



HAL
open science

Reconstruction tridimensionnelle en vision par ordinateur : cas des cameras non etalonnees

Boubakeur Boufama

► **To cite this version:**

Boubakeur Boufama. Reconstruction tridimensionnelle en vision par ordinateur : cas des cameras non etalonnees. Interface homme-machine [cs.HC]. Institut National Polytechnique de Grenoble - INPG, 1994. Français. NNT: . tel-00005080

HAL Id: tel-00005080

<https://theses.hal.science/tel-00005080>

Submitted on 25 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.

THESE

présentée par

Boubakeur BOUFAMA

pour obtenir le grade de DOCTEUR

de l'**INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE**

(Arrêté ministériel du 30 Mars 1992)

Spécialité: **Informatique**

—

**RECONSTRUCTION TRIDIMENSIONNELLE
EN VISION PAR ORDINATEUR : CAS DES
CAMERAS NON ETALONNEES**

—

Date de soutenance : 16 Décembre 1994

Composition du jury :

Président : **Bertrand ZAVIDOVIQUE**

Rapporteurs : **Serge CASTAN**
Marc RICHETIN

Examineurs : **Radu HORAUD**
Roger MOHR
Daphna WEINSHALL

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

Remerciements

Je tiens tout d’abord à remercier l’Institut National de Recherche en Informatique et en Automatique (INRIA) pour son financement qui a permis la préparation de cette thèse.

Je remercie

- Messieurs Bertrand Zavidovique, Serge Castan, Marc Richetin et Madame Daphna Weinshall qui m’ont fait l’honneur de participer au jury de ma thèse.
- Roger Mohr, pour avoir dirigé ma thèse tout en m’accordant une grande liberté dans mon travail, pour les discussions scientifiques que nous avons eues, pour ses conseils qui m’ont guidé le long de cette thèse et pour sa chaleur humaine et relation amicale.
- Radu Horaud, qui m’a fait l’honneur de participer au jury et dont l’intérêt qu’il a toujours porté à mes travaux a été pour moi des plus précieux encouragements.
- Patrick Gros et Françoise Veillon, pour leur relecture de ma thèse, pour leurs conseils et aides lors des difficiles moments de fin de thèse et pour leur amitié chaleureuse.
- toutes les personnes de l’équipe MOVI pour leur amitié, discussions fructueuses et travail en commun. Je tiens à remercier en particulier Long Quan, Pascal Brand, Luce Morin et Danièle Herzog.
- les stagiaires, Jean Claude Cottier et Jérôme Blanc, devenus maintenant amis, qui ont travaillé avec moi et qui ont contribué à cette thèse.
- toutes les personnes (“fumeuses et non fumeuses”) du LIFIA, je n’oublierai certainement pas les effets positifs des pauses cafés et des discussions variées à la cafet du LIFIA.

Table des matières

Introduction	6
1 Géométrie et vision par ordinateur	11
1.1 Géométrie projective	11
1.1.1 Espaces projectifs, coordonnées homogènes	12
1.1.2 Transformations projectives	13
1.1.3 Invariants projectifs	14
1.1.4 Géométrie affine	17
1.1.5 Géométrie euclidienne	17
1.2 Images et géométrie	19
1.2.1 Formation d'image	20
1.2.2 Modèle sténopé	23
1.2.3 Modèle de projection orthographique	24
1.2.4 Modèle de projection perspective faible	25
1.2.5 Modèle de projection para-perspective	27
1.3 Relation géométrique entre les images	27
1.3.1 Relation géométrique entre deux images	28
1.3.2 Relation géométrique entre trois images	29
1.4 Conclusion	30
2 Matrice fondamentale et mise en correspondance de points	31
2.1 Calcul de la matrice fondamentale : nouvelle méthode	32
2.1.1 Aspect géométrique du problème	33
2.1.2 Solution algébrique	35
2.1.3 Calcul de la matrice fondamentale	39
2.1.4 Algorithme de résolution et résultats expérimentaux	40
2.2 Mise en correspondance de points dans deux images	47
2.2.1 Détection des points	48

2.2.2	Mise en correspondance	51
2.3	Conclusion	53
3	Reconstruction 3D à partir d'images non étalonnées	57
3.1	Introduction	57
3.1.1	Etat de l'art	59
3.1.2	Plan du chapitre	61
3.2	Reconstruction avec un modèle affine pour la caméra	62
3.2.1	Reconstruction tridimensionnelle affine	62
3.2.2	Structure euclidienne de quatre points sous l'hypothèse d'une projection perspective faible	63
3.2.3	Reconstruction tridimensionnelle euclidienne	66
3.2.4	Résultats expérimentaux et limites de la méthode	67
3.3	Reconstruction utilisant un modèle sténopé	69
3.3.1	Les équations du problème de la reconstruction	71
3.3.2	Résolution du problème de la reconstruction	72
3.3.3	Résultats expérimentaux	75
3.4	Comparaison et coopération des deux méthodes	83
3.4.1	Comparaison des deux méthodes	84
3.4.2	Un algorithme de coopération	91
3.5	Conclusion	91
4	D'une reconstruction projective à une reconstruction euclidienne	93
4.1	D'une reconstruction projective à une reconstruction euclidienne : aspects théoriques	95
4.1.1	Du projectif à l'anne	95
4.1.2	De l'anne à l'euclidien	98
4.2	D'une reconstruction projective à une reconstruction euclidienne : aspect pratique	100
4.2.1	Contraintes euclidiennes utilisées en pratique	101
4.2.2	Résolution numérique	105
4.3	Expérimentations	105
4.4	Conclusion	111
5	Calcul du mouvement et reconstruction	113
5.1	Principes des méthodes classiques	115
5.2	Les équations du problème	117

5.3	Paramétrer autrement	119
5.3.1	Calcul d'une solution approchée	122
5.3.2	Résumé de la méthode de reconstruction	123
5.4	Expérimentation	124
5.5	Conclusion	128
Conclusion		129
A Détecteur de points d'intérêt		132
A.1	Détecteur de Moravec	132
A.2	Le détecteur de points d'intérêt de Plessey	133
A.3	Amélioration du détecteur de Plessey	135
B Méthode robuste pour trouver la matrice fondamentale		137
Bibliographie de l'auteur		139
Bibliographie		141

Introduction

La vision biologique est l'un des cinq sens que la nature a donné à l'être humain et à la majorité des animaux. Au moyen de nos deux yeux, la vision nous permet d'acquérir des informations visuelles, appelées images, sur l'environnement qui nous entoure. Ces informations sont ensuite transférées au cerveau où elles sont interprétées. Le résultat de cette interprétation peut donner lieu à une prise de décision de la part de l'être humain, lui permettant, par exemple, de se mouvoir sans danger.

D'une manière similaire, la vision artificielle ou la vision par ordinateur vise à donner à une machine cette faculté d'acquérir et d'interpréter les informations visuelles. Là aussi, une décision peut éventuellement être prise en conséquence, par exemple pour permettre la saisie d'un objet par un robot.

Cependant, l'être humain dispose d'un système de vision très performant, dont le mécanisme reste mal connu malgré les progrès réalisés par la psychologie cognitive. De plus, les capacités des traitements cérébraux sont largement supérieures à la capacité de calcul de l'ordinateur le plus performant. Ce dernier reste très modeste, voire dérisoire, comparé au cerveau humain.

La vision par ordinateur ne vise pas la reproduction sur machine de la vision humaine. Elle vise plutôt à arriver par des moyens informatiques à des résultats similaires.

Le processus de la vision comporte en particulier l'acquisition des images et les traitements de celles-ci.

- L'acquisition d'une image consiste à transformer un signal physique d'une nature donnée en un signal d'une autre nature. Pour ce qui nous concerne, nous utilisons les images à *niveaux de gris* obtenues avec des cameras CCD¹. Une image à *niveaux de gris* est codée dans un ordinateur sous forme d'un tableau à deux dimensions. Chaque élément de ce tableau, appelé *pixel*², est un entier codant l'intensité lumineuse perçue en un point. La valeur de l'intensité lumineuse d'un pixel est le résultat d'un long processus opto-électronique, dont le but est de transformer le signal optique de départ en un tableau à valeurs entières.

1. Charge-Coupled Device

2. picture-element

- Les traitements d’une image peuvent être de nature très différente allant du traitement du signal, comme par exemple l’extraction des points d’intérêts tels que les contours, les coins. . . jusqu’à l’interprétation sémantique de ce qui est observé, comme par exemple la description du contenu de la scène avec un langage symbolique. Aussi, les traitements d’une image sont généralement classés en trois niveaux [Mar78]. Un premier niveau, dit bas niveau, concerne le traitement au niveau des pixels. Un deuxième niveau, appelé parfois niveau intermédiaire recouvre les calculs des caractéristiques numériques de la scène comme la reconstruction tridimensionnelle, le calcul du mouvement. . . Et enfin un troisième niveau, dit haut niveau, rassemble ce qui a trait à l’interprétation.

Le travail décrit dans cette thèse concerne principalement la reconstruction tridimensionnelle à partir d’images issues de caméras non étalonnées. Dans la suite, on parlera pour simplifier de reconstruction tridimensionnelle à partir d’images non étalonnées.

Deux sujets liés à la reconstruction ont été traités : le calcul de la matrice fondamentale (géométrie épipolaire) avec la détection et la mise en correspondance de points d’intérêt dans une séquence d’images. Dans la classification donnée ci-dessus, ce travail se situe en grande partie au niveau intermédiaire. Seule la détection des points d’intérêt dans une image se situe au bas niveau.

En considérant l’aspect géométrique, on peut supposer que l’espace qui nous entoure est une représentation tridimensionnelle d’entités physiques. Une image à niveaux de gris est alors considérée comme une représentation bidimensionnelle de cet espace. Il y a donc perte d’information durant le processus de formation d’une image. En particulier, la troisième dimension est perdue. La récupération de cette dernière est le but de la reconstruction tridimensionnelle.

Pour illustrer le problème de la reconstruction, soit une caméra modélisée par un plan (plan image) et un point appelé centre de projection (voir figure 0.1a). Avec ce modèle de caméra, un point de l’espace subit une projection centrale (projection perspective) pour donner naissance à un point de l’image. Malheureusement, cette transformation n’est pas inversible : un point de l’image ne représente pas un point de l’espace mais représente une droite de l’espace. Il est donc impossible de retrouver la troisième dimension d’un point uniquement à partir de sa projection dans une image. Par contre, si on considère un point observé dans deux images différentes, l’ambiguïté disparaît et il devient possible de trouver la position tridimensionnelle du point (voir figure 0.1).

Cependant, l’utilisation de deux images (ou plusieurs images) pose deux nouveaux problèmes qu’il faut résoudre avant de pouvoir reconstruire.

1. *la mise en correspondance* : dans la figure 0.1b nous avons supposé que p' était la projection du même point P de l’espace. Cette information nécessaire n’est

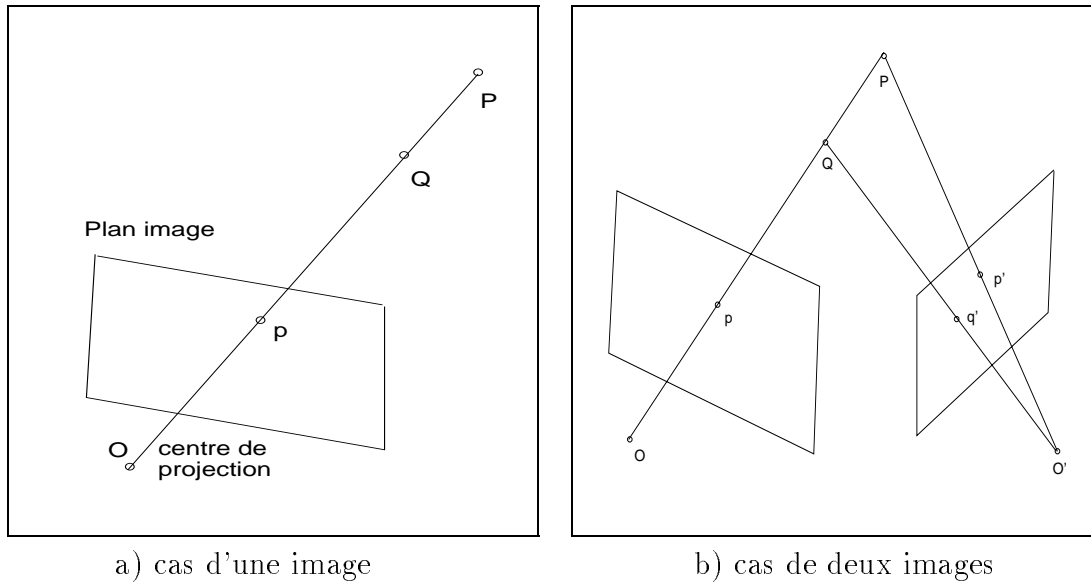


FIG. 0.1 - *Quand on considère une seule image, le point p est la projection du point P de l'espace, mais il est aussi la projection de Q . En utilisant une paire d'images, l'ambiguïté disparaît et il devient possible de retrouver la troisième dimension.*

pas disponible à partir des images, elle doit être trouvée. En d'autres termes, pour un point donné dans une image, nous devons être capable de trouver son homologue dans une autre image pour que la reconstruction devienne possible. Ce problème est appelé mise en correspondance des points dans les images.

2. *l'étalonnage des caméras* : pour pouvoir reconstruire le point P (figure 0.1b), il faut écrire les équations des deux droites de l'espace Op et $O'p'$ dans un repère de référence commun. Le calcul de la relation qui lie les coordonnées tridimensionnelles d'un point aux coordonnées bidimensionnelles de son image s'appelle étalonnage. Ce dernier permet de définir, pour chaque point de l'image, une droite de l'espace dans le repère de référence. La position du point P est alors donnée par l'intersection des deux droites de l'espace Op et $O'p'$. On dit que le point est reconstruit par triangulation (figure 0.1b).

Les deux étapes ci-dessus peuvent être une solution au problème de reconstruction. Cependant, il n'est pas toujours réaliste d'étalonner chaque caméra avant toute opération de reconstruction. Car l'étalonnage d'une caméra est un processus délicat nécessitant une mire parfaitement connue. De plus, il doit être répété chaque fois que la caméra bouge ou change ses caractéristiques. Outre cet inconvénient, il n'est pas toujours possible d'étalonner une caméra, par exemple quand celle-ci est impliquée dans des mouvements.

Dans cette thèse nous proposons des méthodes de reconstruction tridimensionnelle permettant de supprimer le fardeau de l'étalonnage. Nous nous plaçons dans le cas général d'une séquence d'images obtenues avec une ou plusieurs caméras inconnues. Nous entendons par caméras inconnues des caméras dont les caractéristiques géométriques et les positions sont inconnues.

Contributions majeures

Les principales contributions de cette thèse sont :

1. mise en œuvre d'une nouvelle méthode pour calculer la matrice fondamentale (géométrie épipolaire). Cette méthode propose une nouvelle formulation du problème permettant une simplification des équations (section 2.1).
2. étude et réalisation d'un algorithme de mise en correspondance de points dans les images (section 2.2).
3. étude et comparaison de deux méthodes de reconstruction tridimensionnelle. Un algorithme combinant les deux méthodes est proposé.
4. étude théorique et réalisation pratique d'une méthode pour le passage d'une reconstruction projective à une reconstruction euclidienne (chapitre 4).
5. Calcul de la reconstruction euclidienne et du mouvement d'une caméra quand les caractéristiques de la caméra sont connues approximativement (chapitre 5).

Plan de la thèse

Le premier chapitre est consacré à la géométrie et à son lien avec la vision par ordinateur, en particulier au niveau de la formation d'image.

Dans la première partie, nous introduisons les notions fondamentales de la géométrie projective, cette dernière étant la géométrie la plus adaptée à nos problèmes. Les géométries affine et euclidienne sont définies comme étant des restrictions de la géométrie projective. Dans la deuxième partie, nous présentons quelques modèles géométriques utilisés pour modéliser les caméras CCD.

Le deuxième chapitre comporte deux parties. La première partie présente une nouvelle méthode pour le calcul de la matrice fondamentale (géométrie épipolaire). Cette dernière est un concept clé dans le cas d'images non étalonnées. Les expérimentations réalisées avec des données réelles et simulées montrent que, par rapport à la méthode classique, cette méthode permet une meilleure stabilité du calcul. En particulier, quand seules des équations linéaires sont utilisées, notre méthode donne des résultats corrects, ce qui n'est pas le cas pour la méthode classique.

La deuxième partie du chapitre est consacré à un problème plutôt technique : la mise en correspondance des points dans les images. Cette dernière étant nécessaire à la reconstruction, nous avons donc réalisé un outil robuste de mise en correspondance de points.

Dans le troisième chapitre deux méthodes de reconstruction sont présentées. La première méthode utilise un modèle affine et permet une reconstruction euclidienne à partir d'au moins trois images. Le modèle affine n'est pas toujours adapté pour approcher une caméra, aussi les résultats des reconstructions avec des données réelles et simulées montrent les limites de cette méthode. La deuxième méthode utilise un modèle projectif. Elle permet d'obtenir une reconstruction projective, affine ou euclidienne selon qu'on connaît les coordonnées projectives, affines ou euclidiennes de 5 points de l'espace. Le modèle projectif étant une bonne approximation pour une caméra, les résultats montrent la robustesse de la méthode.

Après une comparaison de ces deux méthodes, nous proposons un algorithme permettant leur coopération. La deuxième méthode est itérative mais elle est plus précise. La première méthode est linéaire, et permet d'obtenir une reconstruction proche de la solution. Celle-ci est suffisante pour assurer la convergence de la méthode non linéaire.

Le quatrième chapitre concerne l'utilisation de connaissances a priori sur la scène pour passer d'une reconstruction projective vers une reconstruction euclidienne. Une étude théorique dont le but est de comprendre clairement le lien d'une reconstruction projective avec une reconstruction euclidienne est présentée. Ensuite, une mise en œuvre pratique est proposée.

Le dernier chapitre traite le problème de la reconstruction euclidienne et du calcul du mouvement d'une caméra. En utilisant des valeurs approchées pour les paramètres intrinsèques de la caméra et un nombre de paramètres minimal pour la structure, nous montrons qu'il est possible d'obtenir de bons résultats qualitatifs.

Enfin, la conclusion résume les différents résultats et ouvre quelques perspectives sur la suite de ce travail.

Chapitre 1

Géométrie et vision par ordinateur

La vision par ordinateur vise à donner aux robots la capacité de percevoir l'espace qui les entoure. Ce dernier est tridimensionnel et peut être représenté par des entités géométriques. Les robots sont amenés à bouger, l'espace qui les entoure aussi, la géométrie de cet espace est ainsi dynamique. Il est alors clair que la géométrie fournit à la vision par ordinateur un outil puissant lui permettant d'appréhender ce monde tridimensionnel.

Ce chapitre commence par quelques rappels de géométrie projective. Nous ne visons ni à être exhaustif dans les notions présentées ni à en fournir les démonstrations mais simplement à décrire les résultats fondamentaux utiles. Le lecteur intéressé pourra consulter [SK52] pour des notions plus complètes ou [May90] pour une étude plus avancée sur l'application de la géométrie projective en vision par ordinateur.

La seconde section du chapitre décrit l'apport de la géométrie projective à la vision par ordinateur quant à la formation d'image. Seul l'aspect géométrique de cette dernière est traité ici.

Enfin, la géométrie épipolaire et la notion de matrice fondamentale sont introduites à la section 1.3.

1.1 Géométrie projective

Bien que le monde qui nous entoure soit un monde euclidien, les transformations euclidiennes ne sont pas les mieux adaptées pour modéliser certaines transformations géométriques, comme par exemple la projection perspective.

La géométrie projective possède une propriété intéressante : la dualité. Le principe de dualité en géométrie projective assure l'équivalence entre points et droites, droites et plans... Passer d'un point à une droite ou vice versa ne complique en

rien les calculs ; de plus, la géométrie projective manipule les coordonnées homogènes, ce qui rend l'algèbre linéaire. La géométrie projective est plus générale que la géométrie euclidienne qui n'est qu'un cas particulier de celle-ci. Par conséquent, travailler dans le cas général qui est la géométrie projective offre plus de souplesse.

Outre les arguments cités ci-dessus en faveur de la géométrie projective, la raison principale qui nous incite à utiliser celle-ci en vision est liée à la nature géométrique des systèmes d'acquisition d'images. Ces derniers peuvent être modélisés par des systèmes réalisant une projection perspective du monde tridimensionnel sur un plan. La géométrie projective offre un cadre simple pour représenter une projection perspective, elle est donc la plus adaptée pour le domaine de la vision.

Après une introduction des espaces projectifs, nous donnerons les principales propriétés des transformations projectives, puis la notion d'invariant sera introduite. Enfin, la géométrie affine et la géométrie euclidienne seront définies comme étant des restrictions de la géométrie projective.

1.1.1 Espaces projectifs, coordonnées homogènes

Considérons l'espace $\mathcal{C}^{n+1} \setminus \{(0, \dots, 0)\}$ muni de la relation d'équivalence suivante :

$$\begin{aligned} (x_1, \dots, x_{n+1}) \sim (x'_1, \dots, x'_{n+1}) &\iff \\ \exists \lambda \neq 0 \text{ tel que } (x'_1, \dots, x'_{n+1}) &= \lambda(x_1, \dots, x_{n+1}) \end{aligned} \quad (1.1)$$

L'espace quotient résultant de cette relation d'équivalence définit l'espace projectif de dimension n noté \mathbb{P}^n . Ainsi, les deux $(n+1)$ -uplets (x_1, \dots, x_{n+1}) et $\lambda(x_1, \dots, x_{n+1})$ représentent le même point dans l'espace projectif \mathbb{P}^n et sont appelés coordonnées homogènes de ce point. Le signe \sim est appelé dans la suite égalité projective, c'est une égalité définie à un facteur multiplicatif près.

Quand la dimension de l'espace projectif vaut respectivement 1, 2 et $n-1$, cet espace est généralement appelé *droite*, *plan* et *hyperplan*.

Une injection canonique de \mathcal{C}^n dans \mathbb{P}^n existe et est définie par :

$$\begin{aligned} \mathcal{C}^n &\longrightarrow \mathbb{P}^n \\ (x_1, \dots, x_n) &\longmapsto (x_1, \dots, x_n, 1) \end{aligned}$$

Cette injection n'est pas surjective : elle n'atteint pas les points de \mathbb{P}^n tels que $x_{n+1} = 0$.

Ces points ne sont associés à aucun point usuel de \mathcal{C}^n , on les appelle les *points à l'infini*. Ils peuvent être interprétés comme étant des directions de l'espace \mathcal{C}^n . Cette représentation est reprise par la suite. L'hyperplan d'équation $x_{n+1} = 0$ est appelé *hyperplan de l'infini*.

Dépendance linéaire

Les points P_1, \dots, P_m de \mathbb{P}^n sont *linéairement dépendants* s'il existe un ensemble de scalaires $\lambda_1, \dots, \lambda_m$ non tous nuls tel que :

$$\sum_{i=1}^m \lambda_i P_i = 0$$

Dans le cas contraire, les points sont *linéairement indépendants*.

Base projective

Une base projective de \mathbb{P}^n est formée de $n+2$ points tels qu'aucun sous ensemble de $n+1$ points n'appartienne à un hyperplan. En d'autres termes, tous les sous ensembles de $n+1$ points de cette base sont linéairement indépendants.

La base projective canonique est formée par les points suivants : $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, \dots, 0), \dots, e_{n+1} = (0, 0, \dots, 1)$ et $e_{n+2} = (1, 1, \dots, 1)$. Cela revient à prendre n directions, 1 point pour marquer l'origine e_{n+1} et un point pour marquer l'unité e_{n+2} . Tout point de l'espace projectif peut être représenté comme une combinaison linéaire de n'importe quel sous ensemble de $(n+1)$ points de la base. Par exemple le point X peut s'écrire :

$$X = \sum_{i=1}^{n+1} \lambda_i e_i$$

où les λ_i sont des scalaires

Proposition

Soient x_1, \dots, x_{n+2} ($n+2$) points de \mathbb{P}^n tels qu'aucun sous ensemble de $(n+1)$ points ne soient linéairement dépendants. Ces points forment donc une base de \mathbb{P}^n .

Il existe une matrice de rang maximal W de taille $(n+1) \times (n+1)$ telle que $W e_i = \lambda_i x_i$, $i = 1, \dots, n+2$, où les λ_i sont des scalaires non nuls et où (e_i) est la base canonique de \mathbb{P}^n .

Par conséquent, si x_1, \dots, x_{n+2} et y_1, \dots, y_{n+2} sont deux ensembles de $(n+2)$ points de \mathbb{P}^n formant deux bases projectives, alors il existe une matrice non singulière W de taille $(n+1) \times (n+1)$ telle que $W x_i = \lambda_i y_i$, $i = 1, \dots, n+2$, où les λ_i sont des scalaires non nuls. W est définie à un facteur multiplicatif près.

