme plus complexe. La précision d'extraction de chacune de ces cibles a été calculée théoriquement dans le cas où la localisation est réalisée par la méthode des barycentres. Il s'est avéré que les cibles circulaires sont les plus adaptées, offrant une localisation excellente dans les images, indépendante de l'orientation de la caméra [BA90].

La fabrication de cibles circulaires est très diversifiée [Cla94]. Généralement, une cible circulaire est fabriquée à partir d'adhésifs rétro-réfléchissant. Mais d'autres méthodes existent comme la projection d'une lumière laser ou l'utilisation de balles sphériques [JHT93]. Cette dernière méthode est cependant peu utilisée due à la difficulté d'éclairer de manière uniforme chaque sphère, quelque soit la direction de la prise de vue.

De nombreux algorithmes existent pour localiser des cibles circulaires dans les images ([SCS94] compare les performances de six méthodes d'extraction de cibles). La méthode la plus utilisée consiste à prendre comme centre de la cible le barycentre des pixels qui forment la cible sur une image binarisée [BB82]. Cette méthode voit sa précision augmentée lorsque le barycentre est pondéré par les intensités lumineuses des pixels ou par une fonction des intensités lumineuses [WH94]. Les erreurs de localisation dues à cette méthode ont été étudiées. Ces erreurs proviennent à la fois de la discrétisation du

signal en une image, du bruit de la prise d'images qui se traduit par une imprécisions des intensités lumineuses (modification des poids), ainsi que de l'asymétrie de l'image d'une cible provenant du fait que la cible ne peut être parfaitement centrée sur les pixels [Tri89].

La méthode des barycentres peut aussi être appliquée seulement sur les points de contours de la cible. Les résultats obtenus avec cette méthode sont améliorés lorsque des cibles sont emboîtées les unes dans les autres [OBBA90]. Cet algorithme n'est alors applicable que pour la détection de cibles de tailles importantes (plus de 50 pixels de diamètre).

Une seconde méthode agissant sur les points de contours de la cible à localiser utilise la transformée de Hough, qui fournit directement l'équation de l'ellipse dans l'image [IK87, HC95]. Cependant, de telles méthodes ne peuvent s'appliquer que sur des ellipses de tailles conséquentes (quelques dizaines de pixel de diamètre) pour disposer d'une précision de localisation raisonnable.

Des méthodes faisant intervenir de façon plus forte les niveaux de gris de l'image ont aussi été développée. Ainsi, Förtsner [FG87] propose de localiser le centre d'une cible en réalisant l'intersection de toutes les droites perpendiculaires au gradient. Cette intersection est calculée aux moindres carrés, pondérés par le module du gradient au carré.

Gruen procède à la détection de cibles en réalisant une mise en correspondance entre une image de référence et la cible dans l'image [Gru85]. Cette mise en correspondance est réalisée de manière très précise en recherchant la transformation affine entre les 2 images. Elle a été utilisée avec succès dans [Bey91]. Cette méthode est très similaire à celle que nous avons développée. Néanmoins, Gruen utilise un modèle discret car il est réduit à une image de références, et a donc besoin d'interpoler les intensités lumineuses pour les pixels non entiers. De plus, son modèle est statique (les niveaux de gris sont fixés en début d'optimisation), alors que la méthode développée dans cette partie ajuste les deux niveaux de gris qui forment la cible.

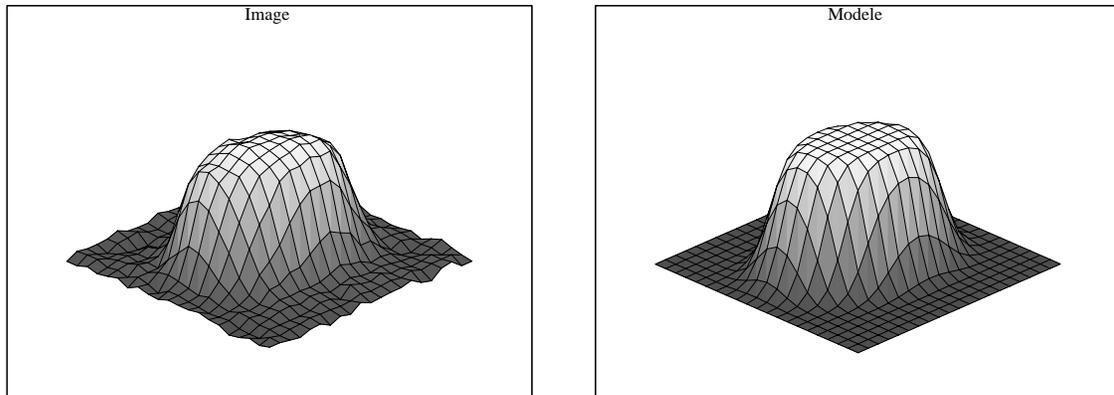
## Modélisation

La cible, dans notre modèle, est centrée à l'origine du repère. Son rayon est égale à  $r$ . Cette valeur est variable pour chaque cible, et est estimée à partir de l'approximation de sa position. Par contre, ce n'est pas un paramètre du modèle, c'est à dire que sa valeur est fixe tout au long des itérations. Le niveau de gris du centre est égal à  $n_c$ , et celui du fond est égal à  $n_f$ . Le dégradé entre ces deux niveau de gris provenant du flou dans l'image est matérialisé par une convolution gaussienne de ces deux niveaux de gris. Le modèle utilisé est donc :

$$I_m(x, y, \mathbf{p}) = g_2 * I'(x, y, \mathbf{p})$$

avec  $I'(x, y, \mathbf{p}) = n_c$  si  $\sqrt{x^2 + y^2} \leq r$ , et  $I'(x) = n_f$  sinon, et  $\mathbf{p} = (n_c, n_f)^\top$ .

Comme le flou gaussien que l'on introduit est le même pour tous les diamètres de la



**Figure 2.3** : Niveau de gris de l'image et dans le modèle d'une cible circulaire. Le modèle a été obtenu avec un flou gaussien d'écart-type  $\sigma = 1$  (voir texte)

cible du fait de sa symétrie, le modèle peut être simplifié en :

$$I_m(x, y, \mathbf{p}) = \begin{cases} n_c & \text{si } \sqrt{x^2 + y^2} \leq r - 3\sigma \\ n_f & \text{si } \sqrt{x^2 + y^2} \geq r + 3\sigma \\ \int_{-\infty}^{+\infty} g_1(\alpha) I''(\sqrt{x^2 + y^2} - \alpha, \mathbf{p}) d\alpha & \text{sinon} \end{cases}$$

où  $g_1(\cdot)$  est la fonction de densité monodimensionnelle de la loi gaussienne,

$$g_1(\alpha) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\alpha^2/(2\sigma^2)}$$

et  $I''(x, \mathbf{p}) = n_c$  si  $|x| \leq r$ , et  $I''(x) = n_f$  sinon.

L'écart-type de cette convolution est fixé pour chaque cible. Il a été montré dans [SCS94] qu'une valeur supérieure à 1.25 n'apportait aucune précision en plus pour les cibles de faibles tailles. L'écart-type de cette convolution est fixé à l'avance pour chaque cible. Il a été montré dans [SCS94] qu'une valeur supérieure à 1.25 n'apportait aucune précision en plus pour les cibles de faibles tailles. Une valeur typiquement utilisée est 1.0.

La figure 2.3(a) montre le graphe des niveaux de gris d'une partie de l'image dans le voisinage d'une cible. La figure (b) montre les niveaux de gris du modèle après optimisation de la transformation affine.

### Initialisation

L'algorithme de Levenberg-Marquardt nécessite, pour chaque cible à localiser, une approximation des paramètres du modèle que l'on veut optimiser. Pour cela, la position  $\mathbf{c}$  approximative du centre de chaque cible est tout d'abord calculée. Cette position a été obtenue après une binarisation de l'image initiale en prenant comme centre de la cible le barycentre des intensités lumineuse de l'image binarisée [BB82, SCS94] :

$$\mathbf{c} = \frac{1}{\sum_{i,j} I_b(i, j)} \begin{pmatrix} \sum_{i,j} i \cdot I_b(i, j) \\ \sum_{i,j} j \cdot I_b(i, j) \end{pmatrix}$$

où  $I_b(i, j)$  est égale à 1 si le niveau de gris du pixel  $(i, j)$  est supérieur à un seuil, et est égale à 0 sinon. Cette méthode a été choisie pour sa simplicité d'implémentation et sa rapidité d'exécution.

L'analyse de l'histogramme des intensités lumineuses de l'image permet de fixer le seuil de la binarisation de l'image [Zho90]. Néanmoins, ces méthodes ne fonctionnent pas sur tous les types d'images. Il faut donc quelques fois imposer un seuil de binarisation.

Une fois le centre de la cible trouvé, on peut trouver 3 points sur le bord de la cible et les affecter à 3 points du modèle. La correspondance de ces 3 points permet de définir l'approximation de la transformation affine  $\mathbf{A}$ .

Le rayon de la cible est égale à la distance qui sépare le centre de la cible du bord de la cible, suivant une direction horizontale. Les niveaux de gris  $n_c$  et  $n_f$  sont respectivement égaux à l'intensité lumineuse de  $\mathbf{c}$  et d'un point qui se trouve en dehors de la cible.

La taille de la fenêtre où l'optimisation est réalisée est adaptative, dépendant de la taille de la cible dans l'image. Le diamètre de la cible dans l'image peut être calculé approximativement. La taille de la fenêtre de corrélation sera 2 fois son diamètre. Ceci permet donc, sans aucune intervention de l'utilisateur, de détecter des cibles de tailles et de luminosité variables.

Le temps de calcul de l'extraction précise d'une cible de cinq pixels de diamètre est de l'ordre de 20ms sur une station SUN SPARC 10.51.

### Tri des bonnes et des mauvaises initialisations

Les points d'approximation des cibles fournis par les barycentres des tâches de l'image binarisée ne correspondent pas toujours à la position d'une cible. Ceci dépend beaucoup du choix du seuil de la binarisation et de la qualité de l'image. Par exemple, les réflexions spéculaires seront considérées comme étant des cibles. Lorsque la configuration des cibles observées est inconnue, il est impossible de s'affranchir de ces points. Par contre, dans de nombreux cas, il est possible de supprimer ces points en utilisant de très faibles contraintes.

La contrainte utilisée est celle de la planarité des cibles. En effet, les ellipses extraites à partir des cibles coplanaires sont semblables, du point de vue de l'orientation du plan ainsi que de la longueur de chacun des deux axes. Les fausses détections peuvent ainsi être éliminées en vérifiant qu'aucune cible voisine est semblable. Cette méthode a été utilisée avec succès sur la mire de la figure 2.10.

### Analyse du biais systématique

Des erreurs systématiques dans l'extraction du centre de l'ellipse apparaissent à cause des cibles elles-mêmes. En effet, la réflexion de la lumière sur les cibles rétro-réfléchissantes dépend de l'orientation de la cible [Zum95]. Ces erreurs ne seront pas estimées dans cette section.

Néanmoins, une erreur systématique plus importante apparaît. En effet, l'algorithme présenté ci-dessus localise le centre d'une ellipse, et non le projeté du centre de la cible.

Ces deux positions sont considérées comme égales, mais ceci n'est en fait qu'une approximation [MZ92]. Il s'avère que la position véritable du projeté du centre de la cible ne peut être connu seulement à partir d'une seule image. Il faut pour localiser sa position des connaissances a priori sur l'orientation de la cible ou disposer une configuration particulière de plusieurs cibles qui permettraient, par construction géométriques, de calculer la position exacte du projeté dans l'image du centre de la cible.

Le but de cette partie est de mesurer la distance entre le centre de l'ellipse et le projeté du centre de la cible. Cette distance varie suivant la taille de la cible et de son orientation par rapport à la caméra. Le calcul effectué passe donc par une simulation, où la cible ainsi que la caméra sont parfaitement connues. Sans perte de généralité, la cible observée est positionnée sur le plan  $(Oxy)$ , centrée à l'origine, et de rayon  $r$ . Son équation est donc

$$\begin{cases} \mathbf{x}^\top \cdot \mathbf{E} \cdot \mathbf{x} = 0 \\ z = 0 \end{cases}$$

avec  $\mathbf{x}^\top = (x \ y \ z \ 1)$  et  $\mathbf{E} = \text{diag}(1, 1, 1, -r^2)$

La matrice perspective caractérisant la caméra est de la forme [HM93, Fau93]

$$\mathbf{M} = \mathbf{I}_c \cdot \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{pmatrix} = (\mathbf{m}_1 \ \mathbf{m}_2 \ \mathbf{m}_3 \ \mathbf{m}_4)$$

où  $\mathbf{I}_c$  est la matrice des paramètres internes,  $\mathbf{R}$  est la matrice de rotation du repère de la caméra par rapport au repère dans lequel est exprimé l'équation de la cible, et  $\mathbf{t}$  est la translation entre ces 2 repères :

$$\mathbf{I}_c = \begin{pmatrix} -k_u f & 0 & u_0 & 0 \\ 0 & k_v f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & -\sin \psi \\ 0 & \sin \psi & \cos \psi \end{pmatrix}$$

$$\mathbf{t}^\top = (t_x \ t_y \ t_z)$$

Ainsi,  $\mathbf{M}$  est paramétrée par

- $f$  la longueur focale de l'objectif, exprimée en mm.
- $k_u$  (resp.  $k_v$ ) le facteur d'échelle vertical (resp. horizontal), en pixel/mm.
- $(u_0 \ v_0)$  est la translation du repère image sur le plan image.
- $\theta$ ,  $\phi$  et  $\psi$  représentent les angles d'Euler de la rotation du repère image par rapport au repère de la scène suivant les axes  $z$ ,  $y$  et  $x$ .
- $(t_x \ t_y \ t_z)$  est la translation entre le repère de la scène et celui de la caméra.

Dans l'image, un point de la cible est observé dans la position  $\mathbf{u} = \mathbf{M} \cdot \mathbf{x}$ , exprimée en coordonnées homogènes. Comme les points de la cible se trouvent dans le plan  $(Oxy)$ , la relation se simplifie pour ces points en

$$\begin{aligned} \mathbf{u} &= \mathbf{M}' \cdot \mathbf{x}' \quad \text{avec} \quad \mathbf{M}' = (\mathbf{m}_1 \ \mathbf{m}_2 \ \mathbf{m}_4) \quad \text{et} \quad \mathbf{x}'^\top = (x \ y \ 1) \\ \text{et} \quad \mathbf{x}'^\top \cdot \mathbf{E}' \cdot \mathbf{x}' &= 0 \quad \text{avec} \quad \mathbf{E}' = \text{diag}(1, 1, -r^2) \end{aligned}$$

La matrice  $\mathbf{K}$  caractérisant l'ellipse observée dans l'image vérifie la relation

$$\mathbf{u}^\top \cdot \mathbf{K} \cdot \mathbf{u} = 0 \iff \mathbf{x}'^\top \cdot \mathbf{M}'^\top \cdot \mathbf{K} \cdot \mathbf{M}' \cdot \mathbf{x}' = 0$$

Ainsi,  $\mathbf{M}'^\top \cdot \mathbf{K} \cdot \mathbf{M}' = \mathbf{E}'$ . La matrice  $\mathbf{M}'$  correspond à une transformation perspective, de centre  $\mathbf{c}$ , du plan  $P_1$  qui comprend la cible sur le plan image. Cette transformation est singulière lorsque le centre de projection  $\mathbf{c}$  se trouve sur  $P_1$ . Si on suppose que l'on ne se trouve pas dans une telle configuration (c'est à dire si le centre de projection de la caméra ne se trouve pas sur le plan  $z = 0$  qui supporte la cible 3D observée, ou encore que l'ellipse observée n'est pas réduite à un segment), on obtient

$$\mathbf{K} = \mathbf{M}'^{-1\top} \cdot \mathbf{E}' \cdot \mathbf{M}'^{-1} = (k_{i,j})_{i,j=1\dots 3}$$

On peut déduire à partir de  $\mathbf{K}$  la position  $(x_e, y_e)$  du centre de l'ellipse dans l'image [Cha90]:

$$\begin{aligned} x_e &= (k_{2,2}k_{1,3} - k_{1,2}k_{2,3}) / (k_{1,2}^2 - k_{1,1}k_{2,2}) \\ y_e &= (k_{1,1}k_{2,3} - k_{1,2}k_{1,3}) / (k_{1,2}^2 - k_{1,1}k_{2,2}) \end{aligned}$$

Par ailleurs, la position  $(x_c, y_c)$  du projeté du centre de la cible est

$$\begin{pmatrix} sx_c \\ sy_c \\ s \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

L'erreur commise en confondant le centre de l'ellipse avec la position du projeté du centre de la cible est donc  $\varepsilon = \sqrt{(x_e - x_c)^2 + (y_e - y_c)^2}$ .

Les graphes de la figure 2.4 montrent le comportement de  $\varepsilon$  en fonction des divers paramètres. Dans chaque graphe, seuls 2 paramètres varient alors que les autres sont fixés. D'une façon générale, nous avons fixé  $u_0$  et  $v_0$  à 256 et  $\theta$  à 0. Ces 3 paramètres ont en effet aucune influence sur la valeur de  $\varepsilon$ . On peut montrer grâce à MAPLE que

$$\begin{aligned} \frac{\partial \varepsilon}{\partial u_0} &= 0 \quad \text{et} \quad \frac{\partial \varepsilon}{\partial v_0} = 0 \\ \frac{\partial \varepsilon}{\partial \theta} &= 0 \end{aligned}$$

La taille des pixels sera de même fixée à  $10\mu\text{m}$ , soit  $k_u = k_v = 100$  pixels/mm. La valeur de la longueur focale  $f$ , lorsque ce paramètre est fixé, est égale à 15mm, ce qui

correspond à un objectif standard. Dans ces conditions, lorsque la cible est disposée à 1m de la caméra et que son rayon est égal à 8mm, la taille de la cible dans l'image est de 24 pixels de diamètre.

La figure 2.4(a) nous montre que l'angle sous lequel est vu la cible a une très grande influence sur la précision du positionnement. En effet, lorsque la cible se trouve sur le plan fronto-parallèle,  $\varepsilon = 0$ . Par contre, lorsque le plan image fait un angle de  $\pi/4$  avec le plan de la cible 3D, l'erreur  $\varepsilon$  approche le 1/20ème de pixel. Le graphe (b) met en valeur le fait que  $\varepsilon$  croît de façon linéaire avec la longueur focale. Le graphe (c) montre que  $\varepsilon$  diminue avec l'éloignement de la cible et que  $\varepsilon$  est aussi sensible au fait que la cible est au centre ou non de l'image. Enfin, le dernier graphe montre que l'erreur en positionnement diminue avec le rayon de la cible.

Ces résultats mettent en évidence qu'il faut être très prudent avec les positions de cibles obtenues. En effet, un biais de 0.05 pixel est du même ordre de grandeur que la précision de la localisation du centre de l'ellipse dans l'image que cette méthode permet d'atteindre. Il n'est donc plus négligeable du tout. Différentes méthodes de correction de ce biais existent suivant la connaissance de la scène observée. Lorsque des cibles appartenant toutes à un même plan se trouvent dans des positions particulières (alignées et équidistantes par exemple), des constructions géométriques permettent de retrouver exactement la projection dans l'image du centre de la cible [CFW95]. Lorsque les paramètres intrinsèques de la caméra sont connus, et si l'orientation du plan sur lequel se trouve la cible est connue, alors la position exacte du projeté dans l'image du centre de la cible peut également être déduite.

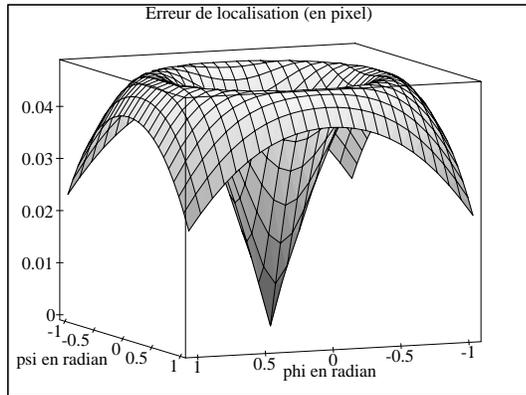
Par la suite, ce biais ne sera pas corrigé car des cibles de seulement 2.5mm de rayon seront utilisées. De plus, la longueur focale de l'objectif utilisé est légèrement plus courte, étant de 12.5mm. La taille dans les images de ces cibles sera de l'ordre de 5 à 6 pixels de diamètre. Toutes ces conditions confondues amènent à un biais majoré par 0.004 pixel, considéré comme négligeable.

## 2.2 Validation de la localisation des coins dans une image

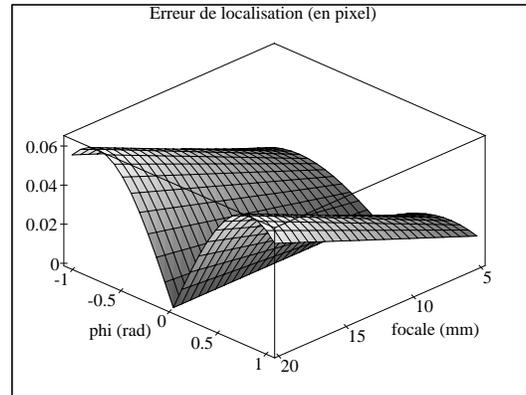
Le principe général de la localisation de points modélisables a été étudié dans la section précédente. Deux cas particuliers ont été détaillés : celui des coins et celui des cibles circulaires. Cette section s'attache à déterminer la précision de localisation des coins dans l'image. La section suivante sera dédiée à la localisation des cibles.

Afin de démontrer la qualité de la localisation des coins dans les images, une série de tests a été effectuée. Chaque test a été réalisé sur les mêmes images. Ces images sont constituées d'une mire plane contenant des carrés noirs sur fond blanc. La figure 2.5 montre des images de la mire de calibrage que nous avons utilisée. Cette mire contient 92 coins. Depuis chaque point de vue, trois positions différentes du plan de la mire (appelées plan 1, plan 2 et plan 3 dans cette section) sont photographiées (figure 2.6). Cette translation a été effectuée à l'aide d'une table millimétrique, ce qui permet la reproductibilité des translations effectuées.

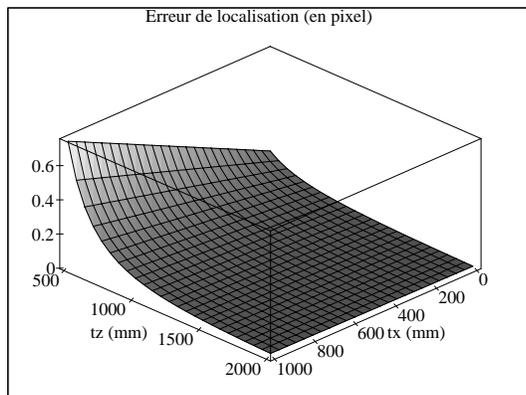
Les prises d'images ont été effectuées pour trois positions différentes de la caméra. Au total, les données obtenues comportent la mise en correspondance de 276 coins entre



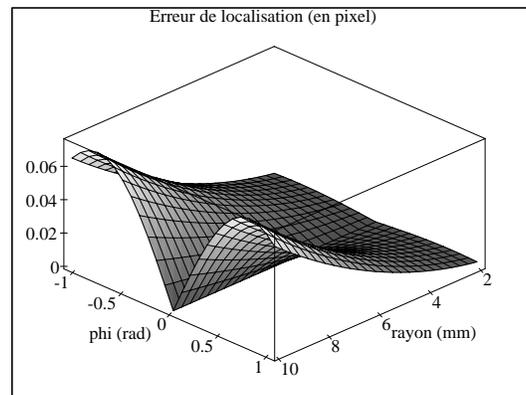
(a)  $\theta = 0$ ,  $t_x = 0\text{mm}$ ,  $t_y = 0\text{mm}$ ,  $t_z = 1000\text{mm}$ ,  $f = 15\text{mm}$ ,  $k_u = 100$ ,  $k_v = 100$ ,  $r = 8\text{mm}$ ,  $u_0 = 256$ ,  $v_0 = 256$ ,  $\phi = -\pi/3 \dots \pi/3$ ,  $\psi = -\pi/3 \dots \pi/3$ . Diamètre maximum de l'ellipse dans l'image : 24 pixels



(b)  $\theta = 0$ ,  $\psi = 0$ ,  $t_x = 0\text{mm}$ ,  $t_y = 0\text{mm}$ ,  $t_z = 1000\text{mm}$ ,  $k_u = 100$ ,  $k_v = 100$ ,  $r = 8\text{mm}$ ,  $u_0 = 256$ ,  $v_0 = 256$ ,  $f = 5 \dots 20\text{mm}$ ,  $\phi = -\pi/3 \dots \pi/3$



(c)  $\theta = 0$ ,  $\phi = \pi/3$ ,  $\psi = 0$ ,  $t_y = 0\text{mm}$ ,  $f = 15\text{mm}$ ,  $k_u = 100$ ,  $k_v = 100$ ,  $r = 8\text{mm}$ ,  $u_0 = 256$ ,  $v_0 = 256$ ,  $t_x = 0 \dots 1000\text{mm}$ ,  $t_z = 500 \dots 2000\text{mm}$



(d)  $\theta = 0$ ,  $\psi = 0$ ,  $t_x = 0\text{mm}$ ,  $t_y = 0\text{mm}$ ,  $t_z = 1000\text{mm}$ ,  $f = 15\text{mm}$ ,  $k_u = 100$ ,  $k_v = 100$ ,  $u_0 = 256$ ,  $v_0 = 256$ ,  $r = 2 \dots 10$ ,  $\phi = -\pi/3 \dots \pi/3$

**Figure 2.4** : Distance, en pixel, entre le centre de l'ellipse observée et la position du projeté du centre de la cible 3D.

3 vues. Les coordonnées homogènes du  $j$ ème coin de la  $i$ ème image seront dénotées par le vecteur  $\mathbf{p}_j^{(i)}$ . Les coins observés forment un volume dans la scène pour chaque vue grâce à la translation en trois positions du plan de la mire. Le volume créé est de taille  $19.8 \times 12.6 \times 6.0 \text{ cm}^3$ . La figure 2.5 montre le premier plan de la mire pour les trois points de vue.

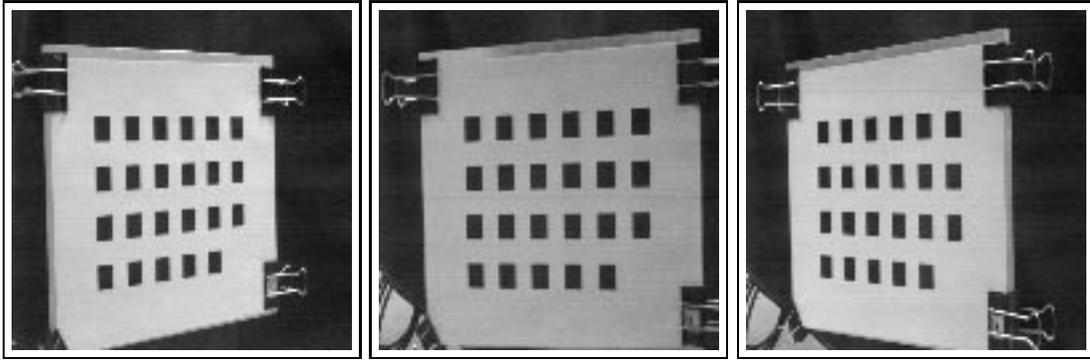


Figure 2.5 : Les 3 vues du premier plan de la mire de calibrage

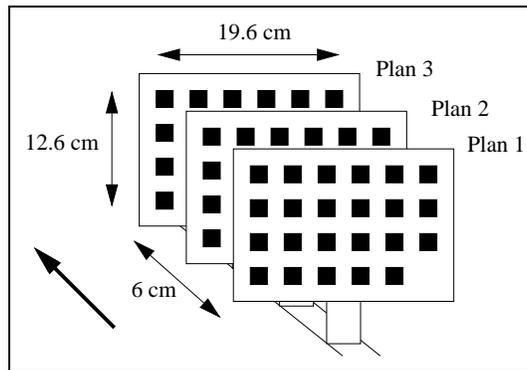


Figure 2.6 : Translation de la mire afin d'obtenir un volume à partir de cette mire plane

La qualité de la localisation des coins est calculée en vérifiant plusieurs critères. La première expérimentation concerne la préservation des alignements de coins dans les images. La seconde expérimentation reconstruit ce volume à l'aide des mises en correspondance sur les 3 vues. La qualité d'extraction des coins peut alors être déduite de la qualité de la reconstruction obtenue. La même expérience est effectuée avec le calcul de la géométrie épipolaire entre deux vues. La dernière expérimentation concerne l'étude de la stabilité des birapports calculés avec les coins.

### 2.2.1 Alignement

Cette mire contenant de nombreux coins alignés, il est vérifié que ces alignements sont préservés dans les images (modèle sténopé pour la caméra). Ainsi, pour chaque ensemble  $\{(x_i, y_i)\}$  de points dans une image correspondant à des points tridimensionnels alignés, une droite est ajustée au sens des moindres carrés. L'équation de chaque droite est de la forme  $ax + by + c = 0$ , avec  $a^2 + b^2 = 1$  comme contrainte.

La moyenne  $\bar{d}$  et l'écart-type  $\sigma$  des distances entre un point dans l'image et la droite ajustée ont été calculés

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n |ax_i + by_i + c|$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (|ax_i + by_i + c| - \bar{d})^2}$$

Ces statistiques ont été calculées sur des alignements horizontaux et verticaux. Le tableau 2.1 montre les résultats obtenus, notamment  $\bar{d} \approx 0.08$  pixels. En supposant que l'extraction des coins est réalisée sans aucun biais (i.e. l'erreur commise sur la localisation des coins est indépendante pour chaque coin), alors la précision en localisation de ces coins est de l'ordre de  $\sqrt{2} \cdot \bar{d} = 0.11$  pixel.

	moyenne des distances $\bar{d}$	écart-type $\sigma$
Vertical	0.07	0.056
Horizontal	0.08	0.069

**Table 2.1** : Qualité des alignements de points

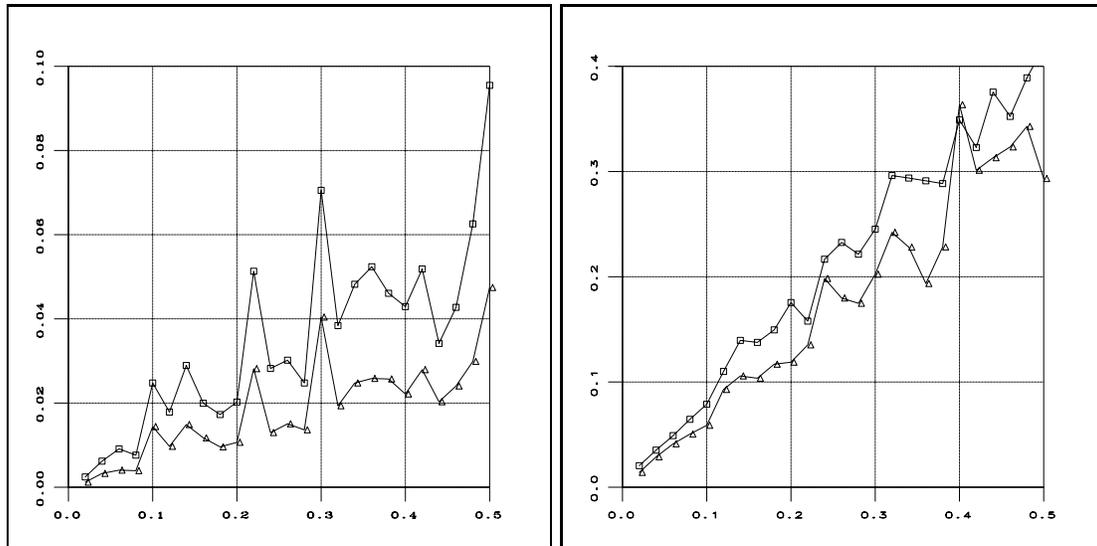
### 2.2.2 Précision de la reconstruction tridimensionnelle

La précision de la reconstruction tridimensionnelle de coins fournit aussi la précision à laquelle les points ont été extraits dans les images. La méthode classique de reconstruction de scène consiste en un étalonnage des caméras suivi par une triangulation. Le calibrage utilise toutes les positions connues des points dans la scène afin d'être stable. La triangulation est alors appliquée pour chaque point observé dans plusieurs images en utilisant les matrices de perspectives calculées dans la phase de calibrage. Le problème majeur de cette méthode est le fait que l'on utilise les données 3D des coins pour le calibrage, et à partir de ces résultats nous les comparons avec une reconstruction. La reconstruction est donc biaisée.