1.1.2 Transformations projectives

Définition

On appelle *transformation projective* ou *collinéation* toute application linéaire en les coordonnées homogènes.

Une transformation projective de \mathbb{P}^n dans \mathbb{P}^m s'exprime donc sous forme matricielle :

$$\begin{aligned}\mathbb{P}^n &\longrightarrow \mathbb{P}^m \\ X &\longmapsto Y = W^t X\end{aligned}$$

où W est une matrice non singulière de taille $(m+1) \times (n+1)$. X et Y étant définis à un facteur multiplicatif près, il en est de même pour W . Une transformation projective de \mathbb{P}^n dans \mathbb{P}^m n'a que $(m+1) \times (n+1) - 1$ degrés de liberté.

Une transformation projective de \mathbb{P}^n dans lui-même est appelé *homographie*. Pour n donné, l'ensemble des *homographies* forment un groupe qui s'appelle aussi *groupe projectif*.

Fixer une transformation projective

Déterminer une *transformation projective* de \mathbb{P}^n dans \mathbb{P}^m revient à déterminer à un facteur multiplicatif près la matrice qui lui est associée, c'est à dire les $(m+1) \times (n+1) - 1$ paramètres indépendants.

Chaque correspondance entre un point de \mathbb{P}^n et son image de \mathbb{P}^m fournit $(m+1)$ équations linéaires en les inconnues. Seules m équations sont indépendantes puisque les points sont exprimés en coordonnées homogènes.

Si $m = n$ (cas des homographies), il faut $(n+2)$ correspondances pour déterminer tous les paramètres de la *transformation projective*. De plus, il est nécessaire que les points utilisés forment une base projective, sinon les équations qui en découlent ne sont pas indépendantes.

Dans le cas où $m \neq n$, le nombre de correspondances suffisant pour déterminer la *transformation projective* n'est plus nécessairement $(n+2)$. Par exemple, pour le cas de la projection de \mathbb{P}^3 dans \mathbb{P}^2 qui est déterminée par une matrice 3×4 avec 11 paramètres indépendants, chaque correspondance entre un point de \mathbb{P}^3 et un point de \mathbb{P}^2 fournit 2 équations indépendantes. Cinq correspondances fournissent 10 équations indépendantes, or il en faut 11.

1.1.3 Invariants projectifs

Un invariant est une quantité numérique calculée à partir d'une certaine configuration. Il a la propriété de rester inchangé même si la configuration de départ a subi une certaine catégorie de transformations.

Pour être plus précis, considérons deux ensembles E et F et un ensemble T de transformations de E dans F . Soit I une fonction des éléments de F . I est un invariant pour T si elle prend la même valeur pour toutes les images d'un élément de E :

$$\forall e \in E, \quad \forall t, t' \in T \quad I(t(e)) = I(t'(e))$$

En particulier, un invariant projectif est un invariant pour les *transformations projectives*.

Nous nous limitons à ces définitions simples, le lecteur intéressé trouvera dans [FMZ90] et [Gro93] une étude détaillée des invariants.

Le birapport

Le *birapport* est l'invariant projectif fondamental. D'autres invariants peuvent s'exprimer en termes de birapports.

Birapport de quatre points alignés

Soient P_1, P_2, P_3 et P_4 quatre points d'une droite projective (espace projectif de dimension 1) et soient respectivement $(\lambda_1, \mu_1), (\lambda_2, \mu_2), (\lambda_3, \mu_3)$ et (λ_4, μ_4) leurs coordonnées homogènes. En posons $\frac{\lambda_i}{\mu_i} = \theta_i$, le birapport des points P_1, P_2, P_3 et P_4 s'écrit sous la forme $[P_1, P_2; P_3, P_4]$, il est défini par :

$$[P_1, P_2; P_3, P_4] = \frac{\theta_1 - \theta_3}{\theta_2 - \theta_3} \div \frac{\theta_1 - \theta_4}{\theta_2 - \theta_4} \quad (1.2)$$

La valeur de ce birapport est indépendante de la représentation de la droite choisie. Ce résultat a été déjà établi par les anciens mathématiciens grecs.

Il peut être étendu en utilisant les conventions suivantes qui traitent le cas des points à l'infini :

$$\frac{\infty}{\infty} = 1, \quad \frac{a}{\infty} = 0, \quad \frac{\infty}{a} = \infty \quad \text{avec } a \in \mathcal{C} \quad (1.3)$$

Ainsi, θ_i de l'équation (1.2) peut valoir éventuellement ∞ si μ_i est nul.

Birapport d'un faisceau de quatre droites

Quatre droites coplanaires et concourantes en un point O définissent un birapport $[l_1, l_2; l_3, l_4]$. Ce birapport n'est autre que le birapport $[A, B; C, D]$ (voir Figure 1.1). où A, B, C et D sont les points d'intersection des quatre droites avec une droite quelconque l ne passant pas par O . Bien sûr, la valeur du birapport $[A, B; C, D]$ est indépendante du choix de l .

Birapport d'un faisceau de quatre plans

De la même manière, dans \mathbb{P}^3 quatre plans p_1, p_2, p_3 et p_4 ayant une droite commune définissent un birapport $[p_1, p_2; p_3, p_4]$ (voir Figure 1.2). Ce birapport n'est autre que le birapport $[l_1, l_2; l_3, l_4]$ de leurs quatre droites d'intersection avec un plan quelconque p ne contenant pas la droite commune.

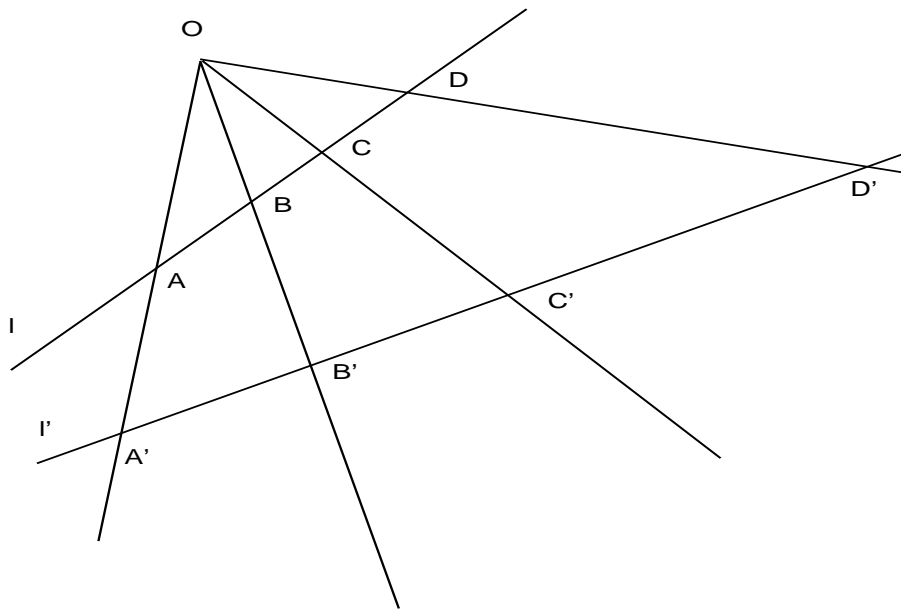


FIG. 1.1 - *Le birapport d'un faisceau de quatre droites*

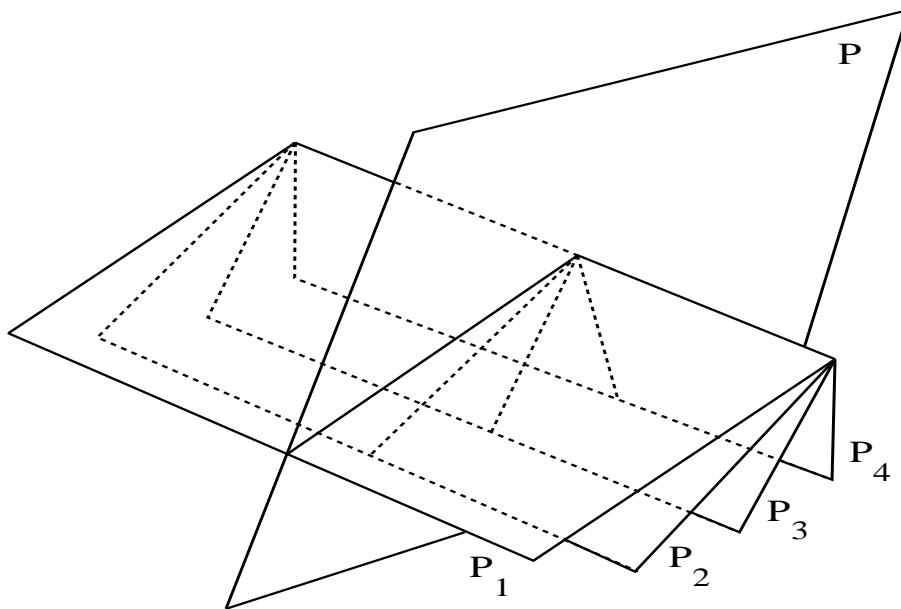


FIG. 1.2 - *Le birapport d'un faisceau de quatre plans*

1.1.4 Géométrie affine

La géométrie affine est une restriction de la géométrie projective. Elle se déduit de la cette dernière en fixant l'hyperplan de l'infini.

Les transformations affines sont les transformations projectives qui laissent l'hyperplan de l'infini globalement invariant. Si on note par IA^n l'espace affine de dimension n , la matrice associée à une transformation affine de IA^n dans IA^m est une matrice non singulière de taille $(m + 1) \times (n + 1)$ dont la dernière ligne a la valeur $(0, \dots, 0, 1)$.

Puisque la géométrie affine est une restriction de la géométrie projective, les invariants projectifs sont aussi des invariants affines. De plus, d'autres invariants plus riches existent pour les transformations affines. Par exemple le parallélisme est un invariant affine.

L'invariant fondamental de la géométrie affine est le rapport des mesures algébriques de trois points alignés. Cet invariant n'est autre que le birapport de trois points alignés et du point à l'infini dans cette direction. Soient A, B, C trois points alignés et D le point à l'infini dans la même direction. Par définition le birapport de ces quatre points est donné par :

$$[A, B; C, D] = \frac{\overline{AC}}{\overline{AD}} \div \frac{\overline{BC}}{\overline{BD}}$$

En utilisant les conventions données par les équations (1.3), on aura :

$$[A, B; C, D] = \frac{\overline{AC}}{\overline{BC}}$$

1.1.5 Géométrie euclidienne

La géométrie euclidienne est une restriction de la géométrie affine. Elle est donc une restriction plus forte de la géométrie projective.

Les transformations euclidiennes sont donc des transformations affines plus restreintes, elles admettent donc plus d'invariants. Les transformations euclidiennes laissent l'hyperplan de l'infini globalement invariant, ainsi que certains points de cet hyperplan.

Nous appelons géométrie euclidienne la géométrie des similitudes et non celle des isométries. L'espace euclidien est alors appelé espace euclidien étendu, c'est à dire, l'espace invariant sous les transformations des similitudes. Cependant, par abus de langage, cet espace est parfois appelé espace euclidien.

Géométrie euclidienne en dimension 2

Les transformations euclidiennes ou similitudes sont les transformations affines qui laissent deux points de la droite de l'infini globalement invariants. Soient I

et J ces deux points, ils sont appelés l'*absolu* ou points cycliques. Les similitudes qui laissent le couple (I, J) dans le même ordre sont appelées similitudes directes, sinon, elles sont dites indirectes.

A partir de la définition de l'*absolu*, le concept d'angle peut être défini. Soit O le point d'intersection de deux droites l_1 et l_2 . L'angle entre ces deux droites peut être calculé en utilisant la formule de Laguerre :

$$\alpha = \widehat{l_1 l_2} = \frac{1}{2i} \log[P_1, P_2; I, J] \quad (1.4)$$

où I et J sont les points cycliques, P_1 et P_2 sont les points d'intersections des deux droites l_1 et l_2 avec la droite de l'infini (voir Figure 1.3).

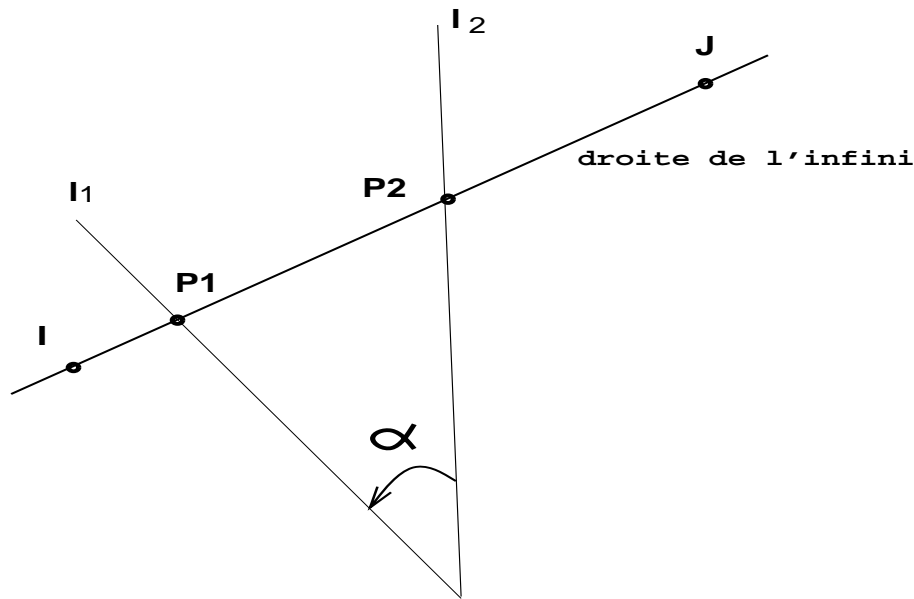


FIG. 1.3 - L'angle en géométrie euclidienne 2D

Géométrie euclidienne en dimension 3

Comme pour le cas 2D, les transformations euclidiennes dans l'espace 3D sont des transformations affines laissant un ensemble de points du plan à l'infini globalement invariant. Cet ensemble de points qui est appelé l'*absolu* est une conique. Généralement on choisit la conique imaginaire notée Ω et appelée *conique absolue* d'équations :

$$\begin{cases} x^2 + y^2 + z^2 = 0 \\ t = 0 \end{cases} \quad (1.5)$$

Deux droites sécantes de l'espace l_1 et l_2 définissent un plan P . Celui-ci coupe la *conique absolue* Ω en deux points I et J . Les deux droites l_1 et l_2 à leur tour coupent

le plan de l'infini en deux points A_1 et A_2 . Les quatre points I , J , A_1 et A_2 sont alignés puisque la *conique absolue* est dans le plan de l'infini. L'angle entre les deux droites l_1 et l_2 est alors défini comme pour le cas 2D par le birapport de ces quatre points (Voir Figure 1.4)

$$\alpha = \widehat{l_1 l_2} = \frac{1}{2i} \log[A_1, A_2; I, J]$$

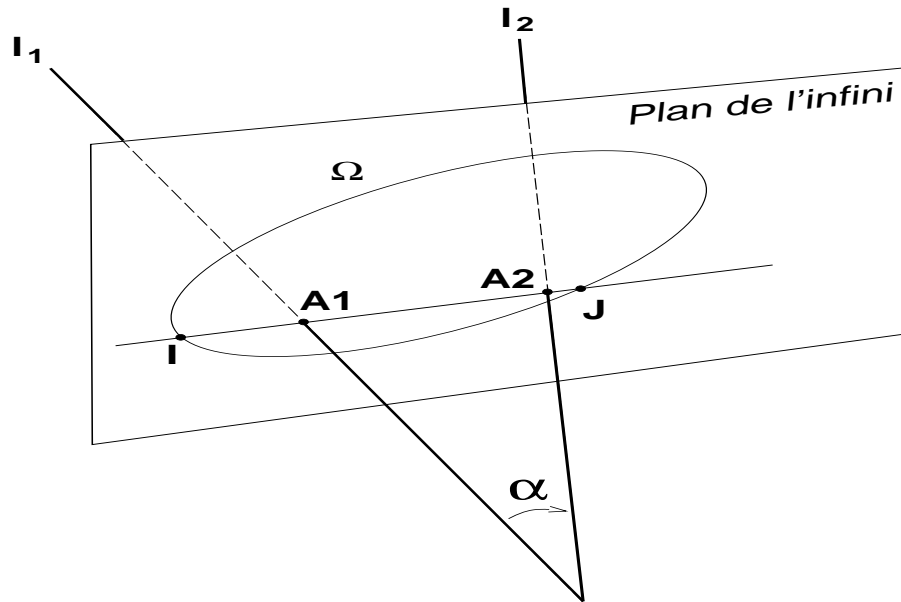


FIG. 1.4 - La notion d'angle en géométrie euclidienne 3D

Il faut noter que la *conique absolue* est un concept utilisé en vision par ordinateur. La connaissance de la projection de celle-ci est équivalente à la connaissance des paramètres intrinsèques d'une caméra [MF92].

1.2 Images et géométrie

Une image à niveaux de gris obtenue avec une caméra CCD est un tableau de dimension 2 dont les éléments sont les pixels. Ce tableau contient un signal résultant d'un processus complexe d'acquisition de données où 3 aspects interviennent : l'aspect optique, l'aspect électronique et enfin l'aspect géométrique. Nous nous limiterons dans ce qui va suivre à l'aspect géométrique de la formation d'image. Pour une étude plus complète du processus d'acquisition d'une image prenant en compte la partie optique et électronique voir [Bey92b].

1.2.1 Formation d'image

En se limitant à l'aspect géométrique, une image à niveaux de gris obtenue avec une caméra est le résultat d'une transformation géométrique. Cette dernière fait passer d'une représentation tridimensionnelle de la scène à une représentation bidimensionnelle de cette même scène.

A une caméra est attaché un repère d'origine O_c appelé *centre de projection*. (O_c, x) et (O_c, y) sont les axes parallèles au plan image, correspondant aux directions des lignes et colonnes des pixels du plan image. L'axe (O_c, z) appelé aussi *axe optique* est orienté vers la scène et est perpendiculaire au plan image. Le repère (O_c, x, y, z) est orthonormé direct (voir Figure 1.5). Pour passer des coordonnées définies dans le repère de la scène aux coordonnées images exprimées en pixels, trois phases sont nécessaires :

1. **Un déplacement tridimensionnel :** les points tridimensionnels exprimés dans un repère de la scène subissent un changement de repère pour passer au repère de la caméra. Ce changement de repère comporte donc 6 paramètres : 3 pour la rotation et 3 pour la translation. Ces paramètres ne sont autre que la position et l'orientation de la caméra, ils sont appelés *paramètres extrinsèques*.
2. **Une projection 3D-2D :** après le changement de repère de la phase précédente, les points tridimensionnels exprimés dans le repère de la caméra sont projetés sur le plan image. Les nouvelles coordonnées ainsi obtenues sont appelées *coordonnées normalisées*.

Il est à noter que cette projection est le plus souvent considérée comme une projection perspective pure (voir paragraphe 1.2.2). Cependant, d'autres modèles peuvent être proposés pour approcher cette transformation.

3. **Changement de coordonnées :** pour passer aux coordonnées pixels, les coordonnées normalisées subissent une transformation affine du plan. Cette dernière, comporte 5 paramètres appelés *paramètres intrinsèques* de la caméra. Cette transformation affine peut s'écrire sous la forme matricielle suivante :

$$A = \begin{pmatrix} \alpha_u & -\alpha_u \cot \theta & u_0 \\ 0 & \alpha_v \sin \theta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.6)$$

où

- α_u et α_v sont les deux facteurs d'échelle le long des axes du plan image. Ils sont fonction de la distance focale f et de la taille des pixels. La distance focale f est la distance du plan image à l'origine du repère caméra. Si la taille d'un pixel image est de $k_u \times k_v$ (pixels non carrés)

en *pixels/mm* et si f est donnée en *millimètres*, alors $\alpha_u = fk_u$ et $\alpha_v = fk_v$.

- u_0 et v_0 sont les coordonnées en pixels du centre image, c'est à dire les coordonnées du point d'intersection de l'axe optique avec le plan image.
- θ est l'angle entre les deux axes du repère image. Cet angle est très proche de $\frac{\pi}{2}$, et en pratique, on suppose qu'il est égal à $\frac{\pi}{2}$. La matrice A n'a alors que 4 paramètres et peut s'écrire sous la forme réduite :

$$A = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.7)$$

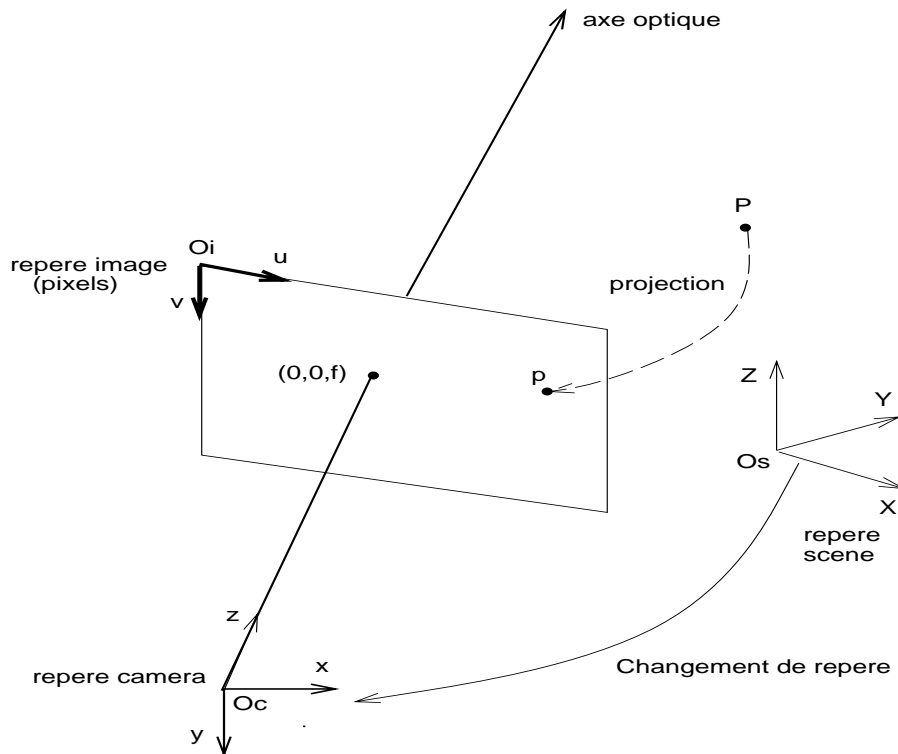


FIG. 1.5 - La formation d'image

Les trois transformations citées ci-dessus que les coordonnées tridimensionnelles des points d'une scène subissent pour arriver aux coordonnées pixels peuvent être codées dans une même matrice M tel que :

$$M = AID$$

où D , I et A sont trois matrices tels que :

- D représente le **déplacement tridimensionnel**.

- I représente la **projection 3D-2D**.
- A représente le **changement de coordonnées**

Les défauts d'optique, la numérisation au niveau des capteurs, et le bruit produit par l'électronique d'une caméra font qu'il n'existe aucun modèle pour représenter exactement une caméra réelle. Tous les modèles proposés ne sont que des idéalizations des systèmes réels d'acquisition d'images.

Plusieurs modèles géométriques sont généralement utilisés pour approcher les caméras : les modèles affines (projection orthographique, perspective faible...) et le modèle projectif (projection perspective). Les modèles affines sont simples. Cependant, ils ne sont pas adaptés à tout type de scène. Le modèle projectif approche mieux une caméra réelle qu'un modèle affine au prix de quelques paramètres en plus. Sans qu'il soit parfait, ce modèle est une bonne approximation pour une caméra.

Tous ces modèles ne sont pas vraiment indépendants. Le modèle projectif approche les modèles affines quand la distance séparant la caméra de la scène devient assez grande comparée à la taille de la scène. On utilisera dans la suite le terme *rapport de profondeur* qui est défini comme étant le rapport de la longueur de la scène le long de l'axe optique avec la distance scène-caméra. Ce rapport devient faible quand le modèle projectif approche les modèles affines pour une caméra. En outre, il est assez fréquent d'être dans des conditions où le *rapport de profondeur* est faible, ce qui donne aux modèles affines un grand intérêt.

Nous présentons ici les principaux modèles utilisés en vision par ordinateur. Ces derniers diffèrent au niveau de la projection 3D-2D, et nous nous limiterons alors à présenter les différentes projections qui définissent ces modèles.

Nous ferons donc abstraction du changement de repère scène-caméra et du passage des coordonnées normalisées aux coordonnées pixels. Cela revient à supposer que les coordonnées 3D sont définies dans le repère caméra, et que les coordonnées 2D sont les coordonnées normalisées.

Nous commençons par le modèle le plus largement utilisé en vision par ordinateur, le modèle projectif, appelé modèle de projection perspective ou modèle sténopé¹.

Ensuite, des modèles plus simples sont proposés, ces derniers pouvant être considérés comme des simplification du modèle sténopé.

Nous reviendrons sur deux de ces modèles au chapitre 3 où les reconstructions utilisant ces modèles sont décrites et comparées.

1. trou d'épingle

1.2.2 Modèle sténopé

Ce modèle (voir figure 1.6) est une bonne approximation pour les projections effectuées par les caméras CCD. De plus, la transformation effectuée est alors une projection perspective pure décrite par une matrice de rang 3, qui appartient aux transformations projectives de \mathbb{P}^3 dans \mathbb{P}^2 [Mor93].

En utilisant les coordonnées homogènes, la projection perspective pure d'un point $P(X, Y, Z, T)$ de l'espace s'écrit sous la forme matricielle :

$$\begin{pmatrix} x \\ y \\ t \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix} \quad (1.8)$$

où \sim dénote l'égalité projective.

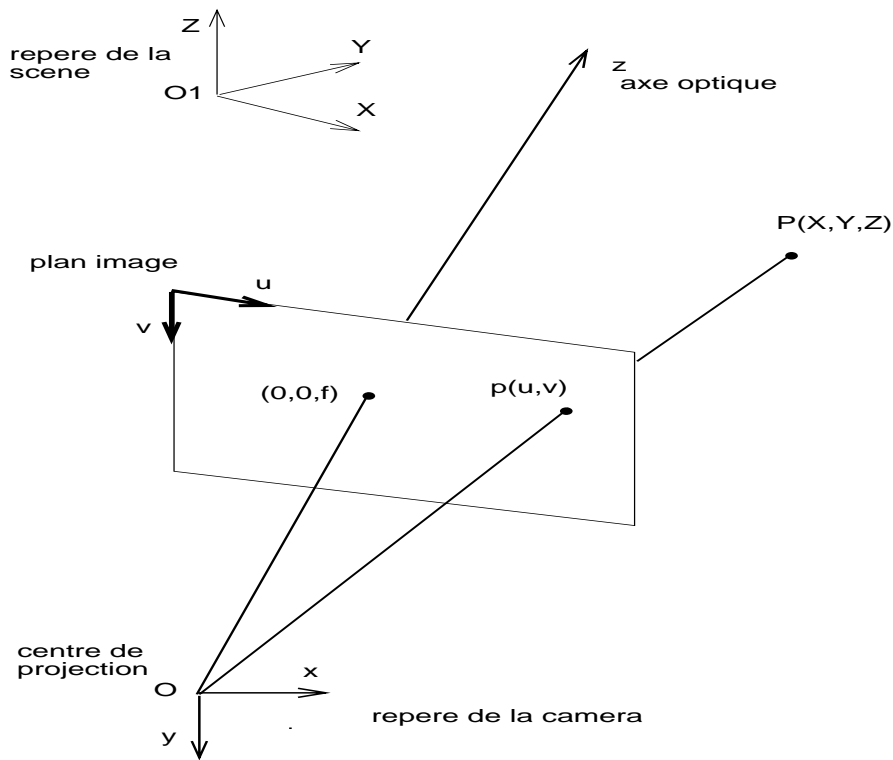


FIG. 1.6 - Le modèle sténopé pour une caméra

Avec les coordonnées non homogènes, les équations de projection perspectives s'écrivent :

$$\begin{cases} u = \frac{X}{Z} \\ v = \frac{Y}{Z} \end{cases} \quad (1.9)$$

Il est à noter que, pour certaines applications, des corrections géométriques sont parfois ajoutées à ce modèle en vue d'obtenir une meilleure précision dans les résultats [Bey91].

1.2.3 Modèle de projection orthographique

C'est le modèle le plus simple pour approcher une caméra, un point de la scène se projette orthogonalement sur l'image (voir Figure 1.7).

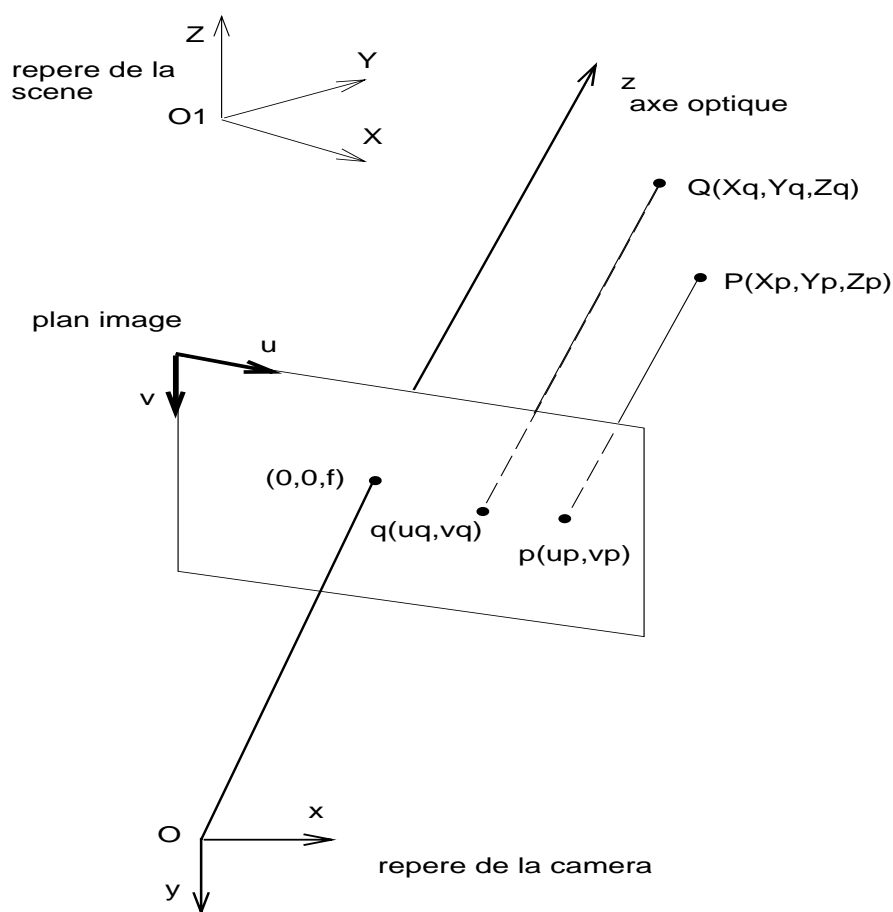


FIG. 1.7 - Le modèle de projection orthographique pour une caméra

Dans ce modèle, les équation de la projection se réduisent à :

$$\begin{cases} u = X \\ v = Y \end{cases} \quad (1.10)$$

où (X, Y, Z) sont les coordonnées tridimensionnelles non homogènes d'un point de la scène, et (u, v) sont les coordonnées non homogènes normalisées de la projection de P .

Quand la distance focale est très grande ou quand la scène se trouve très éloignée de la caméra, la projection orthographique peut être une bonne approximation pour la projection réalisée par une caméra [Tom91].

1.2.4 Modèle de projection perspective faible

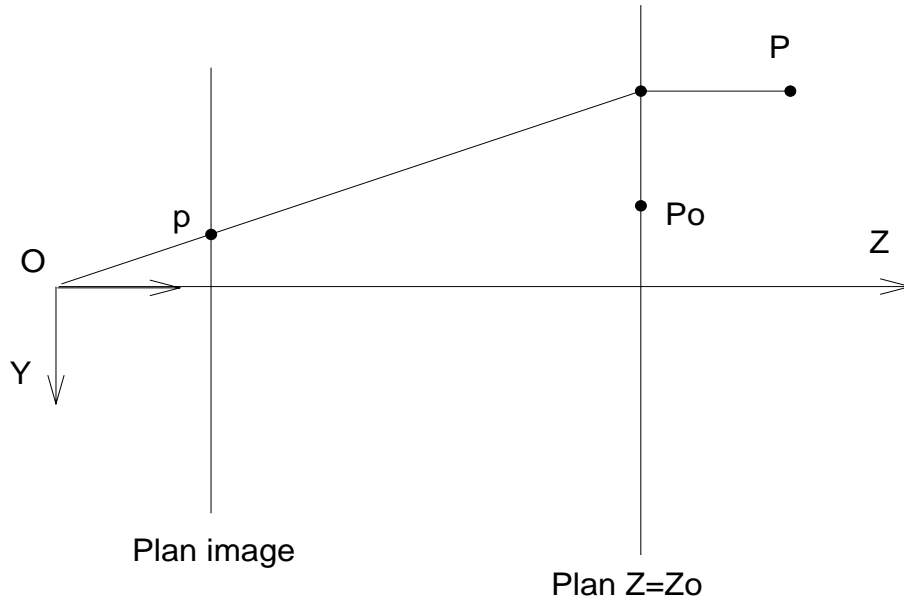


FIG. 1.8 - *Le modèle de projection perspective faible pour une caméra*

Dans ce modèle², on suppose que les points subissent d'abord une projection orthographique sur un plan perpendiculaire à l'axe optique et passant par le centre de gravité des points de la scène observée. Ensuite, ces points, devenus coplanaires, subissent une projection perspective sur le plan image (Fig 1.8).

Par rapport à la projection orthographique, ce modèle à l'avantage de tenir compte de la distance entre la scène observée et la caméra. C'est donc une projection orthographique améliorée. On l'appelle aussi projection orthographique à l'échelle.

Les équations de projection deviennent :

$$\begin{cases} u = \frac{X}{Z_0} \\ v = \frac{Y}{Z_0} \end{cases} \quad (1.11)$$

². en anglais "weak perspective projection model"

où $P_0(X_0, Y_0, Z_0)$ est le centre de gravité des points de la scène (voir figure 1.8).

Ce modèle peut être une bonne approximation pour la caméra, si le *rapport de profondeur* est faible, c'est à dire si la taille de la scène est relativement petite par rapport à la distance entre la caméra et la scène (voir section 3.2).

1.2.5 Modèle de projection para-perspective

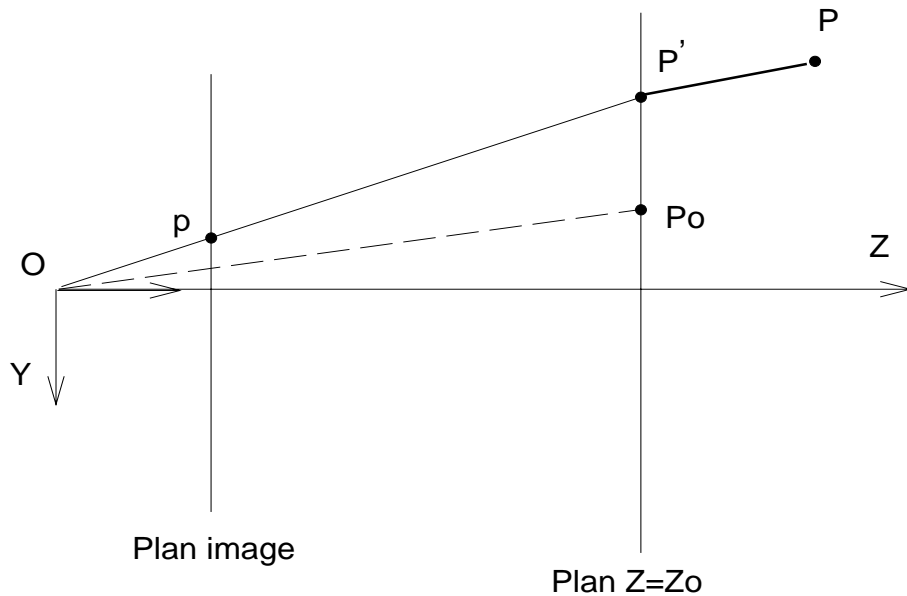


FIG. 1.9 - *Le modèle de projection para-perspective pour une caméra*

Ce modèle est une variante du modèle précédent. Les points subissent d'abord une projection parallèle sur le plan perpendiculaire à l'axe optique et passant par le centre de gravité de la scène P_0 selon la direction de projection définie par \vec{OP}_0 (voir Figure 1.9). Ensuite, ces points subissent une projection perspective sur le plan image.

Si, après la projection parallèle, le point $P(X, Y, Z, T)$ se transforme en $P'(X', Y', Z_0, T)$, les équations de projection sur le plan image s'écrivent :

$$\begin{cases} u = \frac{X'}{Z_0} \\ v = \frac{Y'}{Z_0} \end{cases} \quad (1.12)$$

où (u, v) sont les coordonnées normalisées non homogènes de la projection du point P selon ce modèle.

Par rapport au modèle perspective faible, ce modèle a l'avantage de tenir compte d'une éventuelle translation latérale importante de la scène par rapport à l'axe optique.

1.3 Relation géométrique entre les images

La transformation scène-image n'est inversible que si la scène est plane. Dans ce cas, cette transformation est une homographie de plans qui est par définition

inversible. On peut facilement montrer que deux images d'une scène plane sont liées par une homographie de plans.

Cependant, en général les scènes ne sont pas planes ; par conséquent les transformations scène-images ne sont pas inversibles. En dépit de cela, une relation liant des images prises sous des angles de vue différents existe. Elle est présentée dans les paragraphes qui suivent.

1.3.1 Relation géométrique entre deux images

Considérons deux caméras qui observent une même scène, ce qui est par exemple le cas de la stéréovision. Appelons les deux images produites par ces deux caméras *image gauche* et *image droite* (voir Figure 1.10).

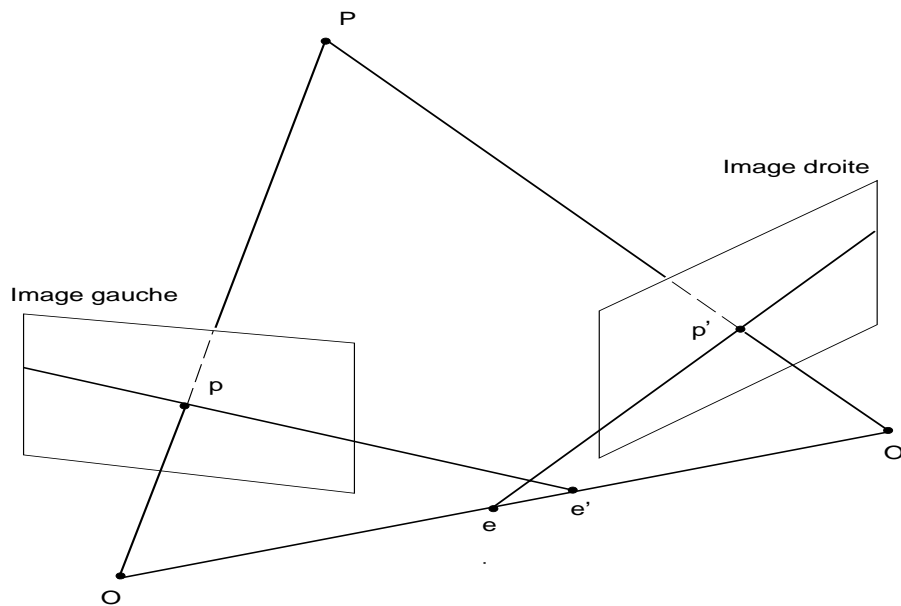


FIG. 1.10 - Relation géométrique entre une paire d'images : géométrie épipolaire

Il est clair (voir la figure 1.10) que le point p de l'*image gauche* représente la projection de tous les points de l'espace appartenant à la droite OP . Les correspondants possibles de p dans l'*image droite* sont contraints de se trouver sur une droite, appelée droite épipolaire de p . Cette dernière n'est autre que la projection de la droite OP dans l'*image droite*.

Toutes les droites épipolaires de l'*image droite* passent par e , intersection de la droite OO' avec le plan de l'*image droite*. e est appelé *épipole* de l'*image droite* par rapport à l'*image gauche*.

De manière symétrique, l'épipole de l'*image gauche* par rapport à l'*image droite* est défini comme étant l'intersection de OO' avec le plan de l'*image gauche*. En d'autres termes, un point P de l'espace définit avec les deux centres de projection

O et O' un plan appelé plan épipolaire. Ce dernier coupe les deux plans images en deux droites, les droites épipolaires. Tous les plans épipolaires passent par la droite OO' et donc par les épipoles e et e' . Cette relation entre deux images est appelée *géométrie épipolaire*. Elle est d'une grande importance. En particulier elle donne une contrainte sur la mise en correspondance : étant donné un point p de l'*image gauche*, ses correspondants possibles dans l'*image droite* ne peuvent que se trouver sur une droite.

La géométrie épipolaire peut être complètement définie par une matrice 3×3 appelée *matrice fondamentale* [Luo92]. Cette dernière transforme un point de l'*image gauche* en une droite de l'*image droite* par la relation :

$$Fp = e'p'$$

où F est la *matrice fondamentale* et Fp la droite épipolaire (ici, la droite est donnée par ses 3 coefficients).

p' appartient à la droite Fp , alors ${}^t p' Fp = 0$. Aussi, par symétrie on a ${}^t p Fp' = 0$.

Nous revenons dans la section 2.1 sur le calcul et les propriétés de cette matrice.

1.3.2 Relation géométrique entre trois images

Considérons maintenant le cas de trois caméras observant une même scène. En prenant les images deux à deux nous avons 3 paires d'images définissant ainsi 3 *matrices fondamentales* F_{12} , F_{13} et F_{23} (voir Figure 1.11). Soit P un point de la scène se projetant en p , p' et p'' dans les images 1, 2 et 3 respectivement.

En considérant la géométrie épipolaire liant la première image avec la troisième, on peut conclure que p'' doit se trouver sur la droite $(F_{13}p)$. De la même manière, en considérant la deuxième image avec la troisième, on peut conclure que p'' doit se trouver sur la droite $(F_{23}p')$. p'' se trouve donc à l'intersection des deux droites épipolaires :

$$p'' = (F_{13}p) \wedge (F_{23}p') \tag{1.13}$$

avec \wedge représentant le produit vectoriel

p , p' et p'' sont donc liés par une relation trilinéaire qui dépend des matrices fondamentales. Cette relation est d'un grand intérêt pour la stéréovision trinoculaire et pour la poursuite de points dans une séquence d'images. Elle permet d'effectuer des prédictions et des vérifications.

Il est à noter que la relation (1.13) devient dégénérée pour les points du plan tri-focal (plan défini par les centres de projection des trois caméras) ou pour le cas de droites épipolaires parallèles.

La trilinearité peut aussi être établie algébriquement [Sha94], sans passer par le calcul de la géométrie épipolaire : elle est calculée directement à partir de 9 points mis en correspondance dans les trois images. Cette façon de procéder évite de conduire aux cas de dégénérescence indiqués ci-dessus.

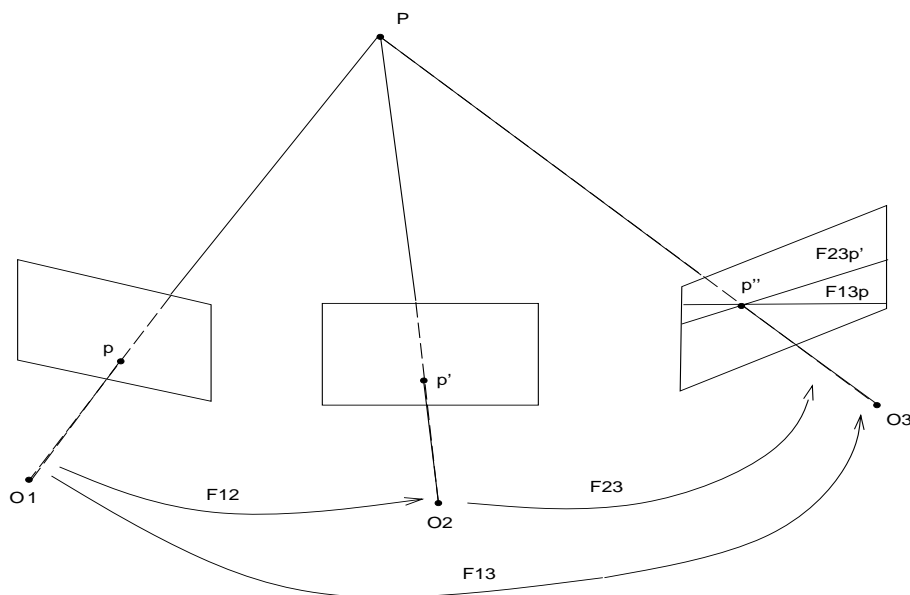


FIG. 1.11 - Relation géométrique entre 3 images : trilinearité

1.4 Conclusion

Dans ce premier chapitre, nous avons présenté quelques notions de géométrie projective utiles pour les prochains chapitres. Quelques modèles pour approcher la projection effectuée par une caméra CCD sont énumérés et décrits. La présentation de ces modèles de caméra a pour but de lever certaines ambiguïtés pour décrire des notions utilisées dans la suite de la thèse.

De ce chapitre, nous retenons en particulier l'importance du rôle joué par la géométrie en vision par ordinateur. Ceci est illustré par l'aspect géométrique de la formation d'image à niveaux de gris présenté dans la section 1.2.1 et par les relations entre images dont la description est basée sur la géométrie.

Chapitre 2

Matrice fondamentale et mise en correspondance de points

La matrice fondamentale décrit une relation géométrique entre deux images prises sous deux angles de vue différents : la géométrie épipolaire. Dans le cas non étalonné (images obtenues avec des caméras non étalonnées), la connaissance de la matrice fondamentale est considérée comme un étalonnage faible¹ pour une paire d'images. Sa relation avec la matrice essentielle introduite antérieurement par Longuet-Higgins [LH81] à été montré dans [Luo92]. En particulier, quand les coordonnées normalisées (section 1.2.1) sont utilisées, la matrice fondamentale coïncide avec la matrice essentielle.

De ce fait, cette matrice joue un rôle central pour résoudre des problèmes de vision tels que le calcul linéaire de la structure projective [Fau92] [Har93], l'auto-calibration [LF92], la poursuite de points dans une séquence d'images (section 2.2) et le calcul du mouvement d'une caméra (chapitre 5).

Dans ce chapitre nous traitons deux sujets qui sont dépendant l'un de l'autre : le calcul de la matrice fondamentale et la mise en correspondance de points.

Nous proposons dans la première partie de ce chapitre une nouvelle expression de la matrice fondamentale permettant son estimation robuste et rapide.

Dans la deuxième partie, le problème de mise en correspondance de points est traité. Un algorithme robuste de mise en correspondance basé sur la matrice fondamentale est proposé.

1. weak calibration

2.1 Calcul de la matrice fondamentale : nouvelle méthode

Soient deux images d'une même scène, par exemple celles de la figure 1.10. Comme dans la section 1.3, les deux images sont appelées image gauche et image droite (simple convention de notre part pour simplifier la présentation).

Soit P un point de l'espace projeté en p et p' sur l'image gauche et l'image droite respectivement. Rappelons que p et p' vérifient la relation :

$${}^t p' F p = 0 \quad (2.1)$$

où F dénote la matrice fondamentale, p et p' sont donnés par leurs coordonnées homogènes. F est une matrice 3×3 de rang 2 [Fau92][HGC92].

La relation (2.1) est une équation linéaire et homogène en les éléments de F ; donc F ne peut être déterminée qu'à un facteur près. On peut par exemple normaliser F (mettre à 1 la somme des carrés des éléments de F).

F peut être calculée à partir des appariements de points au moyen du critère linéaire donné par l'équation (2.1). Cependant, cette façon de faire souffre de deux défauts : l'absence de contrainte sur le rang de F et l'absence de normalisation.

En général le calcul de F se fait en deux étapes : une première estimation de F est obtenue avec le critère linéaire donné par l'équation (2.1). Cette estimation est suivie par un calcul non linéaire utilisant un autre critère. Ce dernier consiste à minimiser la distance des points à leurs droites épipolaires, il n'est autre que l'interprétation géométrique de la matrice fondamentale. Il s'écrit :

$$\min \sum_i d^2(p'_i, F p_i) + d^2(p_i, {}^t F p'_i) \quad (2.2)$$

où $d^2(p'_i, F p_i)$ dénote le carré de la distance euclidienne en pixels du point image p'_i à sa droite épipolaire définie par $(F p_i)$.

Des critères non linéaires plus complexes peuvent être utilisés. On peut se reporter à [Luo92] pour une étude plus complète des méthodes de calcul de F .

Cette description brève de la méthode de calcul de F a pour but d'illustrer la différence profonde avec la méthode que nous proposons ici. En particulier, nous ne cherchons pas à déterminer explicitement la matrice fondamentale mais plutôt la position d'un épipôle dans une des deux images et une homographie de plan entre les deux images. Nous rappelons qu'une homographie de plan est une transformation projective de plan à plan pouvant être décrite par une matrice inversible de taille 3×3 . Avec un astucieux changement de repère dans les images, nous montrons qu'il est possible de fixer 3 des 7 paramètres de la matrice fondamentale et de réduire ainsi à 4 le nombre de paramètres à déterminer.