C'est pourquoi la méthode de Boufama *et al.* a été appliquée [BMV93, Bou94]. Cette méthode reconstruit une scène en connaissant les correspondances de points 2D dans les trois images de la figure 2.5, auxquelles s'ajoutent la connaissance des coordonnées euclidiennes de 5 points dans la scène.

La qualité de la reconstruction obtenue avec cette méthode a été calculée en comparant la distance entre les points reconstruits et leur position théorique connue du fait que la configuration de la mire de calibrage est parfaitement connue. La moyenne des distances entre ces 2 positions est de 0.020cm avec un écart-type de 0.010cm. Ainsi, la reconstruction obtenue est précise au 1/1000<sup>ème</sup> de la taille de la scène.

La précision d'extraction des points dans l'image est obtenue en simulant une reconstruction avec des données synthétiques, pour différentes valeurs d'un bruit blanc gaussien. Les données synthétiques ont été obtenues de façon à être dans une configuration la plus proche possible de celle de l'expérimentation réelle. Ainsi, les points de la mire ont été projetés synthétiquement suivant les matrices perspectives obtenues après calibrage. La méthode de reconstruction a alors été appliquée à chaque point projeté dans l'image. Diverses reconstructions ont été obtenues suivant la valeur de l'écart-type du bruit blanc gaussien qui a été ajouté à chaque point dans les images.



(a) Précision de la reconstruction 3D

(b) Précision de la géométrie épipolaire

**Figure 2.7 :** Calcul, pour des données synthétiques bruitées, de la précision de la reconstruction et de la géométrie épipolaire. Pour différentes valeurs de l'écart-type du bruit blanc gaussien ajouté aux données exactes (en abscisses), valeur de la moyenne (carrés) et de l'écart-type (triangles) de la précision obtenue (en cm pour le graphe (a), et en pixel pour le graphe (b))

Des statistiques ont été faites sur la précision de reconstruction obtenue dans un tel cas. Les résultats, pour un bruit gaussien d'écart-type variant de 0.02 à 0.50, se trouvent figure 2.7(a).

Il est à noter que ce graphe est irrégulier du fait que seule la connaissance de la position dans la scène de 5 points est nécessaire. Cette méthode est donc très sensible à la qualité d'extraction de ces cinq points.

Ce graphe montre que pour une valeur du bruit gaussien d'écart-type variant de 0.1 à 0.2 pixel, nous obtenons, en moyenne et en écart-type, des valeurs identiques aux valeurs calculées dans le cas réel (moyenne  $\approx 0.02$ cm et écart-type  $\approx 0.01$ cm).

### 2.2.3 Précision du calcul de la géométrie épipolaire

Ce troisième test met en avant la précision de la géométrie épipolaire calculée avec de tels indices. La géométrie épipolaire est équivalente à un calibrage projectif des caméras [Fau92, Har92]. Elle représente la relation linéaire qui existe entre 2 images. Elle est représentée par la matrice fondamentale  $\mathbf{F}$ , matrice de taille  $3 \times 3$  et de rang 2. Cette matrice relie 2 points  $\mathbf{u}_1$  et  $\mathbf{u}_2$ , exprimés en coordonnées homogènes, correspondant dans 2 images à un même point de la scène, par la relation  $\mathbf{u}_2^\top \cdot \mathbf{F} \cdot \mathbf{u}_1 = 0$ . Cela signifie que le point  $\mathbf{u}_2$  se trouve sur la ligne épipolaire  $\mathbf{F} \cdot \mathbf{u}_1$ .

L'estimation de la géométrie épipolaire peut être réalisée connaissant au minimum

8 points mis en correspondance entre 2 images [LH81, FLM92]. Néanmoins, l'utilisation d'un nombre plus important de points permet d'obtenir une meilleure stabilité du calcul de la géométrie épipolaire. Ainsi, tous les coins de la mire appartenant aux plans 1 et 3 (cf figure 2.6) ont été utilisés pour calculer la géométrie épipolaire.

La matrice fondamentale  $\mathbf{F}$  reliant les vues 1 et 3 a été calculée. La moyenne  $\bar{d}$  et l'écart-type  $\sigma$  de la distance d'un point du second plan de la vue 3  $\mathbf{p}_j^{(3)}$  à la droite épipolaire associée  $\mathbf{F} \cdot \mathbf{p}_j^{(1)}$  quantifient la qualité de la géométrie épipolaire obtenue :

$$\bar{d} = \frac{1}{n} \sum_{\mathbf{p}_j \in \text{plan 2}} \text{dist} \left( \mathbf{p}_j^{(3)}, \mathbf{F} \cdot \mathbf{p}_j^{(1)} \right)$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{\mathbf{p}_j \in \text{plan 2}} \left( \text{dist} \left( \mathbf{p}_j^{(3)}, \mathbf{F} \cdot \mathbf{p}_j^{(1)} \right) - \bar{d} \right)^2}$$

Cette démarche fournit des résultats non biaisés. En effet, la qualité de la géométrie épipolaire est évaluée à partir de points qui n'ont pas été pris en compte dans le calcul de la matrice fondamentale  $\mathbf{F}$ . Dans ce test particulier,  $\bar{d} = 0.10$  pixel et  $\sigma = 0.088$ . L'erreur d'extraction des coins ne peut être déduite directement de ces statistiques car la matrice fondamentale a été obtenue par une optimisation d'un système non-linéaire

$$\chi^2 = \sum_{\mathbf{p}_j \in \text{plan 1 ou 3}} \left( \text{dist}^2 \left( \mathbf{p}_j^{(3)}, \mathbf{F} \cdot \mathbf{p}_j^{(1)} \right) + \text{dist}^2 \left( \mathbf{p}_j^{(1)}, \mathbf{F}^\top \cdot \mathbf{p}_j^{(3)} \right) \right)$$

en préservant la contrainte  $\text{rang}(\mathbf{F}) = 2$ . Le comportement d'une telle optimisation peut être instable lorsque les données sont bruitées.

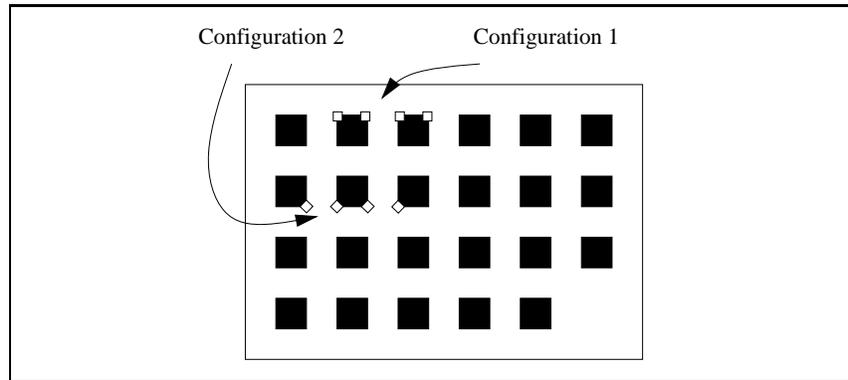
Une simulation est donc nécessaire pour connaître la précision d'extraction des coins dans l'image. La simulation a suivi exactement le même processus que celui de la section précédente. Les résultats de ces simulations sont illustrés dans le graphe de la figure 2.7(b), avec la courbe de la moyenne et de l'écart-type de la distance d'un point à sa droite associée, pour différents valeurs de l'écart-type du bruit gaussien.

Ce graphe est bien plus stable que celui de la figure 2.7(a) car de nombreux points interviennent dans le calcul de la géométrie épipolaire. Les valeurs des statistiques dans le cas réel et dans le cas simulé sont similaires lorsque la valeur du bruit gaussien est de 0.12 pixels ce qui est cohérent avec les deux expérimentations précédentes.

La qualité de la géométrie épipolaire reliant les deux premières vues et de celle reliant les vues 2 et 3 ont également été évaluées. Les précisions obtenues sont :

$$\begin{array}{ll} \text{vues 1 \& 2 :} & \text{moyenne } \bar{d} = 0.13 \text{ pixel} \quad \text{écart-type } \sigma = 0.102 \\ \text{vues 2 \& 3 :} & \text{moyenne } \bar{d} = 0.13 \text{ pixel} \quad \text{écart-type } \sigma = 0.103 \end{array}$$

Ces résultats sont un peu moins bons (à 0.03 pixels près) que ceux obtenus en utilisant les vues 1 et 3. Ceci est dû au fait que le calcul de la matrice fondamentale est d'autant plus stable que les vues sont éloignées. Des simulations identiques réalisées entre les vues 1 et 2, puis entre les vues 2 et 3 ont confirmé les résultats obtenus avec les vues 1 et 3 : la précision de la localisation des coins est de l'ordre de 0.12 pixels.



**Figure 2.8** : Les deux configurations de birappports utilisées pour mettre en évidence le fait que l'erreur commise sur l'extraction des coins est non biaisée

### 2.2.4 Stabilité des birappports

Dans cette sous-section, la qualité des extractions de coins est étudiée en analysant la qualité et la stabilité des birappports calculés avec de tels indices. Cette analyse fournit deux résultats. Tout d'abord, il est mis en évidence que l'extraction de coins n'est pas biaisée, c'est à dire qu'il n'y a pas d'erreur systématique dans leur extraction. Enfin, les résultats précédents sur la précision de leur localisation sont confirmés.

#### Extraction non biaisée

Le birappport de 4 points alignés, ayant comme coordonnées  $a$ ,  $b$ ,  $c$  et  $d$  est égal à

$$K(a, b, c, d) = \frac{a - c}{a - d} \div \frac{b - c}{b - d}$$

Cette valeur est un invariant projectif. Ainsi, elle reste la même quelque soit l'image dans laquelle 4 points alignés de la scène sont projetés. De nombreuses configurations de points ont un birappport égal à  $K(0, 3, 2, 1) = 4$  dans la mire utilisée. Ces birappports correspondent à des points de la mire alignés et équidistants. Afin de montrer que l'extraction des coins est non-biaisé, la stabilité des birappports a été étudiée pour deux configurations de points différentes (cf figure 2.8). La première (resp. seconde) configuration comprend des quadruplets de coins appartenant à deux (resp. trois) carrés différents de la mire. Ainsi, si l'extraction d'un coin contient une erreur systématique due à l'orientation de celui-ci, les distributions des valeurs des birappports ne seront pas les mêmes.

Pour chaque configuration, la moyenne et l'écart-type des valeurs des birappports ont été calculés (cf tableau 2.2). Les distributions obtenues expérimentalement sont identiques quelque soit la configuration choisie. Ainsi, nous pouvons en déduire que l'extraction de coins n'est pas biaisée.

#### Précision à l'aide de birappports

La stabilité du calcul de birappport est fortement liée à la précision d'extraction des coins. Mais il faut s'assurer en premier lieu que l'erreur commise sur le calcul des birap-

ports suit une loi de probabilité normale. Pour le montrer, la moyenne et l'écart-type des valeurs des birapports ont été calculés pour chaque plan de chaque vue. Les résultats de ces statistiques sont donnés dans le tableau 2.3. Chaque distribution est comparée avec une distribution normale à l'aide du test de Kolmogorov-Smirnov [PFTW88]. Étant données deux distributions, ce test fournit la probabilité que ces deux distributions sont issues de la même loi. Les tests réalisés indiquent que la probabilité que la distribution des birapports suit une loi normale est toujours de plus de 50%, et est égale à 90% dans la majorité des cas. La figure 2.9 montre la distribution pour un plan des valeurs des birapports superposée avec la fonction de densité de la loi normale correspondante. Considérer que les birapports calculés suivent une loi normale est donc une hypothèse totalement justifiée.

La fonction  $K$  calculant le birapport peut se linéariser localement autour de la configuration de  $K(0, 3, 2, 1)$ . La relation entre l'écart-type des valeurs de birapports et celui de l'erreur d'extraction des coins peut se déduire du développement de Taylor de  $K$  au premier degré :

$$\sigma_K^2 = \mathbf{J}_K \cdot \mathbf{\Lambda} \cdot \mathbf{J}_K^\top$$

où

- $\sigma_K$  est l'écart-type des valeurs du birapport  $K$ . Pour chaque plan, cette valeur est calculée en considérant tous les birapport  $K_i$  obtenus à partir de 4 coins alignés et consécutifs :

$$\sigma_K^2 = \frac{1}{n-1} \sum_{i=1}^n (K_i - \bar{K})^2 \quad \text{avec} \quad \bar{K} = \frac{1}{n} \sum_{i=1}^n K_i$$

- $\mathbf{J}_K$  est le jacobien de  $K$  :

$$\mathbf{J}_K = \begin{pmatrix} \frac{(b-d)(c-d)}{(a-d)^2(b-c)} & \frac{(a-c)(d-c)}{(a-d)(b-c)^2} & \frac{(b-d)(a-c)}{(a-d)(b-c)^2} & \frac{(a-c)(b-a)}{(a-d)^2(b-c)} \end{pmatrix}$$

$$\text{donc } \mathbf{J}_K(\lambda a, \lambda b, \lambda c, \lambda d) = \frac{1}{\lambda} \mathbf{J}_K(a, b, c, d) \quad \text{et} \quad \|\mathbf{J}_K(0, 3, 2, 1)\|^2 = 80$$

- $\mathbf{\Lambda}$  est la matrice de covariance des 4 points utilisés pour calculer le birapport  $K$ . Comme tous les coins sont extraits sans erreurs systématiques,  $\mathbf{\Lambda} = \sigma^2 \cdot \mathbf{I}_4$ , où  $\sigma$  représente l'écart-type des erreurs d'extraction de coins, et  $\mathbf{I}_4$  est la matrice identité  $4 \times 4$ .

Ainsi, pour des birapports calculés avec 4 points alignés et équidistants

$$\sigma_K^2 = \sigma^2 \cdot \|\mathbf{J}_K\|^2 \quad \implies \quad \sigma = \sigma_K \frac{\lambda}{\sqrt{80}}$$

où  $\lambda$  est la distance dans l'image entre 2 coins. Cette distance est sensiblement constante pour des coins appartenant à la même image car la déformation perspective de nos images est faible.

Le tableau 2.3 montre les résultats obtenus pour chaque plan de chaque vue. La précision atteinte approche donc 1/10ème de pixel. La précision globale, obtenue avec les 486 birapports de la table 2.3 calculés avec quatre points alignés horizontalement, est égale

à 0.12 pixel. En utilisant des birapports calculés avec des points alignés verticalement, la précision obtenue est alors de 0.18 pixel. Le rapport de ces deux précisions est égal à 1.50. Cette valeur est très proche du rapport des facteurs d'échelles de la taille des pixels de la caméra PULNIX-6EX, qui vaut 1.47 (cette valeur a été obtenue par un étalonnage de la caméra).

### 2.2.5 Conclusions

Quatre expérimentations ont été effectuées afin de quantifier la qualité de localisation des points dans les images. Les expérimentations ont vérifié les alignements de coins dans les images, la qualité de la reconstruction tridimensionnelle de coins, la qualité de la géométrie épipolaire, et la stabilité des birapports calculés sur les coins. Les résultats obtenus avec ces expériences sont cohérents. Ils permettent d'affirmer que la précision de la localisation de coins approche le 1/10ème de pixel.

Il faut cependant modérer ces résultats. Ils ont été obtenus avec des coins parfaits dans les images, ne contenant que peu de bruit. De plus, l'extraction de coins devient plus instable lorsque l'angle formant le coin est très aigu ou très obtus. Dans un tel cas, une petite erreur réalisée sur l'orientation de chaque arête formant le coin induit une grande erreur sur la position du coin dans l'image. Ainsi, la précision du 1/10ème de pixel ne peut être atteinte que dans des cas favorables.

## 2.3 Validation de l'extraction des cibles

La deuxième sorte de points modélisables étudiés concerne de la détection du centre de cibles circulaires dans les images. Le but de cette section est la validation de la méthode pour cette entité, ainsi que la quantification des précisions d'extraction atteignables.

Différents tests impliquant le calcul de la géométrie épipolaire, celui de la reconstruction tridimensionnelle de points et la stabilité des birapports permettent de connaître la précision d'extraction des cibles. Chaque critère a été réalisé en utilisant les mêmes images. Les tests ont été effectués avec quatre objectifs différents, qui sont par ordre de qualité décroissante : KINOPTIK 12.5mm, COSMICAR "television lens" de longueur focale 12.5mm, CANON "TV zoom lens J6x12" utilisé avec une longueur focale de l'ordre de 20mm, et enfin un objectif KISANON "ultra wide angle TV lens" de longueur focale 6.5mm.

Ainsi, le gain de précision obtenu lorsque les distorsions sont prises en compte pourra être quantifié en utilisant des méthodes totalement indépendantes du calcul des distorsions.

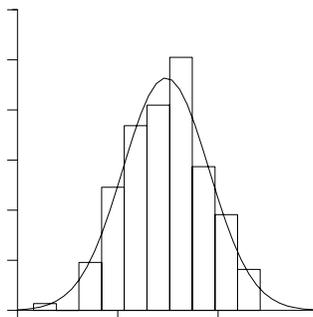
Les images prises sont celles d'une mire contenant 160 cibles circulaires rétro-réfléchissantes. La mire est composée de trois plans. Son volume est de  $33.7 \times 26.2 \times 26.2 \text{ cm}^3$ . La position tridimensionnelle de chaque cible est connue avec une précision de  $20 \mu\text{m}$  dans des conditions standards de température. Pour chaque objectif, deux images de la mire sont prises. La figure 2.10 montre les deux images prise avec l'objectif KINOPTIK 12.5mm.

Image	Numéro de la configuration	Nombre de birapports calculés	Moyenne	Écart-type	Erreur maximale
Image 1	1	90	4.031	0.050	0.145
	2	72	3.971	0.046	0.133
Image 2	1	90	3.990	0.041	0.134
	2	72	4.009	0.039	0.092
Image 3	1	90	4.022	0.053	0.157
	2	72	3.982	0.052	0.157

**Table 2.2 :** Stabilité, pour chaque image, de la valeur du birapport calculés avec 4 coins alignés placés sur 2 carrés consécutif de la mire (configuration 1), ou sur 3 carrés consécutifs (configuration 2) ; voir le texte page 47

Numéro de vue	Plan	Moyenne $\bar{K}$	Écart-type $\sigma_K$	Distance $\lambda$ dans l'image	Précision $\sigma$	Probabilité loi gaussienne
Image 1	1	4.004	0.055	20.12	0.12	0.891
	2	4.005	0.053	19.74	0.12	0.964
	3	4.006	0.064	19.36	0.14	0.997
Image 2	1	3.998	0.043	25.10	0.12	0.818
	2	3.999	0.041	24.16	0.11	0.697
	3	3.997	0.040	23.30	0.10	0.919
Image 3	1	4.003	0.042	19.40	0.09	0.935
	2	4.005	0.066	19.11	0.14	0.551
	3	4.004	0.058	18.83	0.12	0.743

**Table 2.3 :** Statistique sur la distribution des valeurs des birapports obtenus dans chaque plan pour chaque image. La valeur théorique est égale à 4. Chaque valeur a été calculée en utilisant 54 birapports



**Figure 2.9 :** Distribution normalisée des valeurs des birapports dans la seconde vue, superposée avec la fonction de densité de probabilité gaussienne correspondant aux valeurs de la moyenne  $\bar{K}$  et de l'écart-type  $\sigma$  obtenues expérimentalement

Dans la suite de cette section, le centre de la  $j$ ème cible aura les coordonnées homogènes  $\mathbf{u}_{i,j} = (u_{i,j}, v_{i,j}, 1)^\top$  dans la  $i$ ème image. Les coordonnées tridimensionnelles dans la scène de la  $j$ ème cible seront notées  $\mathbf{x}_j = (x_j, y_j, z_j)^\top$ . Ces coordonnées sont fournies par le fabriquant de la mire. La mise en correspondance entre les cibles est réalisée de façon interactive (pour chaque plan de la mire, tous ses points sont mis en correspondance connaissant les correspondants de trois d'entre eux). La position  $\mathbf{u}_{i,j}$  de chaque cible dans l'image sera alors corrigée afin de prendre en compte les distorsions du système d'acquisition. Après correction des distorsions, la nouvelle coordonnée de chaque point sera notée par  $\mathbf{u}_{i,j}^d$  (cf chapitre 1).

### 2.3.1 Qualité de la géométrie épipolaire

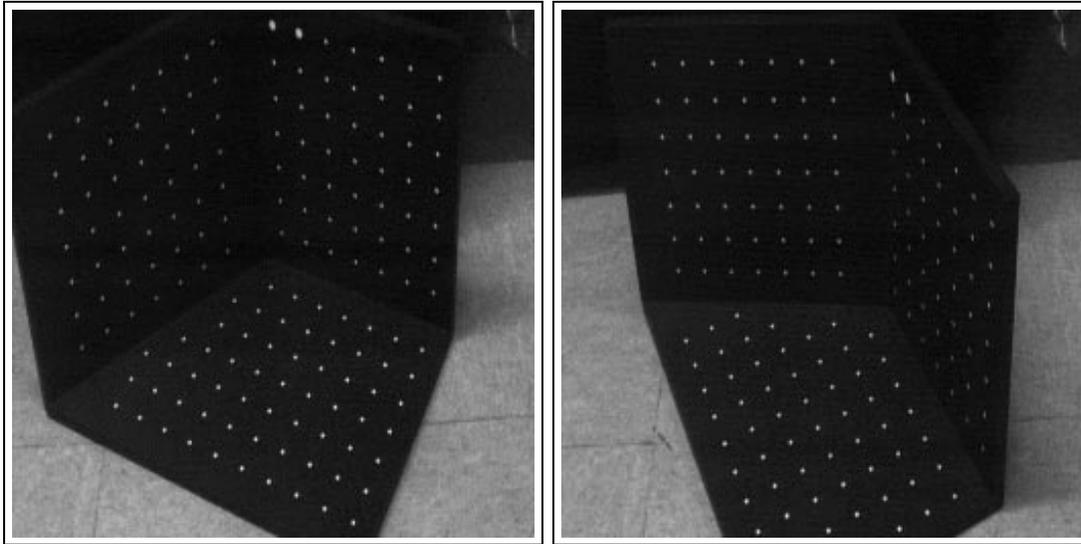
Pour chaque couple d'images, la géométrie épipolaire les reliant a été estimée. Deux estimations  $\mathbf{F}$  et  $\mathbf{F}^d$  de la matrice fondamentale ont été calculées. La matrice  $\mathbf{F}$  a été obtenue avec les cibles d'indice impair  $\mathbf{u}_{i,2j+1}$  dont la position n'a pas été corrigée. La seconde matrice  $\mathbf{F}^d$  a été calculée après la correction des distorsions avec les points  $\mathbf{u}_{i,2j+1}^d$ .

La qualité de  $\mathbf{F}$  (resp.  $\mathbf{F}^d$ ) est estimée à partir de la moyenne, de l'écart-type, du médian et de l'écart-type robuste de la distance entre chaque point d'indice pair de la seconde image et sa ligne épipolaire associée  $dist(\mathbf{u}_{2,2j}, \mathbf{F} \cdot \mathbf{u}_{1,2j})$  (resp.  $dist(\mathbf{u}_{2,2j}^d, \mathbf{F}^d \cdot \mathbf{u}_{1,2j}^d)$ ). Les résultats de ces statistiques sont donnés dans le tableau 2.4. Cette démarche permet d'obtenir des résultats non biaisés car la qualité de la géométrie épipolaire est validée avec des points qui n'ont pas été utilisés pour le calcul de  $\mathbf{F}$  et  $\mathbf{F}^d$ .

L'objectif KINOPTIK se confirme comme un objectif de bonne qualité. En effet, les matrices  $\mathbf{F}$  et  $\mathbf{F}^d$  sont de qualités semblables, ce qui montre que la correction des distorsions est inutile pour cet objectif. La distance entre un point de la seconde image et sa ligne épipolaire associée est de l'ordre de 1/20ème de pixel.

Par contre, les trois autres objectifs testés sont de moins bonne qualité. En effet, la correction des distorsions apportent un effet bénéfique sur l'estimation de la géométrie épipolaire. Le facteur gagné est même de l'ordre de 5 pour l'objectif grand angle KISANON. On peut aussi remarquer que la qualité de la géométrie épipolaire reste constante quelque soit l'objectif utilisé si la correction des distorsions optiques sur la position des cibles est réalisée, ce qui confirme l'intérêt de la méthode proposée au chapitre 1.

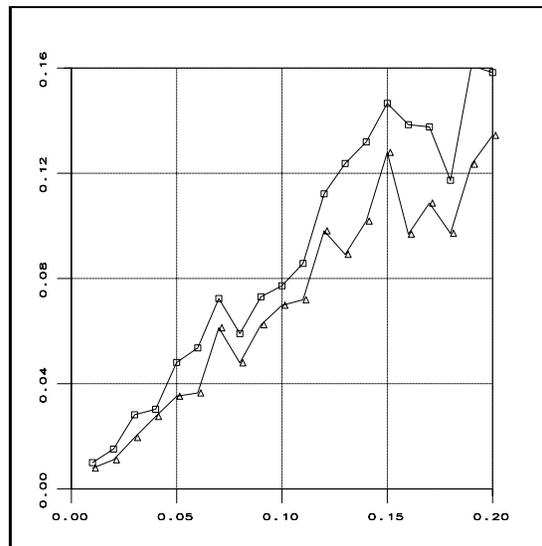
Afin d'obtenir la précision d'extraction des cibles à partir de la qualité de la géométrie épipolaire, les tests synthétiques de la sous-section 2.2.3 ont été réalisés. Les données synthétiques ont été obtenues en projetant les points de la mire suivant les deux matrices perspectives calculées par le calibrage des vues prises avec l'objectif KINOPTIK. La moyenne et l'écart-type des distances entre un point et la ligne épipolaire sont affichés dans le graphe 2.11 pour plusieurs valeurs de l'écart-type du bruit blanc gaussien appliqué aux points projetés synthétiquement. Les résultats des tests synthétiques sont similaires aux résultats obtenus dans le cas réel lorsque le bruit a un écart-type de valeur 0.05. La précision d'extraction des cibles est donc de l'ordre de 1/20ème de pixel.



(a) Première image

(b) Deuxième image

**Figure 2.10 :** Les deux images de la mire contenant des cibles circulaires. Ces images ont été prises avec un objectif KINOPTIK de longueur focale 12.5mm



**Figure 2.11 :** Calcul, pour des données synthétiques bruitées, de la précision de la géométrie épipolaire. Pour différentes valeurs de l'écart-type du bruit blanc gaussien ajouté aux données exactes (en abscisses, de 0.0 à 0.20 pixel), valeur de la moyenne (carrés) et de l'écart-type (triangles) de la précision obtenue. Les ordonnées varient de 0.0 à 0.16 pixel

### 2.3.2 Précision des reconstructions tridimensionnelles

La reconstruction effectuée dans cette sous-section opère par triangulation. Connaissant les coordonnées  $\mathbf{u}_{i,2j+1}$  et  $\mathbf{x}_{2j+1}$  de points d'indice impair, les matrices perspectives  $3 \times 4$   $\mathbf{P}_i$  associée à chaque vue sont calculées en utilisant la méthode de calibrage de Faugeras-Toscani [FT87]. Ensuite, la position de chaque cible d'indice pair est reconstruite par triangulation en les positions  $\mathbf{x}'_{2j} = (x', y', z')^\top$ .

Les erreurs de reconstruction  $dist(\mathbf{x}'_{2j}, \mathbf{x}_{2j})$  ont été calculées, ainsi que les erreurs de reprojection des points reconstruits dans les images. Ces erreurs sont données par les valeurs rms (Root Mean Square) suivant les deux axes de l'image. Chaque point  $\mathbf{x}'_{2j}$  est reprojété dans l'image suivant la matrice perspective  $\mathbf{P}_i$  en un point  $\mathbf{u}'_{i,2j} = (u'_{i,2j}, v'_{i,2j})^\top$ . Les erreurs rms sont alors :

$$\mu x = \sqrt{\frac{1}{2n} \sum_{i=1}^2 \sum_{j=1}^n (u_{i,2j} - u'_{i,2j})^2} \quad \text{et} \quad \mu y = \sqrt{\frac{1}{2n} \sum_{i=1}^2 \sum_{j=1}^n (v_{i,2j} - v'_{i,2j})^2}$$

Les erreurs de reprojection des positions tridimensionnelles des points de la scène, données par le constructeur de la mire, ont été aussi calculées. La moyenne (moy) et l'écart-type ( $\sigma$ ) de ces distances ont été calculées.

Les résultats obtenus pour chaque objectif, avant et après correction des distorsions sont contenus dans le tableau 2.5. Comme lors de l'expérience précédente, l'objectif KINOPTIK s'impose comme un objectif de bonne qualité qui suit très fidèlement le modèle sténopé. La correction des distorsions est inutile avec cet objectif. Par contre, la reconstruction obtenue avec les trois autres objectifs testés est améliorée lorsque les déformations géométriques sont corrigées. Pour l'objectif grand angle KISANON, la précision est améliorée par un facteur 4.

Les corrections appliquées sont homogènes quelque soit la qualité de l'objectif. La moyenne des erreurs en reconstruction 3D varient de 0.12 à 0.16 millimètres, alors que les erreurs de reprojection des points  $\mu x$  varie de 0.028 à 0.035 pixels. La reprojection des coordonnées (fournies par le constructeurs) des points de la mire dans l'image est de l'ordre de 0.06 pixels, ce qui montre la qualité du calibrage effectué ainsi que de l'extraction de points dans les images.

### 2.3.3 Birapports

Les mêmes tests que ceux effectués dans la sous-section 2.2.2 ont été réalisés. Ces tests consistent à calculer des birapports de valeurs 4 dans les images et d'en déduire la précision d'extraction des cibles à partir de la stabilité des birapports.

Pour chaque objectif, la moyenne  $\overline{K}$  et l'écart-type  $\sigma_K$  de la distribution des birapports sont calculés, ainsi que la distance moyenne  $\lambda$  entre deux cibles dans l'image, le tout avant et après correction des distorsions. De ces distributions peut être déduit l'écart-type  $\sigma$  de l'erreur d'extraction des cibles dans le cas où cette erreur suit une loi normale centrée en 0.

Les résultats de ces tests se trouvent dans le tableau 2.6. On peut remarquer que la distribution des erreurs sur les birapports, lorsque les distorsions optiques importantes

Objectif	Correction distorsions	Moyenne	Écart-type	Médian	Écart-type robuste	% outliers
KINOPTIK	non	<b>0.044</b>	<b>0.034</b>	<b>0.033</b>	<b>0.032</b>	1.26
	oui	0.048	<b>0.035</b>	<b>0.035</b>	<b>0.034</b>	1.26
COSMICAR	non	0.063	0.049	0.054	0.046	1.26
	oui	0.049	0.043	<b>0.038</b>	<b>0.030</b>	4.00
CANON	non	0.139	0.097	0.121	0.097	0.00
	oui	0.077	0.049	0.070	0.049	0.00
KISANON	non	0.352	0.270	0.290	0.247	1.51
	oui	0.058	0.047	0.045	0.044	1.51

**Table 2.4 :** Statistique sur les distances points–ligne épipolaire, calculée avant et après correction des distorsions. Les résultats sont donnés en pixel.

Objectif	Correction des distorsions	Qualité de la reconstruction (mm)		Erreur de reprojection (pixel)			
		moy	$\sigma$	$\mu_x$	$\mu_y$	moy	$\sigma$
KINOPTIK	non	<b>0.117</b>	0.065	<b>0.029</b>	<b>0.012</b>	<b>0.059</b>	0.036
	oui	0.122	<b>0.059</b>	0.033	<b>0.014</b>	0.073	0.041
COSMICAR	non	0.200	0.119	0.034	<b>0.014</b>	0.100	0.074
	oui	0.149	0.081	<b>0.029</b>	<b>0.012</b>	0.063	<b>0.035</b>
CANON	non	0.431	0.298	0.084	0.030	0.225	0.158
	oui	0.154	0.105	0.035	<b>0.013</b>	0.072	0.044
KISANON	non	0.555	0.403	0.130	0.075	0.282	0.238
	oui	0.182	0.113	0.039	0.021	0.072	0.050

**Table 2.5 :** Qualité de la reconstruction obtenue avec des cibles circulaires

Objectif de vue	Correction des distorsions	$\bar{K}$	$\sigma_K$	$\lambda$	Probabilité	Précision $\sigma$
KINOPTIK	non	4.000	0.012	27.68	0.857	0.038
	oui	4.000	0.013	27.60	0.835	0.040
COSMICAR	non	4.001	0.012	28.84	0.722	0.039
	oui	4.000	0.013	28.58	0.673	0.042
CANON	non	4.002	0.012	31.21	0.001	<i>ns</i>
	oui	4.000	0.016	30.90	0.600	0.056
KISANON	non	4.004	0.014	24.17	0.013	<i>ns</i>
	oui	4.001	0.017	23.27	0.858	0.045

**Table 2.6 :** Statistique sur la distribution des valeurs des birapports obtenus dans chaque image. La valeur théorique est égale à 4. Chaque valeur a été calculée en utilisant plus de 160 birapports. La probabilité que les erreurs suivent une loi normale centrée est obtenue par le test de Kolmogorov-Smirnov. Lorsque cette probabilité est inférieure à 50%, la valeur de la précision n'est pas significative (*ns*).

ne sont pas corrigées, a une très faible probabilité de suivre une loi normale centrée en 0. Ceci a été vérifié à l'aide du test de Kolmogorov-Smirnov. Ce phénomène est dû au fait que les distorsions optiques apportent une erreur systématique dans la localisation du centre de chaque cible, rendant donc les erreurs sur chaque birapport corrélées les unes aux autres. Dans de tels cas, ce test n'est pas significatif. La précision  $\sigma$  de localisation des cibles est alors étiquetée *ns* dans le tableau 2.6.