Le premier paragraphe de cette section présente l'aspect géométrique du problème. Dans le second paragraphe, notre méthode algébrique pour résoudre le problème est proposée ; nous démontrons en particulier qu'avec 7 couples de points mis en correspondance trois solutions existent pour l'épipole. Nous retrouvons ainsi le résultat déjà trouvé par Sturm [Stu69] mais avec une démonstration plus simple.

Dans le troisième paragraphe nous montrons qu'une matrice fondamentale peut s'écrire comme le produit d'une matrice antisymétrique avec une matrice décrivant une homographie de plans.

Les propriétés d'une matrice fondamentale [FLM92], à savoir le rang et le nombre de paramètres, sont implicites dans notre formulation ; la démonstration est alors directe.

Enfin le dernier paragraphe présente l'algorithme numérique de résolution ainsi que des résultats expérimentaux sur des données réelles et simulées.

La méthode proposée ici présente l'avantage d'une gestion implicite du rang de F et d'une formulation simple, conduisant par ailleurs à un système bien conditionné. La stabilité des résultats remet à l'honneur la méthode linéaire écartée à cause de son instabilité dans la méthode classique.

2.1.1 Aspect géométrique du problème

Soit Π un plan de l'espace défini par 3 points non alignés P_1 , P_2 et P_3 ; soient deux autres points P_4 et P de l'espace, non coplanaires avec les trois points précédents (voir figure 2.1). Cet ensemble de points est observé par deux caméras dont les centres de projection sont respectivement O et O' et qui donnent une image gauche et une image droite.

La droite OP_4 (toujours en observant la figure 2.1) coupe le plan Π en Q_4 . Les deux points de l'espace P_4 et Q_4 se projettent dans l'image gauche sur le même point p_4 . Cependant, dans l'image droite les deux points P_4 et Q_4 se projettent sur deux points différents p'_4 et q'_4 . Ceci est vrai quand le mouvement liant les deux caméras comporte une composante non nulle de translation (effet de la parallaxe). D'une manière similaire, les deux points P et Q sont projetés sur l'image gauche en un point unique p et sur l'image droite en deux points différents p' et q' .

Dans l'image droite, les 2 droites $(p'_4q'_4)$ et $(p'q')$ sont les projections des 2 droites de l'espace (OP_4) et (OP) respectivement. Par définition $(p'_4q'_4)$ et $(p'q')$ sont les 2 droites épipolaires correspondant aux 2 points p_4 et p de l'image gauche. Aussi, l'intersection de ces deux droites donne l'épipole (voir figure 2.1).

Cependant un problème reste non résolu : les localisations de q'_4 et q' dans l'image droite ne sont pas connues, car q'_4 et q' ces sont les projections de deux points virtuels et ne seront donc jamais présents dans les images.

Notons qu'avec 6 points de l'espace dont 4 sont coplanaires, ce problème peut être résolu [Moh93]. Les 4 points coplanaires permettent de trouver une homogra-

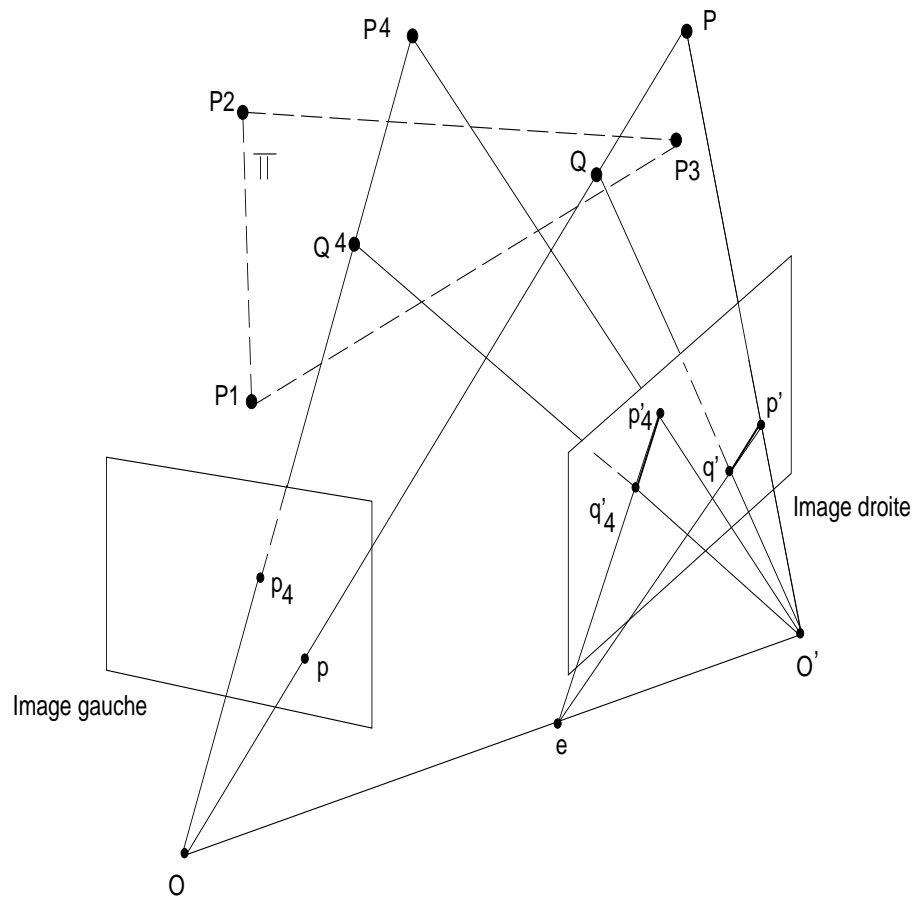


FIG. 2.1 - Construction géométrique de l'épipole dans l'image droite

phie de plans entre les deux images. Cette dernière suffit pour déterminer les projections des points virtuels dans l'image droite.

2.1.2 Solution algébrique

Poser le problème de la géométrie épipolaire comme décrit dans le paragraphe précédent n'est pas une nouveauté en soit [Cou94][LC94]; c'est la solution à ce problème proposée ici qui est originale.

Notre méthode de calcul de la matrice fondamentale suppose donnés les appariements de points dans les deux images ; c'est la seule hypothèse qui sera utilisée dans toute la suite.

Pour simplifier la présentation de cette section, nous utilisons dans toute la suite les mêmes notations que celles de la figure 2.1 ; à savoir : le plan est noté Π , les trois points définissant ce plan P_1 , P_2 et P_3 , leurs projections dans l'image gauche p_1 , p_2 et p_3 et dans l'image droite p'_1 , p'_2 et p'_3 .

Soient P un point quelconque de l'espace et Q le point d'intersection de la droite (OP) avec le plan Π . Supposons qu'il existe un moyen nous permettant de connaître pour chaque point P la position de la projection du point virtuel Q dans l'image droite, c'est à dire q' . Avec de plus la connaissance de l'épipole e , nous avons alors exprimé une contrainte équivalente à celle exprimée par la matrice fondamentale : à un point de l'image gauche p correspond la droite épipolaire dans l'image droite (eq') .

Les projections des points de Π sur l'image gauche et sur l'image droite sont deux homographies du plan qu'on note respectivement H_1 et H_2 . Appelons $H_1(\Pi)$ et $H_2(\Pi)$ les deux projections de Π sur l'image gauche et l'image droite ; comme les homographies sont inversibles, il existe alors une homographie H liant $H_1(\Pi)$ et $H_2(\Pi)$ tel que : $H = H_2 H_1^{-1}$, avec H_1^{-1} homographie inverse de H_1 .

Le problème à résoudre est alors de trouver l'homographie H vérifiant :

$$\forall p \in H_1(\Pi), H(p) = q'$$

ou en notation matricielle :

$$\forall p \in H_1(\Pi), Cp = q' \tag{2.3}$$

où C est une matrice 3×3 décrivant l'homographie H , p est la projection d'un point P de l'espace et q' est la projection du point virtuel Q sur l'image droite. Il est à noter que le point q' peut être la projection d'un point physique quand celui-ci appartient au plan Π (P et Q sont confondus dans ce cas).

Simplifier la forme de C

Pour chaque couple (p, q') de points appariés, la relation (2.3) nous fournit deux équations linéaires et indépendantes en les éléments de C . En d'autres termes, quatre points coplanaires de l'espace dont les projections dans les deux images sont appariées nous permettent de déterminer C . Par contre, sans aucune connaissance a priori il n'est pas possible de détecter la coplanarité depuis une image des points. Cependant, trois points (non colinéaires) de l'espace définissent toujours un plan. Par conséquent, nous pouvons choisir un plan quelconque de l'espace en prenant 3 points de l'image gauche comme étant les projections de 3 points appartenant à ce plan.

Avec nos conventions (voir figure 2.1), la relation (2.3) est vérifiée pour les trois points P_1, P_2 et P_3 , c'est à dire que $Cp_i = p'_i$ pour $i = 1 \dots 3$. Les équations (2.3) peuvent être simplifiées en effectuant un changement de base dans les deux images. Nous passons des coordonnées pixels usuelles aux coordonnées projectives définies dans la base canonique de l'espace projectif de dimension deux \mathbb{P}^2 . Ceci est fait en effectuant une transformation projective du plan qui affecte les coordonnées ${}^t(0, 0, 1)$, ${}^t(1, 0, 0)$, ${}^t(0, 1, 0)$ et ${}^t(1, 1, 1)$ aux points de l'image gauche p_1, p_2, p_3 et v respectivement, avec v un point quelconque de l'image gauche. En procédant de la même manière avec l'image droite, nous effectuons une transformation projective du plan qui affecte les coordonnées ${}^t(0, 0, 1)$, ${}^t(1, 0, 0)$, ${}^t(0, 1, 0)$ et ${}^t(1, 1, 1)$ aux points images p'_1, p'_2, p'_3 et u' respectivement, avec u' un point quelconque de l'image droite (u' n'est pas nécessairement le correspondant de v).

Il faut noter que les quatre points (dans chacune des deux images) ne doivent pas être 3 à 3 colinéaires.

Les homographies de plan à plan sont inversibles. Nous pouvons donc effectuer tous nos calculs avec les nouvelles coordonnées projectives puis revenir aux coordonnées pixels de départ.

Avec ce choix judicieux des nouvelles coordonnées dans les deux images et en utilisant la relation (2.3) pour les trois couples de points (p_1, p'_1) , (p_2, p'_2) et (p_3, p'_3) la matrice C est réduite à la forme :

$$C = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}$$

Cette matrice a trois paramètres dont deux seulement sont indépendants. De plus, ces trois paramètres ne peuvent être nuls. Sinon C ne sera pas inversible et H sera une homographie singulière qui transforme un plan en une droite. Par conséquent nous pouvons fixer un des trois paramètres, par exemple en mettant γ à 1. Dans la suite on utilise la matrice sous sa nouvelle forme réduite :

$$C = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.4)$$

Equation de base liant un couple d'appariement

Nous supposons à présent que C est donnée par (2.4). Soient p un point de l'image gauche avec les coordonnées homogènes ${}^t(x, y, t)$ et p' son correspondant dans l'image droite avec les coordonnées homogènes ${}^t(x', y', t')$. En utilisant la relation (2.3), les coordonnées du point q' (projection du point virtuel Q) dans l'image droite sont données par :

$$q' = Cp = \begin{pmatrix} \alpha x \\ \beta y \\ t \end{pmatrix} \quad (2.5)$$

où seuls α et β sont inconnus.

Soit ${}^t(e_x, e_y, e_t)$ les coordonnées inconnues de l'épipole e de l'image droite.

Par construction géométrique, le point q' appartient à la droite (ep') qui est la droite épipolaire (voir figure 2.1). Ceci s'écrit sous forme d'équation :

$$(e \wedge p') \cdot q' = 0 \quad (2.6)$$

où \wedge dénote le produit vectoriel et \cdot le produit scalaire.

$(e \wedge p')$ est un vecteur de dimension 3 dont les éléments sont les coefficients de la droite passant par les deux points e et p' .

En développant (2.6) et en utilisant les coordonnées de q' données dans (2.5) on obtient l'équation :

$$(e_t y' - t' e_y) \alpha x + (t' e_x - e_t x') \beta y + (e_y x' - y' e_x) t = 0 \quad (2.7)$$

L'équation (2.7) peut être écrite pour chaque appariement (p_i, p'_i) ; les inconnues ici sont au nombre de 5 : α , β , e_x , e_y et e_t . Mais comme (e_x, e_y, e_t) sont des coordonnées homogènes, seules deux de ces trois coordonnées sont indépendantes. Par conséquent, nous avons exactement 4 inconnues indépendantes à déterminer pour trouver l'homographie H et l'épipole e ; au moins quatre points de l'espace n'appartenant pas à Π et dont les projections sont appariées dans les deux images sont alors nécessaires pour trouver une solution aux quatre inconnues.

Il est à noter que même si la scène est coplanaire (tous les points appartiennent à Π), l'équation (2.7) reste valide. Dans ce cas particulier, le vrai épipole ne peut être déterminé car n'importe quel point de l'image pourrait être l'épipole. Cependant l'homographie H vérifiera encore, pour les points de ce plan, la contrainte imposée. Ceci garantit une cohérence au niveau de l'homographie des droites épipolaires, même avec une dégradation du calcul dans le cas d'une scène presque coplanaire.

Nombre de solutions avec 7 couples de points appariés

Avec 7 couples de points appariés, 3 couples sont utilisés pour fixer le plan Π et les quatre couples qui restent permettent de trouver les 4 inconnues de l'équation (2.7) : deux pour l'homographie et 2 pour l'épipole.

Pour distinguer les inconnues posons :

$$\alpha = X_1, \quad \beta = X_2, \quad e_x = X_3 \quad \text{et} \quad e_y = X_4. \quad (2.8)$$

Puisque l'épipole n'a que deux coordonnées indépendantes, on peut fixer l'une de ses coordonnées, par exemple $e_t = 1$. Ceci n'est pas toujours valide si nous voulons calculer correctement les paramètres car e_t peut être nul. Cette hypothèse n'est pas utilisée pour le calcul numérique mais elle est utilisée uniquement pour les démonstrations présentées ici.

En effectuant les substitutions données dans (2.8), pour chaque couple d'appariement (p_i, p'_i) l'équation (2.7) devient :

$$x_i y'_i X_1 X_4 - x_i t'_i X_1 X_3 + y_i t'_i X_2 - y_i x'_i X_2 X_4 + t_i x'_i X_3 - t_i y'_i = 0 \quad (2.9)$$

où (x_i, y_i, t_i) et (x'_i, y'_i, t'_i) sont les coordonnées homogènes connues des points images p_i et p'_i respectivement, avec $i = 1 \dots 4$ (nous supposons ici que les appariements correspondant à $i = 5 \dots 7$ ont été déjà utilisés dans la relation (2.3)).

En théorie, une solution existe pour les inconnues $X_1 \dots X_4$ si quatre équations du type (2.9) sont utilisées.

En posant :

$$\begin{aligned} a_i &= x_i y'_i, \quad b_i = -x_i t'_i, \quad c_i = y_i t'_i, \quad d_i = -y_i x'_i, \quad e_i = t_i x'_i \quad \text{et} \quad f_i = -t_i y'_i, \\ &\text{et} \\ X_1 X_4 &= V1, \quad X_1 X_3 = V2, \quad X_2 = V3, \quad X_2 X_4 = V4 \quad \text{et} \quad X_3 = V5, \end{aligned} \quad (2.10)$$

l'équation (2.9) devient :

$$a_i V_1 + b_i V_2 + c_i V_3 + d_i V_4 + e_i V_5 + f_i = 0. \quad (2.11)$$

Une brève analyse des coefficients a_i, b_i, c_i, d_i, e_i et f_i montre que ces coefficients ne peuvent être tous nuls à la fois.

En utilisant l'équation (2.11) pour quatre couples de points appariés (autres que les 3 couples qui ont servi à réduire la forme de C à celle donnée par (2.4)) nous pouvons trouver une solution linéaire pour les inconnues $V_1 \dots V_4$ en fonction de V_5 (on résout pour 4 inconnues en fonction d'une cinquième).

Les solutions du système (2.11) sont de la forme :

$$V_i = A_i V_5 + B_i \quad (2.12)$$

avec $i = 1 \dots 4$; A_i et B_i connus car ils s'écrivent en fonction des constantes a_i, b_i, c_i, d_i, e_i et f_i qui dépendent des coordonnées homogènes des points dans les images. Un cinquième couple de points appariés est nécessaire si nous voulons une solution linéaire pour toutes les inconnues $V_1 \dots V_5$.

En revenant sur les notations des V_i données par (2.10) et en utilisant (2.12), nous obtenons les équations suivantes :

$$\begin{cases} X_1 X_4 = A_1 X_3 + B_1 \\ X_1 X_3 = A_2 X_3 + B_2 \\ X_2 = A_3 X_3 + B_3 \\ X_2 X_4 = A_4 X_3 + B_4 \end{cases} \quad (2.13)$$

A partir du système ci-dessus, on peut diviser la deuxième équation par la première puis, multiplier l'équation obtenue par la quatrième. Ceci s'écrit :

$$\left(\frac{X_1 X_3}{X_1 X_4}\right) X_2 X_4 = X_3 X_2 = \left(\frac{A_2 X_3 + B_2}{A_1 X_3 + B_1}\right) (A_4 X_3 + B_4)$$

En utilisant l'équation $X_2 = A_3 X_3 + B_3$ et après un calcul simple nous obtenons une équation cubique en l'inconnue X_3 :

$$A_1 A_3 X_3^3 + (A_1 B_3 + B_1 A_3 - A_2 A_4) X_3^2 + (B_1 B_3 - A_2 B_3 - B_2 A_4) X_3 - B_2 B_3 = 0$$

Une telle équation au plus 3 solutions, ce qui montre qu'avec 7 couples d'appariements trois solutions existent pour l'épipole. Ceci rejoint la démonstration de Sturm[Stu69] et celle de Faugeras[FLM92].

Nous avons montré dans ce paragraphe qu'avec 7 couples d'appariements le calcul de l'épipole est non linéaire et admet au plus 3 solutions différentes. Cependant, un huitième appariement rend le calcul de l'épipole linéaire et assure son unicité. A ce stade, nous n'avons pas calculé la matrice fondamentale. Cette dernière se déduit directement de la donnée d'un épipole et d'une homographie de plan entre deux images. Ceci est montré dans le paragraphe suivant.

2.1.3 Calcul de la matrice fondamentale

Supposons que l'épipole dans l'image droite est connu ainsi qu'une homographie de plans entre l'image gauche et l'image droite. Soient $e(e_x, e_y, e_t)$ cet épipole donné par ses coordonnées homogènes et C la matrice 3×3 qui décrit l'homographie écrite sous la forme (2.4).

Appelons F la matrice fondamentale recherchée. F est une matrice 3×3 qui à chaque point $p(x, y, t)$ de l'image gauche associe la droite épipolaire (D_p) donnée par (Fp) qui s'écrit :

$$Fp = \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ t \end{pmatrix} = \begin{pmatrix} F_{11}x + F_{12}y + F_{13}t \\ F_{21}x + F_{22}y + F_{23}t \\ F_{31}x + F_{32}y + F_{33}t \end{pmatrix} \sim \begin{pmatrix} A_p \\ B_p \\ C_p \end{pmatrix} \quad (2.14)$$

A_p , B_p et C_p sont les coefficients de la droite épipolaire dans l'image droite correspondant au point p de l'image gauche. Pour ce même point p , (D_p) correspond

aussi à la droite (eq') (voir figure 2.1) dont l'équation est donnée par ($e \wedge q'$). En utilisant l'équation (2.3) l'équation de (D_p) devient ($e \wedge (Cp)$) qui s'écrit :

$$e \wedge Cp = \begin{pmatrix} te_y - \beta ye_t \\ \alpha xe_t - te_x \\ \beta ye_x - \alpha xe_y \end{pmatrix} = \begin{pmatrix} 0 & -\beta e_t & e_y \\ \alpha e_t & 0 & -e_x \\ -\alpha e_y & \beta e_x & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ t \end{pmatrix} \sim \begin{pmatrix} A_p \\ B_p \\ C_p \end{pmatrix} \quad (2.15)$$

En comparant avec l'équation (2.14), il est clair que la matrice fondamentale F se calcule directement en fonction des coordonnées de e et C :

$$F = \begin{pmatrix} 0 & -\beta e_t & e_y \\ \alpha e_t & 0 & -e_x \\ -\alpha e_y & \beta e_x & 0 \end{pmatrix} = \begin{pmatrix} 0 & -e_t & e_y \\ e_t & 0 & -e_x \\ -e_y & e_x & 0 \end{pmatrix} \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.16)$$

F ainsi définie vérifie l'équation (2.1). De plus elle est par construction une matrice de rang 2 (le déterminant de F est toujours nul). Par conséquent F est formellement la matrice fondamentale décrivant la géométrie épipolaire entre l'image gauche et l'image droite.

La nouvelle expression de F donnée ci-dessus permet de décomposer le problème pour le calcul de la matrice fondamentale. Cette dernière étant le produit d'une matrice antisymétrique par une matrice décrivant une homographie de plans. En d'autres termes, la recherche de la matrice fondamentale est intrinsèquement liée à la recherche d'une homographie de plans.

2.1.4 Algorithme de résolution et résultats expérimentaux

Algorithme utilisé

Le calcul de e et de C (un total de 4 inconnues indépendantes) peut se faire directement en utilisant uniquement les équations linéaires (2.11). Nous appelons cette étape *calcul linéaire* de F , car seules des équations linéaires sont utilisées.

Une amélioration du résultat peut être obtenue en utilisant une deuxième étape non linéaire (2.9). Dans ce dernier cas la résolution linéaire a un rôle d'initialisation pour la résolution non linéaire qui est plus robuste et qui fournit une solution plus précise. Quand cette deuxième étape est utilisée, nous appelons le processus *calcul non linéaire* de F , car des équations non linéaires sont utilisées.

Notre algorithme utilise comme données un ensemble de points appariés dans deux images (au moins 8 appariements) et fonctionne selon le schéma suivant :

- Trois points sont choisis dans l'image gauche, avec leurs correspondants dans l'image droite (en pratique les 3 points choisis sont bien écartés). Après un changement de coordonnées (passage des coordonnées pixels aux coordonnées de la base canonique de \mathbb{P}^2), la forme de C est réduite à celle donnée par (2.4).

- Une première solution est obtenue en résolvant les équations linéaires du type (2.11). Pour ce calcul linéaire nous avons utilisé la méthode de décomposition en valeurs singulières pour sa propriété de robustesse.
- La solution obtenue à l'étape précédente est utilisée comme initialisation d'un processus de résolution non linéaire utilisant les équations du type (2.9). Nous avons utilisé l'algorithme Levenberg-Marquardt [PFTW88] qui est un algorithme bien connu d'estimation aux moindres carré pour les problème non linéaires.

Notons que cette dernière phase n'est pas nécessaire quand seul le *calcul linéaire* de F nous intéresse.

Enfin, on effectue un retour vers les coordonnées de départ (coordonnées pixels); ceci est fait à l'aide des homographies inverses. La matrice F se déduit directement de e et C .

Résultats

Nous présentons ici deux types de résultats : le premier concerne la validation avec des données réelles. Nous montrons (expérimentalement) qu'en utilisant uniquement le critère linéaire (2.11) la solution obtenue est presque aussi correcte que la solution utilisant l'étape non linéaire (2.9), ce qui est loin d'être le cas de la solution linéaire obtenue avec la méthode classique basée sur les équations (2.1) [Luo92]. Le second type de résultats concerne une analyse de stabilité et de précision des calculs avec des données simulées.

Dans toutes nos expériences avec les images réelles, les appariements de points sont obtenus automatiquement avec la méthode décrite dans la section 2.2.

Pour tester la qualité d'une matrice fondamentale F étant donné un ensemble d'appariements (p_i, p'_i) , $i = 1 \dots n$, nous avons utilisé la formule suivante :

$$Q_F = \frac{\sum_{i=1}^n d(p'_i, Fp_i) + d(p_i, {}^tFp'_i)}{2n} \quad (2.17)$$

avec Q_F la qualité de la matrice fondamentale F et d la distance euclidienne d'un point à la droite épipolaire en pixels. Notons qu'ici les points images sont donnés par leurs coordonnées pixels.

En d'autres termes, Q_F est la distance moyenne des points à leurs droites épipolaires.

Tests avec des images réelles :

Nous avons testé notre algorithme avec quatre scènes différentes :

1. Une scène d'intérieur de la "maison" (figure 2.2) : le mouvement (entre les deux images) est presque une translation latérale pure ; c'est à dire que la deuxième image est obtenue en bougeant la caméra latéralement. Dans ce cas l'épipole doit se trouver à l'extérieur de l'image.
2. Une scène d'intérieur de la "route" (figure 2.3) : le mouvement ici est une translation pure le long de l'axe optique de la caméra ; c'est à dire que la deuxième image est obtenue en avançant la caméra vers l'avant. Dans ce cas l'épipole doit se trouver presque au centre de l'image.
3. Une scène d'extérieur de la "bibliothèque" (figure 2.5) : le mouvement est principalement une translation comme pour le premier cas. L'épipole doit se trouver à l'extérieur de l'image
4. Une scène d'extérieur de "Bayard" (figure 2.4) : le mouvement est là aussi principalement une translation (ici une translation oblique). L'épipole doit se trouver à l'extérieur de l'image.

Ces résultats obtenus avec des données réelles pour différents mouvements et différentes scènes prouvent la stabilité du calcul linéaire de F . Ce dernier conditionne la convergence du calcul non linéaire.

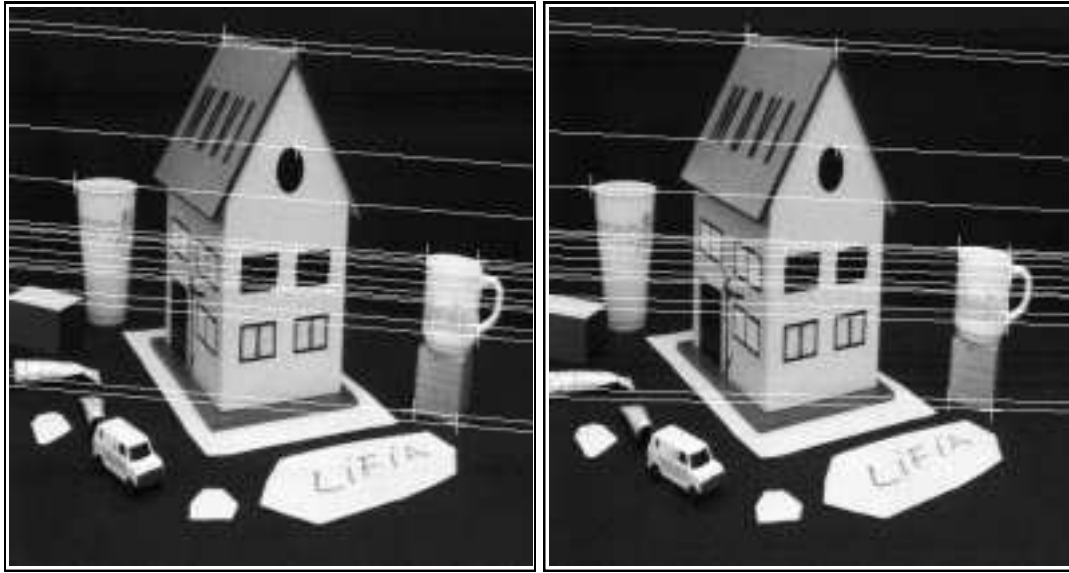
Dans la méthode classique, le calcul linéaire de F est très instable ; la convergence du calcul non linéaire n'est alors plus garantie.

Tests avec des images simulées :

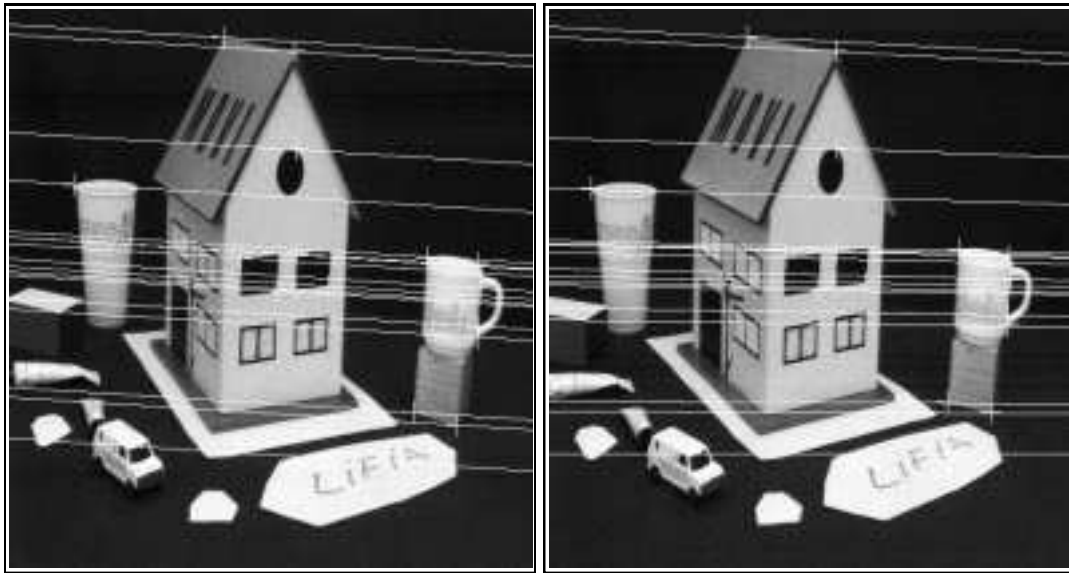
La scène utilisée ici consiste en une soixantaine de points pour un volume de $40cm \times 30cm \times 30cm$. La forme géométrique de cette scène est donnée par la figure 3.2. La caméra simulée a une focale de $12,5mm$ et est située à un mètre de la scène.

Deux séries d'expériences ont été menées ici, où F est toujours calculée avec l'étape non linéaire.

- Dans la première série deux mouvements différents pour la caméra sont utilisés : un mouvement dominé par une translation le long de l'axe optique (épipoles à l'intérieur des images) et un deuxième mouvement dominé par une translation latérale (épipoles en dehors des images). Les matrices fondamentales sont calculées en ajoutant à chaque fois un bruit uniforme d'amplitude différente aux coordonnées images.

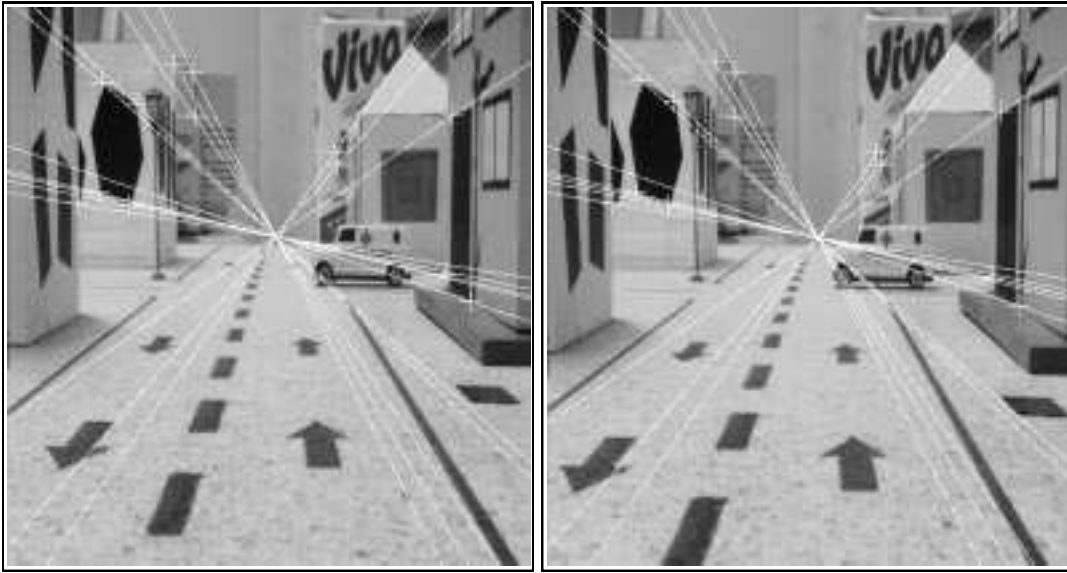


Calcul linéaire de la matrice fondamentale, $Q_F = 0.81$ pixels

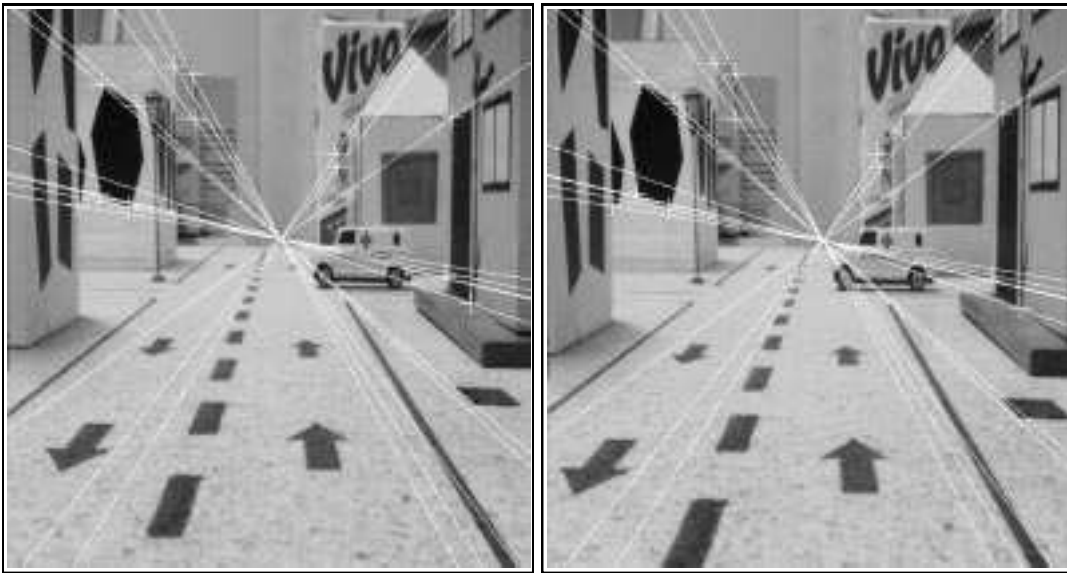


Calcul non linéaire de la matrice fondamentale, $Q_F = 0.43$ pixels

FIG. 2.2 - Résultats du calcul de la géométrie épipolaire (matrice fondamentale) pour une paire d'images d'intérieur "la maison": la caméra a bougé vers le côté. Ici le nombre de points est de 49

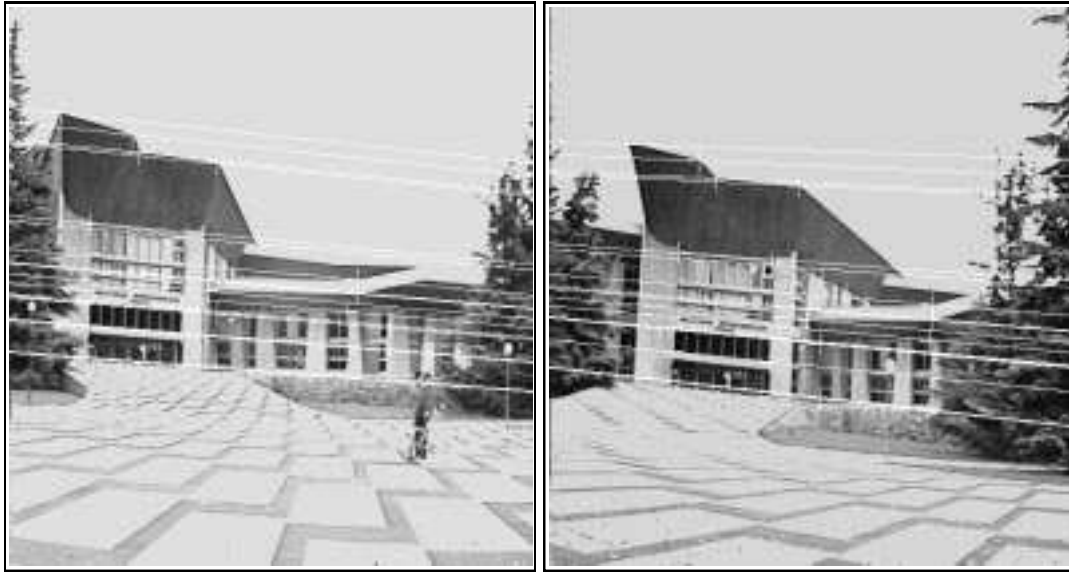


Calcul linéaire de la matrice fondamentale, $Q_F = 0.60$ pixels

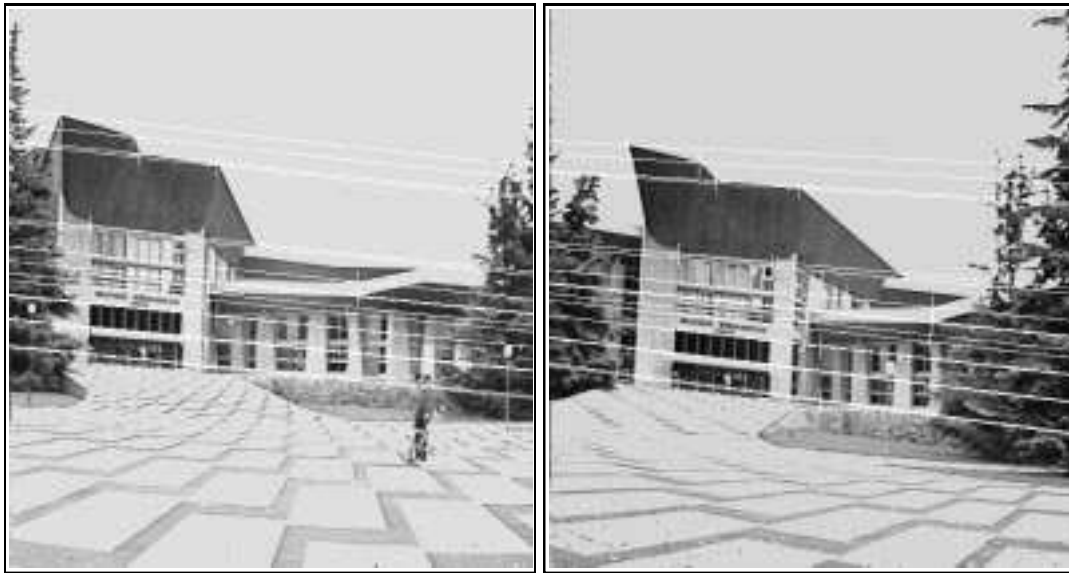


Calcul non linéaire de la matrice fondamentale, $Q_F = 0.53$ pixels

FIG. 2.3 - Résultats du calcul de la géométrie épipolaire (matrice fondamentale) pour une paire d'images d'intérieur "la route": la caméra a bougé vers l'avant. Ici le nombre de points est de 79

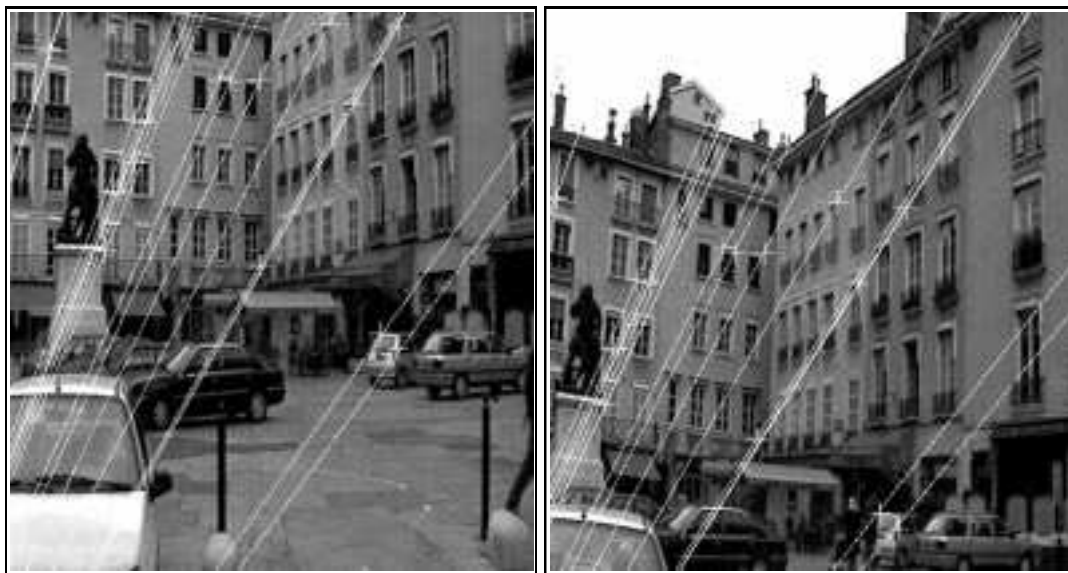


Calcul linéaire de la matrice fondamentale, $Q_F = 0.59$ pixels

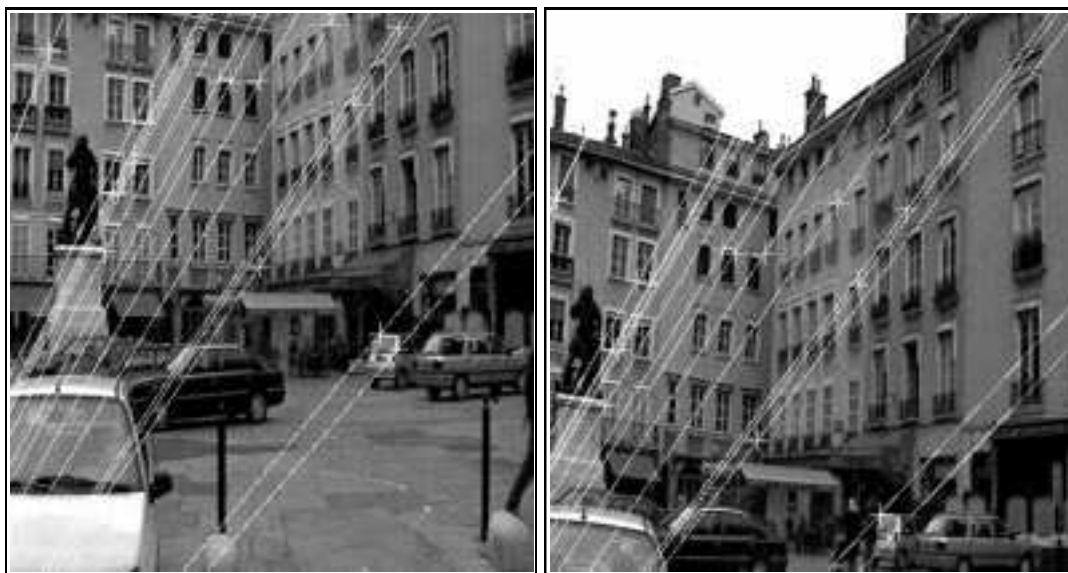


Calcul non linéaire de la matrice fondamentale, $Q_F = 0.21$ pixels

FIG. 2.4 - Résultats du calcul de la géométrie épipolaire (matrice fondamentale) pour une paire d'images d'extérieur "la bibliothèque": la caméra a bougé vers le côté. Ici, le nombre de points est de 21



Calcul linéaire de la matrice fondamentale, $Q_F = 0.47$ pixels



Calcul non linéaire de la matrice fondamentale, $Q_F = 0.26$ pixels

FIG. 2.5 - Résultats du calcul de la géométrie épipolaire (matrice fondamentale) pour une paire d'images d'extérieur "Bayard": la caméra a bougé vers le côté et vers le bas. Ici, le nombre de points est de 25

Dans cette première série d'expériences tous les points (60) sont utilisés pour le calcul de la matrice fondamentale. Les résultats sont donnés par les deux graphes de la figure 2.6.

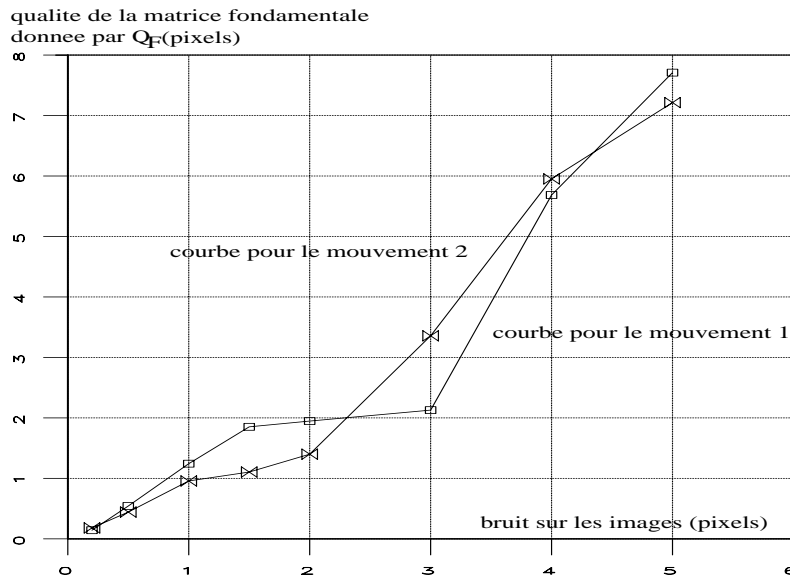


FIG. 2.6 - *Mouvement 1: une translation le long de l'axe optique; Mouvement 2: une translation latérale. Les deux courbes ne sont pas très différentes. De plus, la méthode reste stable pour un bruit d'amplitude inférieure à 3 pixels.*

- La deuxième série consiste à refaire les mêmes expériences que celle de la première série mais en changeant à chaque fois le nombre de points utilisés pour le calcul de F et en gardant un bruit d'amplitude constante. Les résultats sont donnés par les deux graphes de la figure 2.7.

Ces expériences avec des données simulées montrent que le type de mouvement effectué par une caméra n'influe pas sur la stabilité du calcul de F . De plus, quand le nombre de points utilisés pour les calculs diminue, la qualité de F reste acceptable.

2.2 Mise en correspondance de points dans deux images

Les pixels ou points images sont les projections de points physiques de la scène sur le plan image. La mise en correspondance de points dans deux images consiste à identifier dans chacune des deux images la projection des mêmes points physiques. Cette mise en correspondance de points est d'une grande importance pour la reconstruction tridimensionnelle (voir chapitre 3) et le calcul de matrice fondamentale (section précédente), ainsi que pour l'autocalibration [LF92]. C'est, en

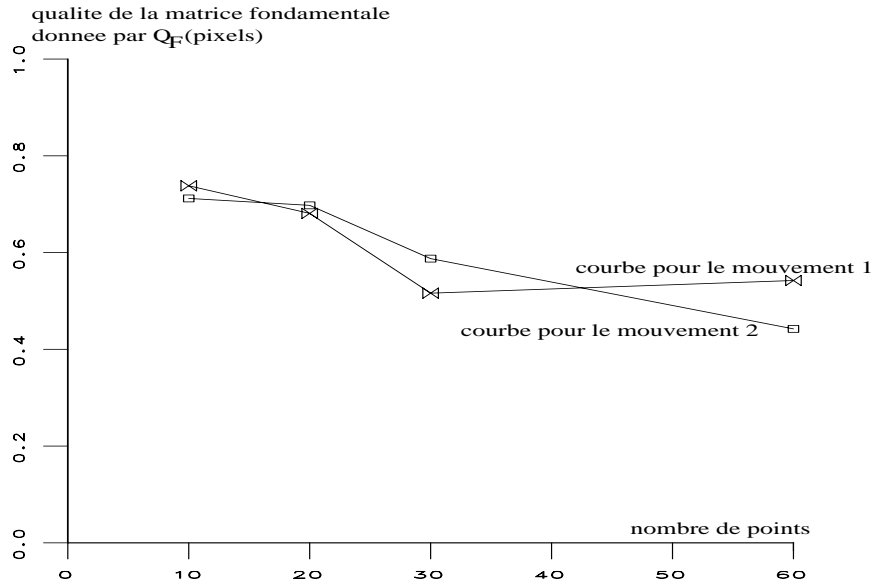


FIG. 2.7 - *Mouvement 1*: une translation le long de l'axe optique; *Mouvement 2*: une translation latérale. Le nombre de points utilisés pour le calcul de F passe de 10 à 60; par contre Q_F est toujours calculé avec les 60 points. Le bruit sur les coordonnées images utilisé ici est constant: c'est un bruit uniforme d'amplitude 0,5 pixels

d'autres termes, un outil de base nécessaire pour des tâches de vision de plus haut niveau.

Pour le cas d'une séquence d'images nous parlerons de poursuite de points. Il s'agit d'être capable de reconnaître les projections des mêmes points d'une image à l'autre dans toute la séquence.

Pour des raisons de clarté, nous limitons cette exposé au cas de deux images; la généralisation à une séquence d'images est directe.

Le problème de la mise en correspondance de points comporte deux parties: la détection ou l'extraction de points, appelés *points d'intérêt*, et leur mise en correspondance dans les deux images.

2.2.1 Détection des points

Les points à détecter dans une image doivent être des points caractéristiques de l'image pour faciliter leur mise en correspondance. Ils sont principalement les coins et les jonctions.