Par contre, lorsque les distorsions optiques sont corrigées, les erreurs systématiques disparaissent. La distribution des erreurs sur les birapports a une forte probabilité de suivre une loi normale centrée en 0. Ce test montre que la précision de localisation des cibles est de l'ordre de 0.04 pixel.

En comparant la stabilité des birapports calculés à partir de quatre cibles alignées horizontalement avec ceux obtenus à partir de cibles alignées verticalement, il s'avère que la localisation des cibles est plus précise suivant l'axe des  $y$  que suivant l'axe des  $x$ . Le facteur de proportionnalité de ces deux précisions est égal à 1.36. Cette valeur est proche du rapport des facteurs d'échelles de la cellule CCD de la caméra utilisée, qui est égal à 1.47.

### 2.3.4 Conclusions

Les expérimentations réalisées dans cette section sont cohérentes et indiquent que la précision de la localisation des cibles est de l'ordre de 1/20ème de pixel. Un très bon calibrage des caméras peut être obtenu avec de telles précisions. Les erreurs de reprojection des coordonnées données par le constructeur de la mire après calibrage sont de l'ordre de 0.06 pixel.

L'objectif KINOPTIK s'est avéré un très bon objectif, ayant très peu de distorsions optiques. Les résultats obtenus avec ou sans correction des distorsions sont très similaires pour cet objectif. Par contre, l'utilisation d'objectifs de moindres qualités nécessite la correction des distorsions lorsque la précision de localisation est requise. Dans la suite de cet travail, seul l'objectif KINOPTIK sera utilisé afin de s'affranchir de la correction des distorsions.

## 2.4 Mise en correspondance de points entre 2 images

La détection de points modélisables est applicable lorsque des points particuliers, comme les coins ou les cibles, sont présents dans une scène. Néanmoins, toutes les scènes ne comportent pas de tels points, et d'autres algorithmes doivent être appliqués pour détecter et suivre dans une séquence un point quelconque de l'image. Un algorithme de mise en correspondance précise de points est développé dans cette section.

De nombreuses méthodes de mise en correspondance de points ont été développées dans la littérature, mais cette section n'étudiera que les techniques utilisant des fonctions de corrélations. La mise en correspondance d'un point  $\mathbf{p}_1$  de la première image  $I_1$  consiste généralement à choisir le point  $\mathbf{p}_2$  de la seconde image  $I_2$  qui optimise le résultat d'une fonction de corrélation. De nombreuses fonctions existent et sont comparées dans [AG92,

FFH<sup>+</sup>92]. Trois grands types de fonctions se dégagent de ces analyses :

1. des fonctions simples qui prennent en compte seulement les différences d'intensités entre deux images. La fonction SSD (Sum of Squared Differences) appartient à cette catégorie.
2. des fonctions intégrant une normalisation par rapport à la moyenne des intensités lumineuses des points de la fenêtre de corrélation. ZSSD (Zero-mean Sum of Squared Differences) en est une. Ces fonctions sont robustes aux différentes conditions d'éclairage et d'ouverture d'iris.
3. des fonctions, comme ZNCC (Zero-mean Normalized Cross-Correlation), normalisées par rapport à l'écart-type des intensités lumineuses des points de la fenêtre de corrélation. Ces fonctions permettent de mettre en correspondance des points entre deux images prises avec deux caméras n'ayant pas la même dynamique de prises d'images.

La géométrie épipolaire permet de contraindre la recherche d'un correspondant le long d'une droite évitant ainsi la recherche dans toute une région de l'image. Cette contrainte fiabilise le résultat. La recherche est également réduite lorsque la contrainte d'ordre est appliquée [Fau93]. Les algorithmes mettant en correspondance directement des lignes épipolaires point à point utilisent des méthodes de programmation dynamique [OK85]. Enfin, une fois la position de  $\mathbf{p}_2$  trouvée, la vérification que  $\mathbf{p}_1$  est bien le correspondant de  $\mathbf{p}_2$  dans  $I_1$  permet d'éliminer certaines fausses mises en correspondance [FFH<sup>+</sup>92].

La taille de la fenêtre de corrélation est généralement fixée pour toute l'image, et est de l'ordre de  $11 \times 11$  pixels. Néanmoins, il est préférable d'avoir une fenêtre de corrélation large pour les points se trouvant sur des régions d'intensités lumineuses homogènes, et plus petite sur des points proches de contours (la disparité est alors rarement autour de ces points). Une fenêtre de taille adaptative suivant la disparité et un modèle du bruit dans l'image est donc préférable [OK92].

De nombreux aspects ne seront pas abordés, notamment celui de la mise en correspondance de points lorsque des réflexions spéculaires sont présentes, ainsi qu'à proximité des contours d'occultations. Les occlusions peuvent être repérées par les accidents de la corrélation [TF90]. Les mises en correspondance de points sont rendues plus robuste sur des parties de l'image proche des occultations en contraignant les fenêtres de corrélation par les contours [LG94], ou en réalisant une corrélation multi-échelle [CTCE93]. La recherche de la transformation affine reliant les intensités lumineuses de deux fenêtres de corrélation permet également, si elle est réalisée à l'aide d'une méthode robuste, la détection des occultations [LMR95].

### 2.4.1 Déformation affine

Les méthodes de mise en correspondance présentées ci-dessus utilisent des mesures de ressemblance d'images. Les deux fenêtres de corrélation comparées sont de même tailles et de même formes. Une telle méthode a un sens lorsque la caméra effectue un

faible mouvement de translation entre les deux images. Mais dans le cas où il y a eu une rotation de la seconde caméra, ou dans le cas d'un changement d'échelle (par zoom) trop important, les résultats se dégradent. En effet, la ressemblance entre les deux fenêtres est trop faible, comme l'illustre la figure 2.12.

Lorsque la transformation entre les deux images est trop importante, la précision de la mise en correspondance est accrue en modifiant la forme de la fenêtre de la corrélation dans la seconde image de façon à ce qu'elle ressemble à celle de la première image. Il est connu que pour l'observation d'un plan, les images sont reliées par une transformation projective de  $\mathbb{P}^2$  dans  $\mathbb{P}^2$  définie par huit paramètres. Localement, une telle transformation peut être approximée par une transformation affine à six paramètres.

Étant donné un point  $(x, y)$  de la première image, son correspondant dans la seconde images se trouve alors en  $(a_1x + a_2y + a_3, a_4x + a_5y + a_6)$ . De plus, les changements d'intensités lumineuses entre les images peuvent être modélisés soit par une transformation affine (modèle utilisé par Ackerman [Ack84]; il prend en compte le changement de dynamique de l'acquisition de l'image), soit simplement par un déplacement radio-métrique  $r$  (modèle utilisé par Gruen [Gru85]). Formellement, ces transformations s'écrivent :

$$\begin{aligned} \text{Ackermann : } I_1(x, y) + n(x, y) &= r + h \cdot I_2(a_1x + a_2y + a_3, a_4x + a_5y + a_6) \\ \text{Gruen : } I_1(x, y) + n(x, y) &= r + I_2(a_1x + a_2y + a_3, a_4x + a_5y + a_6) \end{aligned}$$

où  $n(x, y)$  est la fonction du bruit dû à la prise d'images. Après un développement de Taylor, les paramètres de la transformation affine sont obtenus par la méthode des moindres carrés itératifs.

La méthode développée au sein de cette partie utilise également une transformation affine des coordonnées des points entre les images 1 et 2. Mais les paramètres qui lient les intensités lumineuses des images 1 et 2 ne sont pas recherchés comme lors des optimisations d'Ackermann et de Gruen (paramètres  $r$  et  $h$ ). La mise en correspondance précise de points entre deux images s'effectue en minimisant une fonction de corrélation suivant la position et la forme de la fenêtre de corrélation de la seconde image, définie par transformation affine de la fenêtre de la première image. L'accent sera mis sur la précision de la mise en correspondance de points que l'on peut obtenir. Une méthode similaire a été étudiée par Devernay [DF94] avec comme application la recherche des propriétés différentielles de la scène observée, sans utilisation du flot optique.

La fonction de corrélation qui a été choisie est la fonction ZSSD définie par [AG92]:

$$\begin{aligned} \text{ZSSD}(x, y, dx, dy) &= \sum_{u=-p}^p \sum_{v=-q}^q \\ &\left( \left[ I_1(x + u, y + v) - \bar{I}_1(x, y) \right] - \left[ I_2(x + u + dx, y + v + dy) - \bar{I}_2(x + dx, y + dy) \right] \right)^2 \end{aligned}$$

La notation  $\bar{I}(x, y)$  représente la moyenne des intensités lumineuses de l'image  $I$  sur la fenêtre de corrélation. Afin de prendre en compte la déformation affine, représentée par le vecteur  $\mathbf{a} = (a_1 \dots a_6)^\top$ , de la fenêtre de corrélation, cette mesure est redéfinie par

$$\text{ZSSD}(x, y, \mathbf{a}) = \sum_{u=-p}^p \sum_{v=-q}^q \left( \left[ I_1(x+u, y+v) - \overline{I_1}(x, y) \right] - \left[ I_2(x+u, y+v, \mathbf{a}) - \overline{I_2}(x, y, \mathbf{a}) \right] \right)^2$$

où  $I_2(x, y, \mathbf{a})$  représente l'intensité lumineuse, dans la seconde image, du point  $(a_1x + a_2y + a_3, a_4x + a_5y + a_6)$ . Cette fonction de corrélation a été choisie pour deux raisons :

1. elle est normalisée par rapport à la moyenne, permettant ainsi des changements de luminosité entre deux images. Plusieurs phénomènes peuvent être à l'origine d'un tel changement : changement de l'éclairage, modification de l'ouverture de l'iris, mais également le fait que la luminosité d'un point dépend du point de vue de la prise d'image.
2. cette fonction est pratiquement linéaire en les paramètres de la transformation affine, ce qui facilitera sa minimisation.

### 2.4.2 Contrainte sur la déformation affine

Deux contraintes de natures différentes imposent à la transformation affine (définie par le vecteur  $\mathbf{a}$ ) d'avoir une certaine forme. En premier lieu, la géométrie épipolaire contraint la position de la fenêtre de corrélation dans la seconde image. Cette contrainte doit être plus ou moins souple selon la précision de la géométrie épipolaire dont nous disposons. Afin de prendre en compte cette contrainte, la transformation affine est définie de manière unique connaissant la position  $\mathbf{p}_j$ ,  $j = 1 \dots 3$ , de trois coins de la fenêtre de corrélation dans la seconde image. En notant  $\mathbf{u}_j$  le vecteur directeur de la droite épipolaire associée à  $\mathbf{p}_j$  dans la seconde image,  $\mathbf{q}_j$  un point de cette droite et  $\mathbf{u}_j^\perp$  un vecteur perpendiculaire à  $\mathbf{u}_j$ , trois paramétrisations différentes permettent de prendre en compte la qualité de la géométrie épipolaire :

1. lorsque la géométrie épipolaire est très précise, chaque point  $\mathbf{p}_j$  est contraint de se trouver sur sa ligne épipolaire. Ainsi,  $\mathbf{p}_j$  s'écrit :

$$\mathbf{p}_j = \mathbf{q}_j + \lambda_j \cdot \mathbf{u}_j$$

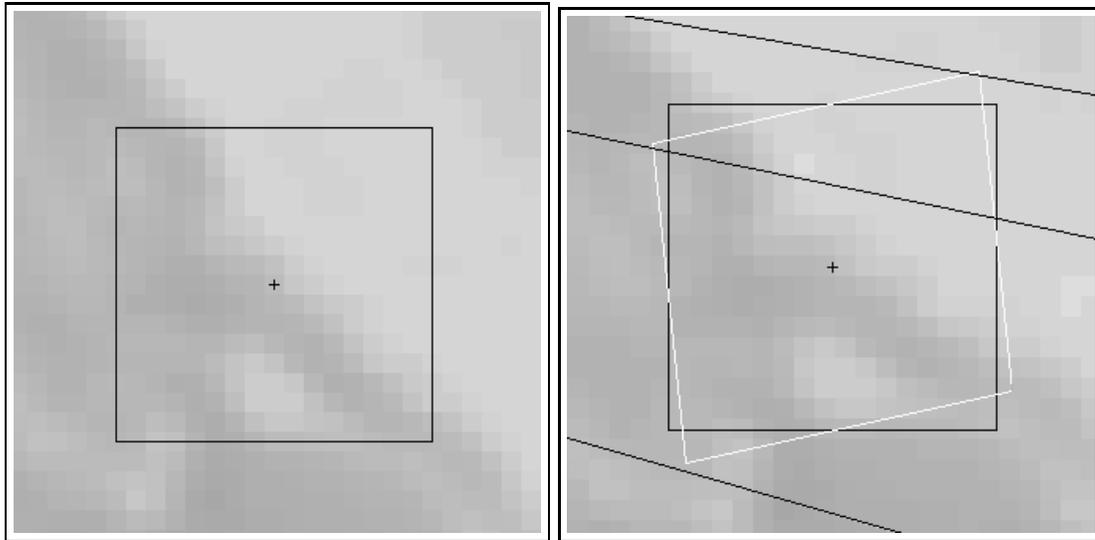
2. lorsque les erreurs sur la géométrie épipolaire ne sont pas négligeables, tout en étant localement homogènes, une tolérance unique est appliquée à chaque point :

$$\mathbf{p}_j = \mathbf{q}_j + \lambda_j \cdot \mathbf{u}_j + \lambda_4 \cdot \mathbf{u}_j^\perp$$

3. en présence d'une géométrie épipolaire imprécise et ayant des erreurs qui dépendent uniquement du point considéré, la tolérance appliquée est spécifique pour chaque point :

$$\mathbf{p}_j = \mathbf{q}_j + \lambda_j \cdot \mathbf{u}_j + \lambda_{j+3} \cdot \mathbf{u}_j^\perp$$

La figure 2.13 illustre les différentes paramétrisations possibles de la transformation affine, à trois, quatre ou six paramètres.



(a) Fenêtre de corrélation dans l'image 1

(b) Fenêtre de corrélation dans l'image 2, avec (en blanc) ou sans (en noir) déformation affine. Les lignes épipolaires associées à trois des sommets de la fenêtre de corrélation sont affichées

**Figure 2.12** : Zoom sur la fenêtre de corrélation dans les deux images, après mise en correspondance utilisant une transformation affine paramétrée par six paramètres

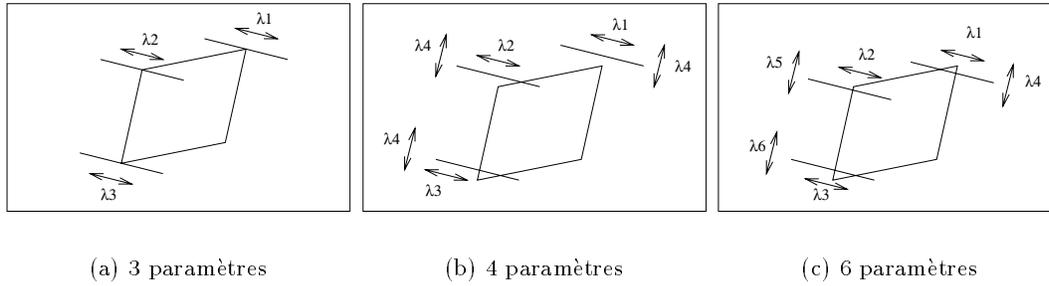
De plus, les points appartenant à un plan observé sous un angle de vue trop important sont très difficiles à mettre en correspondance. Ainsi, en complément à la contrainte épipolaire, une contrainte d'ordre physique est appliquée. Les plans observés ne formeront pas un angle de plus de  $\pi/3$  radians avec le plan image. Ceci se traduit par une contrainte non-linéaire sur les paramètres  $\lambda_j$ .

La figure 2.12(a) montre la fenêtre de corrélation autour d'un point dans la première image. La figure (b) montre la fenêtre de corrélation après déformation affine autour du point mis en correspondance. Les 3 lignes sont les lignes épipolaires associées à trois des sommets de la fenêtre de corrélation. La fenêtre de corrélation sans déformation est également affichée.

La minimisation de  $ZSSD(x, y, \mathbf{a})$ , suivant la paramétrisation  $\lambda_j$ , a été réalisée avec l'algorithme SQP<sup>2</sup> [LZT94, Fle87, GMW89]. Cet algorithme permet la minimisation de systèmes non-linéaires sous contraintes, implémentées par l'ajout de fonctions de pénalités.

---

<sup>2</sup>Sequential Quadratic Problem



**Figure 2.13** : Les 3 paramétrisations différentes de la transformation affine de la fenêtre de corrélation dans la seconde image

### 2.4.3 Initialisation des paramètres

Une estimation des paramètres à optimiser doit être fournie à l'algorithme SQP de minimisation de systèmes. L'initialisation des paramètres se réalise en deux étapes.

Tout d'abord, une approximation du correspondant dans la deuxième image du point  $\mathbf{c}$  est recherchée de façon classique en utilisant la fonction de corrélation ZSSD. Seuls les points se trouvant le long de la ligne épipolaire  $\mathbf{F} \cdot \mathbf{c}$  sont examinés.

Ensuite, les coins de la fenêtre de corrélation dans la seconde image sont déplacés sur leur ligne épipolaire associée. Une fois ces trois points fixés, la transformation affine de la fenêtre de corrélation entre les images 1 et 2 est déterminée.

On peut remarquer que l'initialisation de la transformation affine ne dépend pas du nombre de paramètres  $\lambda_i$  utilisés.

### 2.4.4 Expérimentations

Les tests présentés ci-dessous ont pour but de quantifier la précision de la mise en correspondance que l'on peut atteindre avec une telle méthode. Le premier test simule l'observation d'un plan par deux caméras afin de valider la transformation affine entre les plans vus dans les deux images. Le second test valide la précision de la mise en correspondance qui est de l'ordre de 1/20ème pixel lorsque les deux images sont proches. La stabilité de la méthode en fonction de la luminosité des deux images à mettre en correspondance est également montrée, et une comparaison avec la méthode introduite par Ackermann est donnée.

#### Validation de l'approximation affine

L'approximation de la transformation projective de  $\mathbb{P}^2$  dans  $\mathbb{P}^2$  par une transformation affine est validée en réalisant une simulation. Cette simulation comprend une tête stéréoscopique. Un plan  $P$  est placé devant cette tête. La configuration des caméras (paramètres internes et externes) ainsi que la position du plan sont connues exactement. Ce plan est défini par un vecteur normal  $\mathbf{n}$  et un paramètre  $d$  tel que  $\mathbf{p}$  appartient à  $P$  si et seulement si  $\mathbf{n} \cdot \mathbf{p} = d$ . Sans perte de généralités, le repère de la scène est celui engendré

par la première caméra.

Sous ces conditions idéales, étant donné un point de la première image projeté d'un point du plan, il est toujours possible de connaître son correspondant dans la deuxième image. Soient  $(\mathbf{u}_i; \mathbf{v}_i)$  des couples de correspondants entre les 2 images. Si  $\mathbf{I}_{c1}$  et  $\mathbf{I}_{c2}$  dénotent les matrices des paramètres internes de chacun des caméras, alors

$$\mathbf{u}_i = \mathbf{I}_{c1} \cdot \mathbf{x}_i^{(1)} \quad (2.1)$$

$$\mathbf{v}_i = \mathbf{I}_{c2} \cdot \mathbf{x}_i^{(2)} \quad (2.2)$$

où  $\mathbf{x}_i^{(1)}$  (resp.  $\mathbf{x}_i^{(2)}$ ) est le vecteur des coordonnées, exprimées en coordonnées homogènes dans le repère de la première (resp. deuxième) caméra, du  $i$ ème point de la scène qui se projette dans les images. D'autre part [Fau93],

$$\mathbf{x}_i^{(2)} = \left( \mathbf{R} + \frac{\mathbf{t} \cdot \mathbf{n}^\top}{d} \right) \cdot \mathbf{x}_i^{(1)} \quad (2.3)$$

où  $\mathbf{R}$  représente la matrice de rotation et  $\mathbf{t}$  le vecteur de translation de la seconde caméra par rapport à la première caméra. La relation entre les points correspondants  $\mathbf{x}_i^{(1)}$  et  $\mathbf{x}_i^{(2)}$  peut alors s'obtenir en combinant les équations 2.1, 2.2 et 2.3 :

$$\mathbf{v}_i = \mathbf{I}_{c2} \cdot \left( \mathbf{R} + \frac{\mathbf{t} \cdot \mathbf{n}^\top}{d} \right) \cdot \mathbf{I}_{c1}^{-1} \cdot \mathbf{u}_i = \mathbf{H} \cdot \mathbf{u}_i$$

La transformation projective  $\mathbf{H}$  a été approximée ensuite par une transformation affine  $\mathbf{A}$ .  $\mathbf{A}$  a été obtenue en minimisant, au sens des moindres carrés,

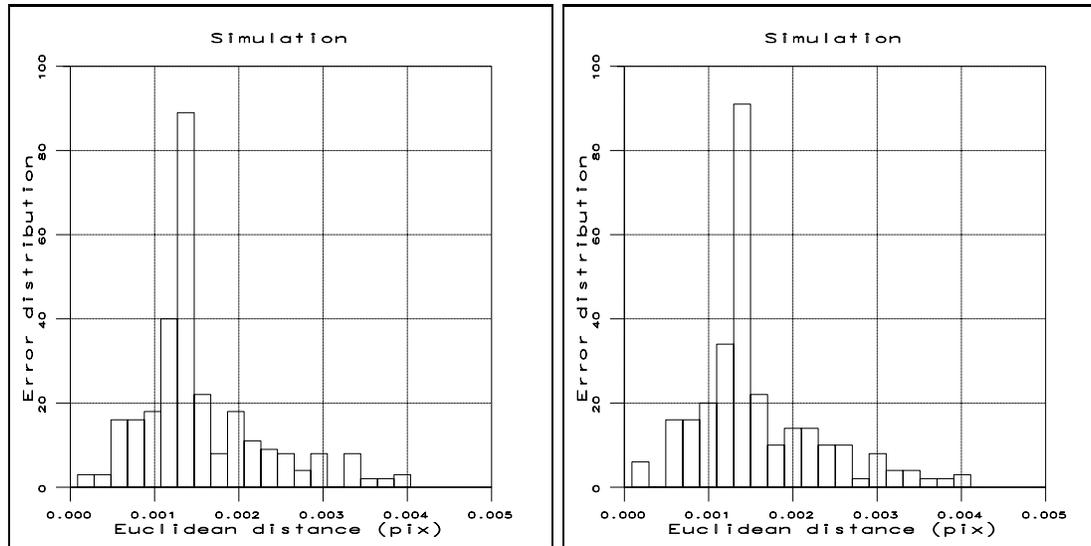
$$\chi^2 = \sum_i dist(\mathbf{v}_i, \mathbf{A} \cdot \mathbf{u}_i)^2$$

Cette simulation a été conduite dans le cas particulier où les deux caméras sont distantes de 10cm, et où leurs axes optiques font un angle de 3 degrés. Les histogrammes des distances  $dist(\mathbf{v}_i, \mathbf{A} \cdot \mathbf{u}_i)$  ont été tracés sur la figure 2.14 pour deux positions différentes du plan. Dans le premier graphe, le plan est parallèle au plan fronto-parallèle. Le second graphe a été obtenu avec une position quelconque du plan. Ces deux graphes montrent que pour ces deux cas particuliers, la transformation affine est une très bonne approximation de l'homographie réelle. En effet, les distances entre  $\mathbf{v}_i$  et  $\mathbf{A} \cdot \mathbf{u}_i$  ne dépassent jamais 0.005 pixels avec une moyenne de 0.002 pixel, ce qui est une précision que nous n'avons pas l'audace d'espérer approcher pour les mises en correspondance.

### Précision des mises en correspondance

La validation de cette méthode de mises en correspondance de points a été réalisée par cinq expérimentations, auxquelles s'ajoutent la comparaison des résultats par rapport à ceux obtenus par la méthode d'Ackermann. Les trois premiers tests s'appuient sur les trois mêmes images (cf figure 2.15). Le mouvement de la caméra entre les images (a) et (b) est essentiellement constitué d'une translation, alors que c'est le mouvement de rotation qui domine entre les images (a) et (c).

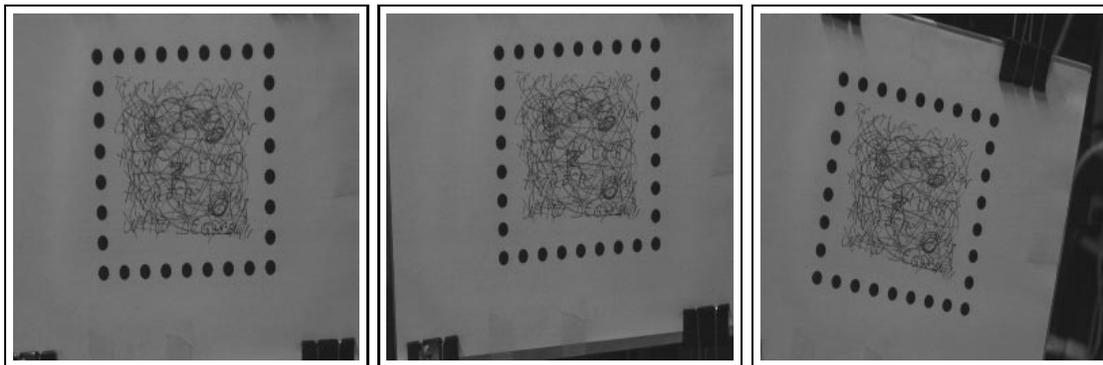
Ces trois images ont été acquises avec la même caméra PULNIX-6EX, équipée d'un objectif KINOPTIK de longueur focale 12.5mm. La scène observée est constituée d'une feuille de papier fixée sur un plan. Cette feuille contient :



(a) Plan parallèle au plan fronto-parallèle

(b) Plan dans une position quelconque

**Figure 2.14** : Distribution des erreurs lorsque une transformation projective est approximée par une transformation affine. Deux positions différentes du plan sont étudiées



(a) image de référence

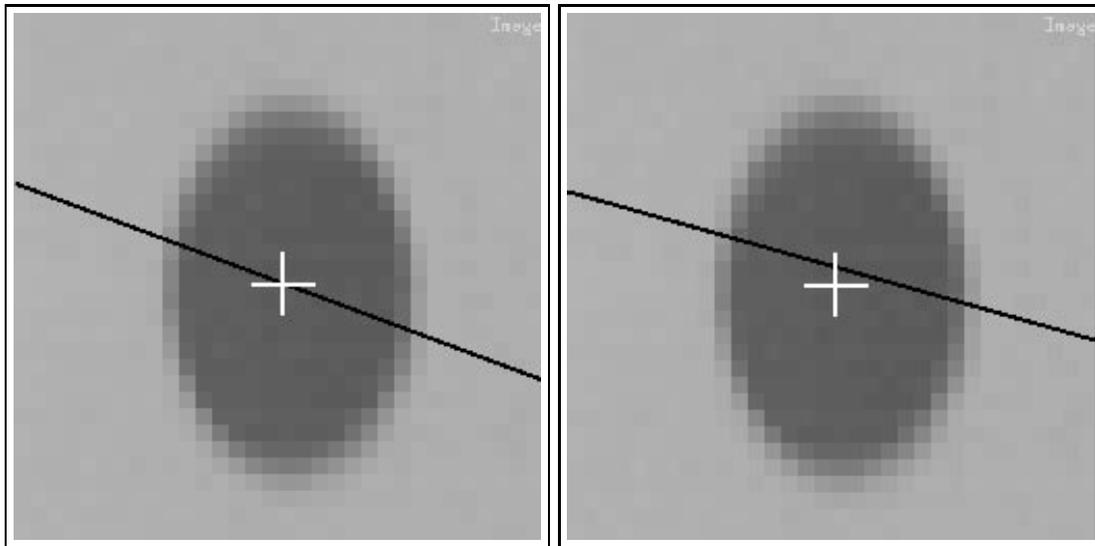
(b) après une translation

(c) après une rotation

**Figure 2.15** : Images utilisées pour la validation de la méthode de mise en correspondance précise de points

- 30 cibles noires sur fond blanc.
- de la texture. Près de 3000 points de cette région ont été mis en correspondance. Ces points forment une grille régulière dans la première image.

Toutes les cibles ont été extraites par le processus de la section 2.1.3, puis mises en correspondance manuellement. Ces cibles permettent le calcul d'une géométrie épipolaire reliant deux de ces images avec une haute précision. En effet, la moyenne des distances



(a) géométrie épipolaire obtenue avec les positions initiales des cibles

(b) géométrie épipolaire obtenue avec les positions bruitées des cibles

**Figure 2.16 :** Exemple de la qualité des géométries épipolaires calculées entre les images (a) et (b). La croix blanche représente le centre d'une cible, et la ligne noire sa ligne épipolaire associée

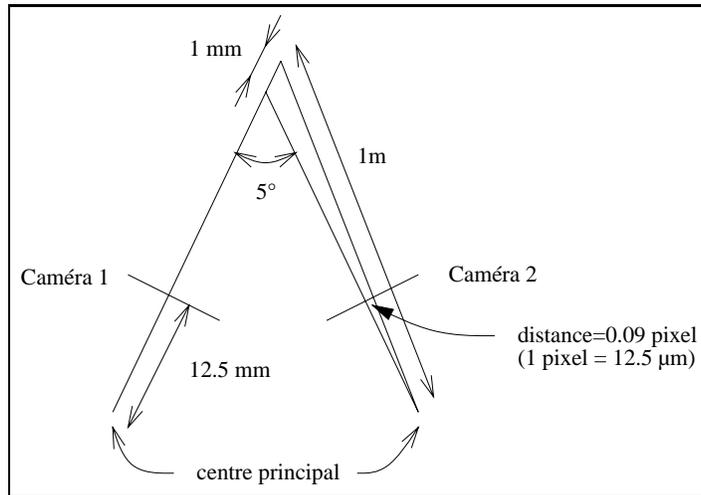
entre un point de l'image et sa ligne épipolaire est de l'ordre de 0.02 pixel. Il est à noter que comme la scène est plane, la solution obtenue n'est pas unique : une infinité de géométries épipolaires peuvent relier ces deux vues.

Une telle précision de la géométrie épipolaire ne peut cependant pas être obtenue dans le cas général. Ainsi, d'autres estimations, de moins bonnes qualités, de la matrice fondamentale ont été calculées à partir des cibles dont les positions ont été bruitées par un bruit blanc gaussien, d'écart-type égal à 1. La distance moyenne entre un point et sa ligne épipolaire est alors de l'ordre de 0.5 pixel. La figure 2.16 montre par un exemple la qualité des géométries épipolaires obtenues.

L'influence des différents paramètres de l'algorithme sur la précision des mises en correspondance a été étudiée en calculant les correspondants pour différentes valeurs des paramètres ci-dessous :

- la paramétrisation de la transformation affine utilisant 3, 4 ou 6 paramètres.
- la taille de la fenêtre de corrélation. Les fenêtres de corrélation utilisées ont pour taille  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ,  $11 \times 11$ ,  $13 \times 13$  et  $15 \times 15$  pixels.
- la qualité de la géométrie épipolaire.

Un point  $(x^a \ y^a)$  de l'image (a) sera mis en correspondance en  $(x^b \ y^b)$  dans l'image (b), et en  $(x^c \ y^c)$  dans l'image (c).