Si aucune information sur les scènes n'est disponible, ce qui est généralement le cas, alors seules les variations d'intensité de l'image peuvent aider à détecter les points d'intérêt. Ce problème a suscité un grand nombre de travaux [Mor77]

[MY87] [HS88] [DG90] [Roh92]. Cependant il n'a pas été complètement résolu, comme certains problèmes de bas niveau. Même si des détecteurs de points d'intérêt récents [DB93] [OG93] donnent de meilleurs résultats, aucun d'entre eux n'est complètement fiable.

Deux détecteurs de points d'intérêt ont été explorés :

- Le premier, est un détecteur de coins, c'est à dire qu'il détecte les points d'intérêt de type coin. Il prend en entrée une image à niveaux de gris et fonctionne en trois étapes :
 1. extraction des contours puis approximations de ceux-ci par des fonctions B-splines,
 2. extraction des extremums locaux de ces fonctions B-splines,
 3. ces extremums ne sont pas précis (approximations successives) et sont généralement situés à quelques pixels du vrai coin. Une méthode inspirée de [DB93] et développée au sein de notre équipe par P. Brand [BM94] est alors utilisée pour améliorer la précision. Cette méthode est basée sur l'approximation d'un coin par un modèle théorique dont les paramètres sont les coordonnées du sommet, l'angle entre les deux segments formant le coin et les niveaux de gris à l'intérieur et à l'extérieur du coin.

La localisation des coins ainsi détectés a une précision sub-pixel (voir Figure 2.8). Cependant, ce détecteur présente deux inconvénients : il est coûteux en temps de calcul et ne fonctionne que sur des scènes polyédriques.

- Le second, appelé détecteur de points d'intérêt, est plus général. Il détecte des points situés dans des régions fortement texturées (riches en information), en particulier les coins. Son principe est inspiré de [Mor77] et [HS88]. Pour quantifier la richesse en information d'un point, une région centrée autour de celui-ci est considérée. Cette région est en général une fenêtre carrée qu'on appelle masque (voir annexe A pour les détails techniques de ce détecteur).

Par rapport à [HS88], nous avons apporté une amélioration au niveau de la précision de localisation des points détectés en effectuant la détection à deux échelles. Une première détection est réalisée avec un grand masque, ce qui permet de ne pas détecter des points d'intérêt dus au bruit. Cependant, la précision de localisation est mauvaise quand les masques sont grands. Une deuxième étape avec un masque plus petit permet d'améliorer la précision des localisations des points déjà détectés. Notons que cette deuxième étape est rapide car elle ne cherche pas à détecter de nouveaux points d'intérêt.

En remarquant qu'en général les points d'intérêt se trouvent sur les contours ou à proximité de ceux-ci, nous avons donc limité la recherche de notre détecteur aux points des contours. Ceci permet de gagner en rapidité au prix

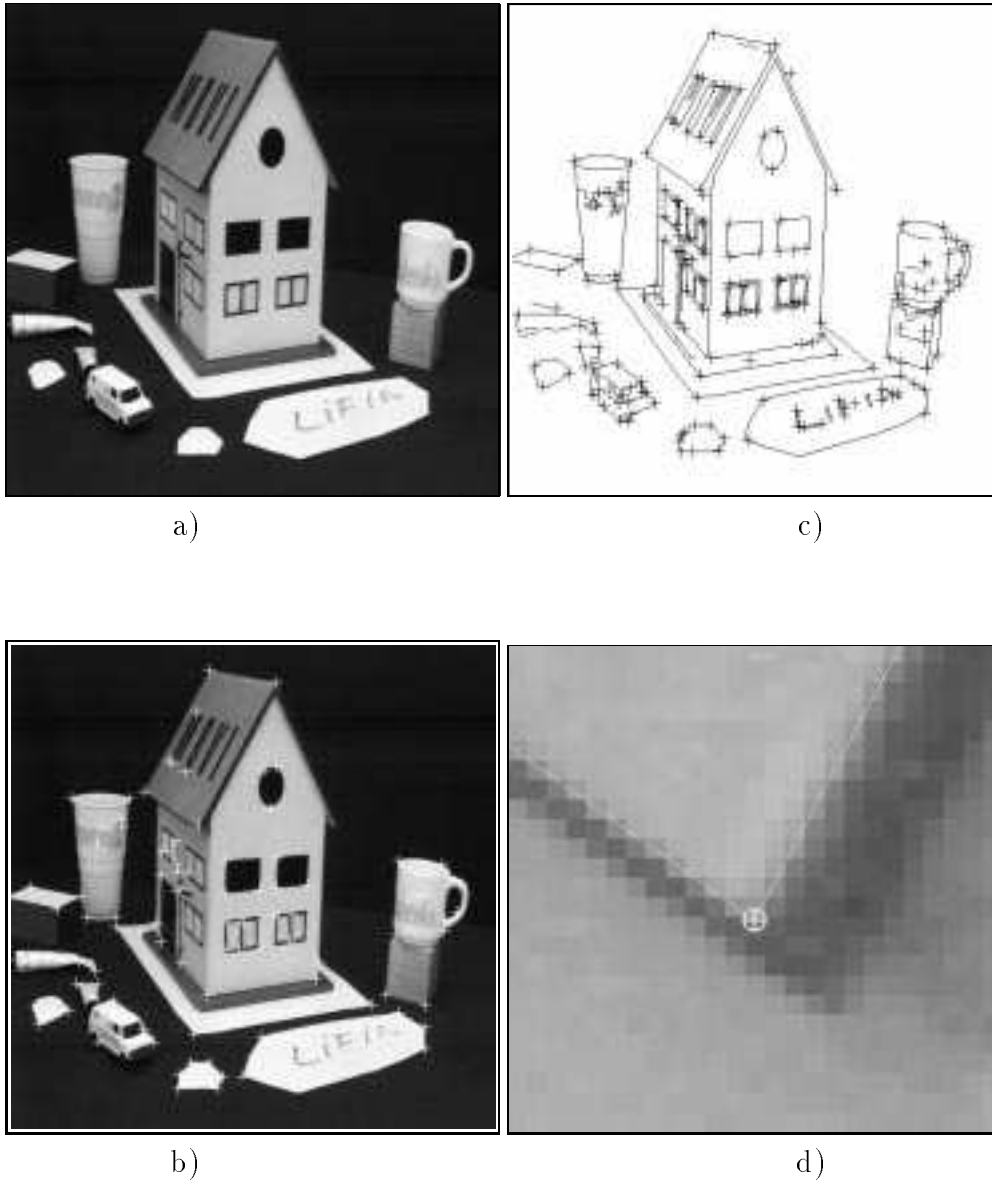


FIG. 2.8 - a) Une vue de la scène ; b) Les contours approchés par des B-splines avec les extremums de courbure ; c) les coins ; d) un agrandissement sur un coin.

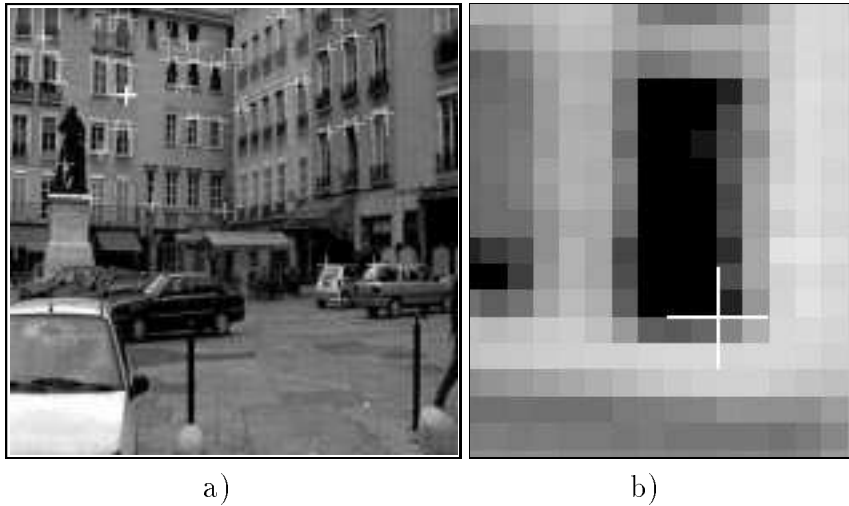


FIG. 2.9 - a) Points d'intérêt détectés sur une image d'extérieur (avec le second détecteur) – b) Un agrandissement sur un point d'intérêt : le point marqué avec une croix plus épaisse sur l'image a)

de quelques points non détectés.

Notre détecteur prend en entrée une image à niveaux de gris et fonctionne selon le schéma suivant :

1. Extraction des contours.
2. Détection des points d'intérêt en utilisant un grand masque.
3. Amélioration de la précision des localisations des points détectés en utilisant cette fois un masque plus petit.

Pour ce deuxième détecteur, la précision de la localisation des points d'intérêt détectés est moins bonne que celle du premier détecteur. (voir Figure 2.9). Cependant, ce détecteur présente deux avantages : il est plus rapide (moins coûteux en temps de calcul) que le précédent et fonctionne sur des scènes plus complexes (voir figure 2.9).

2.2.2 Mise en correspondance

La mise en correspondance de points entre deux images est le plus souvent basée sur le principe suivant : un point de la première image représente le même point physique qu'un autre point de la seconde image (on dit aussi que les deux points se correspondent) si les deux points se “ressemblent”. Cette ressemblance doit prendre en compte les voisinages des points à cause du bruit dans les images, des occultations, du changement de point de vue de prise d'image...

En absence de toute connaissance a priori, la ressemblance entre deux portions de deux images d'une même scène est quantifiée grâce à la mesure de *corrélacion* qui n'est autre qu'une mesure de ressemblance. Cette mesure de corrélacion est calculée entre deux fenêtres, en général carrées et centrées sur les points à mettre en correspondance.

Différentes mesures de corrélacion sont utilisées en vision par ordinateur [AG92]. Après des expériences menées sur les performances de ces différentes corrélacions, nous avons retenu, pour sa robustesse, la corrélacion normalisée, appelée *ZNCC*². La corrélacion normalisée entre le point $p(x, y)$ d'une image et le point $p'(x', y')$ d'une autre image est donnée par la formule :

$$c(p, p') = \frac{\sum_{i=-n}^n \sum_{j=-n}^n (I(x+i, y+j) - \overline{I(x, y)}) \cdot (I'(x'+i, y'+j) - \overline{I'(x', y')})}{\sqrt{\sum \sum (I(x+i, y+j) - \overline{I(x, y)})^2 \cdot \sum \sum (I'(x'+i, y'+j) - \overline{I'(x', y')})^2}}$$

où $I(x, y)$ est l'intensité au point (x, y) et $\overline{I(x, y)}$ est la moyenne des intensités sur la fenêtre de taille $(2n+1) \times (2n+1)$ centrée en (x, y) .

La mise en correspondance revient donc à chercher pour un point donné de la première image le point de la seconde image ayant la meilleure réponse de corrélacion.

Deux cas de figure peuvent se présenter : dans le premier cas, les deux images sont relativement éloignées, c'est à dire que les deux caméras ayant servi à prendre les deux images se trouvent loin l'une par rapport à l'autre. La mise en correspondance dans ce cas est assez difficile et assez coûteuse en temps, car la différence entre les deux images peut être importante jusqu'à rendre la mise en correspondance impossible. Le deuxième cas de figure suppose des images suffisamment proches, ce qui préserve la ressemblance entre les mêmes régions de l'image et facilite ainsi la mise en correspondance.

Nous nous plaçons par la suite dans le deuxième cas, et présentons deux algorithmes de mise en correspondance de points que nous avons réalisés et testés sur des images d'intérieur et d'extérieur.

Nos deux algorithmes de mise en correspondance fonctionnent en trois étapes :

1. détection des points d'intérêt dans les deux images par une des deux méthodes décrites au paragraphe 2.2.1.
2. pour chaque point d'intérêt de la première image, rechercher parmi les points d'intérêt de la seconde image, et dans les limites d'une fenêtre appelée fenêtre de recherche, le candidat donnant la meilleure réponse de corrélacion. La taille de la fenêtre de recherche est déterminée en fonction des conditions expérimentales. Le candidat ainsi trouvé n'est admis définitivement que si sa

2. Zero mean Normalized Cross-Correlation

réponse de corrélation dépasse un certain seuil dont la valeur est obtenue expérimentalement. A cette étape, on choisit un seuil de corrélation très sévère, ce qui limite considérablement le risque d'erreur, même si un tel seuillage se traduit par un nombre de mises en correspondance assez faible en général.

3. Les points mis en correspondance à l'étape précédente nous permettent de calculer la matrice fondamentale en utilisant la méthode décrite dans la section 2.1 (voir annexe B pour un calcul statistique robuste permettant de trouver la meilleure matrice fondamentale). Les mises en correspondance rejetées à la première phase à cause d'un seuillage sévère sont maintenant en majorité récupérées en utilisant un seuil moins fort. En effet la géométrie épipolaire contraint la mise en correspondance, puisque pour chaque point de la première image elle donne la droite le long de laquelle doit se trouver le correspondant dans la seconde image. Ainsi, notre recherche de mise en correspondance passe d'une complexité bidimensionnelle à une complexité unidimensionnelle, ce qui est une réduction considérable de l'espace de recherche (voir Figure 2.10). En pratique, le correspondant d'un point dans la seconde image n'est jamais situé exactement sur la droite épipolaire qui lui correspond. La recherche dans la seconde image est alors faite dans une bande autour de la droite épipolaire. La taille de cette bande dépend de la précision de la géométrie épipolaire et est généralement déterminée expérimentalement. Pour notre cas nous avons utilisé une bande de 2 pixels. En d'autres termes, le correspondant peut s'éloigner de la droite épipolaire de ± 1 pixel.

Nos deux algorithmes de mise en correspondance ne diffèrent qu'au niveau de la détection des points. Le premier utilise le détecteur de coins, alors que le second utilise le détecteur de points d'intérêt. Lorsque les scènes traitées sont simples, contenant des objets polyédriques, le premier algorithme est le plus adapté car donnant une précision sub-pixel (voir figure 2.8).

Le second, plus général, donne des résultats satisfaisants pour une grande variété de scènes, en particulier pour les scènes d'extérieur (voir figure 2.4). Mais ce deuxième algorithme de mise en correspondance est moins précis dans les localisations et donc dans les mises en correspondance.

2.3 Conclusion

Dans la première partie de ce chapitre, nous avons proposé une nouvelle méthode pour calculer la matrice fondamentale. Par rapport à la méthode classique, nous posons le problème différemment en cherchant à déterminer la position d'un épipole ainsi qu'une homographie de plans entre les deux images. Nous dérivons alors une nouvelle expression pour la matrice fondamentale : celle-ci s'écrit comme un produit d'une matrice antisymétrique par une homographie de plans.

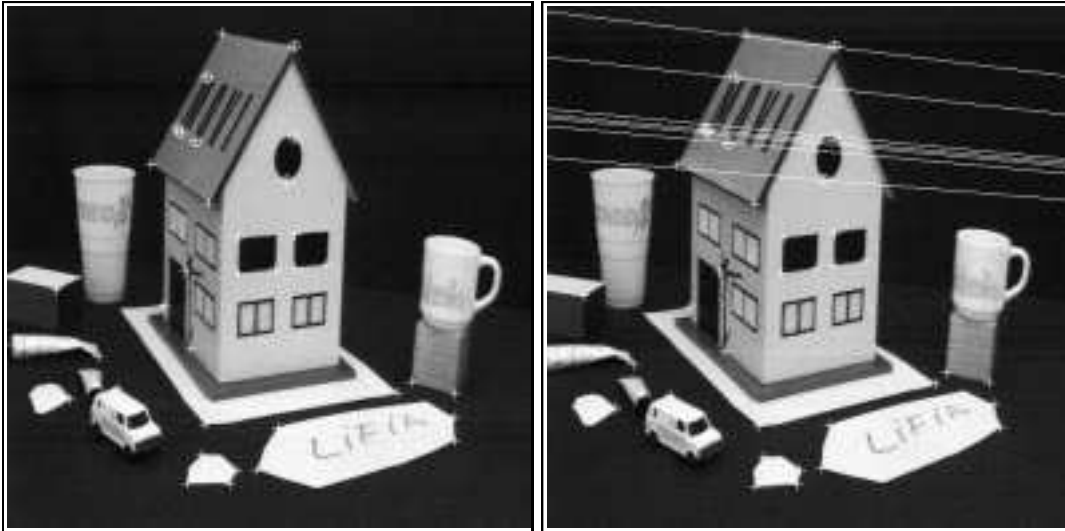


FIG. 2.10 - Deux images de la séquence de la maison : les droites épipolaires réduisent l'espace de recherche pour la mise en correspondance. Les points entourés par des cercles sont des points rejetés à la première phase de mise en correspondance mais sont récupérés avec l'utilisation de la géométrie épipolaire

Cette nouvelle formulation permet de réduire le nombre de paramètres à estimer, qui passe de 8 à 4. Les équations utilisées sont plus simples, conduisant à une stabilité dans les calculs. Les tests réalisés avec des données réelles et des données simulées ont validé la méthode. En particulier, le calcul utilisant uniquement les équations linéaires donne d'excellents résultats, garantissant alors une convergence rapide de l'étape non linéaire quand celle-ci est utilisée à la suite du calcul linéaire.

Notons qu'en absence de l'étalonnage, la matrice fondamentale joue un rôle clé en stéréovision. Elle permet la reconstruction projective d'une scène [Fau92] [HGC92] [Sha92], facilite la mise en correspondance et permet de calculer le mouvement entre deux caméras quand les paramètres intrinsèques de celles-ci sont connus. Dans ce dernier cas, la matrice fondamentale coïncide avec la matrice essentielle qui contient la translation et la rotation du mouvement entre les deux caméras ayant servi à obtenir les deux images.

Dans la deuxième partie de ce chapitre, nous avons traité le problème de la mise en correspondance. Ce dernier est une phase nécessaire pour la reconstruction tridimensionnelle, celle-ci étant le sujet principal de la thèse. Nous nous sommes donc intéressés à ce problème qui, malgré sa simplicité apparente, reste un problème délicat et difficile à mettre en œuvre en pratique. Nous avons présenté deux algorithmes de mise en correspondance de points que nous avons développé et réalisés. Ces deux algorithmes sont basés sur la géométrie épipolaire, et sont destinés à deux types d'applications.

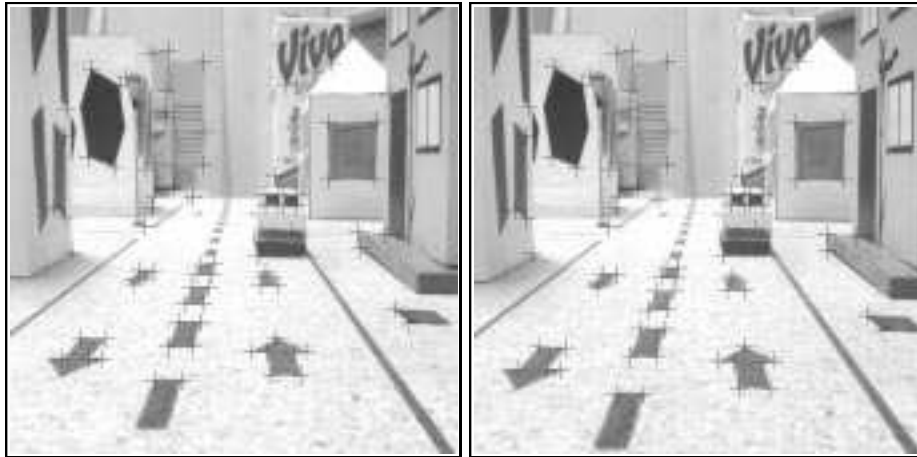


image 1

image 2

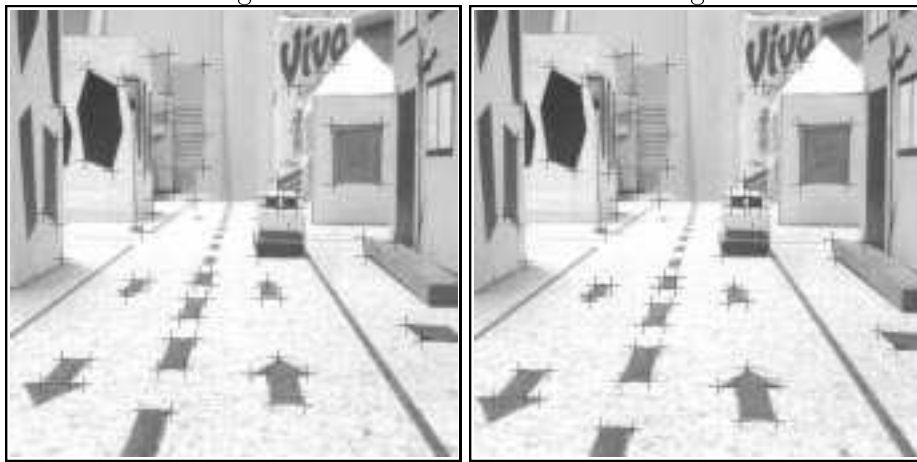


image 3

image 4



image 5

FIG. 2.11 - *La poursuite de points dans une séquence de 5 images*

Le premier algorithme utilise un détecteur de coins avec une précision sub-pixel. Cet algorithme est coûteux ; il peut être utilisé pour des applications où une grande précision est requise. Cependant, il ne traite que des scènes simples, les scènes principalement polyédriques.

Le second algorithme est plus général. Il utilise un détecteur de points d'intérêt. Ces derniers sont des points caractéristiques dans une image et sont plus généraux que les coins qui ne sont qu'un cas particulier de points d'intérêt. Ce deuxième algorithme est adapté à une large variété de scènes. Il est moins coûteux en temps de calcul ; de plus il permet de traiter des scènes d'intérieur et des scènes d'extérieur complexes. Seul inconvénient de ce détecteur, la précision des localisations des points d'intérêt reste de l'ordre du pixel. Ce qui peut être gênant pour certaines applications, par exemple, pour une reconstruction dont les contraintes de précision exigent une localisation sub-pixel dans les images.

Chapitre 3

Reconstruction 3D à partir d'images non étalonnées

3.1 Introduction

Ce chapitre concerne la reconstruction tridimensionnelle de points à partir de leurs observations dans les images. Sans aucune information supplémentaire et en utilisant uniquement les images bidimensionnelles, on parle de *reconstruction tridimensionnelle à partir d'images non étalonnées*. La reconstruction est souvent nécessaire pour des applications robotiques ; c'est l'information qui permet de percevoir le monde tridimensionnel. Il est clair qu'en l'absence de toute information a priori un point de l'image correspond à une infinité de points de la scène (la droite de l'espace qui passe par ce point image). Retrouver alors l'information tridimensionnelle à partir d'une seule image est théoriquement impossible. En outre, il n'existe aucun invariant [BWR90] pouvant caractériser une configuration quelconque de points de l'espace tridimensionnel. Au moins deux images prises sous deux angles de vue différents sont alors nécessaires pour retrouver la structure tridimensionnelle d'une scène. Le problème à résoudre est le suivant : étant donné un ensemble de points tridimensionnels observés par une caméra (ou par plusieurs caméras) sous différents angles de vue (voir Figure 3.1), en supposant les points déjà appariés dans les différentes images, comment obtenir l'information tridimensionnelle relative à ces points ?

Ce problème peut être résolu en deux étapes : d'abord étalonner les caméras puis reconstruire la scène. La reconstruction est alors un problème classique déjà traité par de nombreux auteurs [Aya89][Fau93][HM93]. Cependant, l'étalonnage d'une caméra est un processus délicat nécessitant l'usage d'une mire avec des points parfaitement connus dans l'espace [FT87]. Outre cet inconvénient, il n'est pas toujours possible d'étalonner une caméra, par exemple quand celle-ci est impliquée dans des mouvements.

Une direction de recherche est apparue voici 3 ans : retrouver la structure d'une

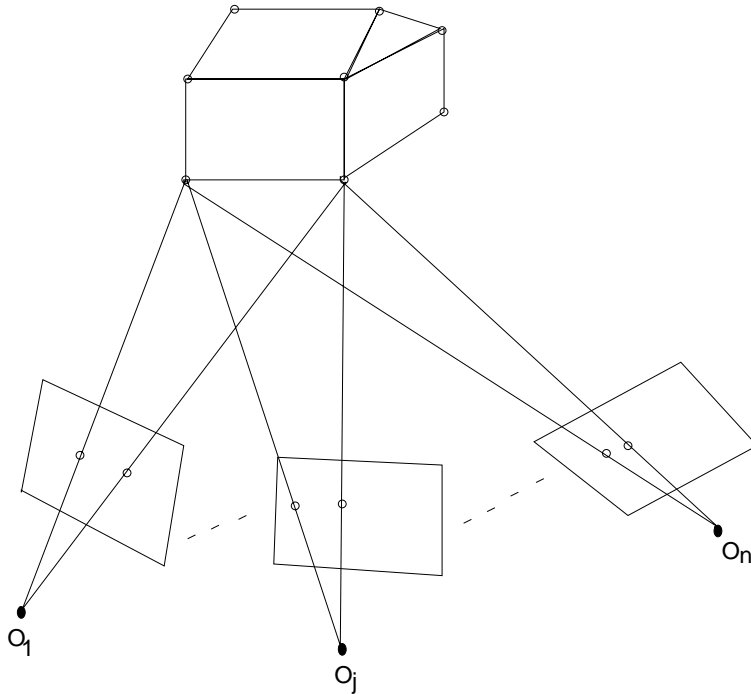


FIG. 3.1 - Une scène tridimensionnelle observée sous différents angles de vue

scène sans passer explicitement par l'étalonnage de la caméra. Ce chapitre rentre dans ce cadre de “reconstruction tridimensionnelle sans étalonnage de la caméra”.