**Figure 2.17 :** Déplacement du point mis en correspondance dans la seconde image en fonction d'un changement de profondeur du point de la scène. Attention : la figure n'est pas à l'échelle

Il est à noter que la mise en correspondance de points entre les images (a) et (b) concerne près de 3000 points de la zone texturée. Par contre, les points mis en correspondance entre les images (a) et (c) n'ont été obtenus que pour 375 points d'intérêts de l'image (a) car la corrélation n'accroche pas lorsqu'il y a trop de rotation. En effet, l'initialisation de la position d'un correspondant est réalisée par une fonction de corrélation classique, mal adaptée aux images en rotation. D'autres méthodes d'initialisations doivent être utilisées dans de tels configurations, comme par exemple celles utilisant des invariants locaux dans l'image [SM96].

Remarque : la qualité de la validation dépend de la qualité de la planarité de la surface observée. Ce facteur n'est pas négligeable. Par exemple, des simulations ont montré qu'une erreur de profondeur de 1mm dans la première vue provoque un déplacement de 0.09 pixel du correspondant dans la seconde image (dans cette simulation, la longueur focale des objectifs est de 12.5mm, la distance des caméras au point observé est de 1m, la taille des pixels est de  $12.5\mu\text{m}$ , et les lignes de vues forment un angle de 5 degrés – cf figure 2.17). C'est pourquoi un faible biais sera observé dans les tests.

**Évaluation des alignements :** les points mis en correspondance entre les images (a) et (b) possèdent deux propriétés :

1. ils forment des alignements dans l'image (a)
2. ils correspondent à la projection de points coplanaires de la scène.

Ainsi, les mises en correspondance dans l'image (b) doivent conserver ces alignements. Pour chaque ensemble de points alignés  $\{(x_i^a, y_i^a)\}$  dans l'image (a), une droite est ajustée aux points correspondants  $\{(x_i^b, y_i^b)\}$  dans l'image (b). L'ajustement par la méthode des

moindres carrés médians permet la suppression des erreurs grossières de mise en correspondance, et donc de conserver les points corrects (cf annexe A). Différentes statistiques sur les distances droites ajustées – point correct ont été calculées. Tous les résultats se trouvent dans les tables B.1 à B.6 de l’annexe B. Ces tests n’ont pas pu être réalisés sur l’image (c) car nous ne disposons pas d’une grille (seuls les points d’intérêts ont été mis en correspondance).

Le graphe de la figure 2.18(a) montre la variation de la moyenne et de l’écart-type de ces distances calculées sur les points corrects en fonction de la taille de la fenêtre de corrélation. Les mises en correspondance ont été obtenues en utilisant 4 paramètres pour la transformation affine, ainsi que l’estimation imprécise de la matrice fondamentale. L’histogramme de la figure 2.18(b) montre la distribution de ces distances algébriques obtenues avec une fenêtre de corrélation de taille  $15 \times 15$  pixels.

L’analyse des résultats contenus dans les tables B.1 à B.6 montre que l’ajustement de chaque droite est non biaisée, la moyenne des distances algébriques étant nulle. De plus, les résultats obtenus avec 4 ou 6 paramètres pour la transformation affine sont équivalents quelque soit la qualité de la géométrie épipolaire employée.

Les mises en correspondance obtenues en utilisant 3 paramètres pour la transformation affine sont bien alignées lorsque la géométrie épipolaire utilisée est précise, mais la perte de précision, par rapport aux résultats obtenus avec 4 ou 6 paramètres, est de l’ordre d’un facteur 2 lorsque la géométrie épipolaire bruitée est utilisée.

L’influence de la taille de la fenêtre de corrélation est aussi significative. Le gain de précision est de l’ordre d’un facteur 3 entre les tailles  $5 \times 5$  et  $15 \times 15$  pixels, passant de 0.06 à 0.02 pixel en distance moyenne. De plus, la quantité d’erreurs grossières diminue fortement (de 20% de points erronés à 2.3 %). Ceci est dû au fait que l’estimation initiale de la position d’un point se base sur moins d’informations lorsque la fenêtre de corrélation est petite, et est donc moins fiable.

Il faut cependant prendre beaucoup de précautions lors de la lecture de ces résultats. En effet, il ne concerne que la précision du positionnement d’un point relativement à ces voisins. Cette précision n’est pas absolue. Des biais peuvent apparaître, notamment à cause de la géométrie épipolaire.

**Évaluation des positions :** les homographies  $\mathbf{H}^b$  et  $\mathbf{H}^c$  reliant les plans images (a) et (b), et (a) et (c) ont été estimées à l’aide des cibles. La précision des homographies obtenues est de l’ordre de 0.04 pixel. Ainsi, pour un point  $(x^a \ y^a)$  de l’image (a), une estimation de la position de ses correspondants dans les images (b) et (c) est donnée par

$$\begin{pmatrix} s'x' \\ s'y' \\ s'' \end{pmatrix} = \mathbf{H}^b \begin{pmatrix} x^a \\ y^a \\ 1 \end{pmatrix} \quad \begin{pmatrix} s''x'' \\ s''y'' \\ s'' \end{pmatrix} = \mathbf{H}^c \begin{pmatrix} x^a \\ y^a \\ 1 \end{pmatrix}$$

Les tables B.7 à B.12 de l’annexe B regroupent les statistiques robustes des distances entre les points  $(x^b \ y^b)$  et  $(x' \ y')$ , ainsi qu’entre les points  $(x^c \ y^c)$  et  $(x'' \ y'')$ . Ces résultats montrent qu’un biais dans la position de chaque point apparaît lorsque 3 paramètres définissent la transformation affine entre les deux images. Ce biais est dû à la géométrie

épipolaire utilisée. Ce biais est faible (0.10 pixel) lorsque la géométrie épipolaire utilisée est précise, mais il augmente avec l'imprécision de la géométrie épipolaire (0.25 pixel entre les images (a) et (b), et 0.60 pixel entre les images (a) et (c)). Les erreurs systématiques suivant l'axe des  $x$  sont bien plus importantes que suivant l'axe des  $y$  car les lignes épipolaires sont pratiquement parallèle à l'axe des  $y$ .

Le biais de localisation des points disparaît pratiquement lorsque les correspondants sont recherchés avec une paramétrisation utilisant 4 ou 6 paramètres. Ce biais est alors de l'ordre de 0.02 pixel suivant l'axe des  $x$ . Cette valeur ne s'annule pas car la surface observée n'est pas parfaitement plane. Par contre, la valeur du biais ne dépend pas de la taille de la fenêtre de corrélation ou de la paramétrisation de la transformation affine.

Le pourcentage d'erreurs grossières approche 50% pour les mises en correspondance de points entre les images (a) et (c). Ce nombre élevé est dû à la mauvaise initialisation des positions. Néanmoins, une fois que l'initialisation est bonne, les points sont bien repositionnés par notre algorithme.

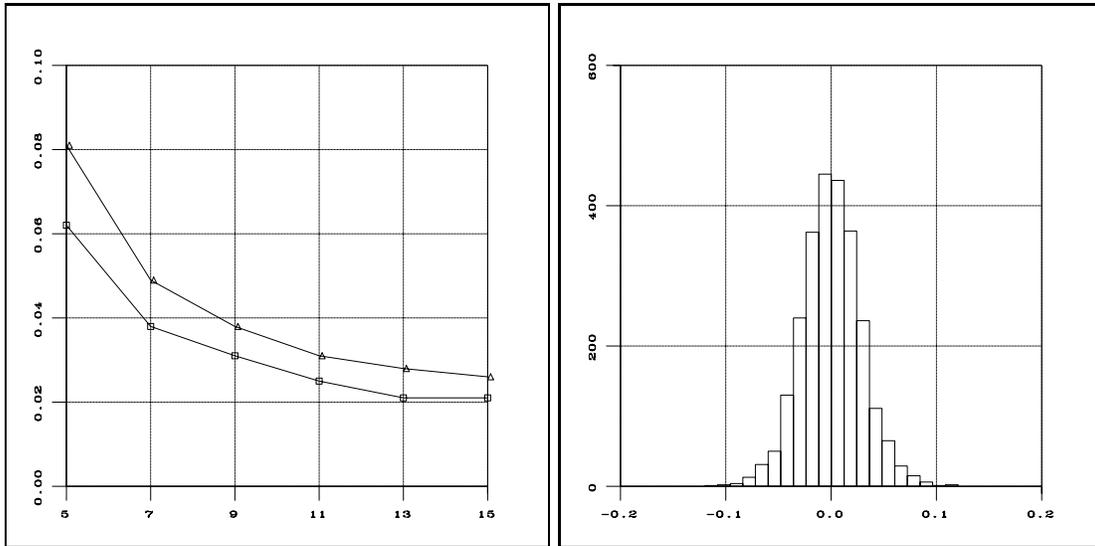
Les résultats obtenus avec une transformation affine ayant 4 ou 6 paramètres sont équivalents, avec un petit avantage pour l'utilisation de 4 paramètres. Ainsi, cette méthode sera préférée car elle converge plus facilement, et elle est un peu plus rapide, ayant moins de paramètres à optimiser.

Entre les images (a) et (b), le pourcentage d'erreurs grossières se comporte comme pour lors du test utilisant les alignements de points : il diminue d'autant plus que la taille des fenêtres de corrélation augmente. Il en va de même pour la précision de la localisation des points qui passe de 0.14 pixel avec une fenêtre de corrélation de taille  $5 \times 5$ , à 0.04 pixel avec une fenêtre  $15 \times 15$  lorsque le mouvement est translationnel. Par contre, en présence de rotation, la précision atteinte est égale à 0.08 pixel. Cette différence de précision est due aux changements plus importants de l'image dans le cas de la rotation que lors d'une translation. Il est à noter également que la qualité de la géométrie épipolaire n'influence pas la qualité de la mise en correspondance lorsque 4 ou 6 paramètres définissent la transformation affine.

Le graphe 2.19(a) montre la proportion de points erronés et le comportement de la précision de localisation des points mis en correspondance entre les images (a) et (b) en fonction de la taille de la fenêtre de corrélation. Dans cet exemple, la transformation affine est définie par 4 paramètres, et la géométrie épipolaire utilisée est bruitée. L'histogramme de la figure 2.19(b) montre la distribution des distances entre les points trouvés et leurs positions attendues, en utilisant une fenêtre de corrélation de taille  $15 \times 15$  pixels.

**Symétrie des mises en correspondance :** Ce test vérifie la cohérence des résultats obtenus entre l'image (a) et les images (b) et (c), avec les mises en correspondance d'un point de (b) ou (c) dans l'image (a). Lors de ce test, les positions calculées sont :

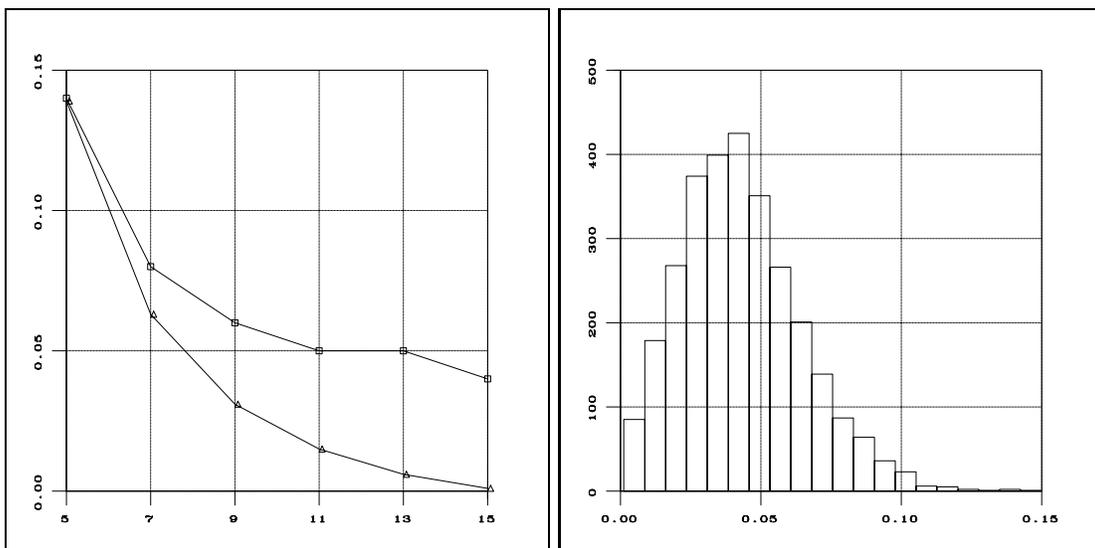
- $(x^a \ y^a)$  de (a) est mis en correspondance en  $(x^b \ y^b)$  dans (b), qui est mis en correspondance avec  $(x'^a \ y'^a)$  dans (a)
- $(x^a \ y^a)$  de (a) est mis en correspondance en  $(x^c \ y^c)$  dans (c), qui est mis en correspondance avec  $(x''^a \ y''^a)$  dans (a)



(a) Moyenne (carrés) et écart-type (triangles) des distances (en pixels) points-droites, en fonction de la taille de la fenêtre

(b) Histogramme des distances algébriques, exprimées en pixel, entre un point de l'image (b) et la droite ajustée avec ces voisins

**Figure 2.18** : Qualité des alignements de points mis en correspondance dans l'image (b) (voir le texte page 64)



(a) Distances (carrés: en pixels) et proportion d'erreurs grossières (triangles) en fonction de la taille de la fenêtre de corrélation

(b) Histogramme des erreurs de positionnement exprimées en pixels (épipolaire imprécise, fenêtre de corrélation de taille  $15 \times 15$ )

**Figure 2.19** : Qualité des positionnements des points mis en correspondance entre les images (a) et (b) obtenues avec une transformation affine utilisant 4 paramètres

La dernière colonne des tables B.7 à B.12 est appelée vérification duale, et comprend la moyenne robuste des distances

$$\frac{\sqrt{(x^a - x'^a)^2 + (y^a - y'^a)^2}}{\sqrt{(x^a - x''^a)^2 + (y^a - y''^a)^2}}$$

Les résultats obtenus sont conformes à ceux attendus et atteignent, lorsque les tailles de fenêtre de corrélation sont de l'ordre de  $15 \times 15$  pixels, une valeur de 0.02 pixel.

**Stabilité des mises en correspondance :** Ce test évalue la stabilité des mises en correspondance calculées sur une séquence de cinq images. La séquence utilisée est celle la figure 3.4. Étant donné les points d'intérêts  $(x_i^1, y_i^1)$  de la première image, ces points sont mis en correspondance dans la seconde image, qui sont eux-mêmes mis en correspondance dans la troisième image... Le schéma général de cette expérimentation est :

$$\begin{array}{c} \begin{pmatrix} x_i^1 \\ y_i^1 \end{pmatrix} \xrightarrow{1 \text{ et } 2} \begin{pmatrix} x_i^2 \\ y_i^2 \end{pmatrix} \xrightarrow{2 \text{ et } 3} \begin{pmatrix} x_i^3 \\ y_i^3 \end{pmatrix} \xrightarrow{3 \text{ et } 4} \begin{pmatrix} x_i^4 \\ y_i^4 \end{pmatrix} \xrightarrow{4 \text{ et } 5} \begin{pmatrix} x_i^5 \\ y_i^5 \end{pmatrix} \\ \begin{pmatrix} x_i^1 \\ y_i^1 \end{pmatrix} \xrightarrow{\text{mise en correspondance entre 1 et 5}} \begin{pmatrix} x_i^{5'} \\ y_i^{5'} \end{pmatrix} \end{array}$$

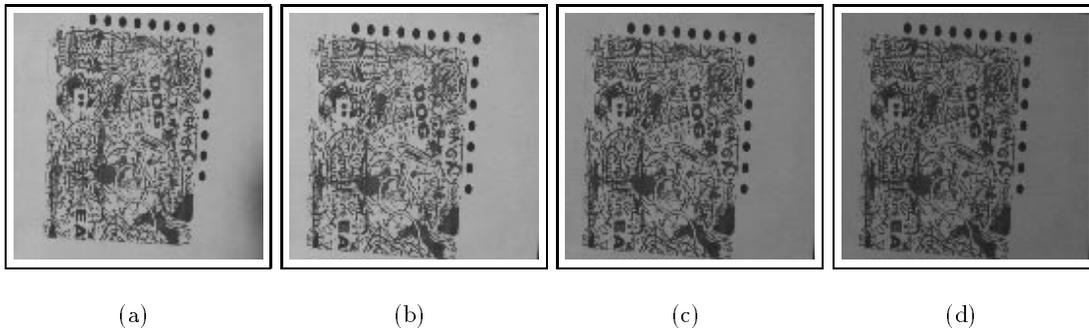
Les erreurs de mises en correspondance lors de ces tests ont été calculées et sont caractérisées par les statistiques (calculées de façon robuste) suivantes :

$$\begin{aligned} \text{moyenne des distances entre } (x_i^5, y_i^5) \text{ et } (x_i^{5'}, y_i^{5'}) &= 0.05 \text{ pixel.} \\ \text{écart-type des distances entre } (x_i^5, y_i^5) \text{ et } (x_i^{5'}, y_i^{5'}) &= 0.05 \text{ pixel.} \\ \text{moyenne de } x_i^5 - x_i^{5'} &= 0.003 \text{ pixel.} \\ \text{moyenne de } y_i^5 - y_i^{5'} &= -0.012 \text{ pixel.} \end{aligned}$$

Il apparaît que ces erreurs ne sont pas biaisées, les moyennes algébriques des distances suivant chaque axe étant de l'ordre de 1/100ème de pixel. Ces tests confirment que la précision des mises en correspondance est égal à 0.05 pixel lorsque les conditions de prises d'images sont bonnes.

**Résistance aux changements de luminosité :** Les expérimentations de cette partie montrent la stabilité de la mise en correspondance de points entre deux images prises avec des ouvertures d'iris différentes. Les images sur lesquelles les tests ont été effectués sont celles de la figure 2.20. Le mouvement de la caméra entre les vues (a) et (b) est principalement un mouvement de translation. L'ouverture d'iris de l'objectif reste inchangée entre ces deux images. Les images (c) et (d) sont prises selon le même point de vue que celui de l'image (b). Seule l'ouverture de l'iris a été modifiée. Le niveaux de gris moyen de la partie texturée de ces images est de 125 (sur 255 possibles) pour l'image (a), 126 pour l'image (b), 95 pour l'image (c) et 78 pour l'image (d).

Ces tests consistent dans la mise en correspondances de 1466 points d'intérêts de l'image 2.20(a) dans les images (b), (c) et (d). La qualité de la mise en correspondance est calculée en utilisant l'homographie entre deux de ces images, la scène observée étant



**Figure 2.20 :** Images pour calculer la robustesse de la mise en correspondance de points en fonction de la luminosité de l'images

	(a) vers (b)	(a) vers (c)	(a) vers (d)
moyenne	0.04	0.06	0.09
écart-type	0.02	0.04	0.06
% d'erreur	0.41	4.73	4.11

**Table 2.7 :** Statistiques sur la qualité des erreurs (en pixel) de la mise en correspondance de points en fonction de la luminosité de la scène (l'image 2.20(b) est la plus claire, et l'image (d) est la plus foncée)

plane. La configuration de la méthode de mise en correspondance utilisée dans ces tests est celle qui a fourni les meilleurs résultats lors des expérimentations précédentes : 4 paramètres définissent la déformation affine de la fenêtre de corrélation, la géométrie épipolaire est de bonne qualité, et la fenêtre de corrélation est de taille  $15 \times 15$  pixels.

La table 2.7 contient les statistiques (moyenne, écart-type et pourcentage d'erreurs grossières) des erreurs de positionnement entre un point trouvé par la mise en correspondance et sa position estimée par l'homographie. Ces statistiques sont exprimées en pixel. Il apparaît que les résultats se dégradent quelque peu lorsque la mise en correspondance est effectuée entre une image claire et une image sombre, avec une précision de moyenne 0.09 pixel (contre 0.04 lorsqu'il n'y a pas de changement de luminosité).

### Comparaison avec la méthode d'Ackermann

La méthode introduite par Ackermann [Ack84] pour la mise en correspondance de points entre deux images  $I_1$  et  $I_2$  procède par la recherche de 8 paramètres définissant une transformation affine de la fenêtre de corrélation ainsi qu'une transformation affine des changements d'intensités lumineuses entre les deux images :

$$I_1(x, y) = r + h \cdot I_2(a_1x + a_2y + a_3, a_4x + a_5y + a_6)$$

La recherche des paramètres de ce modèle est réalisée par l'approximation de cette estimation par un développement de Taylor au premier degré.

Les mêmes expérimentations que celles réalisées dans les parties précédentes ont été effectuées avec cette méthode de mise en correspondance de points afin de la comparer avec la méthode développée dans ce mémoire. Il est apparu que :

- Lorsque les fenêtres de corrélation sont de même tailles, les précisions des mises en correspondance sont sensiblement équivalentes, avec un avantage non significatif de 0.01 pixel pour la méthode développée dans cette partie.

Néanmoins, la méthode d’Ackermann produit un plus grand nombre d’erreurs grossières (1.13% contre 0.14% pour notre méthode). Et même si elle ne nécessite pas la connaissance de la géométrie épipolaire, il est important d’initialiser la transformation affine en utilisant cette contrainte (cf sous-section 2.4.3) pour ne pas encore accroître par un facteur 8 la proportion des erreurs grossières, qui passe alors à 8.25%.

- La méthode développée par Ackermann est plus stable aux changements de luminosité des deux images à mettre en correspondance. Sur les tests effectués sur les images de la figure 2.20, la dégradation de la qualité de la mise en correspondance est de seulement 0.01 pixel, contre 0.05 pixel avec notre méthode.

### 2.4.5 Résumé des expérimentations

À l’aide d’une scène plane, différents tests ont pu être menés afin de valider la précision de la mise en correspondance, suivant différents paramètres : taille de la fenêtre de corrélation, qualité de la géométrie épipolaire, image en rotation ou non. Les principaux enseignements de ces tests sont :

- la précision que l’on peut atteindre par cette méthode est de 0.05 pixel. Cette précision est atteinte sur deux images proches (sans rotation) en utilisant une transformation affine définie par 4 paramètres (figure 2.13), et une fenêtre de corrélation de taille supérieure à  $11 \times 11$  pixels. Sur deux images en rotation, la précision obtenue est de l’ordre de 0.08 pixel.
- lorsque la géométrie épipolaire n’est pas suffisamment précise, il est indispensable d’utiliser 4 ou 6 paramètres pour définir la transformation affine (en effet, l’utilisation de seulement 3 paramètres impose à la fenêtre de corrélation de se trouver sur les lignes épipolaires). Cependant, la qualité de la géométrie épipolaire n’influe pas sur la qualité des mises en correspondance lorsque 4 ou 6 paramètres définissent la transformation affine.
- les résultats obtenus avec une paramétrisation à 4 ou 6 variables de la transformation affine sont équivalents. Néanmoins, l’optimisation à 4 paramètres étant plus facile à réaliser, ce sera la solution retenue.
- la qualité des mises en correspondance se dégradent un peu en fonction de la différence de luminosité des deux images. Ainsi, l’erreur atteint 0.09 pixels (perte de précision de l’ordre de 0.04 pixel) lorsque la différence de la luminosité des images est de 50 (soit 20% de la plage totale des intensités lumineuses, ce qui est relativement important).

## 2.5 Conclusions

Ce chapitre a montré deux méthodes d'extraction précise de points dans les images. La première méthode concerne les points caractérisés dans la scène par un modèle. L'algorithme général a été appliqué à deux types de points particuliers : les coins de l'image et les cibles circulaires. Les précisions de positionnement de ces entités sont de l'ordre de 1/10ème de pixel pour les coins, et 1/20ème de pixel pour les cibles. Les tests effectués sur la qualité de localisation des cibles ont également montré que l'objectif KINOPTIK est un objectif qui permet d'avoir un système d'acquisition d'images bien modélisé par le modèle sténopé. Par contre, les distorsions optiques doivent être corrigées lorsque les objectifs COSMICAR, KISANON et le zoom CANON sont utilisés.

L'extraction et la mise en correspondance de points ne possédant pas de modèles ont été étudiées. Les mises en correspondance sont réalisées par la minimisation d'une fonction de corrélation suivant une déformation affine de la fenêtre de corrélation dans la seconde image. Des expérimentations ont validé cette méthode comme pouvant atteindre une précision de 1/20ème de pixel avec une fenêtre de corrélation de taille  $15 \times 15$  pixels.



## Chapitre 3

# Reconstruction tridimensionnelle

Ce chapitre met en œuvre des techniques de reconstruction tridimensionnelle de points étant donnée une séquence d'images afin de mettre en évidence les gains de reconstruction que l'on peut attendre des précisions obtenues précédemment. La précision de la reconstruction sera étudiée suivant la précision de l'extraction de points dans les images.

La première section de ce chapitre traite de la reconstruction de points étant données deux images, puis de la fusion des informations recueillies lorsque l'on dispose d'une séquence d'images. Un bref état de l'art des différents types de reconstruction (projective, affine ou euclidienne) d'une scène est tout d'abord réalisé, qui conduit à la solution retenue : une reconstruction projective utilisant des cibles afin de retrouver la structure euclidienne de la scène. La mise en correspondance grossière de points est alors détaillée car c'est la base de notre processus. La reconstruction projective de points, utilisant une décomposition des matrices perspectives suivant la matrice fondamentale qui relie deux vues, est ensuite décrite, tout comme le passage d'une reconstruction projective à une reconstruction euclidienne.

Dans la seconde section, la reconstruction de surface à l'aide d'une triangulation de Delaunay des points dans les images est étudiée. Elle s'effectue en deux étapes. La première étape consiste à reconstruire la surface étant donnée deux vues. Dans la seconde étape, différentes heuristiques sont étudiées afin de compléter itérativement une surface en prenant en compte de nouvelles vues. Trois cas peuvent se présenter :

- une nouvelle estimation de la position d'un point est donnée. Il faut l'intégrer à la surface existante.
- Un nouveau point s'insère dans la surface. La surface est alors mieux échantillonnée.
- Un point qui n'avait pas été observé précédemment à cause d'une occultation par des objets de la scène complète la surface en lui ajoutant de nouvelle partie.

Enfin, des expérimentations mettant en évidence la qualité de la reconstruction obtenue en fonction de la précision de la localisation de points dans les images, ainsi que la qualité de la surface reconstruite terminent ce chapitre.

### 3.1 Reconstruction de points

Différents types de reconstruction des points de position euclidienne (exprimée en coordonnées homogènes dans le repère de la scène)  $\mathbf{X}_i = (X_i \ Y_i \ Z_i \ 1)^\top$  sont réalisables. Elles dépendent des connaissances que l'observateur a de la scène et de son système d'acquisition d'images.

La reconstruction *projective* d'un point  $\mathbf{X}_i^p = (X_i^p \ Y_i^p \ Z_i^p \ T_i^p)^\top$  fournit des coordonnées tridimensionnelles qui sont égales aux coordonnées recherchées à une transformation projective  $\mathbf{P}^p$  de  $\mathbb{P}^3$  dans  $\mathbb{P}^3$  près. Cette matrice est de taille  $4 \times 4$ , définie à un facteur près (elle a donc 15 degrés de liberté), et relie  $\mathbf{X}_i$  et  $\mathbf{X}_i^p$  par :

$$s_i \mathbf{X}_i = \mathbf{P}^p \cdot \mathbf{X}_i^p$$

où  $s_i$  représente le facteur d'échelle.

Cette reconstruction peut être obtenue connaissant seulement la géométrie épipolaire reliant deux vues. Rappelons que la mise en correspondance de huit points entre deux vues permet de calculer la matrice fondamentale  $\mathbf{F}$  représentant la géométrie épipolaire. En fixant les coordonnées homogènes de cinq points non-coplanaires 4 à 4 de la scène, la reconstruction projective est obtenue soit en explicitant les matrices perspectives de la prise d'images [Fau92], soit en optimisant directement la valeur des matrices perspectives ainsi que les coordonnées projectives recherchées des points de la scène de façon à minimiser les erreurs de reprojections dans les images [BMV93, Bou94]. Lorsque trois vues d'une même scène sont observées, des contraintes sur la position dans la scène de chaque point apparaissent, permettant également la reconstruction projective de ces points [PMC94].

Les deux premières méthodes précédentes ont le désavantage de se reposer sur seulement cinq points de la scène en considérant leurs mises en correspondance entre les deux images comme exactes. En décomposant les matrices perspectives de chaque caméra en fonction de la matrice fondamentale  $\mathbf{F}$ , la reconstruction projective est obtenue en donnant la même importance à tous les points de la scène [Har94, BZM94], prenant en compte tous les points observés. Une fois les matrices perspectives déterminées, une reconstruction projective est déduite. Cette méthode est détaillée dans la section 3.1.2.

D'autres méthodes utilisant les birapports de droites ou de plans conduisent également à ce type de reconstructions. Le lecteur pourra se reporter à [RCF95] qui compare les performances de différentes méthodes de reconstruction projective.

La reconstruction *affine*  $\mathbf{X}_i^a = (X_i^a \ Y_i^a \ Z_i^a \ 1)^\top$  fournit des coordonnées 3D de points, à une transformation affine près représentée par la matrice à 12 degrés de liberté  $\mathbf{P}^a$  :

$$\mathbf{X}_i = \mathbf{P}^a \cdot \mathbf{X}_i^a = \begin{pmatrix} - & - & - & - \\ - & - & - & - \\ - & - & - & - \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \mathbf{X}_i^a$$

Le passage d'une reconstruction projective à une reconstruction affine est réalisé connaissant la position du plan de l'infini  $\Pi$  dans la scène. Généralement, au moins trois

prises en correspondance de points qui sont la projection de points de  $\Pi$  [KvD89, Qua93] sont connues. Ces points sont localisés à partir des lignes de fuite dans le cas particulier d'un déplacement translationnel de la caméra [MGDP93]. Dans le cas général, la localisation du plan de l'infini est réalisée par l'introduction de connaissances sur la scène, comme le fait que certaines droites sont parallèles (elles s'intersectent donc à un point de l'infini) ou qu'un point est le milieu d'un segment (la position du point de l'infini le long de cette droite est déduit à l'aide d'un birapport). Il apparaît que ce type de reconstruction est suffisant pour la navigation d'un robot, pour la détection d'obstacles ou pour circuler au milieu d'un couloir [BZM94].

La reconstruction *euclidienne*  $\mathbf{X}_i^e = (X_i^e \ Y_i^e \ Z_i^e \ 1)^\top$  est la plus complète et la plus proche de l'environnement qui nous entoure. Elle est définie à une transformation euclidienne près, déterminée par une rotation et une translation du repère de la reconstruction. Plusieurs hypothèses sont nécessaires pour obtenir une telle reconstruction. La méthode la plus couramment utilisée consiste à calibrer une tête stéréoscopique d'un robot permettant alors, par simple triangulation, la reconstruction de la scène dans le repère du robot [ZFD95].

Néanmoins, d'autres méthodes permettent d'atteindre ce résultat sans calibrage proprement dit. À partir d'une reconstruction projective, le passage à une reconstruction euclidienne peut s'effectuer en utilisant certaines informations a-priori connues sur la scène : un point particulier appartient au plan  $(Oxy)$ , deux segments sont parallèles, deux points forment un segment parallèle à un axe du repère, deux points se trouvent à une distance  $d$  donnée, ou encore deux vecteurs sont orthogonaux [Bou94].

Deux hypothèses permettent également d'obtenir une reconstruction euclidienne : lorsque les paramètres intrinsèques de la caméra sont connus [TH84], ou lorsque la scène est observée en au moins trois points de vue par la même caméra [Har93].

Le contexte dans lequel est placé cette partie est celui de la reconstruction d'une scène à partir d'une seule caméra non étalonnée. Le mouvement de la caméra entre deux vues n'est pas connu. Néanmoins, il sera supposé comme faible afin de faciliter le suivi de points le long de la séquence d'images. La reconstruction de points est effectuée afin de permettre la réalisation de deux applications : la navigation d'un bras robotisé au sein d'une scène, puis la prise d'un objet par une main articulée montée sur ce bras [BLMB<sup>+</sup>95]. Cette dernière tâche nécessite la reconstruction euclidienne de l'objet à saisir. La solution retenue est donc de poser quelques cibles autour de la scène à reconstruire afin de retrouver la structure euclidienne de l'objet à partir de sa structure projective.

La structure de cette section est la suivante : le premier point développe la mise en correspondance de points entre deux images. Seuls les points d'intérêts sont appariés. Ensuite, la reconstruction projective de points utilisant une décomposition des matrices perspectives suivant la matrice fondamentale est détaillée. Le passage à la reconstruction euclidienne fait l'objet de la sous-section suivante. Puis, la reconstruction classique utilisant un étalonnage de la caméra par la méthode de Faugeras-Toscani [FT87] est détaillée dans le but de la comparer à la méthodologie précédente. Enfin, une méthode faisant contribuer ces deux méthodes de reconstruction de scène (par une reconstruction projective suivie d'un passage à la structure euclidienne, ou par triangulation utilisant

un étalonnage des caméras) est détaillée.

### 3.1.1 Mises en correspondance de points

La mise en correspondance de points entre deux images est l'outil de base permettant la reconstruction tridimensionnelle de ces points. Trois facteurs caractérisent une bonne mise en correspondance de points : le nombre de points appariés, le pourcentage d'erreurs grossières et la précision de la localisation des points mis en correspondance.

La mise en correspondance utilisée dans ce travail a été réalisée par J.C. Cottier, inspirée d'un travail de Zhang *et al* [ZDFL94]. La mise en correspondance réalisée n'est pas dense, c'est à dire qu'elle ne concerne pas tous les points de l'image, pour des raisons de temps de calcul mais aussi de fiabilité des points appariés. En effet, la mise en correspondance d'un point est d'autant plus fiable que les intensités lumineuses dans l'image autour de ce point ne sont pas homogènes, c'est à dire que cette région de l'image est riche en information. Ainsi, seuls les points d'intérêts fournis par le détecteur de Pleysey [HS88, Bou94] sont mis en correspondance.

Cette méthode procède en deux phases. Tout d'abord, après avoir extrait les points d'intérêts des deux images, ces points sont appariés en utilisant une fonction de corrélation. La géométrie épipolaire étant encore inconnue, aucune contrainte limite la recherche des appariements. De façon à éviter au maximum le nombre de points mal appariés, les seuils utilisés (comme par exemple la réponse minimale d'une fonction de corrélation) sont élevés. Ces premières mises en correspondance permettent le calcul de la géométrie épipolaire reliant ces deux vues. La géométrie épipolaire est obtenue par une minimisation aux moindres carrés médians (annexe A) de façon à être robuste aux fausses mises en correspondance.

La connaissance de la géométrie épipolaire fiabilise l'appariement de points entre deux images en rajoutant une contrainte. La mise en correspondance des points d'intérêts est donc de nouveau effectuée, mais en prenant des seuils plus laxistes afin d'obtenir un plus grand nombre d'appariements, et en considérant la contrainte de la géométrie épipolaire.

Étant donné un couple d'images, ce processus fournit, dans sa globalité, quelques points mis en correspondance (typiquement 200), auxquels s'ajoutent la matrice fondamentale  $\mathbf{F}$  reliant ces deux vues, ainsi que l'épipole  $\mathbf{e} = (e_x, e_y, e_t)^\top$  dans la première image et celle dans la seconde image  $\mathbf{e}' = (e'_x, e'_y, e'_t)^\top$ , exprimées en coordonnées homogènes dans  $\mathbb{P}^2$ .

La précision des mises en correspondance effectuées ci-dessus est de l'ordre de 1 à 2 pixels. Elle tient à la faible précision de la détection des points d'intérêts. La méthode énoncée au chapitre 2, réalisant une mise en correspondance par déformation affine des fenêtres de corrélation, permet de repositionner de façon plus précise les points dans la seconde image. À partir de ces mises en correspondance plus précises, la matrice fondamentale reliant les deux vues est recalculée afin d'obtenir une meilleure estimation de la géométrie épipolaire.

### 3.1.2 Reconstruction projective de points

La méthode de reconstruction projective de points, à partir de deux vues, détaillée dans cette sous-section consiste à décomposer les deux matrices perspectives de la caméra en fonction de la matrice fondamentale.

Étant donnés deux couples de matrices perspectives  $(\mathbf{P}_1, \mathbf{P}_2)$  et  $(\mathbf{P}'_1, \mathbf{P}'_2)$ , ces deux couples possèdent la même géométrie épipolaire si et seulement si il existe une matrice  $\mathbf{H}$   $4 \times 4$  non singulière telle que [HGC92]

$$\mathbf{P}_1 = \mathbf{P}'_1 \cdot \mathbf{H} \quad \text{et} \quad \mathbf{P}_2 = \mathbf{P}'_2 \cdot \mathbf{H}$$

La reconstruction tridimensionnelle de points obtenue à partir de chacun de ces couples est identique à une homographie près. Les matrices perspectives  $\mathbf{P}$  et  $\mathbf{P}'$  associées à deux vues particulières peuvent donc s'écrire, sans perte de généralités, [HGC92]

$$\mathbf{P} = (\mathbf{I}_3 \mid \mathbf{0}_3) = \begin{pmatrix} \mathbf{p}_1^\top \\ \mathbf{p}_2^\top \\ \mathbf{p}_3^\top \end{pmatrix} \quad \text{et} \quad \mathbf{P}' = (\mathbf{M}' \mid \mathbf{t}') = \begin{pmatrix} \mathbf{p}'_1^\top \\ \mathbf{p}'_2^\top \\ \mathbf{p}'_3^\top \end{pmatrix} \quad (3.1)$$

où  $\mathbf{I}_3$  est la matrice identité  $3 \times 3$  et  $\mathbf{0}_3$  est le vecteur nul de taille 3.  $\mathbf{P}'$  est non singulière, et dans le cas d'une projection perspective,  $\mathbf{M}'$  est de rang 3.

D'après le choix de  $\mathbf{P}$ , le repère de la scène est celui défini par la première caméra, dont le centre  $\mathbf{C}$  se trouve dans la position  $(0, 0, 0, 1)^\top$ . D'une part,  $\mathbf{C}$  se projette en  $\mathbf{e}'$  dans la seconde image. D'autre part, en développant la projection de  $\mathbf{C}$ ,  $\mathbf{P}' \cdot \mathbf{C} = \mathbf{t}'$ . Ainsi,  $\mathbf{t}' = \lambda \mathbf{e}'$  définissant  $\mathbf{t}'$  à un facteur d'échelle près [BZM94].

De plus, on peut montrer d'une part que  $\mathbf{F}$  se factorise par

$$\mathbf{F} = [\mathbf{t}]_{\times} \cdot \mathbf{M} \quad (3.2)$$

où  $\mathbf{t}$  est un vecteur de taille 3 unique défini à un facteur d'échelle près, et  $\mathbf{M}$  est une matrice non singulière. De plus, si  $\mathbf{M}$  est solution de l'équation 3.2, alors il en est de même pour toutes matrices de la forme  $\mathbf{M} + \mathbf{t} \cdot \mathbf{a}^\top$ ,  $\mathbf{a}$  étant un vecteur quelconque de taille 3. Comme  $\mathbf{e}'$  est le noyau de  $\mathbf{F}^\top$ ,  $\mathbf{t} = \mathbf{e}'$ . D'autre part, on peut montrer que  $\mathbf{F}$  est une matrice fondamentale reliant les deux vues représentées par les matrices  $\mathbf{P}$  et  $\mathbf{P}'$  si et seulement si la matrice  $\mathbf{P}'^\top \cdot \mathbf{F} \cdot \mathbf{P}$  est une matrice du produit vectoriel [Har94]. Or, la décomposition  $\mathbf{F} = [\mathbf{e}']_{\times} \cdot \mathbf{M}'$  vérifie cette contrainte.

Ainsi, la matrice  $\mathbf{P}'$  est définie avec 4 degrés de liberté (un facteur d'échelle définissant  $\mathbf{t}'$  et trois facteurs d'échelle contenus dans  $\mathbf{a}$  pour définir  $\mathbf{M}$ ).

Une fois les deux matrices perspectives  $\mathbf{P}_1$  et  $\mathbf{P}_2$  déterminées, la reconstruction tridimensionnelle d'un point  $\mathbf{X}$  (exprimé en coordonnées homogènes) observé dans les deux images en les points  $\mathbf{x}_1 = (x_1, y_1)^\top$  et  $\mathbf{x}_2 = (x_2, y_2)^\top$  se réduit à la résolution du système linéaire suivant :

$$\mathbf{A} \cdot \mathbf{X} = \begin{pmatrix} \mathbf{p}_1^\top - x_1 \mathbf{p}_3^\top \\ \mathbf{p}_2^\top - y_1 \mathbf{p}_3^\top \\ \mathbf{p}'_1^\top - x_2 \mathbf{p}'_3^\top \\ \mathbf{p}'_2^\top - y_2 \mathbf{p}'_3^\top \end{pmatrix} \cdot \mathbf{X} = \mathbf{0} \quad (3.3)$$

Ce système est habituellement résolu en décomposant la matrice  $\mathbf{A}$  en valeurs singulières [PFTW88]

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^\top$$

où  $\mathbf{U}$  est une matrice  $4 \times 4$  orthogonale,  $\mathbf{S} = \text{diag}(s_1 \dots s_4)$  est une matrice  $4 \times 4$  diagonale dont les éléments sont les valeurs singulières de  $\mathbf{A}$ , et  $\mathbf{V}$  est une matrice orthogonale de taille  $4 \times 4$  dont la  $i$ ème colonne contient le vecteur propre correspondant à la valeur singulière  $s_i$ . La colonne  $i$  de  $\mathbf{V}$  correspondant à la plus petite valeur singulière  $s_i$  est donc une solution du système d'équations 3.3.

Utilisée directement, la solution de ce système n'est pas optimale. Elle n'a pas de sens physique car elle s'exprime en fonction de coordonnées projectives et non euclidiennes. De plus, la résolution du système d'équation 3.3 n'est pas invariante par transformation projective, c'est à dire si le couple de matrices perspectives  $(\mathbf{P}, \mathbf{P}')$  conduit à la reconstruction d'un point en  $\mathbf{X}$ , alors le couple  $(\mathbf{P} \cdot \mathbf{H}, \mathbf{P}' \cdot \mathbf{H})$  ne conduit pas nécessairement à la reconstruction de ce même point en  $\mathbf{H}^{-1} \cdot \mathbf{X}$  ( $\mathbf{H}$  représente une homographie de  $\mathbb{P}^3$  dans  $\mathbb{P}^3$ ).

La méthode la plus couramment utilisée pour obtenir une solution invariante projectivement consiste à affiner la solution estimée par la résolution du système linéaire 3.3 en minimisant les erreurs de reprojection dans les images. L'expression à minimiser est :

$$\left(x_1 - \frac{\mathbf{p}_1^\top \mathbf{X}}{\mathbf{p}_3^\top \mathbf{X}}\right)^2 + \left(y_1 - \frac{\mathbf{p}_2^\top \mathbf{X}}{\mathbf{p}_3^\top \mathbf{X}}\right)^2 + \left(x_2 - \frac{\mathbf{p}'_1^\top \mathbf{X}}{\mathbf{p}'_3^\top \mathbf{X}}\right)^2 + \left(y_2 - \frac{\mathbf{p}'_2^\top \mathbf{X}}{\mathbf{p}'_3^\top \mathbf{X}}\right)^2 \quad (3.4)$$

Cette équation est non-linéaire en  $\mathbf{X}$ . Il faut donc utiliser des techniques standards de minimisation itérative [ZFD95]. Hartley va même plus loin en minimisant cette équation avec comme entrée les positions tridimensionnelles  $\mathbf{X}$  des points de la scène ainsi que la matrice  $\mathbf{P}'$  [Har93]. Le système à minimiser devient alors hautement non-linéaire. Hartley le résout par l'algorithme de Levenberg-Marquardt, modifié de façon à prendre en compte la structure de blocs de la matrice du jacobien.

Afin d'éviter les problèmes de non-convergence lors de la résolution de 3.4, la méthode utilisée par la suite est une méthode alternative également développée par Hartley [HS94]. Cette méthode consiste à replacer les deux points  $\mathbf{x}_1$  et  $\mathbf{x}_2$  vus dans l'image exactement sur les mêmes lignes épipolaires. Ceci s'exprime par la recherche du minimum global d'un polynôme de degré 6. Une fois les points replacés, une triangulation classique par la résolution du système 3.3 produit alors une solution exacte (c'est à dire sans minimisation par les moindres carrés), qui est invariante projectivement.

### 3.1.3 Passage d'une reconstruction projective à une reconstruction euclidienne

Le passage de la structure projective trouvée par l'algorithme de la sous-section précédente à une reconstruction euclidienne est effectuée connaissant la position tridimensionnelle, dans le repère de la scène, de  $n$  points  $\mathbf{p} = (x, y, z, t)^\top$ , exprimés en coordonnées homogènes. Ces points seront appelés par la suite *les points de la base*. Ils sont identifiés par des cibles déposées tout autour de la scène. La structure projective  $\mathbf{p}' = (x', y', z', t')^\top$  de ces points est également déterminée. Le passage de la reconstruction projective vers



Une fois les positions tridimensionnelles  $\mathbf{p}$  exprimées dans un repère euclidien, une matrice de covariance  $\mathbf{C}$  peut leur être associée à partir de l'équation 3.3 et de la matrice  $\mathbf{W}$ . La matrice  $\mathbf{C}$  s'exprime en fonction de la précision de l'extraction de points d'intérêts dans les images, que nous avons fixé à 2 pixels.

### 3.1.4 Reconstruction euclidienne avec un étalonnage des caméras

Comme la scène comporte des cibles dont les positions tridimensionnelles sont connues, chaque vue peut être calibrée par la méthode de Faugeras-Toscani [FT87]. La reconstruction euclidienne des points est alors immédiate en procédant par triangulation. Les résultats obtenus par calibrage seront comparés à ceux obtenus par une reconstruction projective suivi d'un passage à des coordonnées euclidiennes.

### 3.1.5 En l'absence des points de la base

En l'absence de points de la base pour quelques vues, les points reconstruits précédemment sur l'objet peuvent être utilisés pour retrouver la structure euclidienne d'une reconstruction réalisée sur un nouveau couple de vue. Les deux types de reconstructions sont applicables avec ces points : soit en calculant la structure projective suivi d'un passage à des informations euclidiennes (sous-sections 3.1.2 et 3.1.3), soit en calibrant directement par la méthode de Faugeras-Toscani sur la structure euclidienne estimée lors des étapes précédentes (sous-section 3.1.4).

Une reconstruction utilisant conjointement ces deux méthodes a été testée. Elle consiste à :

- reconstruire la scène en calculant tout d'abord sa structure projective lorsque les points de la base sont visibles.
- reconstruire la scène par un calibrage des deux caméras lorsque les points de la base ne sont pas visibles.

## 3.2 Reconstruction de surface

Pour la prise d'un objet par une main articulée, la reconstruction tridimensionnelle de points de l'objet est utile. Mais cette information n'est pas suffisante. Il faut en plus connaître certaines propriétés de la surface de l'objet, comme la direction des normales à l'objet. Ainsi, dans cette section, la reconstruction de points réalisée précédemment est complétée par une reconstruction de surface des objets.

Plusieurs méthodes de reconstruction de surface peuvent s'appliquer suivant les caractéristiques de la scène observée. Les propriétés différentielles de la luminosité d'une surface sont prises en compte dans les algorithmes de type *shape from shading* [Hor86]. Différents modèles de luminosité de surface (comme par exemple les modèles lambertiens) permettent la reconstruction de surface. Cette méthode ne peut cependant pas

être appliquée directement que sur des objets relativement simple. La difficulté de cette méthode réside dans le choix du modèle de luminosité qui dépend des objets de la scène.

Lorsque les objets observés ont des formes courbes, les contours d'occultations permettent la reconstruction de surface. En effet, les lignes de vues sont perpendiculaires aux contours occultants, engendrant ainsi les directions des normales à la surface en ces points [Vai90, Zha93, CB92]. Cependant, cette méthode s'adapte mal à la reconstruction de surface plane qui ne comporte pas de contours d'occultations proprement dit. De plus, la détection automatique de ces contours n'est pas aisée.

Enfin, une surface peut être ajustée sur des positions tridimensionnelles des points de la scène. Plusieurs modèles de surface ont été testés dans la littérature, comme par exemple les super-ellipsoïdes [WF92]. La surface reconstruite par cette méthode présente l'avantage d'être très régulière. Néanmoins, le choix du modèle à utiliser se pose, et tous les objets ne sont pas si facilement modélisables. De plus, dans le cas d'une scène comportant différents objets, il faut tout d'abord classifier les points pour différencier les objets, puis appliquer cette méthode à chaque objet.

La méthode développée dans cette section s'apparente plus à la méthode décrite dans [FBMB90] pour la reconstruction de volume. Dans cet article, une triangulation de Delaunay est réalisée sur les positions des points d'intérêts de la scène, produisant de nombreux tétraèdres. Néanmoins, certains d'entre eux n'ont pas de sens physique dans la scène (notamment pour les parties concaves des objets). Ils sont supprimés en vérifiant la cohérence qu'ils ont avec les différentes lignes de vues (si un point est observé derrière un tétraèdre, ce tétraèdre est supprimé).

L'approche développée par Zhengyou-Zhang [Zha94] consiste en la fusion de deux surfaces. Ces surfaces sont obtenues à partir de points mis en correspondance dans les images de façon dense. Ainsi, il y a de nombreux points régulièrement disposés dans les images, contrairement aux données que nous utilisons dans ce chapitre.

Une seconde différence avec notre méthode réside dans les hypothèses de travail. En effet, les deux surfaces sur lesquelles travaillent cet auteur sont obtenues à partir d'une tête stéréoscopique montée sur un robot. La position des deux têtes stéréoscopiques l'une par rapport à l'autre est inconnue. De plus, aucune mise en correspondance n'est réalisée entre les images provenant de positions différentes du robot. Ainsi, les deux surfaces sont reconstruites dans le repère de la première caméra de la tête stéréoscopique. Elles ne sont donc pas reconstruites dans le même repère. Zhengyou-Zhang fusionne ces deux surfaces par la recherche de la meilleure transformation euclidienne (rotation et translation) qui fait les coïncider.

La méthode pour la reconstruction de surface présentée dans cette section s'appuie sur les positions tridimensionnelles des points reconstruits par une des méthodes détaillées dans la section précédente. La position dans les images des points reconstruits est également utilisée. De façon à ce que cet algorithme fonctionne sur un grand nombre de scène, aucune modélisation des objets observés n'est utilisée. La surface reconstruite à partir de deux images consiste en une reprojection de la triangulation de Delaunay [Dev92] des projections des points de la scène dans la seconde image. La surface reconstruite est donc

une surface polygonale.

Lorsqu'un nombre plus important de vues est observé, plusieurs surfaces sont engendrées en utilisant différents couples d'images. Ces surfaces sont ensuite fusionnées. La fusion de deux surfaces se déroule en trois étapes. Chaque point de la seconde surface est traité suivant son statut : il appartient également à la première surface, il peut s'intégrer dans une facette triangulaire de la première surface, ou il se trouve en dehors de la première surface.

### 3.2.1 Création d'une surface à partir de deux vues

Étant données deux vues consécutives  $V_{i-1}$  et  $V_i$ , la reconstruction euclidienne d'un point  $\mathbf{P}_j$  dans la scène observé dans ces deux images aux points  $\mathbf{u}_j$  et  $\mathbf{u}'_j$  est considérée comme acquise. La surface polygonale qui approxime tous les points reconstruits est obtenue en projetant la topologie triangulaire obtenue par triangulation de Delaunay [Dev92] sur les points de la vue  $V_i$ .

Sur cette surface, un arc peut être erroné pour deux raisons :

- il relie deux points appartenant à deux objets différents. Deux cas peuvent alors se produire :
  1. la distribution des points d'intérêts qui sont reconstruits n'est pas assez dense, et ne permet pas de distinguer deux objets différents. La longueur de cet arc est alors *standard* par rapport aux autres arcs de ce graphe. Un tel arc sera conservé.
  2. l'échantillonnage des points d'intérêts est plus important que la distance séparant deux objets. L'arc en question est alors plus long que les autres arcs.
- un des deux points de cet arc a été mal reconstruit du fait d'une erreur dans les mises en correspondance de points entre deux images. Cet arc a alors une longueur disproportionnée par rapport aux autres arcs.

Ainsi, les arcs dont la longueur est supérieure à un seuil sont supprimés. La valeur de ce seuil peut être fournie par l'utilisateur lorsque la distance entre les divers objets de la scène est connue. Dans le cas général, une estimation de ce seuil peut être calculée automatiquement en réalisant un processus des moindres carrés médians sur les longueurs des arcs (cf annexe A).

À la suite de cette démarche, certains arcs ne sont plus associés à des facettes triangulaires. Ils seront appelés par la suite des arcs *pendants*. Ces arcs sont alors supprimés. Ces deux suppressions d'arcs successives peuvent entraîner l'isolement dans la surface de points (ils ne sont plus reliés à la surface). Ces points sont également supprimés de la surface. Cette phase a pour but de supprimer les points mal mis en correspondances, ainsi que les points trop isolés dans l'espace ne permettant ainsi pas d'améliorer la reconstruction de la surface.

Après avoir supprimé ces points, la surface de l'objet est recalculée par une nouvelle triangulation de Delaunay. L'algorithme complet 3.1 résume la méthode de la reconstruction de surface à partir de deux vues. La figure 3.1 montre un exemple de reconstruction

---



---

Entrées : Positions tridimensionnelles euclidiennes  $P_j$   
Mises en correspondance  $u_j$  et  $u'_j$  dans deux vues de ces points  
Sortie : Surface triangulaire

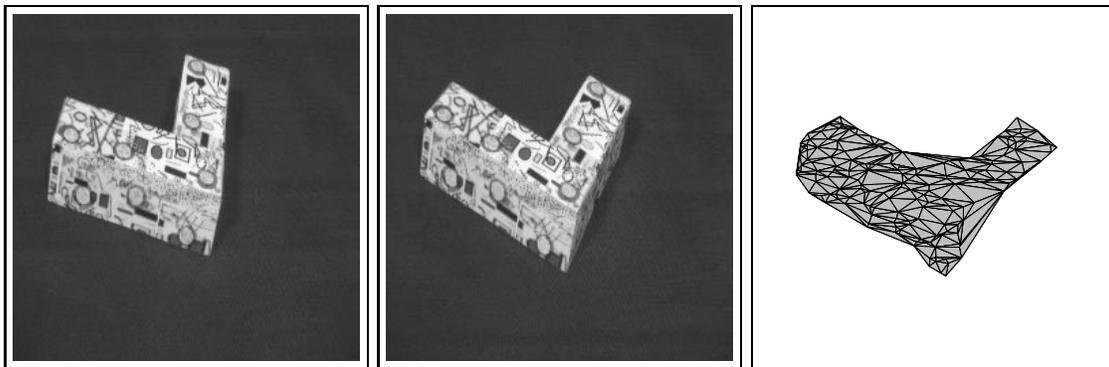
Triangulation de Delaunay dans l'image des points  $u'_j$   
Projection de la topologie obtenue afin d'obtenir une première approximation de la surface tridimensionnelle  
Suppression des arcs trop longs par les moindres carrés médians  
Suppression des arcs pendants  
Suppression des points isolés, c'est à dire sans voisins  
Triangulation de Delaunay des points  $u'_j$  restant  
La surface triangulaire est obtenue en reprojétant dans l'espace la topologie obtenue après suppression des arcs trop long

---



---

**Algorithme 3.1 :** Génération d'une surface à partir de deux vues



(a) Image 1

(b) Image 2

(c) Surface

**Figure 3.1 :** Étant données les deux images (a) et (b), la surface (c) est reconstruite

de surface étant données deux vues d'un objet. Nous rappelons ici qu'aucune hypothèse de planarité de l'objet observé n'est faite.

### 3.2.2 Fusion de deux surfaces

La sous-section précédente a montré comment une surface tridimensionnelle peut être obtenue à partir de deux vues. Cette sous-section détaille la fusion de la surface  $S$  obtenue à partir des vues  $V_1$  à  $V_n$  avec la surface  $S'$  déduite des deux vues  $V_n$  et  $V_{n+1}$  afin d'obtenir la reconstruction de surface à partir des vues  $V_1$  à  $V_{n+1}$ .

La surface  $S$  est considérée comme la plus exacte car provenant d'un plus grand

nombre de vues. La fusion des surfaces  $S$  et  $S'$  revient donc à inclure dans  $S$  les données contenues dans  $S'$ . Chaque point  $\mathbf{p}'$  (correspondant à l'estimation de la position d'un point  $\mathbf{P}$  de la scène) de  $S'$  est étudié afin de l'intégrer dans  $S$ . Deux cas peuvent se produire :

- la position du point  $\mathbf{P}$  de la scène est estimée en  $\mathbf{p}'$  dans la surface  $S'$ , mais également en  $\mathbf{p}$  dans  $S$ . Il faut donc fusionner ces deux estimations afin d'en obtenir une meilleure. Ce cas est appelé le cas des *points communs* dans la suite de cette partie.
- le point  $\mathbf{P}$  est nouvellement reconstruit, et aucune estimation de sa position n'est présente dans  $S$ . Un tel cas peut se produire pour différentes raisons :
  1. il était invisible dans la vue  $V_{n-1}$  car occulté par un objet de la scène. Il n'a donc pu être mis en correspondance entre les vues  $V_{n-1}$  et  $V_n$ .
  2. bien que visible en  $\mathbf{u}_{n-1}$  dans la vue  $V_{n-1}$ , le point  $\mathbf{u}_{n-1}$  n'est pas un point d'intérêt.
  3. bien que visible en  $\mathbf{u}_{n-1}$  dans la vue  $V_{n-1}$ , et bien que  $\mathbf{u}_{n-1}$  soit un point d'intérêt, ce point n'a pu être mis en correspondance dans la vue  $V_n$ .

Le cas 1. concerne généralement des points de  $S'$  qui se trouve à l'extérieur de la surface  $S$ , c'est à dire qu'il n'est inclus dans aucune facette de  $S$ . Les cas 2. et 3. caractérisent le plus souvent des points qui peuvent trouver une place dans la surface  $S$ .

Ces trois cas sont maintenant étudiés : les points communs aux deux surfaces, les points de  $S'$  dont la position est au cœur de  $S$  (points de  $S'$  à l'intérieur de  $S$ ), et enfin les points de  $S'$  qui se trouvent à l'extérieur de  $S$ .

### Points communs

La position d'un point de la scène  $\mathbf{P}$  peut avoir deux estimations différentes  $\mathbf{p}$  et  $\mathbf{p}'$  suivant les reconstructions effectuées (celle de  $S$  ou celle de  $S'$ ). À chacune de ces estimations est associée une matrice de covariance  $\mathbf{C}$  et  $\mathbf{C}'$ . La fusion de ces deux estimations donne une meilleure approximation de la position tridimensionnelle de ce point [DW87] [ZF90] :

$$\begin{aligned}\hat{\mathbf{p}} &= (\mathbf{C}^{-1} + \mathbf{C}'^{-1})^{-1} (\mathbf{C}^{-1} \cdot \mathbf{p} + \mathbf{C}'^{-1} \cdot \mathbf{p}') \\ \hat{\mathbf{C}} &= (\mathbf{C}^{-1} + \mathbf{C}'^{-1})^{-1}\end{aligned}$$

Une telle fusion n'est significative que lorsque les deux estimations initiales  $\mathbf{p}$  et  $\mathbf{p}'$  sont obtenues avec de bonnes mises en correspondances. En faisant l'hypothèse que les estimations  $\mathbf{p}$  et  $\mathbf{p}'$  suivent une loi normale, le test du consensus bayésien permet de s'assurer que ces deux estimations sont cohérentes avec la loi normale de moyenne  $\hat{\mathbf{p}}$  et de covariance  $\hat{\mathbf{C}}$  [DW87] :

$$\frac{1}{2}(\mathbf{p} - \mathbf{p}')^\top (\mathbf{C}^{-1} + \mathbf{C}'^{-1})^{-1} (\mathbf{p} - \mathbf{p}') \leq 6.25 \quad (3.6)$$

Dans le cas tridimensionnel, la valeur 6.25 correspond à prendre des ellipsoïdes de confiance ayant une probabilité de 90%.

Si la relation 3.6 n'est pas vérifiée, alors une des estimations a été obtenue avec une mise en correspondance erronée. Afin de décider laquelle de ces deux estimations est erronée, l'heuristique suivante a été appliquée :

- soit  $\mathbf{p}$  a été obtenu en utilisant seulement deux vues. L'estimation  $\mathbf{p}$  n'est donc pas plus fiable que l'estimation  $\mathbf{p}'$ . Les deux points sont alors supprimés des surfaces  $S$  et  $S'$  (il est préférable de supprimer un point exact que de conserver une valeur erronée).
- soit  $\mathbf{p}$  a été obtenu en utilisant plus de deux vues. Cette estimation est donc fiable. Alors seul le point  $\mathbf{p}'$  est supprimé de  $S'$ .

### Points de $S'$ à l'intérieur de $S$

Pour chaque point  $\mathbf{p}'$  de  $S'$ , il est vérifié si ce point se trouve ou non à l'intérieur d'une facette de  $S$  (c'est à dire un triangle de la surface). Pour cela, les facettes  $f$  proche de  $\mathbf{p}$  sont parcourues, et il est vérifié que :

- le projeté de  $\mathbf{p}'$  sur le plan généré par  $f$ , se trouve dans  $f$ .
- $\mathbf{p}'$  est suffisamment proche de  $f$ . Cette distance est fixée soit manuellement par l'utilisateur, soit elle peut être calculée en fonction des précisions de reconstruction des points (utilisation des matrices de covariance de  $\mathbf{p}'$  et des trois points formant la facette  $f$ ).

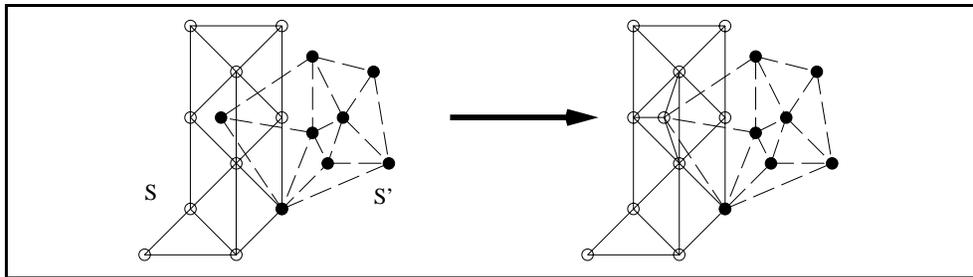
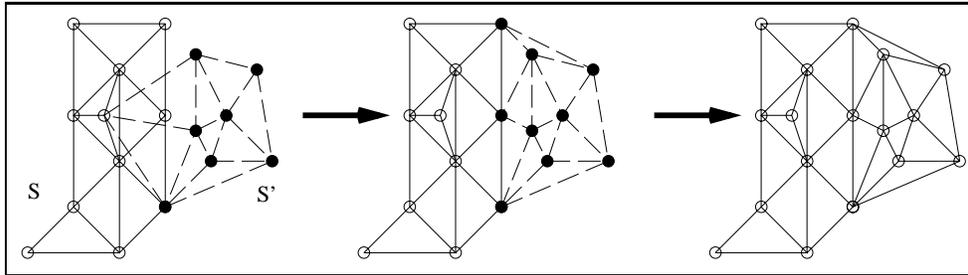
Dans ce cas, le point  $\mathbf{p}'$  de  $S'$  est intégré dans  $S$ , et il est relié au trois points de la facette  $f$ . La figure 3.2(a) illustre ce cas.

### Points de $S'$ à l'extérieur de $S$

Seuls les points de  $S'$  qui n'appartiennent pas à  $S$ , et qui ne se trouve pas dans une facette de  $S$ , n'ont pas encore été traités. Afin de les intégrer au mieux à  $S$ , une nouvelle surface  $S''$  est créée. Elle relie certains points de  $S$  aux points de  $S'$  qui n'ont pas encore été traités.

Tous les points  $\mathbf{p}'_1$  de  $S'$  non encore traités sont retenus. Dans  $S'$ , ces points  $\mathbf{p}'_1$  sont reliés à un certain nombre de points  $\mathbf{p}'_2$ . Parmi ces derniers, nous ne considérons que ceux qui ont été inclus dans  $S$ . Deux cas se profilent :

1. l'arc  $(\mathbf{p}'_1 ; \mathbf{p}'_2)$  n'intersecte aucun arc de  $S$ . Alors le point  $\mathbf{p}'_2$  est sélectionné pour appartenir à  $S''$  (il est à noter que ce cas ne peut apparaître que pour les points  $\mathbf{p}'_2$  qui étaient déjà compris dans  $S$ ).
2. Dans le cas contraire, les deux points de  $S$  qui forment l'arc intersectant  $(\mathbf{p}'_1 ; \mathbf{p}'_2)$  au plus proche de  $\mathbf{p}'_1$  sont inclus dans  $S''$ .