La reconstruction tridimensionnelle d'un ensemble de points informe sur la structure géométrique de ce même ensemble. Selon le contexte et sans relation avec le modèle de la caméra utilisé, nous parlerons de :

- **reconstruction projective** quand les points sont définis dans un repère projectif. La reconstruction est alors définie à une transformation projective près. Une telle reconstruction ne contient donc pas d'information critique (absence de la notion d'angle, de longueur, parallélisme...).
- **reconstruction affine** quand les points sont définis dans un repère affine. La reconstruction est alors définie à une transformation affine près. Une telle reconstruction contient des informations affines (rapport de longueurs, parallélisme).
- **reconstruction euclidienne** quand les points sont définis dans un repère euclidien. La reconstruction est alors définie à une similitude près. Cette dernière reconstruction est la plus riche ; elle contient les informations que l'homme a l'habitude de manipuler. Nous parlerons dans ce cas de reconstruction euclidienne, car le facteur d'échelle absent ici ne pose pas de problème particulier.

Sans spécification du type de reconstruction, le terme *reconstruction tridimensionnelle* (ou *structure tridimensionnelle*) est générique et concerne les trois différentes reconstructions citées ci-dessus. Par abus de langage, les termes *reconstruction* et *structure* sont parfois utilisés pour désigner la *reconstruction tridimensionnelle*.

3.1.1 Etat de l'art

Quand les seules données qu'on peut utiliser sont les observations dans les images, le problème de la reconstruction tridimensionnelle est complexe. De nombreux travaux ont été menés pour retrouver la structure tridimensionnelle d'une scène sans étalonnage de la caméra. Le point commun de tous ces travaux est l'utilisation d'un repère relatif dans la scène.

L'un des premiers travaux en ce sens est celui de K nderink et Doorn [KvD89]. Ces derniers ont montr  que sous l'hypoth se d'une projection parall le (transformation affine), la structure affine d'une sc ne pouvait  tre calcul e   partir d'au moins deux images. Ils ont utilis  quatre points arbitraires de la sc ne comme rep re affine (les points doivent  tre non coplanaires), la structure affine de tous les autres points  tant alors calcul e relativement   ce rep re.

Sous l'hypoth se d'une projection orthographique, Tomasi et Kanade [TK91] calculaient en m me temps la structure de la sc ne et la rotation de la caméra. Les mesures image sont d'abord rang es sous forme d'une matrice. Ensuite, en utilisant la D composition en Valeurs Singuli res, cette matrice est d compos e (factoris e) en deux matrices : l'une contenant l'information sur la structure tridimensionnelle et l'autre l'information sur le mouvement de la caméra (rotation). La connaissance des param tres internes de la caméra est n cessaire pour cette m thode, la structure ainsi obtenue est euclidienne.

Une autre m thode de reconstruction propos e par D. Weinshall [Wei93], consiste   calculer en premier lieu la structure affine avec la m thode de K nderink et Doorn. Ensuite, sous l'hypoth se d'un mod le de projection perspective faible pour la caméra, le Gramien¹ de quatre points non coplanaires de la sc ne est calcul . La structure euclidienne de ces quatre points est alors extraite de ce Gramien (plus de d tails seront donn s dans la section 3.2). Enfin, une transformation affine permet de passer de la structure affine d j  calcul e   une structure euclidienne. Cette m thode a l'avantage de se passer de la connaissance des param tres internes de la caméra. Seule hypoth se utilis e : les pixels de l'images sont suppos s carr s ($\alpha_u = \alpha_v$ dans les notations du chapitre 1).

Une variante de la m thode de Tomasi et Kanade a  t  propos e par Poelman et Kanade [PK94]. Ces derniers ont utilis  une projection paraperspective au lieu d'une projection orthographique. La qualit  de leurs reconstructions est meilleure

1. appel e aussi matrice de Gram d fini dans le paragraphe 3.2.2

que celle de Tomasi et Kanade. Ceci provient du fait que la projection paraperspective est une meilleure approximation pour les caméras CCD (voir chapitre 1). Outre cet avantage, la projection paraperspective permet de tenir compte des variations de profondeur. C'est-à-dire, que la caméra (ou la scène) peut effectuer des translations le long de l'axe optique, ce qui n'était pas permis avec l'hypothèse de projection orthographique.

L'approximation de la transformation scène-caméra par une transformation affine simplifie le calcul de la reconstruction. Cependant, cette hypothèse peut être non justifiée, rendant alors la structure calculée complètement fautive dans certains cas. En effet, si la scène est suffisamment éloignée (ou si la focale est très longue), cette approximation peut être acceptable; dans le cas contraire elle ne peut être utilisée (voir section 3.2 pour plus de détails).

Une meilleure approximation de la transformation scène-caméra est la projection perspective (transformation projective). Cette dernière est plus largement utilisée et est plus proche d'une caméra réelle; elle reste valide pour tout type de scène. Traiter le problème de la reconstruction d'une manière plus correcte a incité de nombreux chercheurs à abandonner la géométrie affine au profit de la géométrie projective.

Une première approche réaliste dans ce sens a été proposée par Sparr [SN90] [Spa91]. Ce dernier a développé un descripteur pour la structure affine tridimensionnelle d'un ensemble de points. Ce descripteur est indépendant de la base affine utilisée, et ne dépend que des propriétés intrinsèques de la structure. En outre, par un théorème, Sparr montre que ce descripteur est relié avec la projection perspective des points, permettant ainsi de retrouver la structure affine relative de l'ensemble des points. Cette approche est très intéressante quand seule l'information affine est requise pour une application donnée. Cependant, la résolution du système hautement non linéaire pour retrouver une telle structure reste un problème majeur limitant son utilisation.

Mohr et al [MA90] [MM91] ont été parmi les pionniers à utiliser les propriétés de la géométrie projective pour éviter un étalonnage explicite de la caméra. Leur approche est basée sur la reprojection des points images sur deux plans de la scène. Ils ont utilisé la propriété d'invariance des coordonnées projectives dans un plan après projection. Six points connus de la scène sont utilisés comme points de référence, les autres points sont reconstruits ou "positionnés" relativement à ces six points.

Plus récemment, Faugeras [Fau92] a développé une méthode linéaire pour la reconstruction projective uniquement à partir de points mis en correspondance dans deux images. Il choisit 5 points arbitraires de la scène (non 4 à 4 coplanaires) comme base projective et leur affecte les coordonnées $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$ et $(1, 1, 1, 1)$. Grâce à ce choix de coordonnées, et grâce à la mise en correspondance dans les deux images, seules deux parmi les 22 inconnues

des matrices de projection restent à déterminer. Pour ce faire l'auteur suppose que les épipoles sont connus. Ces derniers sont intrinsèquement liés aux matrices de projection des deux images, et permettent ainsi de déterminer entièrement les deux matrices de projection. Connaissant ces dernières, les points sont alors reconstruits. Cette reconstruction est projective; aussi est-elle définie à une transformation projective de l'espace tridimensionnel près. Elle ne contient aucune information métrique, ni même affine.

Vers la même époque, Hartley et al [HGC92] convergèrent vers un résultat similaire, une reconstruction projective. Depuis, d'autres chercheurs [Sha92] [PMC94] ont proposé des méthodes linéaires basées aussi sur la géométrie épipolaire pour le calcul de la reconstruction projective d'une scène. Toutes ces méthodes peuvent être considérées comme similaires, car utilisant les mêmes propriétés de la géométrie projective.

3.1.2 Plan du chapitre

Nous nous proposons dans ce chapitre d'étudier deux méthodes de reconstruction tridimensionnelle à partir d'une séquence d'images non étalonnées. La première méthode développée par D. Weinshall [Wei93] suppose un modèle affine pour la caméra, le processus de la reconstruction est alors linéaire. Les limites de cette méthode pour les cas défavorables au modèle affine n'ont pas été développées par son auteur. Nous ferons donc une étude expérimentale sur les limites d'application de cette méthode pour le cas de la reconstruction euclidienne.

La deuxième méthode utilise un modèle projectif pour la caméra : la projection perspective. Cette méthode non linéaire a été développée au sein de notre équipe où nous avons contribué avec d'autres personnes à sa mise en œuvre [MQVB92].

L'aspect numérique de cette méthode est bien développé; en particulier, beaucoup d'importance sera donnée à la précision de la reconstruction et à la stabilité de la méthode.

Les deux méthodes sont ensuite comparées dans différentes conditions de prise d'images. Enfin, un algorithme combinant ces deux méthodes est proposé.

Dans la section suivante, nous présentons la méthode de reconstruction linéaire. Le premier paragraphe montre comment à partir de deux images la structure affine peut être obtenue en résolvant un système d'équations linéaires. Le deuxième paragraphe est consacré au calcul de la structure euclidienne de quatre points observés dans au moins 3 images. Dans le troisième paragraphe, les différentes étapes de la méthode de reconstruction euclidienne de tous les points de la scène sont présentées. Enfin, le dernier paragraphe sera consacré aux résultats expérimentaux.

La section 3.3 présente notre méthode de reconstruction sous l'hypothèse d'un modèle projectif pour la caméra. Dans le premier paragraphe, le problème de la

reconstruction est d'abord posé. Les équations utilisées pour la reconstruction sont ensuite présentées. Le deuxième paragraphe expose une façon de résoudre le système d'équations non linéaires et le dernier paragraphe présente les résultats expérimentaux.

La dernière section comporte d'abord une étude comparative des deux méthodes présentées ; ensuite un algorithme combinant les deux méthodes de reconstruction est présenté.

3.2 Reconstruction avec un modèle affine pour la caméra

Une reconstruction utilisant un modèle affine (la transformation scène-image est approchée par une transformation affine) est naturellement plus simple qu'une reconstruction utilisant le modèle projectif, la transformation scène-image possédant moins de paramètres dans ce cas.

Il existe toute une famille de modèles dits affines : projection parallèle, projection orthographique, projection perspective faible, projection para-perspective... Ces modèles partagent tous la même propriété affine de la transformation scène-image. Cependant, ils diffèrent au niveau de la décomposition de cette transformation scène-image (voir paragraphe 1.2.1).

Il est à noter que parmi les méthodes de reconstruction décrites dans le paragraphe 3.1.1, la méthode présentée ici est la seule qui permet d'obtenir l'information euclidienne sans connaissance des paramètres internes de la caméra. Elle fonctionne en trois phases : la reconstruction affine est d'abord obtenue sous l'hypothèse d'une projection parallèle. Puis sous l'hypothèse d'une projection perspective faible la structure euclidienne de quatre points est calculée. Enfin, la reconstruction affine est transformée en une reconstruction euclidienne.

Puisque les transformations affines laissent les points de l'infini invariants, nous pouvons utiliser les coordonnées non homogènes des points. Ceci n'interdit pas l'utilisation des coordonnées homogènes, qui par ailleurs n'apportent rien de plus pour le cas affine.

Par conséquent, nous utilisons dans le reste de cette section les coordonnées non homogènes.

3.2.1 Reconstruction tridimensionnelle affine

Soient quatre points non coplanaires de l'espace Q_0, Q_1, Q_2 et Q_3 . Ces points forment une base affine et on peut leur affecter respectivement les coordonnées affines suivantes : $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$ et $(0, 0, 1)$

3.2. RECONSTRUCTION AVEC UN MODÈLE AFFINE POUR LA CAMÉRA63

Par définition, les coordonnées de tout autre point de l'espace Q , défini dans cette même base, peuvent s'exprimer en une combinaison linéaire des coordonnées de Q_1 , Q_2 et Q_3 .

$$Q = \lambda_1 Q_1 + \lambda_2 Q_2 + \lambda_3 Q_3$$

où λ_1 , λ_2 et λ_3 sont des scalaires.

Supposons maintenant que les points Q_0 , Q_1 , Q_2 , Q_3 et P soient projetés respectivement en $q_0(x_0, y_0)$, $q_1(x_1, y_1)$, $q_2(x_2, y_2)$, $q_3(x_3, y_3)$ et $q(x, y)$ sur l'image. Sans perte de généralité, on peut supposer que $x_0 = 0$ et $y_0 = 0$. En d'autres termes, on peut prendre q_0 comme origine dans l'image (pour se ramener à ce cas, il suffit de remplacer tous les x_i par $x_i - x_0$ et tous les y_i par $y_i - y_0$).

Puisque la transformation scène-image est une transformation affine (modèle affine), nous aurons donc pour la projection du point Q :

$$x = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 \quad \text{et} \quad y = \lambda_1 y_1 + \lambda_2 y_2 + \lambda_3 y_3$$

Un point comporte 3 inconnues alors que l'image d'un point nous fournit deux équations linéaires en les 3 inconnues. Deux images sont donc suffisantes pour calculer la structure affine d'une scène (pour un point observé dans deux images, nous avons 4 équations linéaires en les 3 inconnues). Le calcul de cette structure revient à résoudre un système d'équations linéaires.

3.2.2 Structure euclidienne de quatre points sous l'hypothèse d'une projection perspective faible

Considérons trois points P_1 , P_2 et P_3 de l'espace donnés par leurs coordonnées euclidiennes (X_1, Y_1, Z_1) , (X_2, Y_2, Z_2) et (X_3, Y_3, Z_3) respectivement. Ces coordonnées sont définies dans un repère euclidien quelconque d'origine P_0 (P_1 , P_2 , P_3 et P_0 doivent être non coplanaires).

Nous appelons *Gramien* des points P_0 , P_1 , P_2 et P_3 la matrice symétrique de taille 3×3 suivante :

$$G = \begin{pmatrix} P_0 \vec{P}_1 \cdot P_0 \vec{P}_1 & P_0 \vec{P}_1 \cdot P_0 \vec{P}_2 & P_0 \vec{P}_1 \cdot P_0 \vec{P}_3 \\ P_0 \vec{P}_2 \cdot P_0 \vec{P}_1 & P_0 \vec{P}_2 \cdot P_0 \vec{P}_2 & P_0 \vec{P}_2 \cdot P_0 \vec{P}_3 \\ P_0 \vec{P}_3 \cdot P_0 \vec{P}_1 & P_0 \vec{P}_3 \cdot P_0 \vec{P}_2 & P_0 \vec{P}_3 \cdot P_0 \vec{P}_3 \end{pmatrix}$$

où \cdot représente le produit scalaire entre vecteurs.

Cette matrice contient toutes les informations sur la géométrie tridimensionnelle euclidienne de ces quatre points. Les éléments diagonaux de G contiennent les carrés des modules (longueurs) des vecteurs $P_0 \vec{P}_1$, $P_0 \vec{P}_2$ et $P_0 \vec{P}_3$ tandis que les autres éléments nous permettent d'extraire les différents angles entre ces mêmes vecteurs.

On peut remarquer que les coordonnées des points P_i , $i = 1 \dots 3$, sont aussi les coordonnées des vecteurs $P_0 \vec{P}_i$. Pour simplifier, nous utilisons dans la suite G sous

la forme :

$$G = \begin{pmatrix} {}^tP_1P_1 & {}^tP_1P_2 & {}^tP_1P_3 \\ {}^tP_2P_1 & {}^tP_2P_2 & {}^tP_2P_3 \\ {}^tP_3P_1 & {}^tP_3P_2 & {}^tP_3P_3 \end{pmatrix}$$

Il est à noter que la matrice G est invariante par toute transformation euclidienne (rotation et translation). Connaissant G , on peut donc calculer les coordonnées euclidiennes tridimensionnelles de trois points P_1 , P_2 et P_3 dans un repère relatif d'origine P_0 (P_0 a comme coordonnées $(0, 0, 0)$).

Calcul de G

Sous l'hypothèse d'une projection perspective faible pour la transformation scène-caméra (voir paragraphe 1.2.4), nous cherchons ici à calculer G à partir des observations de P_0 , P_1 , P_2 et P_3 dans les images.

Soient $p_0(x_0, y_0)$, $p_1(x_1, y_1)$, $p_2(x_2, y_2)$ et $p_3(x_3, y_3)$ les projections dans une image des points P_0 , P_1 , P_2 et P_3 respectivement.

En suivant le modèle de projection perspective faible, un point de l'espace $P(X, Y, Z)$ se projette sur l'image en un point $p(x, y)$ selon la relation suivante :

$$\begin{aligned} x &= \alpha_u \frac{r_{11}X + r_{12}Y + r_{13}Z + T_x}{T_z} + u_0 \\ y &= \alpha_v \frac{r_{21}X + r_{22}Y + r_{23}Z + T_y}{T_z} + v_0 \end{aligned} \quad (3.1)$$

où (T_x, T_y, T_z) sont les coordonnées de P_0 dans le repère attaché à la caméra (vecteur translation) et $R = \{r_{ij}\}_{i,j=1}^3$ la matrice rotation qui fait passer du repère scène au repère attaché à la caméra. α_u , α_v , u_0 et v_0 sont les paramètres intrinsèques de la caméra présentés dans le paragraphe 1.2.1.

$P_0(0, 0, 0)$ se projette alors en $p_0(x_0, y_0)$ avec :

$$\begin{aligned} x_0 &= \alpha_u \frac{T_x}{T_z} + u_0 \\ y_0 &= \alpha_v \frac{T_y}{T_z} + v_0 \end{aligned}$$

En posant $x' = x - x_0$ et $y' = y - y_0$, nous avons :

$$\begin{aligned} x' &= \alpha_u \frac{r_{11}X + r_{12}Y + r_{13}Z}{T_z} \\ y' &= \alpha_v \frac{r_{21}X + r_{22}Y + r_{23}Z}{T_z} \end{aligned} \quad (3.2)$$

Supposons maintenant que $\alpha_u = \alpha_v = \alpha$, ou à défaut, supposons la connaissance du rapport $\frac{\alpha_u}{\alpha_v}$, ce qui revient à faire une transformation dans les coordonnées image.

3.2. RECONSTRUCTION AVEC UN MODÈLE AFFINE POUR LA CAMÉRA65

En posant $\frac{\alpha}{T_z} = k$, les équations (3.2) deviennent :

$$\begin{aligned} x' &= k \cdot (r_{11}X + r_{12}Y + r_{13}Z) \\ y' &= k \cdot (r_{21}X + r_{22}Y + r_{23}Z) \end{aligned} \quad (3.3)$$

En regroupant les équations (3.3) pour les trois points P_1, P_2 et P_3 sous forme matricielle, nous avons :

$$\begin{aligned} k \cdot \begin{pmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \end{pmatrix} \begin{pmatrix} r_{11} \\ r_{12} \\ r_{13} \end{pmatrix} &= \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} \\ \text{et} \\ k \cdot \begin{pmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \end{pmatrix} \begin{pmatrix} r_{21} \\ r_{22} \\ r_{23} \end{pmatrix} &= \begin{pmatrix} y'_1 \\ y'_2 \\ y'_3 \end{pmatrix} \end{aligned} \quad (3.4)$$

Posons :

$$S = \begin{pmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \end{pmatrix} \quad R_1 = \begin{pmatrix} r_{11} \\ r_{12} \\ r_{13} \end{pmatrix} \quad R_2 = \begin{pmatrix} r_{21} \\ r_{22} \\ r_{23} \end{pmatrix} \quad X = \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} \quad \text{et} \quad Y = \begin{pmatrix} y'_1 \\ y'_2 \\ y'_3 \end{pmatrix}$$

Nous pouvons maintenant écrire les relations matricielles (3.4) sous la forme condensée :

$$R_1 = \frac{1}{k} S^{-1} X \quad \text{et} \quad R_2 = \frac{1}{k} S^{-1} Y \quad (3.5)$$

où S^{-1} est la matrice inverse de S qui existe toujours puisque les trois points P_1, P_2 et P_3 ne peuvent pas être coplanaires avec P_0 .

Une matrice de rotation est, par définition, une matrice orthonormée, $R^{-1} = {}^tR$, ce qui implique :

$$\begin{aligned} {}^tR_1 R_2 &= 0 \\ \text{et} \\ {}^tR_1 R_1 &= {}^tR_2 R_2 = 1 . \end{aligned}$$

En utilisant les égalités données dans (3.5) nous obtenons deux contraintes sur S

$$\begin{aligned} {}^tX ({}^tSS)^{-1} Y &= 0 \\ {}^tX ({}^tSS)^{-1} X - {}^tY ({}^tSS)^{-1} Y &= 0 . \end{aligned} \quad (3.6)$$

Notons que la matrice tSS n'est autre que le *Gramien* des quatre points P_0, P_1, P_2 et P_3 .

$${}^tSS = \begin{pmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \end{pmatrix} \begin{pmatrix} X_1 & X_2 & X_3 \\ Y_1 & Y_2 & Y_3 \\ Z_1 & Z_2 & Z_3 \end{pmatrix} = \begin{pmatrix} {}^tP_1 P_1 & {}^tP_1 P_2 & {}^tP_1 P_3 \\ {}^tP_2 P_1 & {}^tP_2 P_2 & {}^tP_2 P_3 \\ {}^tP_3 P_1 & {}^tP_3 P_2 & {}^tP_3 P_3 \end{pmatrix} \quad (3.7)$$

En posant $A = G^{-1} = (S {}^tS)^{-1}$, les équations (3.6) deviennent :

$$\begin{aligned} {}^t\mathbf{X}A\mathbf{Y} &= 0 \\ {}^t\mathbf{X}A\mathbf{X} - {}^t\mathbf{Y}A\mathbf{Y} &= 0 \end{aligned} \tag{3.8}$$

D'après les équations (3.8), A ne peut être définie qu'à un facteur multiplicatif près. De plus, la matrice A est symétrique, elle n'a donc que 5 paramètres indépendants. Une image fournissant deux équations linéaires en les éléments de A , au moins trois images sont nécessaires pour calculer le *Gramien* de 4 points sous l'hypothèse d'une projection perspective faible.

Une fois ce *Gramien* calculé, il faut trouver S vérifiant : ${}^tSS = G$. Pour cette décomposition on peut utiliser la méthode de factorisation de Choleski.

Notons que cette décomposition n'est possible que si la matrice G est définie positive. Si cette condition n'est pas vérifiée alors le Gramien ne peut être utilisé. Ce cas peut se produire quand le modèle de projection perspective faible pour la caméra n'est pas valide (voir section 3.4).

Remarque

Les équations (3.8) ne dépendent pas du facteur k . Par conséquent, les paramètres intrinsèques de la caméra peuvent changer d'une image à une autre ; seuls les rapports des deux facteurs d'échelle α_u et α_v doivent être connus pour chaque image. En pratique, ce rapport est facile à calculer. De plus, il est généralement constant même pour une caméra à focale variable (zoom) [Luo92].

3.2.3 Reconstruction tridimensionnelle euclidienne

A présent nous sommes capables de calculer la structure affine d'une scène (paragraphe 3.2.1) ainsi que la structure euclidienne de 4 points de cette même scène (paragraphe 3.2.2).

Supposons, pour une scène donnée, que la reconstruction affine a été calculée. Cette dernière est liée à la reconstruction euclidienne de la même scène par une transformation affine (voir section 4.1.2). En d'autres termes, si P_a est un point donné par ses coordonnées affines, P_e le même point donné par ses coordonnées euclidiennes et T_{af} la matrice décrivant la transformation affine, alors nous avons la relation

$${}^tP_e = T_{af} {}^tP_a.$$

Quand P_e et P_a sont connus, la relation ci-dessus fournit 3 équations linéaires en les inconnues de la transformation affine T_{af} .

Les coordonnées euclidiennes de 4 points peuvent être calculées. De plus leurs coordonnées affines sont données (reconstruction affine de la scène). Aussi, le calcul

de T_{af} est possible ; cette dernière matrice a 12 paramètres.

En résumé, la méthode présentée dans cette section permet d'obtenir une reconstruction tridimensionnelle euclidienne d'une scène à partir d'au moins trois images. Cette reconstruction est obtenue en trois étapes :

1. calcul la structure affine de la scène en utilisant la méthode décrite dans le paragraphe 3.2.1. Quatre points de la scène sont utilisés comme base affine, les autres points sont reconstruits relativement à cette base affine,
2. calcul du *Gramien* de quatre points, ce qui permet d'avoir leurs coordonnées euclidiennes (en pratique on prend les quatre points de la base affine),
3. calcul de la transformation affine qui permet de passer d'une reconstruction affine à une reconstruction euclidienne.