(a) Ajout d'un point de  $S'$  qui se trouve à l'intérieur de  $S$ (b) Ajout dans la surface  $S$  de tous les points de  $S'$  qui se trouve en dehors de  $S$ 

**Figure 3.2 :** Fusion des surfaces  $S$  et  $S'$  pour deux types de points. Les points entourés en noir appartiennent à  $S'$ , les autres à  $S$ . Les arcs affichés en pointillés sont des arcs de  $S'$ , les autres appartenant à  $S$  (attention, certains arcs appartiennent aux deux surfaces)

La projection dans la dernière vue de tous les points sélectionnés dans  $S''$  est réalisée. La surface  $S''$  triangulaire correspondant à une triangulation de Delaunay de ces points est calculée. La surface obtenue à partir des vues  $V_1$  à  $V_{n+1}$  est alors définie par la fusion de  $S$  et de  $S''$ . La figure 3.2(b) illustre l'intégration des points de  $S'$  se trouvant à l'extérieur de  $S$ .

Une fois ces 3 étapes réalisées, les liens trop longs, puis les arcs pendants, et enfin les points isolés sont éliminés. La table 3.2 résume l'algorithme pour la création d'une surface tridimensionnelle à partir de points reconstruits.

### 3.3 Expérimentations

Deux évaluations différentes sont réalisées dans cette section :

- une évaluation quantitative de la reconstruction euclidienne des points de la scène. Les trois méthodes décrites dans la section 3.1 sont comparées.
- une évaluation qualitative de la reconstruction de la surface obtenue par la méthode développée dans la section 3.2.

---



---

```

Pour chaque vue  $V_i, i = 2 \dots n$ , faire
    création de la surface triangulaire  $S'$  pour la reconstruction
        effectuée entre les vues  $V_{i-1}$  et  $V_i$  (cf algorithme 3.1)
    Si  $i = 2$  alors  $S = S'$ 
    Sinon
        fusionner dans  $S$  les estimations communes à  $S$  et  $S'$ 
        inclure dans  $S$  les points de  $S'$  dont la position
            se trouve dans une facette triangulaire de  $S$ 
        ajouter les autres points de  $S'$  dans  $S$ 
        élimination des arcs trop longs, pendants, ainsi
            que des points isolés de  $S$ 

```

---



---

### Algorithme 3.2 : Création de surface

La robustesse aux erreurs grossières de mise en correspondance de ces algorithmes sera ainsi prouvée.

#### 3.3.1 Évaluation quantitative de la reconstruction de points

##### Description de la configuration

Deux séquences d'images ont été prises dans le but d'évaluer quantitativement la précision de la reconstruction euclidienne que l'on obtient. Dans ces deux séquences, un seul objet est observé. Cet objet possède une structure tridimensionnelle connue afin de pouvoir l'évaluer. La différence entre ces deux séquences est le nombre et la qualité de la connaissance de la position tridimensionnelle des points définissant le repère de la scène.

La première séquence est appelée la séquence du  $L$ . Elle est composée de 9 images réalisée avec une caméra tournant autour d'un objet de bois ayant la forme d'un  $L$ . La figure 3.3 montre la première et la dernière image de cette séquence. Une forte texture a été plaquée contre cet objet afin d'extraire de nombreux points d'intérêts dans les images et de faciliter la phase de mise en correspondance entre deux images de cette séquence.

Les points de la base (leurs coordonnées euclidiennes sont connues afin de déduire d'une structure projective la structure euclidienne de la scène) sont peu nombreux. Ce sont le centre des cibles circulaires rétro-réfléchissantes posées sur le  $L$ . Au maximum 11 points de la base sont observés simultanément entre deux images (en général, seulement 8 points sont vus depuis deux points de vues consécutifs). De plus, la mesure de ces points dans l'espace euclidien a été obtenue avec une règle. Ces mesures ne sont donc pas précises (de l'ordre de 1mm). Cet objet est contenu dans un volume de taille  $73 \times 73 \times 23 \text{ mm}^3$ .

Sur cette scène, les points appartenant à trois plans ont été reconstruits. L'angle entre deux de ces trois plans est égal à 90 degrés.

La figure 3.4 montre la première et la dernière image des 11 images de la seconde séquence. Cette scène est constituée d'une mire de calibrage contenant des cibles circulaires. La mire comprend trois plans faisant un angle de 90 degrés entre les plans 1 et 3, et entre les plans 2 et 3. L'angle que font les plans 1 et 2 est de 120 degrés. Sur chacun de ces plans est disposée une feuille texturée. Ce sont les points de cette feuille qui sont reconstruits.

Le passage de la reconstruction projective à une reconstruction euclidienne est réalisée grâce aux points de la mire (ce sont les points de la base de la scène). Ces points sont nombreux (environ une centaine) et sont extraits avec une précision de l'ordre de 1/20ème de pixel. De plus, les positions tridimensionnelles de ces points sont connues très précisément (l'erreur est au maximum égale à  $20\mu m$ ). Cette mire est contenue dans un volume de dimension  $337 \times 262 \times 262 \text{ mm}^3$ .

Les tests réalisés concernent la reconstruction euclidienne de points texturés dans la scène. Les points de la base ont été extraits manuellement dans les images. Les appariements entre les points de la base vus dans les images et leur position euclidienne est également réalisée par l'utilisateur. Par contre, la reconstruction des points de la scène s'est faite automatiquement. Les tableaux 3.1 à 3.6 montrent la qualité des reconstructions obtenues. Différentes configurations et algorithmes ont été utilisés :

- La colonne *fréquence-base* indique la fréquence de l'utilisation des points de la base. Ainsi, la valeur 1 signifie que les points de la base sont utilisés dans toutes les images. Au contraire, fréquence=4 signifie que les coordonnées euclidiennes des points de la base ne sont utilisées seulement 1 image sur 4.
- La colonne *plan* identifie le plan qui a été reconstruit. Les points appartenant à chaque plan ont été sélectionnés manuellement. L'ajustement d'un plan sur un ensemble de points sélectionnés est réalisé de façon à être robuste aux erreurs grossières, en utilisant la méthode des moindres carrés médians (cf annexe A).
- Deux précisions des mises en correspondance ont été utilisées : une précision pixelle obtenue par une mise en correspondance classique (sous-section 3.1.1), et une précision subpixelle obtenue par la méthode développée dans le chapitre précédent. Les résultats correspondant à la précision pixelle (resp. subpixelle) se trouve dans la colonne *imprécis* (resp. *précis*) des tables 3.1 et 3.2.

Dans chacune de ces configurations, les évaluations de la qualité de la reconstruction sont basées sur deux critères : la qualité de la planarité dans l'espace euclidien des points reconstruits pour chaque plan des objets, et les erreurs angulaires entre deux plans reconstruits.

Le choix des points appartenant à chaque plan est réalisé manuellement. Pour cela, tous les points reconstruits sont projetés dans la dernière image de la séquence. Les points qui se superposent avec un plan donné sont considérés. Une fois les points appartenant à un plan isolé, un plan est ajusté par la méthode des moindres carrés médians (cf annexe A). La qualité de ce plan est donnée par la moyenne robuste et l'écart-type des distances entre chaque point et le plan ajusté. Le pourcentage de points dont la distance

au plan ne suit pas la loi normale est également fourni (pourcentage d'erreurs grossières). Ces résultats sont exprimés en mm dans les tables 3.1 et 3.2.

L'angle entre deux plans ajustés est alors calculé par produit scalaire entre les normales des deux plans. Les erreurs, en degrés, entre les angles calculés et ceux attendus sont donnés en tables 3.3 à 3.6.

La reconstruction tridimensionnelle des points est obtenue par les trois méthodes détaillées dans la première section de ce chapitre. Ces méthodes sont dénotées dans les tableaux de résultats par :

1. *décomposition* : chaque matrice perspective est estimée, à une transformation projective près, par la décomposition avec la matrice fondamentale. La reconstruction projective obtenue est alors transformée en une reconstruction euclidienne en utilisant les points de la base (sous-sections 3.1.2 et 3.1.3).
2. *calibrage* : les points dont les coordonnées euclidiennes ont été estimées servent à calibrer les caméras par la méthode de Faugeras-Toscani. La reconstruction de points est alors obtenue par simple triangulation (sous-section 3.1.4).
3. *mixte* : les deux méthodes précédentes sont utilisées conjointement. Lorsque les points de la base sont disponibles, la reconstruction est obtenue par décomposition des matrices perspectives. Au contraire, si les points de la base sont ignorés, la méthode utilisée s'appuie sur un calibrage des caméras obtenu avec les points précédemment reconstruits (sous-section 3.1.5).

Les résultats de la qualité de la reconstruction obtenue suivant la configuration choisie (qualité de localisation des points dans les images, méthode de reconstruction et fréquence d'utilisation des points de la base) sont à présent discutés et s'appuient sur les résultats donnés par les tables 3.1 à 3.6.

### Qualité en fonction de la précision

Lorsque les points de la scène sont détectés avec une précision grossière, la précision de la reconstruction de la scène du  $L$  apparaît meilleure que pour la séquence de la mire. En effet, la moyenne des distances entre les points et le plan ajusté sur ces points est de l'ordre de 0.80mm pour le  $L$ , et de l'ordre de 1.50mm pour la scène de la mire. De fait, il s'agit de valeurs absolues qu'il faut relativiser par rapport à la taille de chaque objet. Ainsi, il s'avère que la précision de la reconstruction du  $L$  est de l'ordre de 5/1000ème de la taille de la scène (en prenant en compte que seule une moitié de l'image contient l'objet), et de 4/1000ème pour la mire. Cette petite différence de précision provient de la qualité de la localisation du repère de la scène dans le second cas.

Comme le lecteur pouvait s'y attendre, la qualité de la reconstruction s'accroît avec la précision d'extraction des points dans les images. Ainsi, lorsque la mise en correspondance de points est réalisée de façon précise, la qualité de la reconstruction est de 0.30mm pour le  $L$  (soit 2/1000ème de la taille de la scène), et de 0.15mm pour la mire (soit 1/2500ème de la taille de la scène). Le facteur de précision gagnée est de 3 pour la séquence du  $L$  et

est de 10 pour la séquence de la mire. Ces facteurs de gain sont très différents pour ces deux reconstructions et nécessitent une petite analyse suivant la précision des points de la base et celle des points dans les images que sont à disposition :

- points imprécis et base imprécise : les erreurs de reconstruction proviennent principalement des erreurs de localisation des points à reconstruire. En effet, le passage à l'eulidien (obtenu à l'aide des points de la base), même s'il est calculé avec des points dont les localisations tridimensionnelles sont imprécises, reste relativement proche de la réalité avec des erreurs se compensant : le passage à l'eulidien est calculée par la méthode des moindres carrés avec de nombreux points, alors que la reconstruction d'un point de la scène n'est basée que sur deux points de l'image.
- points imprécis et base précise : le passage à l'eulidien est parfait. Ainsi, les erreurs de reconstruction proviennent presque exclusivement de l'erreur sur la localisation des points à reconstruire. La qualité de la reconstruction obtenue est donc relativement proche de celle obtenue dans le cas précédemment cité.
- points précis et base imprécise : les erreurs de reconstruction proviennent principalement du passage à l'eulidien. Même si le passage d'une structure projective à une structure euclidienne conserve la planarité, la fusion de plusieurs vues corrèle les reconstructions les unes aux autres, et ainsi perturbent la bonne planarité des points.
- points précis et base précise : les erreurs de reconstruction sont minorées.

Ainsi, la localisation précise dans les images de points de la scène associée à une bonne connaissance des positions tridimensionnelles des points de la base permettent de gagner un facteur 10 dans la précision de la reconstruction en ce qui concerne la planarité des différentes parties de la scène.

Les erreurs angulaires entre les différents plans reconstruits sont stables quelque soit la précision de la localisation des points dans les images. Ceci est dû au fait que les erreurs angulaires caractérisent la qualité du passage d'une structure projective à une structure euclidienne. Dans le cas où les erreurs de reconstruction ne sont pas biaisées (c'est à dire qu'elles ne proviennent que des erreurs de localisation des points dans les images et en aucun cas d'un passage approximatif à une structure euclidienne), les erreurs angulaires doivent même être de moyenne nulle et d'écart-type décroissant avec le nombre de points utilisés. C'est pour cela que la précision de la localisation des points n'influe pas sur les valeurs des tables 3.3 à 3.6.

### Qualité en fonction de la méthode de reconstruction

Il apparaît que la reconstruction euclidienne obtenue après un calibrage des caméras est un peu meilleure que celle obtenue par une décomposition des matrices perspectives à l'aide de la matrice fondamentale. Deux raisons expliquent cette différence de qualité :

1. La précision de la matrice fondamentale est souvent caractérisée par la moyenne des distances entre les points et leur ligne épipolaire associée. Cette précision est excellente lorsque la localisation des points dans l'image servant au calcul de la matrice fondamentale est bonne. Néanmoins, la position des épipoles dans les images reste très instable. Cette instabilité est la source d'une imprécision dans la reconstruction projective de points.

En ce qui concerne le calibrage d'une caméra, celui-ci est excellent lorsque la localisation des points est très bonne. Ainsi, les deux reconstructions (par calibrage ou utilisant une reconstruction projective) n'ont pas les mêmes sources d'erreurs.

2. La reconstruction euclidienne par calibrage s'effectue en deux étapes : étalonnage des caméras suivi d'une triangulation. Les erreurs de reconstruction proviennent donc des erreurs de localisation dans l'image des points à reconstruire ainsi que des erreurs effectuées sur la position des points de la base.

La reconstruction euclidienne utilisant une reconstruction projective comprend une étape supplémentaire. Ainsi, les erreurs de reconstruction ont trois sources : l'erreur de l'estimation de la géométrie épipolaire, l'erreur sur la position des points à reconstruire et l'erreur due au passage à une structure euclidienne utilisant les points de la base.

### Qualité en fonction de *fréquence-base*

L'analyse de la précision de la reconstruction en fonction de la fréquence d'utilisation des points de la base procure également des enseignements. Tout d'abord, il est confirmé que les erreurs angulaires sont principalement sensibles aux erreurs réalisées lors du passage à une structure euclidienne. En effet, sur la scène de la mire où ce passage est très précis, les erreurs angulaires entre les différents plans reconstruits sont faibles, quelque soit la fréquence d'utilisation de la base. Seule une reconstruction dégénère : celle utilisant une reconstruction projective, avec  $\text{fréquence}=9$ .

Par contre, en ce qui concerne la séquence du  $L$  où le passage à une structure euclidienne est bien plus délicat, les erreurs angulaires dégénèrent rapidement, et ceci quelque soit la précision d'extraction des points à reconstruire.

Le second point concerne la qualité de la planarité des points reconstruits. Dans la séquence de la mire, les résultats se dégradent avec l'augmentation de la valeur de la fréquence. Par contre, pour la séquence du  $L$ , ces résultats sont inexploitable en raison des erreurs angulaires trop importantes qui apparaissent dès que la fréquence est différente de 1. En fait, la reconstruction de points à partir des deux premières images n'est pas assez précise. Ainsi, la reconstruction obtenue après plusieurs vues sans utilisation de la base qui permettait de se replacer dans le repère de la scène dégénère rapidement.

### 3.3.2 Évaluation qualitative de la reconstruction de surface

#### Description

La validation de la méthode de la reconstruction de surface est réalisée de manière qualitative sur la séquence de six images représentées sur les figures 3.5 et 3.6. Cette scène est reconstruite itérativement à chaque nouvelle vue observée. La reconstruction tridimensionnelle des points qui est à la base de la reconstruction de surface est réalisée dans cet exemple particulier par simple triangulation après un calibrage de chaque caméra. Cependant, il faut noter que qualitativement, les différents résultats obtenus suivant la méthode de reconstruction employée sont identiques.

La scène est composée de quatre objets principaux texturés afin d'obtenir de nombreux points d'intérêts sur ces objets, et de faciliter le suivi de points de ces objets le long d'une séquence d'images. Trois objets possèdent une forme polyédrique, et le quatrième objet au centre de la scène a une forme quelconque. La rotation de la caméra est réalisée autour de ce dernier objet qui est le centre d'intérêt de la reconstruction effectuée, étant l'objet le plus difficile à reconstruire.

Des cibles circulaires ont été posées sur le sol et sur les trois objets polyédriques. Elles définissent le repère de la scène, permettant d'obtenir une reconstruction euclidienne. Leurs positions tridimensionnelles sont approximativement connues, avec une précision de l'ordre de 1 à 2mm (la position de ces points a été mesurée à l'aide d'une simple règle).

Les mises en correspondance des points de la scène sont réalisées de manière grossière, avec une précision de 1 à 2 pixels. Cette qualité des mises en correspondance est suffisante pour des expérimentations qualitatives. L'élimination des fausses mises en correspondance est réalisée par le calcul de la géométrie épipolaire (sous-section 3.1.1), par l'élimination de points isolés car trop éloignés d'autres points de la surface (sous-section 3.2.1) et par la fusion de deux estimations de la position tridimensionnelle de points (sous-section 3.2.2). En complément, seuls les points d'intérêts qui sont mis en correspondance dans au moins trois images sont reconstruits, les autres n'étant pas sélectionnés.

#### Analyse des résultats

Les surfaces obtenues sont affichées sur les figures 3.5 et 3.6. La première remarque sur la reconstruction obtenue est l'absence du sol sur lequel repose les objets. Cela vient du fait que le sol possède des niveaux de gris très homogène. Ainsi, très peu de points d'intérêts sont extraits sur cette partie de la scène. Le sol n'est pas reconstruit, les points étant éliminés car leur échantillonnage n'est pas assez important (ces points sont isolés sur la surface). Néanmoins, quelques points du sol sont tout de même visible, par exemple sur la figure 3.5(e), au pied de l'objet central. Ce point est par la suite éliminé car trop isolé (figure 3.6(d)).

Dans le cas d'un sol texturé, sa reconstruction s'effectue au même ordre que celle des objets proprement dit de la scène. Le sol et les objets ne sont alors pas dissociés, et la reconstruction consiste en une seule surface comprenant toute la scène observée.

L'objet polyédrique le plus à gauche de la scène n'est pas reconstruit à partir des images 1 et 2 car il disparaît pratiquement dans la vue 3. Ainsi, aucun de ces points ne peut être mis en correspondance dans au moins 3 images à partir de la vue 1. Sa reconstruction apparaît en fait seulement à partir de la vue 5, figure 3.6(b) (en effet, dans les vues 7 et 8 qui ne sont pas affichées, cet objet est bien visible).

Ce phénomène s'applique également à l'objet du fond de la scène qui est trop occulté par l'objet central au début de la séquence. Il est donc impossible de suivre des points le long des trois premières images de la séquence. Le suivi de points de cet objet s'effectue correctement une fois que les vues sont prises depuis le côté droit de la scène (figure 3.5(e)). Cependant, la partie reconstruite de cet objet à partir des trois premières images est restreinte. Il faut attendre la fusion de points complémentaires à cet objet qui sont observés par la suite. Ces points sont alors fusionnés avec la reconstruction déjà existante, ce qui complète la reconstruction de l'objet du fond de la scène (figures 3.5(e), 3.5(g) et 3.6(b)).

La reconstruction itérative est également mise en évidence sur l'objet le plus à droite de la scène avec la reconstruction du côté droit, invisible au début de la séquence. Il faut d'ailleurs noter qu'une imperfection, due à la suppression des arcs trop longs, dans la reconstruction obtenue après trois vues est corrigée : le trou dans cet objet sur la surface de la figure 3.5(e) est bouché sur la surface de la figure 3.5(g).

La reconstruction de la surface est itérative et se complète avec l'apport de nouvelles vues. Ceci amène une question fondamentale pour toute application utilisant ces résultats : la reconstruction de la scène est elle complète, ou bien certaines parties de la scène n'ont elles pas été observées (à cause d'occultations par exemple) ? Ainsi, cette reconstruction de surface, bien que plus complète qu'une simple reconstruction de points, est encore insuffisante pour des applications automatiques ne faisant appel à aucune hypothèse. Les zones libres ainsi que les zones non observées de la scène doivent être connues pour de telles applications.

En ce qui concerne l'objet central à la scène, sa reconstruction encore grossière obtenue à partir des vues 1 et 2 est complétée par la suite, grâce à un meilleur échantillonnage des points sur cet objet. Un trou sur le sommet de cet objet est notamment corrigé. Cet échantillonnage plus fin des points sur cet objet est dû à l'apparition et à la disparition des points d'intérêts dans les images. Le niveau de cet échantillonnage peut être guidée par l'utilisateur pour certaines applications nécessitant une précision de reconstruction différente suivant les objets (par exemple un objet à saisir doit être reconstruit de façon plus précise qu'un obstacle dans la scène). Cette manipulation est facile à réaliser car le nombre de points d'intérêts dans les images dépend d'un seuil dans le détecteur de Plessey.

La partie concave de cet objet est relativement bien rendue grâce à la suppression des arcs trop long par rapport aux autres arcs. Néanmoins, ceci reste une faiblesse de cette reconstruction de surface. Un meilleur rendu des parties concaves des objets reconstruits passe par l'utilisation d'informations supplémentaires, comme les contours des objets. Néanmoins, de telles informations sont en pratique difficiles à extraire.

### 3.4 Conclusions

Dans ce chapitre, la relation entre la précision de la reconstruction de points et la précision que l'observateur a de la scène a été mise à jour. Pour cela, des expérimentations ont été effectuées, vérifiant la planarité et les erreurs angulaires des différentes parties des objets de la scène. Il est naturellement apparu que la précision de localisation des points joue un très grand rôle dans la précision de la reconstruction obtenue. Mais même si c'est la source d'erreur la plus fréquente, il ne faut cependant pas négliger les erreurs de reconstruction provenant de la connaissance, précise ou non, des positions tridimensionnelles des points permettant d'obtenir une reconstruction euclidienne.

Lorsque la reconstruction s'effectue à partir d'une séquence d'images et qu'une estimation initiale correcte est disponible dès la première reconstruction, la qualité de la reconstruction en fusionnant de nouvelles vues se dégrade peu, même en l'absence des points de la base.

Lorsque les conditions idéales sont réunies (localisation et connaissance précises des points de la mire), la précision de la localisation tridimensionnelle atteinte est de l'ordre de 1/2500ème de la taille de la scène, sur des points n'ayant pas de signal particulier (ils sont simplement fortement texturés).

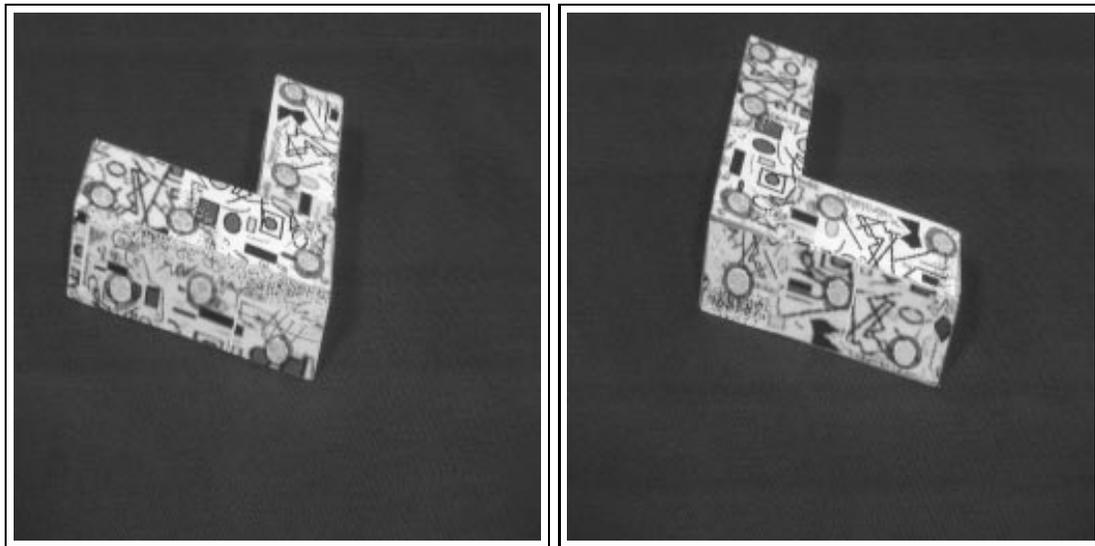
D'autre part, une méthode de reconstruction de surface, basée sur une triangulation de Delaunay des points dans l'image, a été décrite. Étant données deux images ainsi que les positions tridimensionnelles de points reconstruits à partir de ces images, la reconstruction de surface des objets de la scène est obtenue en reprojétant dans la scène la topologie de la triangulation de Delaunay réalisée sur les points de l'image. L'intégration de nouvelles vues dans cette surface comprend trois étapes suivant les points à traiter considérés :

- intégration d'une nouvelle estimation de la position d'un point de la scène : cette étape permet d'accroître la précision de la reconstruction de ces points.
- intégration d'un point au centre d'une facette de la surface : un meilleur échantillonnage de la surface résulte de cette étape.
- intégration de points qui se trouvent en dehors de la surface : des parties jusque là invisibles, le plus souvent dû es à des occultations, de la scène sont intégrées.

Une expérimentation a été menée et a été évaluée qualitativement. Ainsi, il a été montré que la reconstruction itérative de la surface est correctement effectuée, et que les parties concaves des objets sont bien restituées lorsqu'elles sont suffisamment importantes.

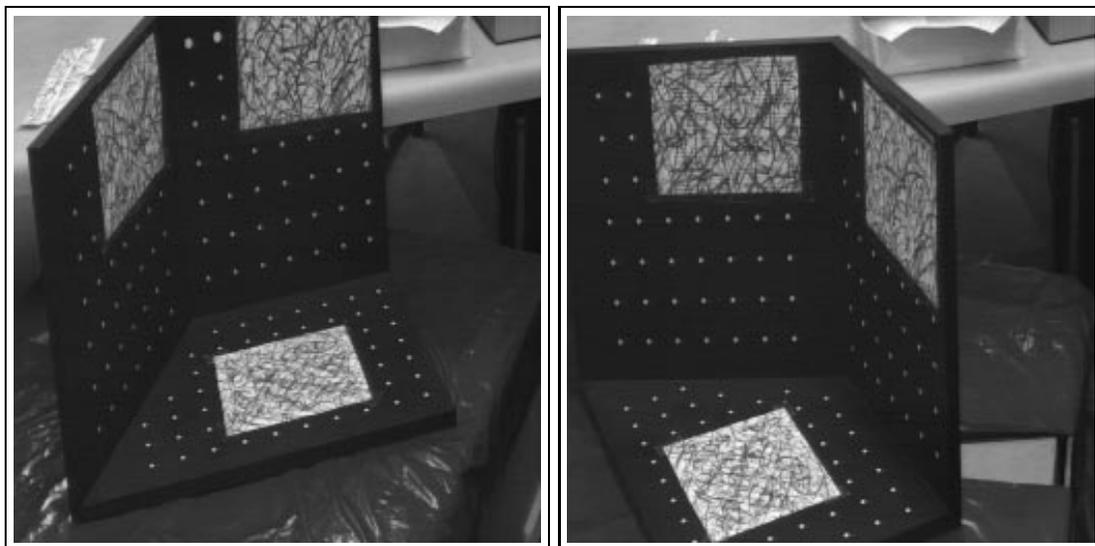
Néanmoins, il est apparu qu'une telle reconstruction itérative est insuffisante pour des applications automatiques ne faisant aucune hypothèse sur la scène. En plus de cette reconstruction de surface, la gestion des zones libres et des parties inconnues de la scène doit être effectuée.

Les parties concaves des objets peuvent également être mieux détectées par l'utilisation d'informations complémentaires, comme par exemple les contours dans les images. Cependant, de telles informations restent difficiles à extraire dans le cas général.



(a) Image 1

(b) Image 9

**Figure 3.3** : La première et la dernière image de la séquence du  $L$ 

(a) Image 1

(b) Image 11

**Figure 3.4** : La première et la dernière image de la séquence de la mire

Méthode	Fréq. base	Plan	Imprécis			Précis		
			Moyenne	Écart type	% d'erreur	Moyenne	Écart type	% d'erreur
Mixte	1	1	0.97	0.75	6.95	0.78	0.67	13.37
		2	0.98	0.81	11.86	0.65	0.64	23.66
		3	1.50	1.16	4.13	0.42	0.34	14.54
	4	1	0.81	0.62	5.72	0.45	0.45	30.00
		2	0.51	0.43	15.50	0.06	0.05	8.47
		3	1.07	0.83	6.17	0.21	0.16	8.47
	9	1	0.60	0.44	5.08	0.08	0.07	12.87
		2	0.42	0.33	8.33	0.03	0.03	17.96
		3	0.68	0.52	2.90	0.04	0.03	28,84
Calibrage	1	1	0.72	0.53	7.22	0.30	0.23	4.92
		2	0.89	0.76	16.43	0.56	0.41	8.00
		3	0.74	0.61	5.65	0.34	0.31	18.86
	4	1	0.66	0.52	8.22	0.27	0.21	11.40
		2	0.76	0.57	16.91	0.43	0.38	18.55
		3	2.21	1.65	2.19	0.30	0.28	38.88
	9	1	0.54	0.40	5.37	0.07	0.06	18.18
		2	0.87	0.66	4.65	0.62	0.54	5.51
		3	0.64	0.47	3.44	0.18	0.19	23.40
Décompo.	1	1	0.97	0.75	6.95	0.78	0.67	13.37
		2	0.98	0.81	11.86	0.65	0.64	23.66
		3	1.50	1.16	4.13	0.42	0.34	14.54
	4	1	0.93	0.70	4.04	0.28	0.21	5.02
		2	0.96	0.78	5.37	0.07	0.05	12.93
		3	0.99	0.75	8.42	1.17	0.93	1.66
	9	1	0.84	0.63	3.54	0.08	0.05	3.18
		2	0.95	0.75	9.48	0.05	3.91	9.52
		3	1.06	0.83	3.98	0.61	0.55	3.77

**Table 3.1** : Scène du  $L$  – Évaluation de la planarité des points reconstruits

Méthode	Fréq. base	Plan	Imprécis			Précis		
			Moyenne	Écart type	% d'erreur	Moyenne	Écart type	% d'erreur
Mixte	1	1	2.43	1.92	7.38	0.30	0.24	11.89
		2	2.93	2.30	6.63	0.16	0.12	1.20
		3	2.08	1.61	12.07	0.07	0.05	12.50
	4	1	2.06	1.58	7.07	0.18	0.13	12.59
		2	3.13	2.42	3.18	0.17	0.12	1.20
		3	1.54	1.18	4.01	0.08	0.06	14.55
	9	1	1.99	1.49	7.61	0.18	0.15	12.41
		2	2.07	1.64	8.42	0.12	0.08	1.20
		3	1.33	1.05	7.97	0.09	0.07	10.71
Calibrage	1	1	1.65	1.27	8.82	0.16	0.13	13.98
		2	1.58	1.22	3.72	0.16	0.11	0.0
		3	1.20	0.92	5.77	0.08	0.06	12.73
	4	1	1.83	1.44	7.37	0.15	0.12	14.68
		2	1.56	1.24	5.58	0.15	0.11	2.41
		3	1.26	0.96	6.25	0.08	0.06	12.72
	9	1	2.23	1.65	6.90	0.15	0.12	15.38
		2	1.61	1.26	4.61	0.16	0.11	1.20
		3	1.25	0.95	7.51	0.08	0.05	12.72
Décompo.	1	1	2.43	1.92	7.38	0.30	0.24	11.89
		2	2.93	2.30	6.63	0.16	0.12	1.20
		3	2.08	1.61	12.07	0.07	0.05	12.50
	4	1	6.71	5.03	1.22	0.15	0.13	17.48
		2	14.14	9.90	1.35	0.16	0.12	7.22
		3	15.06	10.53	0.36	0.11	0.09	14.03
	9	1	8.54	6.92	6.58	0.51	0.58	40.75
		2	6.71	5.39	5.04	3.35	2.59	0.57
		3	6.35	4.78	4.98	0.88	0.78	14.70

**Table 3.2** : Scène de la mire – Évaluation de la planarité des points reconstruits

Fréquence base	Couple	Méthode		
		Mixte	Calibrage	Décompo.
1	1-2	1.40	0.33	1.40
	1-3	1.44	7.19	1.44
	2-3	1.39	2.58	1.39
4	1-2	0.07	0.39	2.28
	1-3	4.82	4.65	7.21
	2-3	0.86	0.41	1.03
9	1-2	2.02	3.31	4.96
	1-3	11.22	3.58	7.72
	2-3	0.30	4.98	2.55

**Table 3.3 :** Scène du  $L$  – Précision pixelle – Qualité des angles entre les différents plans reconstruits (erreurs en degré)

Fréquence base	Couple	Méthode		
		Mixte	Calibrage	Décompo.
1	1-2	1.96	0.30	1.96
	1-3	21.29	5.90	21.29
	2-3	6.20	0.15	6.20
4	1-2	0.83	6.45	0.70
	1-3	0.21	61.55	2.40
	2-3	2.69	4.27	5.40
9	1-2	35.45	4.57	1.03
	1-3	16.62	77.47	6.11
	2-3	49.21	1.08	2.26

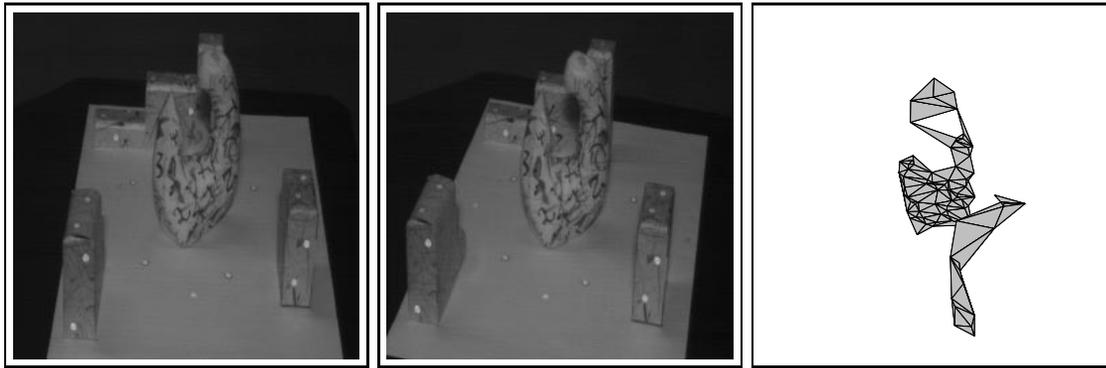
**Table 3.4 :** Scène du  $L$  – Précision subpixelle – Qualité des angles entre les différents plans reconstruits (erreurs en degré)

Fréquence base	Couple	Méthode		
		Mixte	Calibrage	Décompo.
1	1-2	0.18	0.36	0.18
	1-3	0.27	0.07	0.27
	2-3	0.48	0.22	0.48
4	1-2	0.60	0.50	3.40
	1-3	0.72	0.11	10.13
	2-3	0.99	0.13	26.15
9	1-2	1.22	0.24	8.39
	1-3	0.93	0.00	1.83
	2-3	0.48	0.76	0.54

**Table 3.5 :** Scène de la mire – Précision pixelle – Qualité des angles entre les différents plans reconstruits (erreurs en degré)

Fréquence base	Couple	Méthode		
		Mixte	Calibrage	Décompo.
1	1-2	0.51	0.16	0.51
	1-3	0.68	0.02	0.68
	2-3	0.01	0.12	0.01
4	1-2	0.03	0.23	0.39
	1-3	0.02	0.00	0.45
	2-3	0.10	0.08	0.19
9	1-2	0.14	0.30	17.91
	1-3	0.15	0.01	9.22
	2-3	0.23	0.08	10.74

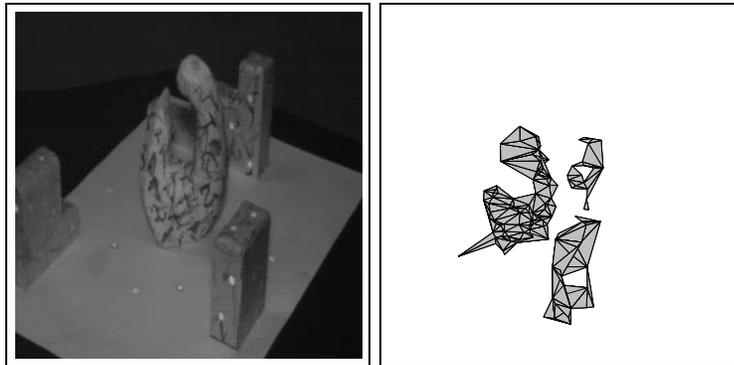
**Table 3.6 :** Scène de la mire – Précision subpixelle – Qualité des angles entre les différents plans reconstruits (erreurs en degré)



(a) Image 1

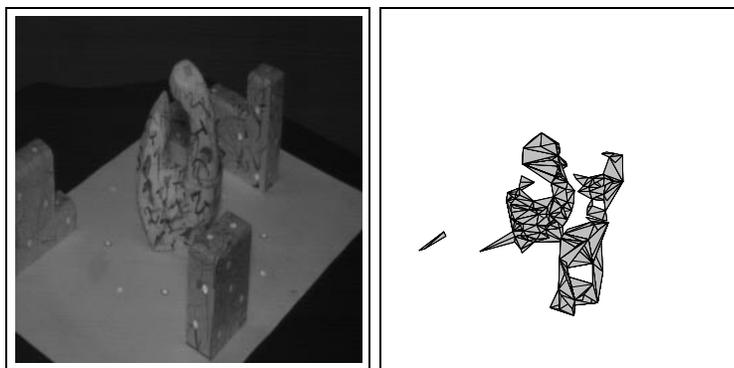
(b) Image 2

(c) Reconstruction de 1 à 2



(d) Image 3

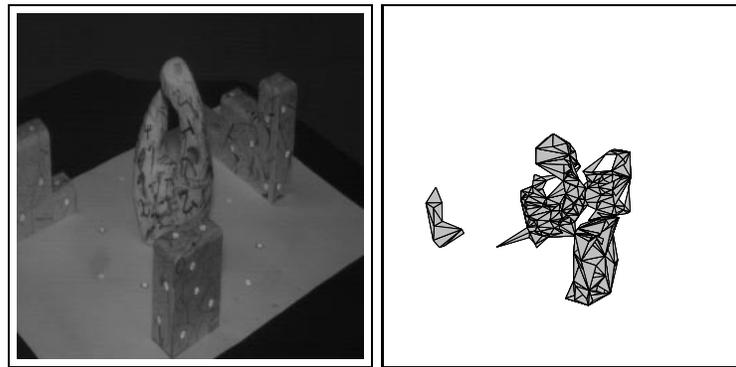
(e) Reconstruction de 1 à 3



(f) Image 4

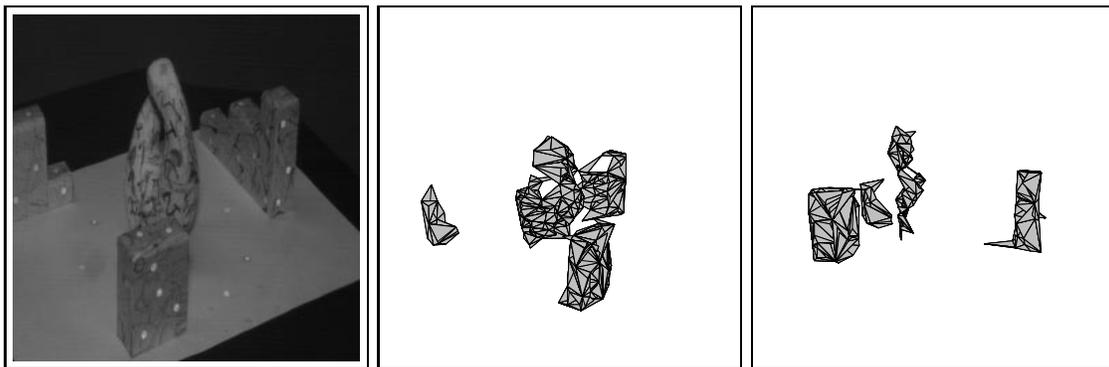
(g) Reconstruction de 1 à 4

**Figure 3.5** : Reconstruction de surfaces : la suite page suivante



(a) Image 5

(b) Reconstruction de 1 à 5



(c) Image 6

(d) Reconstruction de 1 à 6

(e) Reconstruction de 1 à 6,  
sous une autre vue**Figure 3.6** : Six vues d'une scène, et la surface reconstruite itérativement



# Conclusion

Ce travail s'est articulé autour du thème de la précision et des mesures qui peuvent être effectuées à partir d'images, au niveau de la localisation d'entités dans les images et de mesures tridimensionnelles de scènes. Plusieurs points ont été abordés sur ce thème.

Dans le premier chapitre, une méthode de correction géométrique des distorsions de la prise d'images par rapport au modèle sténopé généralement utilisé a été étudiée. Elle permet de se ramener au cas théorique pour des objectifs présentant des types de distorsions tout à fait quelconques. La méthode utilisée est simple à mettre en œuvre, et nécessite peu de matériels : une mire plane dont seuls les alignements de points sont connus. Aucune information tridimensionnelle est nécessaire. De nombreux tests ont montré que la méthode est stable dans le temps et dans l'espace avec une précision de 0.02 pixel. Cette correction de distorsion a été appliquée dans le second chapitre pour le repositionnement du centre de cibles circulaires. La validation de la correction des distorsions réalisée au chapitre 1 a été confirmée par différentes expérimentations complémentaires utilisant des données indépendantes de celles utilisées lors du calcul des distorsions. Il s'est avéré que la correction des distorsions doit être effectuée lors d'applications particulières utilisant des données extraites avec une haute précision dans l'image.

Cette méthode de correction des distorsions a également été utilisée par la société ITMI dans le cadre d'une application industrielle comprenant la comparaison de quatre objectifs de courte focale. Les résultats obtenus durant leurs expérimentations sont cohérents avec ceux présentés dans ce mémoire.

Le second chapitre a abordé la principale cause des erreurs de mesure : les erreurs de localisation de points dans les images. Deux types de points ont été étudiés : les points dont le signal dans l'image est modélisable et les points ayant une forte texture. La localisation du premier type de points s'appuie sur un modèle, et se concrétise par la recherche du meilleur ajustement entre ce modèle et l'image. Cet algorithme général a été appliqué à deux entités particulières : les coins et les cibles circulaires. De nombreux tests ont montré que la localisation de coins est précise à 1/10ème de pixel, et celle des cibles à 1/20ème de pixel.

La localisation de points généraux s'applique à des points dont la texture dans l'image est importante, mais dont le signal n'est pas particulier. Elle est réalisée par une mise en correspondance précise de points entre deux images, utilisant une fonction de corrélation, et prenant en compte une déformation affine de la fenêtre de corrélation. L'hypothèse sous-jacente à la déformation choisie est d'avoir une surface localement plane autour du point considéré. Des expérimentations ont validé une précision de l'ordre de 1/20ème de

pixel sur des scènes planes.

Le dernier chapitre a mis en évidence la qualité des mesures sur une scène en fonction de la précision de la localisation de points dans les images, et également en fonction du degré de confiance des positions tridimensionnelles de points permettant le passage d'une reconstruction projective vers une reconstruction euclidienne. Il s'est avéré que lorsque toutes les bonnes conditions sont réunies (localisations précises de points et bonne connaissance des données tridimensionnelles des points de la base), alors la précision de la reconstruction que l'on peut obtenir est de l'ordre de 1/2500ème de la taille de la scène.

Enfin, une méthode de reconstruction de surface, procédant par la reprojexion dans la scène de la topologie d'une triangulation de Delaunay obtenue dans les images, a été développée. L'évaluation qualitative des surfaces reconstruites a été réalisée sur une scène comportant quatre objets, dont un non-polygonal. Le bon déroulement de la fusion de nouvelles vues et le bon rendu des parties concaves des objets ont été concluants.

Ces deux caractéristiques de la reconstruction d'une scène, que sont la précision et la reconstruction d'une surface, permettent d'envisager dans l'avenir la prise par une main articulée d'un objet reconstruit par cette méthode.

Durant tout ce travail, l'utilisation de la méthode des moindres carrés médians a permis d'obtenir une grande robustesse lors des recherches de modèles face aux erreurs grossières. L'utilisation de cette méthode s'est notamment avérée indispensable pour la reconstruction de surfaces.

Les travaux abordés dans ce mémoire suscitent quelques développements qu'il serait intéressant de réaliser :

- Au niveau de la correction des distorsions optiques de la prise d'images, quelques améliorations mineures peuvent être apportées dans le cas de systèmes à forte distorsion. Dans un tel cas, l'hypothèse d'avoir des distorsions constante localement peut s'avérer insuffisante. L'utilisation de fonctions d'interpolations plus complexes que l'interpolation bi-linéaire peut s'avérer utile.
- La localisation de coins en  $Y$  (comportant trois branches) peut s'effectuer par l'algorithme de localisation de points modélisables développé au chapitre 2. Néanmoins, cette application n'est pas sans problème. La caractérisation du modèle est plus complexe, et s'ajuste généralement moins bien avec l'image.
- La mise en correspondance précise de points qui a été validée dans le chapitre 2 n'a pas tenu compte des éventuelles occultations. Une généralisation de cet algorithme paraît donc nécessaire afin de prendre en compte ce phénomène qui apparaît fréquemment.
- La reconstruction de la surface telle qu'elle est décrite mérite quelques approfondissements, notamment pour la détection des parties concaves des objets. L'utilisation des contours des objets dans l'image ainsi que des points observés derrière certaines facettes doit permettre une meilleure détection des concavités.

Plusieurs perspectives de travail s'offrent à la suite de cette thèse, tant au niveau théorique qu'au niveau expérimental :

- Aucun banc d'expérimentations complet permettant la validation d'algorithmes quelques soient la configuration des prises de vues et les objets observés existe à l'heure actuelle. Il est pourtant important de prévoir la réaction de chaque méthode suivant le type d'images à disposition, et ceci de façon systématique. Par exemple, dans cette thèse, aucune réponse n'a été apportée sur la dégradation de la qualité des mises en correspondance en fonction de la courbure de la surface observée. La création d'un banc d'expérimentations apparaît donc comme une tâche utile pour l'ensemble de la communauté de la vision par ordinateur.
- Dans ce mémoire, la précision de la reconstruction a été étudiée en fonction de la méthode de la reconstruction et de la précision de la localisation des points dans les images. Cependant, le déplacement de la caméra n'a pas été pris en compte, alors qu'il est un facteur important dans la précision de la reconstruction. De fait, le mouvement de la caméra était imposé par l'utilisateur. Un point important semble donc le calcul du meilleur mouvement de la caméra, à la fois pour augmenter la précision de la reconstruction et également pour compléter la reconstruction de la scène. De plus, un critère pour l'arrêt de la reconstruction d'une scène (par exemple lorsque la précision de la reconstruction des points a atteint sa limite, ou lorsque la scène entière a été reconstruite) est également nécessaire.
- La reconstruction de la surface a été décrite dans l'univers euclidien. Cependant, une reconstruction projective ou affine suffit à certaines applications. La généralisation de la reconstruction de surfaces à partir d'une reconstruction de points plus générale qu'une reconstruction euclidienne est une voie de recherche.



# Annexes



## Annexe A

# Ajustement robuste de modèle

L'ajustement d'un modèle sur des données est un outil qui a été utilisé tout au long de cette thèse. Par exemple, l'estimation d'une droite (chapitre 1) ou celle d'une homographie de  $\mathbb{P}^3$  dans  $\mathbb{P}^3$  (chapitre 3) ont été effectuées. Dans le cas général, l'ajustement d'un modèle sur  $n$  données  $(\mathbf{x}_i, y_i)$  consiste à minimiser les valeurs des résidus  $r_i = y_i - f(\mathbf{x}_i, \Theta)$ .

Un modèle est habituellement estimés par la méthode des moindres carrés qui consiste à minimiser

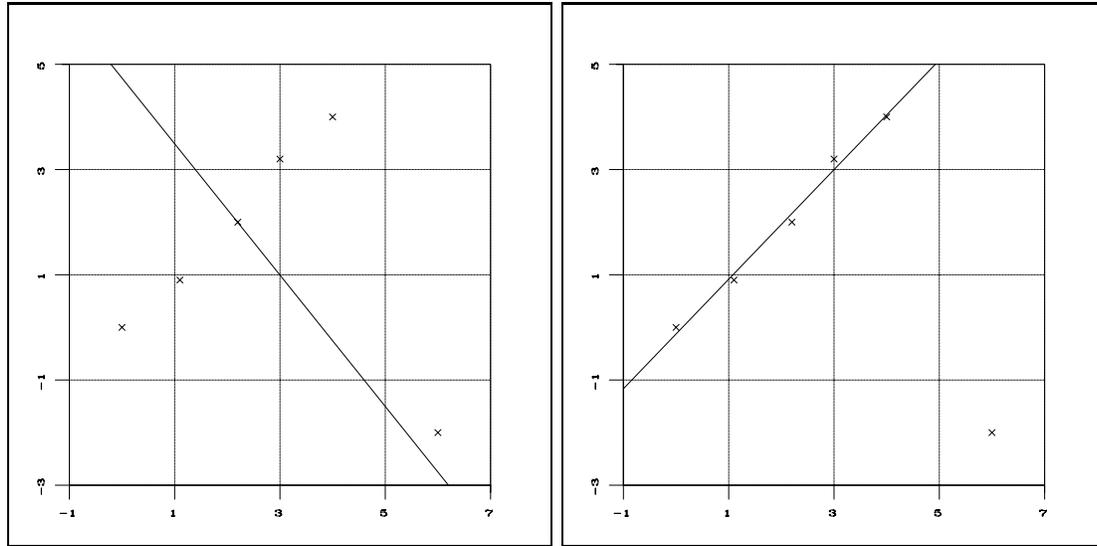
$$\chi^2 = \min_{\Theta} \sum_{i=1}^n r_i^2$$

Cette minimisation est simple à réaliser. De plus, la solution obtenue est optimale pour des données dont les erreurs suit une loi normale de moyenne nulle, de même écart-type, et qui ne sont pas corrélées les unes aux autres [PFTW88].

Malheureusement, même s'il est admis généralement que la majorité des  $n$  données  $(\mathbf{x}_i, y_i)$  ont une erreur qui suivent cette loi, il y a très fréquemment quelques éléments qui sont en dehors de cette distribution et qui sont totalement erronés. Ceci est d'autant plus vrai que ces données sont obtenues par des processus de vision par ordinateur qui ne sont pas toujours fiables. L'exemple le plus frappant dans ce domaine est celui de la mise en correspondance de points entre deux images : aboutir à 90% d'appariements corrects est considéré comme remarquable.

Or, l'ajustement d'un modèle par la méthode des moindres carrés n'est pas robuste aux données erronées. Ainsi, un seul couple  $(x_j, y_j)$  erroné peut faire échouer l'estimation d'un modèle. De plus, il est ensuite impossible d'isoler ce couple [För87]. La figure A.1(a) illustre ce défaut des moindres carrés sur un exemple : l'estimation d'une droite étant donné six points dont un seul est mal positionné.

Cette annexe introduit deux algorithmes d'estimation robuste de modèle. Dans un premier temps, la méthode des moindres carrés médians est détaillée car elle est appliquée dans cette thèse. Ensuite, les M-estimateurs sont abordés car ils sont également souvent utilisés dans le domaine de la vision par ordinateur.



(a) Ajustement d'une droite avec la méthode des lagrangiens [Fau93] (basée sur les moindres carrés)

(b) Ajustement d'une droite avec la méthode robuste des moindres carrés médians

**Figure A.1 :** Ajustement d'une droite  $ax + by + c = 0$  de sorte que  $a^2 + b^2 = 1$  en utilisant la méthode des lagrangiens ou une méthode robuste

## A.1 Moindres carrés médians

La méthode des moindres carrés médians permet l'ajustement de modèle robuste aux erreurs grossières. Cette méthode est basée sur les mêmes concepts que l'algorithme RANSAC <sup>1</sup> [FB80], et se décompose en deux étapes :

1. élimination des données qui ne suivent pas le modèle recherché.
2. recherche des paramètres du modèle par la méthode des moindres carrés, appliquée seulement aux données vérifiant le modèle.

L'élimination des données erronées est réalisée par une simulation de Monte-Carlo. Soit  $p$  le nombre minimal d'informations permettant de déterminer les paramètres du modèle. Par exemple,  $p = 2$  pour ajuster une droite, et  $p = 6$  lors du calibrage d'une caméra car 6 points sont nécessaires pour effectuer cette tâche. La simulation de Monte-Carlo consiste à prendre tous les  $p$ -uplets de données  $\{(\mathbf{x}_i, y_i)\}_j$  et de calculer les paramètres  $\Theta_j$  du modèle correspondant à ce jeu de données. Pour chaque  $\Theta_j$  est alors calculé le médian  $m_j$  des résidus au carré  $r_{i,j}^2 = (y_i - f(\mathbf{x}_i, \Theta_j))^2$ . Le modèle  $\Theta_k$  tel que  $k = \underset{j}{\operatorname{argmin}} m_j$  est alors considéré comme obtenu sans aucune donnée erronée. Ces paramètres sont donc une bonne approximation des véritables paramètres du modèle recherché.

<sup>1</sup>RANSAC = RANdom SAmple Consensus

Une estimation  $\hat{\sigma}$ , robuste aux erreurs grossières, de l'écart-type des résidus obtenus avec le modèle  $\Theta_k$  est alors calculée par [RL87, MMRK91, MM91] :

$$\hat{\sigma} = 1.4826 \left(1 + \frac{5}{n-p}\right) \sqrt{\text{med } r_{i,k}^2}$$

Cette estimation de l'écart-type de la distribution des résidus est fiable sous l'hypothèse que la distribution des erreurs sur les données suit une loi normale. En effet, la valeur 1.4826 compense le biais induit par le médian en présence de bruit gaussien (son inverse est en fait la moitié de la valeur de l'interquartile de la distribution gaussienne de moyenne 0 et d'écart-type 1). Cette estimation robuste  $\hat{\sigma}$  peut tolérer jusqu'à 50% de données ne suivant pas le modèle recherché.

La seconde phase consiste à affiner le modèle trouvé en utilisant toutes les données  $(\mathbf{x}_i, y_i)$  fiables. Pour cela, il est associé un poids  $w_i$  à chaque couple  $(\mathbf{x}_i, y_i)$  défini par :

$$w_i = \begin{cases} 1 & \text{si } |r_{i,k}| \leq c\hat{\sigma} \text{ (cette donnée est conforme au modèle recherché)} \\ 0 & \text{sinon (cette donnée est erronée)} \end{cases}$$

La constante  $c$  est généralement égale à 2.5, ce qui correspond à prendre près de 99% des données dont l'erreur suit une loi normale. Cette constante est choisie de sorte à prendre le maximum de données exactes tout en minimisant le nombre de couples  $(\mathbf{x}_i, y_i)$  ne suivant pas le modèle recherché et ayant un poids  $w_i$  égal à 1. La valeur des paramètres du modèle ajusté est alors obtenue par minimisation aux moindres carrés pondérés (cela revient à réaliser une minimisation par moindres carrés seulement avec les données fiables).

Le nombre de  $p$ -uplets à tester est égal à  $C_n^p$ , ce qui est trop important et limite l'algorithme. Cependant, si l'on s'autorise une probabilité d'erreur égale à  $Q$ , il suffit alors de prendre aléatoirement seulement  $q$   $p$ -uplets, où  $q$  est défini par

$$Q \leq [1 - (1 - \varepsilon)^p]^q$$

$\varepsilon$  étant la proportion de données erronées ( $\varepsilon < 0.5$ ). On peut remarquer que  $q$  ne dépend pas du nombre de données à ajuster.

L'algorithme A.1 résume cette méthodologie pour l'estimation robuste d'un modèle. Il est à noter que cette méthode supporte jusqu'à 50% de données erronées.

## A.2 M-estimateurs

L'ajustement d'une fonction à des données utilisant la méthode des M-estimateurs est une généralisation de la méthode des moindres carrés. Cela s'exprime par la minimisation de

$$\chi^2 = \min_{\Theta} \sum_{i=1}^n \rho(r_i) \tag{A.1}$$

où  $\rho(\cdot)$  est une fonction dépendant du M-estimateur choisi. La minimisation de l'équation A.1 peut se résoudre par une minimisation itérative aux moindres carrés pondérés :

$$\chi^2 = \min_{\Theta} \sum_i w_i^2 \cdot (y_i - f(\mathbf{x}_i, \Theta))^2$$

---



---

Entrée :  $\{(x_i, y_i)\}$ , les  $n$  données à ajuster  
 Sortie :  $\Theta$ , les paramètres ajustant au mieux les données précédentes

Calculer $q$ , nombre de $p$ -uplets à tirer Faire $q$ fois <table border="1" style="border-collapse: collapse; margin-left: 20px;"> <tr> <td style="padding: 5px;">           Tirer aléatoirement un <math>p</math>-uplet parmi les <math>n</math> données <math>\{(x_i, y_i)\}</math>            Calculer les paramètres <math>\Theta_j</math> qui ajuste au mieux les données de ce <math>p</math>-uplet            Calculer le médian <math>m_j</math> des carrés des résidus obtenus avec ce modèle         </td> </tr> </table> Choisir $j$ tel que $m_j = \min m_k$ Calculer $\hat{\sigma}$ ainsi que les poids $w_i$ Calculer $\Theta$ par les moindres carrés pondérés	Tirer aléatoirement un $p$ -uplet parmi les $n$ données $\{(x_i, y_i)\}$ Calculer les paramètres $\Theta_j$ qui ajuste au mieux les données de ce $p$ -uplet Calculer le médian $m_j$ des carrés des résidus obtenus avec ce modèle
Tirer aléatoirement un $p$ -uplet parmi les $n$ données $\{(x_i, y_i)\}$ Calculer les paramètres $\Theta_j$ qui ajuste au mieux les données de ce $p$ -uplet Calculer le médian $m_j$ des carrés des résidus obtenus avec ce modèle	

---



---

**Algorithme A.1 :** Ajustement d'un modèle par les moindres carrés médians

Les poids choisis sont alors de la forme [MMRK91]

$$w_i = \frac{1}{r_i} \frac{d\rho(r_i)}{dr_i}$$

À chaque pas d'itérations, les paramètres du modèle sont estimés et s'il n'y a pas convergence, une nouvelle estimation est effectuée en modifiant les poids avec le nouveau modèle obtenu. La mesure de la qualité de l'ajustement utilisée comme critère d'arrêt lors de l'itération  $l$  est

$$\varepsilon^{(l)} = \sqrt{\frac{\sum w_i r_i^2}{\sum w_i}}$$

La convergence de l'algorithme est atteinte lorsque  $|\varepsilon^{(l)} - \varepsilon^{(l-1)}| < 0.001$  [MMRK91, JMB91]. L'algorithme A.2 présente la méthode des M-estimateurs dans son ensemble.

Différentes fonctions  $\rho(\cdot)$  peuvent être utilisées. La fonction la plus répandue est la fonction de Tukey qui s'écrit [HJL<sup>+</sup>89]

$$\rho(r_i) = \begin{cases} \frac{1}{6} \left[ 1 - \left( 1 - (r_i/c\hat{\sigma})^2 \right)^3 \right] & \text{si } |r_i| \leq c\hat{\sigma} \\ \frac{1}{6} & \text{sinon} \end{cases}$$

où  $c$  est une constante. Sa valeur est généralement égale à 4.685. Celle-ci permet d'optimiser les performances de cette estimation dans le cas d'un bruit gaussien blanc (cette valeur a été obtenue par simulation de Monte-Carlo sur des données synthétiques).

Cette fonction  $\rho(\cdot)$  dépend également de la valeur de l'estimation robuste  $\hat{\sigma}$  de l'écart-type de la distribution des résidus. Cette estimation est égale à l'expression <sup>2</sup> [HRRS86,

---

<sup>2</sup>MAD = Mean Absolute Deviation

---



---

Entrée :  $\{(\mathbf{x}_i, y_i)\}$ , les  $n$  données à ajuster  
 Sortie :  $\Theta$ , les paramètres ajustant au mieux les données précédentes

```

  Initialisation des paramètres  $\Theta$  du modèle
  Tant que il n'y a pas convergence
    Calcul des poids  $w_i$  de chaque données  $(\mathbf{x}_i, y_i)$  en fonction
      des paramètres du modèle  $\Theta$ 
    Estimation du modèle  $\Theta$  par les moindres carrés pondérés
    Calcul du critère de qualité  $\varepsilon^{(l)}$ 
  
```

---



---

**Algorithme A.2** : Ajustement d'un modèle par les M-estimateurs

JMB91]

$$\hat{\sigma} = 1.4826 \text{ MAD}(r_i) = 1.4826 \underset{i}{\text{med}} |r_i - \underset{i}{\text{med}} r_i|$$

Ce choix de  $\rho(\cdot)$  permet de minimiser l'effet, dans la minimisation de l'équation A.1, des données erronées en limitant la valeur de  $\rho(r_i)$  pour les résidus trop importants. Ceci rend l'estimation de modèle robuste aux erreurs grossières, contrairement à la méthode des moindres carrés. Les poids associés à la fonction de Tukey sont définis par

$$w_i = \begin{cases} [1 - (r_i/c\hat{\sigma})^2]^2 & \text{si } |r_i| \leq c\hat{\sigma} \\ 0 & \text{sinon} \end{cases}$$

Cet estimateur a été appliqué avec succès dans plusieurs processus de vision par ordinateur, par exemple pour le calcul de pose [HJL<sup>+</sup>89] ou pour la classification [JMB91]. L'avantage qu'il possède par rapport à la méthode des moindres carrés médians est la rapidité d'exécution. En effet, il n'y a pas de simulation de Monte-Carlo qui est très coûteuse en temps de calcul.

Néanmoins, son principal défaut réside dans la difficulté pour trouver le minimum global de l'équation A.1. Lorsque la minimisation est effectuée itérativement en utilisant la méthode des moindres carrés pondérés, cette difficulté se traduit par la recherche d'une bonne approximation du modèle, et donc par l'initialisation des poids  $w_i$ . Même s'il est prouvé que la méthode converge toujours pour la fonction de Tukey, elle peut néanmoins converger vers un minimum local [BBW88, RL90]. L'article [KH94] montre que pour le calcul de pose, la solution obtenue est parfois très bonne malgré les fausses mises en correspondance, mais elle peut également être totalement erronée lorsque la solution initiale est trop éloignée de la véritable solution.

Généralement, l'initialisation des poids est donnée par  $w_i = 1$ . Elle n'est donc pas robuste aux fausses données. Meer *et al.* [MMRK91] proposent d'initialiser les valeurs des paramètres par la méthode des moindres carrés médians, puis d'affiner l'ajustement obtenue par un M-estimateur. L'exécution de ce processus est alors fortement ralentie.



## Annexe B

# Résultats de mises en correspondance précises

Cette annexe présente tous les tests réalisés au chapitre 2, dans la sous-section 2.4.4 page 60. Ces expérimentations ont eu pour but de valider la méthode de mise en correspondance précise de points entre deux images.

Les tableaux B.1 à B.6 présentent la qualité de la conservation des alignements de points dans l'image (b) (se reporter à la figure 2.15 qui montre les trois images utilisées). Pour chaque tableau, la taille de la fenêtre de corrélation est fixe (de  $5 \times 5$  pixels pour la table B.1 à  $15 \times 15$  pixels pour la table B.6).

Pour chacune des deux géométries épipolaires (précise ou bruitée), et suivant la paramétrisation de la transformation affine utilisée (3, 4 ou 6 paramètres), les statistiques suivantes ont été calculées :

- distance algébrique moyenne d'un point à la droite ajustée.
- distance moyenne d'un point à la droite ajustée.
- écart-type des distances.
- pourcentage des erreurs grossières. Cette valeur est fournie par la méthode des moindres carrés médians.

Les distances ne sont pas calculées sur les points erronés afin de ne pas perturber les statistiques.

La seconde série de tableaux (B.7 à B.12) contient les précisions de positionnement des points trouvés. Dans chaque tableau, la taille de la fenêtre de corrélation est fixée. Suivant la géométrie épipolaire utilisée (précise ou bruitée), la paramétrisation de la transformation affine (3, 4 ou 6) ainsi que le couple d'images considéré (images (a) et (b) pour la translation, et images (a) et (c) pour la rotation), les statistiques suivantes sont fournies :

- la moyenne, suivant les axes  $x$  et  $y$  de l'image, des différences entre les coordonnées obtenues et celles attendues (moyenne signée).
- la moyenne, suivant les axes  $x$  et  $y$ , de la valeur absolue des différences entre les coordonnées obtenues et celles attendues.
- la moyenne des distances entre les points obtenus et ceux prédits par les homographies  $\mathbf{H}^b$  et  $\mathbf{H}^c$ .
- l'écart-type de ces distances
- le pourcentage d'erreurs grossières

Toutes ces statistiques ont également été calculées sans prendre en compte les points erronés.

Enfin, la dernière colonne des tables B.7 à B.12 comprend les résultats des tests de la *vérification duale* (lire le texte page 66).

Géométrie épipolaire	Param	Distance algébrique moyenne	Distance moyenne	Écart type	% d'erreurs
Précise	3	0.00	0.05	0.07	13.36
	4	0.00	0.06	0.07	16.51
	6	0.00	0.08	0.10	18.16
Bruitée	3	0.00	0.12	0.17	21.34
	4	0.00	0.06	0.08	19.22
	6	0.00	0.08	0.10	20.52

**Table B.1** : Test d'alignements : taille de la fenêtre de corrélation :  $5 \times 5$

Géométrie épipolaire	Param	Distance algébrique moyenne	Distance moyenne	Écart type	% d'erreurs
Précise	3	0.00	0.04	0.05	8.25
	4	0.00	0.04	0.05	8.92
	6	0.00	0.04	0.06	12.70
Bruitée	3	0.00	0.08	0.10	17.80
	4	0.00	0.04	0.05	9.95
	6	0.00	0.04	0.06	12.58

**Table B.2** : Test d'alignements : taille de la fenêtre de corrélation :  $7 \times 7$

Géométrie épipolaire	Param	Distance algébrique moyenne	Distance moyenne	Écart type	% d'erreurs
Précise	3	0.00	0.03	0.04	7.07
	4	0.00	0.03	0.04	6.05
	6	0.00	0.03	0.04	8.41
Bruitée	3	0.00	0.06	0.08	16.27
	4	0.00	0.03	0.04	6.48
	6	0.00	0.03	0.04	8.57

**Table B.3** : Test d'alignements : taille de la fenêtre de corrélation :  $9 \times 9$

Géométrie épipolaire	Param	Distance algébrique moyenne	Distance moyenne	Écart type	% d'erreurs
Précise	3	0.00	0.03	0.03	5.35
	4	0.00	0.02	0.03	4.64
	6	0.00	0.03	0.03	5.90
Bruitée	3	0.00	0.05	0.07	15.72
	4	0.00	0.02	0.03	5.62
	6	0.00	0.03	0.03	5.82

Table B.4 : Test d'alignements : taille de la fenêtre de corrélation :  $11 \times 11$ 

Géométrie épipolaire	Param	Distance algébrique moyenne	Distance moyenne	Écart type	% d'erreurs
Précise	3	0.00	0.02	0.03	4.44
	4	0.00	0.02	0.03	3.18
	6	0.00	0.02	0.03	4.36
Bruitée	3	0.00	0.04	0.05	16.08
	4	0.00	0.02	0.03	3.22
	6	0.00	0.02	0.03	4.32

Table B.5 : Test d'alignements : taille de la fenêtre de corrélation :  $13 \times 13$ 

Géométrie épipolaire	Param	Distance algébrique moyenne	Distance Moyenne	Écart type	% d'erreurs
Précise	3	0.00	0.02	0.03	3.97
	4	0.00	0.02	0.02	1.76
	6	0.00	0.02	0.03	3.02
Bruitée	3	0.00	0.04	0.05	15.56
	4	0.00	0.02	0.03	2.31
	6	0.00	0.02	0.03	3.11

Table B.6 : Test d'alignements : taille de la fenêtre de corrélation :  $15 \times 15$

Géométrie épipolaire	Param	Moyenne signée		Moyenne			Écart type	% d'erreurs grossières	Vérif. duale
		en $x$	en $y$	en $x$	en $y$	dist			
Translation									
Précise	3	-0.09	-0.02	0.10	0.06	0.14	0.07	6.57	0.02
	4	-0.02	0.00	0.07	0.06	0.11	0.07	13.98	0.04
	6	-0.02	0.00	0.09	0.07	0.14	0.11	16.03	0.06
Bruitée	3	0.25	-0.05	0.31	0.14	0.42	0.31	10.68	0.03
	4	-0.02	0.00	0.07	0.06	0.11	0.08	17.40	0.05
	6	-0.02	0.00	0.09	0.08	0.15	0.11	18.64	0.07
Rotation									
Précise	3	-0.11	0.07	0.11	0.08	0.15	0.08	48.53	0.14
	4	-0.03	0.06	0.08	0.08	0.13	0.09	48.80	0.12
	6	-0.02	0.06	0.14	0.11	0.18	0.15	49.06	0.22
Bruitée	3	0.61	0.05	0.65	0.17	0.58	0.24	32.27	0.23
	4	-0.02	0.05	0.08	0.08	0.12	0.07	34.67	0.13
	6	-0.01	0.05	0.13	0.10	0.17	0.11	34.13	0.24

**Table B.7** : Test de positionnement : taille de la fenêtre de corrélation :  $5 \times 5$

Géométrie épipolaire	Param	Moyenne signée		Moyenne			Écart type	% d'erreurs grossières	Vérif. duale
		en $x$	en $y$	en $x$	en $y$	dist			
Translation									
Précise	3	-0.10	-0.02	0.10	0.05	0.13	0.05	1.92	0.01
	4	-0.02	0.00	0.05	0.04	0.08	0.05	6.30	0.03
	6	-0.02	0.00	0.06	0.05	0.09	0.06	9.49	0.04
Bruitée	3	0.26	-0.05	0.32	0.11	0.40	0.28	2.94	0.02
	4	-0.02	0.00	0.05	0.04	0.08	0.05	6.71	0.03
	6	-0.02	0.01	0.06	0.05	0.09	0.06	9.38	0.04
Rotation									
Précise	3	-0.11	0.07	0.11	0.07	0.14	0.06	45.87	0.11
	4	-0.03	0.06	0.06	0.07	0.10	0.06	46.67	0.11
	6	-0.03	0.06	0.07	0.08	0.12	0.07	46.40	0.11
Bruitée	3	0.63	0.05	0.64	0.14	0.56	0.19	23.20	0.10
	4	-0.03	0.06	0.06	0.07	0.10	0.05	22.40	0.05
	6	-0.02	0.06	0.07	0.08	0.12	0.07	22.93	0.12

**Table B.8** : Test de positionnement : taille de la fenêtre de corrélation :  $7 \times 7$

Géométrie épipolaire	Param	Moyenne signée		Moyenne			Écart type	% d'erreurs grossières	Vérif. duale
		en $x$	en $y$	en $x$	en $y$	dist			
Translation									
Précise	3	-0.10	-0.02	0.10	0.04	0.12	0.05	0.58	0.01
	4	-0.02	0.00	0.04	0.03	0.06	0.04	3.12	0.02
	6	-0.02	0.01	0.04	0.04	0.07	0.04	5.79	0.03
Bruitée	3	0.27	-0.04	0.33	0.10	0.38	0.27	0.96	0.02
	4	-0.02	0.00	0.04	0.03	0.06	0.04	3.32	0.02
	6	-0.02	0.01	0.04	0.03	0.07	0.04	5.69	0.03
Rotation									
Précise	3	-0.11	0.07	0.11	0.07	0.14	0.06	45.33	0.11
	4	-0.03	0.06	0.05	0.07	0.09	0.04	46.13	0.08
	6	-0.03	0.06	0.06	0.07	0.10	0.05	46.13	0.11
Bruitée	3	0.63	0.05	0.64	0.13	0.56	0.20	19.20	0.11
	4	-0.03	0.06	0.04	0.07	0.09	0.04	20.26	0.09
	6	-0.02	0.06	0.05	0.07	0.10	0.05	19.73	0.04

Table B.9 : Test de positionnement : taille de la fenêtre de corrélation :  $9 \times 9$ 

Géométrie épipolaire	Param	Moyenne signée		Moyenne			Écart type	% d'erreurs grossières	Vérif. duale
		en $x$	en $y$	en $x$	en $y$	dist			
Translation									
Précise	3	-0.10	-0.02	0.10	0.04	0.12	0.04	0.27	0.01
	4	-0.02	0.00	0.03	0.03	0.05	0.03	1.51	0.02
	6	-0.02	0.01	0.04	0.03	0.06	0.03	3.53	0.02
Bruitée	3	0.28	-0.05	0.33	0.09	0.37	0.27	0.48	0.02
	4	-0.02	0.00	0.03	0.03	0.05	0.03	1.47	0.02
	6	-0.02	0.01	0.04	0.03	0.06	0.03	3.29	0.02
Rotation									
Précise	3	-0.11	0.07	0.11	0.07	0.14	0.05	45.07	0.08
	4	-0.03	0.06	0.04	0.06	0.08	0.03	46.13	0.06
	6	-0.03	0.06	0.05	0.07	0.09	0.04	46.40	0.08
Bruitée	3	0.64	0.05	0.64	0.12	0.54	0.17	20.00	0.03
	4	-0.03	0.06	0.04	0.06	0.08	0.03	20.53	0.02
	6	-0.03	0.06	0.04	0.07	0.09	0.04	20.00	0.03

Table B.10 : Test de positionnement : taille de la fenêtre de corrélation :  $11 \times 11$

Géométrie épipolaire	Param	Moyenne signée		Moyenne			Écart type	% d'erreurs grossières	Vérif. duale
		en $x$	en $y$	en $x$	en $y$	dist			
Translation									
Précise	3	-0.10	-0.02	0.10	0.04	0.11	0.04	0.31	0.01
	4	-0.02	0.00	0.03	0.03	0.05	0.02	0.65	0.02
	6	-0.02	0.01	0.03	0.03	0.05	0.03	1.68	0.02
Bruitée	3	0.28	-0.05	0.33	0.09	0.37	0.26	0.21	0.02
	4	-0.02	0.00	0.03	0.03	0.05	0.02	0.68	0.02
	6	-0.02	0.01	0.03	0.03	0.05	0.03	1.71	0.02
Rotation									
Précise	3	-0.11	0.07	0.11	0.07	0.14	0.05	45.60	0.07
	4	-0.03	0.06	0.04	0.06	0.08	0.03	45.33	0.07
	6	-0.03	0.06	0.04	0.07	0.09	0.03	45.60	0.06
Bruitée	3	0.63	0.06	0.65	0.11	0.55	0.18	19.73	0.03
	4	-0.03	0.06	0.04	0.06	0.08	0.03	20.00	0.02
	6	-0.03	0.07	0.04	0.07	0.09	0.03	20.00	0.03

**Table B.11** : Test de positionnement : taille de la fenêtre de corrélation :  $13 \times 13$

Géométrie épipolaire	Param	Moyenne signée		Moyenne			Écart type	% d'erreurs grossières	Vérif. duale
		en $x$	en $y$	en $x$	en $y$	dist			
Translation									
Précise	3	-0.10	-0.02	0.10	0.04	0.11	0.04	0.07	0.01
	4	-0.02	0.01	0.03	0.02	0.04	0.02	0.14	0.02
	6	-0.02	0.01	0.03	0.03	0.04	0.02	0.96	0.02
Bruitée	3	0.27	-0.05	0.33	0.08	0.36	0.26	0.14	0.02
	4	-0.02	0.01	0.03	0.03	0.04	0.02	0.17	0.02
	6	-0.02	0.01	0.03	0.03	0.04	0.02	0.96	0.02
Rotation									
Précise	3	-0.11	0.07	0.11	0.07	0.14	0.05	46.67	0.06
	4	-0.03	0.06	0.03	0.07	0.08	0.03	46.40	0.05
	6	-0.03	0.07	0.04	0.07	0.09	0.03	46.13	0.06
Bruitée	3	0.64	0.06	0.68	0.11	0.54	0.16	22.13	0.04
	4	-0.03	0.06	0.03	0.06	0.08	0.03	22.13	0.02
	6	-0.03	0.07	0.04	0.07	0.08	0.03	22.40	0.03

**Table B.12** : Test de positionnement : taille de la fenêtre de corrélation :  $15 \times 15$



# Bibliographie

- [Ack84] F. Ackermann. Digital image correlation : performance and potential application in photogrammetry. *Photogrammetric Record*, 64(11):429–439, October 1984.
- [AG92] P. Aschwanden and W. Guggenbühl. Experimental results from a comparative study on correlation-type registration algorithms. In Förstner and Ruwiedel, editors, *Robust Computer Vision*, pages 268–282. Wichmann, 1992.
- [BA90] C.B. Bose and I. Amir. Design of fiducials for accurate registration using machine vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(12):1196–1200, December 1990.
- [BB82] D.H. Ballard and C.M. Brown. *Computer Vision*. Prentice Hall, 1982.
- [BBW88] P.J. Besl, J.B. Birch, and L.T. Watson. Robust Window Operators. In *Proceedings of the 2nd International Conference on Computer Vision, Tampa, Florida, USA*, pages 591–600, 1988.
- [BC92] R. A. Boie and I. J. Cox. An analysis of camera noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):671–674, June 1992.
- [BdPS94] P. Bobet, S. de Paoli, and T. Skordas. Projet IARES S/S Perception. Tâche 4.4: Calibrage de Capteurs de Perception - Complément d'Étude. Technical Report 2h3209/IARES/DT-T4.4-017/V1.0/TS, ITMI APTOR, September 1994.
- [Bea78] P.R. Beaudet. Rotationally invariant image operators. In *Proceedings of the 4th International Joint Conference on Pattern Recognition, Tokyo*, pages 579–583, 1978.
- [Bey91] H. Beyer. Photogrammetric on-line inspection for car crash analysis. In *Proceedings of First Australian Photogrammetric Conference*, Sydney, Australia, 1991.
- [Bey92a] H.A. Beyer. Accurate calibration of CCD cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 96–101, 1992.
- [Bey92b] H.A. Beyer. *Geometric and Radiometric Analysis of a CCD-Camera Based Photogrammetric Close-Range System*. PhD thesis, ETH-Zurich, 1992.

- [Bey95] H.A. Beyer. Digital photogrammetry in industrial applications. In *ISPRS Intercommission Workshop: Sensors, Algorithms and Systems*, pages 373–378, Zurich, Switzerland, March 1995.
- [BLMB<sup>+</sup>95] C. Bard, C. Laugier, C. Milsi-Bellier, J. Troccaz, B. Triggs, and G. Vercelli. Achieving dextrous grasping by integrating planning and vision based sensing. *The International Journal of Robotics Research*, 1995. to appear.
- [BMV93] B. Boufama, R. Mohr, and F. Veillon. Euclidean constraints for uncalibrated reconstruction. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, pages 466–470, May 1993.
- [Bou94] B. Boufama. *Reconstruction Tridimensionnelle en Vision par Ordinateur : cas des caméra non étalonnées*. Thèse de doctorat, Institut National Polytechnique de Grenoble, December 1994.
- [BP91] D.A. Butler and P.K. Pierson. A distortion-correction scheme for industrial machine-vision applications. *IEEE Transactions on Robotics and Automation*, 7(4):546–551, August 1991.
- [Bro66] D.C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering and Remote Sensing*, pages 444–462, May 1966.
- [BZM94] P. Beardsley, A. Zisserman, and D. Murray. Sequential Update of Projective and Affine Structure from Motion. Technical Report 2012/94, University of Oxford, Oxford, United Kingdom, August 1994.
- [CB92] R. Cipolla and A. Blake. Surface Shape from the Deformation of Apparent Contours. *International Journal of Computer Vision*, 9(2):83–112, 1992.
- [CFW95] D.H. Coe, J. Barry Fallon, and R.L. West. Finding the center : using incidence to recover geometric features from single, monocular views. In *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, Xian, China*, pages 372–375, April 1995.
- [CG93] A. Cumani and A. Guiducci. A new camera calibration method for high accuracy non-contact metrology. *Pattern Recognition Letters*, 14(5):415–419, May 1993.
- [Cha90] F. Chaumette. *La relation vision-commande: théorie et application à des tâches robotiques*. PhD thesis, Université de Rennes I, July 1990.
- [Cla94] T.A. Clarke. An analysis of the properties of targets used in digital close range photogrammetric measurement. In Sabry F. El-Hakim, editor, *Videometrics III*, pages 251–262. SPIE 2350, November 1994.
- [CLSC92] G. Champleboux, S. Lavallé, P. Sautot, and P. Cinquin. Accurate calibration of cameras and range imaging sensors : the NPBS method. In *Proceedings of IEEE International Conference on Robotics and Automation, Nice, France*, pages 1552–1557, May 1992.

- [CR89] F. Chaumette and P. Rives. Modélisation et calibration d'une caméra. In *Actes du 7ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Paris, France*, pages 527–536, November 1989.
- [CTCE93] W.S. Ching, P.S. Toh, K.L. Chan, and M.H. Er. Robust vergence with concurrent detection of occlusion and specular highlights. In *Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany*, pages 384–394, Berlin, 1993.
- [DB93] R. Deriche and T. Blaszk. Recovering and characterizing image features using an efficient model based approach. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 530–535, June 1993.
- [Dev92] O. Devillers. Robust and Efficient Implementation of the Delaunay Tree. Technical Report 1619, INRIA, February 1992.
- [DF94] F. Devernay and O.D. Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pages 208–213. IEEE Computer Society Press, June 1994.
- [DG90] R. Deriche and G. Giraudon. Accurate corner detection: an analytical study. In *Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japan*, 1990.
- [DG93] R. Deriche and G. Giraudon. A computational approach for corner and vertex detection. *International Journal of Computer Vision*, 10(2):101–124, 1993.
- [DW87] H. F. Durrant-Whyte. Consistent integration and propagation of disparate sensor observations. *International Journal of Computer Vision*, 6(3):3–24, 1987.
- [EVF93] R. Enciso, T. Viéville, and O. Faugeras. Approximation du Changement de Focale et de Mise au Point par une Transformation à Trois Paramètres. Technical Report 2071, INRIA, France, October 1993.
- [Fai75] W. Faig. Calibration of close-range photogrammetric systems: mathematical formulation. *Photogrammetric Engineering and Remote Sensing*, 41(12):1479–1486, December 1975.
- [Fau92] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.
- [Fau93] O. Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint*. Artificial intelligence. M.I.T. Press, Cambridge, MA, 1993.

- [FB80] M.A. Fischler and R.C. Bolles. Random Sample Consensus: a Paradigm for Model Fitting with Applications to Image Analysis. Technical report, SRI International, Menlo Park, CA, 1980.
- [FBMB90] O.D. Faugeras, E. Le Bras-Mehlman, and J.D. Boissonnat. Representing stereo data with the delaunay triangulation. *Artificial Intelligence*, 44:41–87, 1990.
- [FFH<sup>+</sup>92] O. Faugeras, P. Fua, B. Hotz, R. Ma, L. Robert, M. Thonnat, and Z. Zhang. Quantitative and qualitative comparisons of some area and feature-based stereo algorithms. In W. Forstner, R. Haralick, and B. Radig, editors, *Robust Computer Vision*, pages 1–26, February 1992. Comparisons of area and feature-based stereo algorithms.
- [FG87] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Intercommission conference on fast processing of photogrammetric data, Interlaken, Switzerland*, pages 281–305, June 1987.
- [Fle87] R. Fletcher. *Practical Methods of Optimization*. John Wiley and Sons, second edition, 1987.
- [FLM92] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera Self-Calibration: Theory and Experiments. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 321–334. Springer-Verlag, May 1992.
- [För87] W. Förstner. Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision. *Computer Vision, Graphics and Image Processing*, 40:273–310, 1987.
- [FT87] O.D. Faugeras and G. Toscani. Camera calibration for 3D computer vision. In *Proceedings of International Workshop on Machine Vision and Machine Intelligence, Tokyo, Japan*, 1987.
- [FvDFH91] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes. *Computer Graphics: Principles and Practice*. Addison Wesley, November 1991.
- [GMW89] P. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Harcourt Brace Jovanovich, Publishers, 1989.
- [Gos89] A. Goshtasby. Correction of image deformation from lens distortion using Bezier patches. *Computer Vision, Graphics and Image Processing*, 47:385–394, 1989.
- [Gru85] A.W. Gruen. Adaptive least squares correlation: a powerful image matching technique. *S. Afr. Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3):175–187, 1985.

- [GS91] A. Gruen and D. Stallmann. High accuracy edge matching with an extension of the MPGC-matching algorithm. In *Industrial vision metrology, Winnipeg*, pages 42–55. SPIE 1526, 1991.
- [GT90] W.I. Grosky and L.A. Tamburino. A unified approach to the linear camera calibration problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):663–671, July 1990.
- [Har92] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 579–587. Springer-Verlag, 1992.
- [Har93] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 187–202, October 1993.
- [Har94] R.I. Hartley. Projective reconstruction and invariants from multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(10):1036–1041, October 1994.
- [Har95] R. Hartley. In Defence of the 8-point Algorithm. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1064–1070, June 1995.
- [Hav91] D. I. Havelock. The topology of locales and its effects on position uncertainty. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):380–386, April 1991.
- [HC95] C. Ho and L. Chen. A fast ellipse/circle detector using geometric symmetry. *Pattern Recognition*, 28(1):117–124, January 1995.
- [HGC92] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.
- [HJF89] M-H. Han, D. Jang, and J. Foster. Identification of cornerpoints of two-dimensional images using a line search method. *Pattern Recognition*, 22(1):13–20, 1989.
- [HJL<sup>+</sup>89] R. M. Haralick, H. Joo, C. Lee, X. Zhuang, V.G. Vaidya, and M. B. Kim. Pose estimation from corresponding point data. *IEEE Transactions on Systems, Man and Cybernetics*, 6(19):1426–1446, November/December 1989.
- [HK92] G. Healey and R. Kondepudy. CCD camera calibration and noise estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 90–95, 1992.
- [HM93] R. Horaud and O. Monga. *Vision par Ordinateur: Outils Fondamentaux*. Editions Hermès, Paris, 1993.

- [Hor86] B.K.P. Horn. *Robot Vision*. The MIT Press, 1986.
- [HRRS86] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics : the Approach Based on Influence Functions*. Wiley series in probability and mathematical. John Wiley and Sons, New York, 1986.
- [HS88] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.
- [HS90] Y.P. Hung and S.W. Shieh. When should we consider lens distortion in camera calibration. In *Machine Vision and Applications*, pages 367–370, November 1990.
- [HS94] R. Hartley and P. Sturm. Triangulation. In *Proceedings of ARPA Image Understanding Workshop*, pages 957–966, 1994.
- [HSV90] R. Horaud, T. Skordas, and F. Veillon. Finding geometric and relational structures in an image. In *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, Lecture Notes in Computer Science, pages 374–384. Springer-Verlag, April 1990.
- [IK87] J. Illingworth and J. Kittler. The adaptative hough transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):690–698, 1987.
- [JHT93] J. Jansa, Y. Huang, and J.C. Trinder. Problems of precise target location and camera orientation in digital close-range photogrammetry. In *SPIE, vol. 2067, Videometrics II*, pages 151–161, 1993.
- [JMB91] J.M. Jolion, P. Meer, and S. Bataouche. Robust clustering with applications in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):791–802, August 1991.
- [KH94] R. Kumar and A.R. Hanson. Robust methods for estimating pose and a sensivity analysis. *Computer Vision, Graphics and Image Processing: Image Understanding*, 60(3):313–342, November 1994.
- [KvD89] J.J. Koenderink and A. J. van Doorn. Affine structure from motion. Technical report, Utrecht University, Utrecht, The Netherlands, October 1989.
- [LG94] J.L. Lotti and G. Giraudon. Adaptive window algorithm for aerial image stereo. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, pages 701–703, October 1994.
- [LH81] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. In *Nature*, volume 293, pages 133–135. XX, September 1981.
- [LMR95] Z. D. Lan, R. Mohr, and P. Remagnino. Robust matching by partial correlation. In *Proceedings of the sixth British Machine Vision Conference, Birmingham, England*, September 1995.

- [LS90] J.C. Li and T. Schenk. An accurate camera calibration for the aerial image analysis. In *Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, New Jersey, USA*, pages 207–209, 1990.
- [LT88] R.K. Lenz and R.Y. Tsai. Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):713–720, September 1988.
- [LZT94] C.T. Lawrence, J.L. Zhou, and A.L. Tits. User’s Guide for CFSQP Version 2.0. Technical Report tr-94-16, University of Maryland, University of Maryland, MD, February 1994.
- [MBK81] H.A. Martins, J.R. Birk, and R.B. Kelley. Camera models based on data from two calibration planes. *Computer Graphics and Image Processing*, 17:173–180, 1981.
- [MGDP93] T. Moons, L. Van Gool, M. Van Diest, and E. Pauwels. Affine reconstruction from perspective image pairs. In *Proceeding of the DARPA–ESPRIT workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 249–266, October 1993.
- [MK94] H. Maas and T. Kernsten. Digital close-range photogrammetry for dimensional checking and control in shipyard. In Sabry F. El-Hakim, editor, *Videometric III*, pages 108–114. SPIE 2350, November 1994.
- [MM91] D. Mintz and P. Meer. Robust estimators in computer vision: an introduction to least median of squares regression. In Y.A. Feldman and A. Bruckstein, editors, *Artificial Intelligence and Computer Vision*, pages 61–70. Elsevier Science, 1991.
- [MMRK91] P. Meer, D. Mintz, A. Rosenfeld, and D.Y. Kim. Robust regression methods for computer vision: a review. *International Journal of Computer Vision*, 6(1):59–70, 1991.
- [MN90] R. Mehrotra and S. Nichani. Corner detection. *Pattern Recognition*, 23(11):1223–1233, November 1990.
- [Mor77] H.P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, USA*, page 584, August 1977.
- [MY87] G. Medioni and Y. Yasumoto. Corner detection and curve representation using cubic B-splines. In *Computer Vision, Graphics and Image Processing*, volume 39, pages 267–278. 1, 1987.
- [MZ92] J. Mundy and A. Zisserman. Projective geometry for machine vision. In J. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, chapter 23, pages 463–519. MIT Press, 1992.

- [Nag83] H.H. Nagel. Displacement vectors derived from second order intensity variations in image sequences. *Computer Vision, Graphics and Image Processing*, 21:85–117, 1983.
- [NSNI92] Y. Nomura, M. Sagara, H. Naruse, and A. Ide. Simple calibration algorithm for high-distortion-lens camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(11):1095–1099, November 1992.
- [OBBA90] L. O’Gorman, A. M. Bruckstein, C. B. Bose, and I. Amir. Subpixel registration using a concentric ring fiducial. In *Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, New Jersey, USA*, pages 249–253, June 1990.
- [OG93] C.M. Orange and F.C.A Groen. Model based corner detection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 690–691, June 1993.
- [OK85] Y. Ohta and T. Kanade. Stereo by intra and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154, 1985.
- [OK92] M. Okutomi and T. Kanade. A locally adaptive window for signal matching. *International Journal of Computer Vision*, 7(2):143–162, 1992.
- [Pen87] A. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pami-9(4):523–531, July 1987.
- [Pen91] M.A. Penna. Camera calibration : a quick and easy way to determine the scale factor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(12):1240–1245, December 1991.
- [Peu92] B. Peuchot. Camera virtual equivalent model - 0.01 pixel detectors. In *14th Annual International Conference IEEE EMBS. Satellite Symposium on 3D Advanced Image Processing in Medicine, Rennes.*, pages 41–45, November 1992.
- [Peu94] B. Peuchot. Utilisation de détecteurs subpixels dans la modélisation d’une caméra. In *Actes du 9ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Paris, France*, pages 691–695, January 1994.
- [PFTW88] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling W.T. *Numerical Recipes in C*. Cambridge University Press, 1988.
- [PMC94] J. Ponce, D. H. Marimont, and T. A. Cass. Analytical methods for uncalibrated stereo and motion reconstruction. In J. O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 463–470, May 1994.
- [Pra78] W.K. Pratt. *Digital Image Processing*. Wiley-Interscience, 1978.

- [Qua93] L. Quan. Affine stereo calibration for relative affine shape reconstruction. In *Proceedings of the fourth British Machine Vision Conference, Surrey, England*, pages 659–668, 1993.
- [RCF95] C. Rothwell, G. Csurka, and O. Faugeras. A comparison of projective reconstruction methods for pairs of views. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 932–937, June 1995.
- [RDDL94] S. Rémy, M. Dhome, N. Daucher, and J.T. Lapresté. Estimation de la distorsion radiale d'un système optique. In *Actes du 9ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Paris, France*, pages 99–108, January 1994.
- [RK82] A. Rosenfeld and A.C. Kak. *Digital Picture Processing*. Academic, New York, 1982.
- [RL87] P.J. Rousseeuw and A.M. Leroy. *Robust regression and outlier detection*, volume XIV of *Wiley*. J.Wiley and Sons, New York, 1987.
- [RL90] G. Roth and M.D. Levine. Segmentation of geometric signals using robust fitting. In *Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, New Jersey, USA*, pages 826–831, 1990.
- [Roh92] K. Rohr. Recognizing corners by fitting parametric models. *International Journal of Computer Vision*, 9(3):213–230, December 1992.
- [RSB89] K. Rangarajan, M. Shah, and D. Van Brackle. Optimal corner detector. *Computer Vision, Graphics and Image Processing*, 48:230–245, November 1989.
- [SAKS94] K. Sohn, W.E. Alexander, J.H. Kim, and W.E. Snyder. A constrained regularization approach to robust corner detection. *IEEE Transactions on Systems, Man and Cybernetics*, 24(5):820–828, May 1994.
- [SCS94] M.R. Shortis, T.A. Clarke, and T. Short. A comparison of some techniques for the subpixel location of discrete target image. In Sabry F. El-Hakim, editor, *Videometric III*, pages 239–250. SPIE 2350, November 1994.
- [SHB93] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision*. Chapman and Hall, 1993.
- [SK52] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford Science Publication, 1952.
- [Sla80] C.C. Slama, editor. *Manual of Photogrammetry, fourth edition*. American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.

- [SM96] C. Schmid and R. Mohr. Mise en correspondance par invariants locaux. In *Actes du 10ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Rennes, France, 1996*. soumis.
- [SS90] A. Singh and M. Shneier. Grey level corner detection : a generalization and a robust real time implementation. *Computer Vision, Graphics and Image Processing*, 51:54–69, July 1990.
- [SS94] U. Seeger and R. Seeger. Fast corner detection in grey-level images. *Pattern Recognition Letters*, 15:669–675, July 1994.
- [TF90] P. Toh and A. Forrest. Occulsion dedection in early vision. In *Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japan*, pages 126–132, December 1990.
- [TH84] R.Y. Tsai and T.S. Huang. Uniqueness and estimation of 3D motion parameters of rigid bodies with curved surfaces. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 6, pages 13–27. IEEE Computer Society Press, 1984.
- [Tri89] J.C. Trinder. Precision of digital target location. *Photogrammetric Engineering and Remote Sensing*, 55(6):883–886, June 1989.
- [Tsa86] R.Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Miami Beach, Florida, USA*, pages 364–374, 1986.
- [Vai90] R. Vaillant. *Géométrie différentielle et vision par ordinateur : détection et reconstruction des contours d'occultation de surfaces*. PhD thesis, Université de Paris-Sud, Orsay, France, December 1990.
- [WC90] G.A.W. West and T.A. Clarke. A survey and examination of subpixel measurement techniques. In *Close-Range Photogrammetry Meets Machine Vision*, pages 456–463. SPIE, vol 1395, 1990.
- [WCH90] J. Weng, P. Cohen, and M. Herniou. Calibration of stereo cameras using a non-linear distortion model. In *Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, New Jersey, USA*, pages 246–253, 1990.
- [WCH92] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accurate evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, October 1992.
- [WF92] P. Whaite and F.P. Ferrie. Uncertain Views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 3–9, 1992.
- [WH94] P. Weckesser and G. Hetzel. Photogrammetric calibration methods for an active stereo vision system. In *Intelligent Robotics System, Grenoble, France*, pages 326–333, July 1994.

- [WM91] G.Q. Wei and S.D. Ma. Two plane camera calibration: a unified model. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, USA*, pages 133–138, 1991.
- [ZDFL94] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. Rapport de recherche 2273, INRIA, May 1994.
- [ZF90] Z. Zhang and O. Faugeras. Building a 3D world representation with a mobile robot: 3D line segment representation and integration. In *Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, New Jersey, USA*, pages 38–42, June 1990.
- [ZFD95] Z. Zhang, O. Faugeras, and R. Deriche. Calibrating a binocular stereo through projective reconstruction using both a calibration object and the environment. In *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, Xian, China*, pages 253–260, April 1995.
- [Zha93] C.S. Zhao. *Reconstruction de Surfaces Tridimensionnelles en Vision par Ordinateur*. PhD thesis, Institut National Polytechnique de Grenoble, December 1993.
- [Zha94] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, 1994.
- [Zho90] H.B. Zhou. Object points detection in a photogrammetric test field. In *SPIE vol 1395, Close-Range Photogrammetry Meets Machine Vision*, pages 1227–1233, 1990.
- [Zum95] R. Zumbrunn. Systematic pointing errors with retro-reflective targets. In *ISPRS Intercommission Workshop : Sensors, Algorithms and Systems*, pages 60–65, Zurich, Switzerland, March 1995.