Cette méthode de reconstruction a deux avantages : elle permet une reconstruction sans étalonnage de la caméra avec un calcul linéaire (rien que des équations linéaires). Cependant, elle ne garantit pas toujours une reconstruction correcte, en particulier quand le *rapport de profondeur* devient grand (la notion de *rapport de profondeur* a été définie à dans le paragraphe 1.2.1). Le paragraphe suivant est consacré à l'étude du comportement de la méthode en fonction de ce *rapport de profondeur*.

3.2.4 Résultats expérimentaux et limites de la méthode

Nous présentons ici des expérimentations avec des données simulées qui visent l'étude du comportement de la méthode en fonction du *rapport de profondeur*. Des résultats avec des données réelles sont présentés dans la section 3.4.

La scène simulée est constituée d'une soixantaine de points ayant la forme d'une maison (voir Figure 3.2). Cette scène a un volume de $30cm \times 40cm \times 30cm$, la distance maximale le long de l'axe optique de la caméra virtuelle est de $30cm$. C'est cette dernière distance qui sera utilisée pour calculer le *rapport de profondeur*. La même scène (Figure 3.2) est reconstruite plusieurs fois en utilisant quatre images obtenues avec une caméra virtuelle d'une focale de $12,5mm$. Les différentes reconstructions correspondent à des *rapports de profondeur* différents (la caméra est déplacée le long de son axe optique). Les quatre images utilisées pour chaque reconstruction sont obtenues en tournant la caméra virtuelle autour de la scène comme sur la figure 3.1.

Le tableau 3.1 donne les résultats sur la précision des reconstructions en fonction des distances scène-caméra. Dans ce tableau, les notations suivantes sont utilisées :

- *Dis_cam* : la distance scène-caméra.

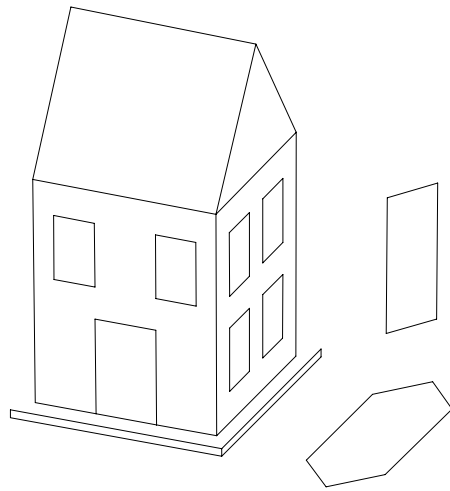


FIG. 3.2 - *Géométrie de la scène simulée : les points utilisés sont les coins de cette scène*

- Rp : le rapport de profondeur.
- ΔX : l'erreur moyenne en X pour tous les points de la scène. Pour un point donné cette erreur n'est autre que la différence entre la coordonnée en X exacte et celle reconstruite.
- ΔY et ΔZ représentent la même chose que ΔX pour les coordonnées en Y et en Z respectivement.
- Er_moy : l'erreur moyenne pour tous les points de la scène. L'erreur d'un point reconstruit $P(X_r^p, Y_r^p, Z_r^p)$ est calculée en utilisant la distance euclidienne D_p définie par :

$$D_p = \sqrt{(X_e^p - X_r^p)^2 + (Y_e^p - Y_r^p)^2 + (Z_e^p - Z_r^p)^2}$$

où (X_e^p, Y_e^p, Z_e^p) sont les coordonnées exactes du point.

- dis_max : la plus grande distance séparant deux points de la scène. Ici, elle vaut $58,3cm$; elle est utilisée pour avoir une idée de l'erreur relative.

Les résultats du tableau 3.1 sont aussi résumés sous forme d'un graphe sur la figure 3.3. Notons que nous n'avons pas eu de résultat pour un rapport de profondeur supérieur à 1 pour cette scène. La matrice Gramien calculée depuis les images n'était pas définie positive, ce qui rendait impossible l'extraction de la structure euclidienne des points depuis cette matrice.

Ces résultats montrent que la précision de la reconstruction se dégrade d'une manière linéaire en fonction du rapport de profondeur, ce qui est un bon aspect de la méthode.

De plus, la reconstruction reste qualitativement acceptable pour certaines applications qui ne nécessitent pas une grande précision (par exemple pour la reconnaissance d'objets). Nous allons voir aussi qu'une telle reconstruction peut être utilisée comme initialisation à une méthode de reconstruction plus complexe (section 3.4.2).

3.3 Reconstruction utilisant un modèle sténopé

Le modèle de projection perspective pure (modèle sténopé) est plus proche d'une caméra réelle que le modèle de projection perspective faible. Même si une caméra réelle ne suit pas un modèle de projection perspective parfait, ce dernier reste une bonne approximation. Des modèles prenant en compte les distorsions optiques sont utilisés par certains chercheurs [Tsa87][Bey92a][WCH92], au prix d'une complexité supérieure. Les améliorations obtenues ne justifient pas toujours le prix payé par la complexité engendrée par la prise en compte de ces distorsions [MBB93].

Dis_cam (cm)	Rp	ΔX (cm)	ΔY (cm)	ΔZ (cm)	Er_moy (cm) (cm)	Er_moy relative $Er_moy/dis_max(\%)$
600	0,05	0,048556	0,100161	0,033132	0,127343	0,22
300	0,1	0,101679	0,192141	0,069255	0,250486	0,43
200	0,15	0,159231	0,275766	0,109966	0,370413	0,64
150	0,2	0,221599	0,350728	0,155012	0,487852	0,84
100	0,3	0,353883	0,476308	0,261929	0,717690	1,23
80	0,375	0,504791	0,565918	0,391103	0,947793	1,63
50	0,6	0,849132	0,662586	0,701891	1,431256	2,47
40	0,75	1,064246	0,729925	0,940098	1,791447	3,09
35	0,857	1,250279	0,782252	1,129043	2,083504	3,60
30	1,0	1,501748	0,932731	1,379789	2,494919	4,30

TAB. 3.1 - Les erreurs de reconstruction sous l'hypothèse de projection perspective faible en fonction du rapport de profondeur avec des données simulées exactes.

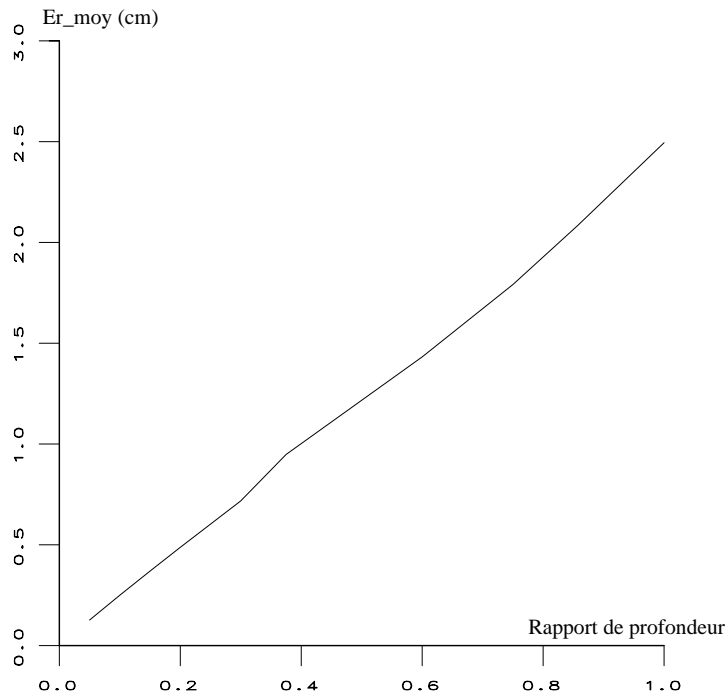


FIG. 3.3 - L'erreur moyenne de la reconstruction en fonction du rapport de profondeur

Néanmoins, ces modèles restent indispensables dans certains cas comme pour des applications de mesure de haute précision[Bey91].

Nous supposons donc dans cette section que le modèle de la caméra suit une projection perspective parfaite (Figure 1.6). En d'autres termes, la transformation scène-image est une transformation projective. Aussi, les coordonnées homogènes sont utilisées dans cette section.

Bien que les méthodes de reconstruction linéaire proposées par Faugeras, Hartley, Shashua et Ponce soient théoriquement correctes, leur mise en œuvre pratique reste très difficile, voire impossible quand les données manipulées sont entachées d'erreurs. Une méthode numérique utilisant des mesures redondantes est proposée ici; c'est une alternative à ces méthodes linéaires.

Nous nous inspirons des méthodes déjà utilisées par les photogrammètres, et qui consistent à considérer quelques points de la scène comme étant des points connus.

Les étapes suivantes sont nécessaires pour obtenir une reconstruction tridimensionnelle :

1. Une séquence d'images (au moins deux) d'une scène rigide est prise à l'aide d'une caméra (voir figure 3.1).
2. Des points d'intérêt sont extraits et mis en correspondance dans la séquence, en utilisant la méthode décrite dans la section 2.2.
3. Cinq points non 4 à 4 coplanaires sont choisis comme base pour un repère relatif et se voient donc affecter des coordonnées. Ces dernières peuvent être euclidiennes, affines ou projectives.
4. L'ensemble des observations (coordonnées images) est traduit en un système d'équations non linéaires dont la résolution donne directement la structure tridimensionnelle de la scène et les matrices de projections.

3.3.1 Les équations du problème de la reconstruction

Considérons v images ($v \geq 2$) d'une scène rigide comprenant p points. Certains points apparaissent et disparaissent d'une image à l'autre. Cependant pour pouvoir être reconstruit, un point doit apparaître dans deux images au moins. Pour simplifier l'exposé de la méthode, les points sont supposés apparaître dans toutes les images, ce qui nous donne un total de $p \times v$ points images.

Notons les points de la scène $P_i, i = 1, \dots, p$. Pour chaque image j , le point P_i de coordonnées homogènes $(x_i, y_i, z_i, t_i)^T$ se projette en p_{ij} de coordonnées homogènes $(u_{ij}, v_{ij}, w_{ij})^T$ et de coordonnées image $(U_{ij}, V_{ij})^T$. Soit M_j la matrice de projection (de taille 3×4) correspondant à la $j^{\text{ème}}$ image.

Nous avons pour les coordonnées homogènes la relation qui découle immédiatement du modèle sténopé :

$$\rho_{ij}p_{ij} = M_j P_i, \quad i = 1, \dots, p, \quad j = 1, \dots, v \quad (3.9)$$

où ρ_{ij} est un facteur d'échelle inconnu (différent pour chaque point image). L'équation (3.9) peut être écrite sous une autre forme plus utilisée dans la littérature :

$$\begin{cases} U_{ij} = \frac{m_{11}^{(j)} x_i + m_{12}^{(j)} y_i + m_{13}^{(j)} z_i + m_{14}^{(j)} t_i}{m_{31}^{(j)} x_i + m_{32}^{(j)} y_i + m_{33}^{(j)} z_i + m_{34}^{(j)} t_i} \\ V_{ij} = \frac{m_{21}^{(j)} x_i + m_{22}^{(j)} y_i + m_{23}^{(j)} z_i + m_{24}^{(j)} t_i}{m_{31}^{(j)} x_i + m_{32}^{(j)} y_i + m_{33}^{(j)} z_i + m_{34}^{(j)} t_i} \end{cases} \quad (3.10)$$

où les $m_{kl}^{(j)}$, $k = 1 \dots 3$, $l = 1 \dots 4$; sont les éléments de la matrice de projection de la $j^{\text{ème}}$ image.

Ces équations expriment tout simplement la colinéarité des points tridimensionnels avec leurs correspondants sur le plan image et les centres de projection.

Comme nous avons p points et v images, ceci nous donne un total de $2 \times p \times v$ équations. Nous avons $11 \times v$ inconnues pour les matrices de projection et $3 \times p$ pour les points de l'espace. Dès que le nombre d'équations est plus grand ou égal que celui des inconnues, le système (3.9) peut être résolu en théorie.

3.3.2 Résolution du problème de la reconstruction

Résolution du système non linéaire

Le système (3.9) tel qu'il est posé, sans aucune contrainte additionnelle, n'admet pas de solution unique. En effet, si (M_j) et (P_i) sont solutions du système, il en est de même pour $M_j W^{-1}$ et $W P_i$ où W est une matrice 4×4 inversible quelconque.

Comme une transformation projective de l'espace correspond à une matrice 4×4 inversible (voir Chapitre 1), l'interprétation géométrique de cette ambiguïté correspond à l'existence d'une famille de reconstructions qui ne diffèrent que par des transformations projectives de l'espace.

Par conséquent, pour assurer l'unicité de la solution, une base quelconque peut être choisie dans l'espace projectif IP^3 , ce qui revient à choisir 5 points indépendants (non quatre à quatre coplanaires) pour définir une telle base. Il est à noter que le type de reconstruction dépend des coordonnées des 5 points utilisés comme base. Par exemple quand ces coordonnées sont euclidiennes, la reconstruction le sera aussi.

Le problème décrit par les équations ci-dessus peut être formulé comme un problème d'estimation de paramètres. Nous avons à estimer un ensemble des pa-

ramètres, ici les m_{ij} et les P_i , étant données des mesures non exactes, ici les coordonnées images.

Supposons que les mesures soient obtenues avec une matrice de covariance C et appelons Q le vecteur de tous les paramètres avec q_k un de ses éléments. Soit U le vecteur de toutes les mesures U_{ij} et V_{ij} avec u_l un de ses éléments.

Supposons de plus que la relation entre les mesures u_l et les paramètres q_k est linéaire, c'est à dire, $U = AQ$ (dans le cas non linéaire, il est possible de linéariser les équations en utilisant un développement de Taylor). L'estimation du maximum de vraisemblance des paramètres est alors le vecteur Q minimisant la distance de Mahalanobis, qui n'est autre que le critère des moindres carrés suivant :

$$\chi^2 = (U - AQ)^t C^{-1} (U - AQ) \quad (3.11)$$

Les coordonnées images sont bruitées et les sources de ce bruit sont multiples ; en particulier, le modèle de projection perspective n'est pas correct en présence de distorsion optique. Cette dernière est de l'ordre du $\frac{1}{2}$ pixel selon nos estimations, ce qui est en accord avec les données du fabricant (Kinoptics). Une autre source d'erreurs non négligeables est celle de la localisation des points dans les images. Contrairement à celles dues à la distorsion, ces erreurs ne sont pas corrélées. Nous pourrions considérer dans la suite que nos principales erreurs sont non corrélées, vu que les erreurs dues aux mesures dans les images sont les plus dominantes et les plus difficiles à estimer.

Par conséquent la matrice de covariance est une matrice diagonale dont les éléments ne sont autres que les variances σ_{ij}^2 . Ainsi, dans notre cas, l'équation (3.11) conduit à la minimisation de la somme suivante :

$$\chi^2 = \sum_{ij} \left(\frac{U_{ij} - \frac{m_{11}^{(j)} x_i + m_{12}^{(j)} y_i + m_{13}^{(j)} z_i + m_{14}^{(j)} t_i}{m_{31}^{(j)} x_i + m_{32}^{(j)} y_i + m_{33}^{(j)} z_i + m_{34}^{(j)} t_i}}{\sigma_{ij}} \right)^2 + \sum_{ij} \left(\frac{V_{ij} - \frac{m_{21}^{(j)} x_i + m_{22}^{(j)} y_i + m_{23}^{(j)} z_i + m_{24}^{(j)} t_i}{m_{31}^{(j)} x_i + m_{32}^{(j)} y_i + m_{33}^{(j)} z_i + m_{34}^{(j)} t_i}}{\sigma_{ij}} \right)^2 \quad (3.12)$$

Résolution numérique

Le problème de la reconstruction est défini par le système d'équations (3.10), ce qui revient à minimiser la somme donnée par (3.12).

Les coordonnées homogènes sont définies à un facteur multiplicatif près. Il faut donc utiliser une contrainte pour chaque point. On peut par exemple fixer la quatrième coordonnée de tous les points à 1. Cependant, pour être plus général, et prendre en compte les points à l'infini dans le cas d'une reconstruction projective, nous avons utilisé pour chaque point (x_i, y_i, z_i, t_i) la contrainte suivante :

$$x_i^2 + y_i^2 + z_i^2 + t_i^2 = 1 \quad (3.13)$$

Les matrices de projection M_i sont aussi définies à un facteur multiplicatif près (cas de toutes les collinéations). En général les $m_{34}^{(i)}$ ne sont jamais nuls : dans le cas euclidien, les $m_{34}^{(i)}$ correspondent à la composante en Z du vecteur translation entre le repère de la caméra et celui de la scène. Si l'origine du repère de la scène n'est pas dans le plan $Z = 0$ du repère caméra, alors les $m_{34}^{(i)}$ ne peuvent être nuls. C'est toujours le cas pour la reconstruction relative, puisqu'on reconstruit dans un repère lié à la scène. Dans le cas projectif, si on suppose que le point $(0, 0, 0, 1)$, qui est l'origine d'une base projective, ne se projette pas à l'infini sur les images, alors les $m_{34}^{(i)}$ ne pourront être nuls. Ceci est toujours vrai car seuls les points visibles d'une image sont traités.

Nous avons donc choisi de mettre les derniers éléments des M_i (les $m_{34}^{(i)}$) à 1.

Simplification des équations

Minimiser la somme donnée en (3.12) est un problème difficile (rapport de deux fonctions de degré deux). Nous avons simplifié la complexité de ce problème en effectuant la transformation des équations (3.10) en :

$$\begin{cases} U_{ij}(m_{31}^{(j)}x_i + m_{32}^{(j)}y_i + m_{33}^{(j)}z_i + m_{34}^{(j)}t_i) = m_{11}^{(j)}x_i + m_{12}^{(j)}y_i + m_{13}^{(j)}z_i + m_{14}^{(j)}t_i \\ V_{ij}(m_{31}^{(j)}x_i + m_{32}^{(j)}y_i + m_{33}^{(j)}z_i + m_{34}^{(j)}t_i) = m_{21}^{(j)}x_i + m_{22}^{(j)}y_i + m_{23}^{(j)}z_i + m_{24}^{(j)}t_i \end{cases} \quad (3.14)$$

Il est à noter que cette transformation n'est valide que si les dénominateurs de (3.10) sont non nuls, ce qui est forcément le cas puisque aucun point de l'image n'est à l'infini (si un point de l'image est à l'infini alors il n'est pas visible).

Ainsi, nous obtenons des équations polynomiales de degré deux plus simples à manipuler. La somme à minimiser devient :

$$\begin{aligned} \chi^2 = \sum_{ij} & \left(\frac{U_{ij}(m_{31}^{(j)}x_i + m_{32}^{(j)}y_i + m_{33}^{(j)}z_i + m_{34}^{(j)}t_i) - (m_{11}^{(j)}x_i + m_{12}^{(j)}y_i + m_{13}^{(j)}z_i + m_{14}^{(j)}t_i)}{\sigma_{ij}} \right)^2 + \\ & \sum_{ij} \left(\frac{V_{ij}(m_{31}^{(j)}x_i + m_{32}^{(j)}y_i + m_{33}^{(j)}z_i + m_{34}^{(j)}t_i) - (m_{21}^{(j)}x_i + m_{22}^{(j)}y_i + m_{23}^{(j)}z_i + m_{24}^{(j)}t_i)}{\sigma_{ij}} \right)^2 \end{aligned} \quad (3.15)$$

Le problème à résoudre consiste donc à minimiser la somme suivante :

$$\sum_{k=1}^{2 \times p \times v + p} \left(\frac{f_k(x_{ij}, y_{ij}; x_i, y_i, z_i, t_i, m_{11}^{(j)}, \dots, m_{34}^{(j)})}{\sigma_k} \right)^2 \quad (3.16)$$

où σ_k est l'écart type de la $k^{\text{ème}}$ mesure, $f_k(\cdot)$ sont les fonctions non linéaires définies dans (3.14) et

$$(x_i, y_i, z_i, t_i, m_{11}^{(j)}, \dots, m_{33}^{(j)}) \text{ pour } i = 1, \dots, v, j = 1, \dots, p;$$

sont les paramètres du problème.

Nous pouvons éventuellement utiliser la solution trouvée en minimisant la somme (3.15) pour initialiser un autre processus qui effectue la minimisation de (3.12). Cependant, nous avons remarqué en pratique qu’une autre minimisation n’apporte aucune amélioration à nos résultats. Ceci peut s’expliquer par le fait que la quantité $(m_{31}^{(j)}x_i + m_{32}^{(j)}y_i + m_{33}^{(j)}z_i + m_{34}^{(j)}t_i)$ n’approche pas zéro. Par conséquent, minimiser la somme (3.12) revient à minimiser la somme (3.15). Une autre raison possible est la complexité des équations (3.10).

De nombreuses méthodes existent pour résoudre ce type de problème qui n’est autre que la résolution d’un système d’équations non linéaires au sens des moindres carrés. L’algorithme de Levenberg-Marquardt [PFTW88] est un algorithme bien connu pour les problèmes non linéaires. De nombreux chercheurs en photogrammétrie [Bey92b] et en vision [Har93] l’ont utilisé avec beaucoup de succès pour résoudre des problèmes similaires au notre. Aussi, nous avons utilisé l’algorithme de Levenberg-Marquardt pour notre cas.

3.3.3 Résultats expérimentaux

La méthode décrite ci-dessus est utilisée pour obtenir la reconstruction tridimensionnelle de trois scènes différentes : une scène d’intérieur, une scène d’extérieur et une scène de la mire.

- *Scène d’intérieur “maison”* (figure 3.4) : cinq images de cette scène sont utilisées pour cette expérience. La caméra est située à environ $1m$ de la scène dont le volume est d’environ $40cm \times 40cm \times 40cm$. Les cinq images sont obtenues en tournant la caméra autour de la scène (configuration similaire à celle de la figure 3.1).

Soixante treize points ont été détectés et poursuivis avec l’algorithme décrit dans la section 2.2 (pour la poursuite des points nous avons utilisé plus de cinq images pour faciliter la mise en correspondance ; les images étaient proches l’une de l’autre). Certains points disparaissent tandis que d’autres apparaissent comme le montre la figure 3.4. Ceci ne pose aucun problème pour l’algorithme de reconstruction puisque pour pouvoir être reconstruit, un point doit apparaître dans deux images au moins ; ce qui est le cas pour notre séquence.

Les 73 points de cette scène sont reconstruits relativement à une base de 5 points pris dans la scène. Ces 5 points sont marqués par x sur la figure 3.5 “vue d’ensemble”. Les coordonnées de ces derniers sont euclidiennes, mesurées avec une règle millimétrique. Le résultat de la reconstruction est donc de type euclidien comme le montre la figure 3.5. Pour illustrer le résultat, nous avons joint les points reconstruits par des segments de droite.



image 1



image 2



image 3



image 4



image 5

FIG. 3.4 - Les cinq images de la séquence “maison” avec les points d’intérêt détectés et mis en correspondance : notons qu’à la troisième image de nouveaux points apparaissent, puis à la cinquième image, d’autres disparaissent

Cette reconstruction est qualitativement excellente, la différence entre les coordonnées calculées et celles mesurées (avec notre règle millimétrique!) ne dépassant pas le seuil de $1,5mm$.

Remarque

Dans la figure 3.5 certains points ont des régions de confiance (parallélépipèdes) très grandes, ce qui correspond à de grandes incertitudes sur ces points. Un tel résultat peut être simplement due à des erreurs de localisation dans les images (précision des points d'intérêt). Cependant, en vérifiant (avec des agrandissement d'images), nous avons constaté que certains de ces points étaient bien localisés dans les images, mais qu'ils étaient isolés et loin des 5 points de la base. Comme nous ne manipulons pas des données exactes (bruit dans les images, absence d'un modèle exact pour la caméra et incertitudes sur les mesures des 5 points de la base), le processus de minimisation consiste à minimiser d'une manière globale la somme donnée par (3.15), c'est à dire qu'il cherche les paramètres du modèle qui satisfont le mieux la majorité des mesures. Ainsi, les paramètres trouvés pour le modèle (modèle de projection perspective pour notre cas) vont bien pour la partie de la scène où il y a la majorité des points ; les points isolés sont alors pénalisés.

Une solution possible à ce problème serait de pondérer différemment les équations (3.15) en donnant plus de poids aux points isolés. Malheureusement, ce type de manipulation peut être désastreux si pour de tels points la localisation est peu précise, la qualité des localisations des points étant difficile à estimer automatiquement.

- *Scène d'extérieur "Bayard"* (figure 3.6): trois images de cette scène sont utilisées pour cette expérience. Ces images sont obtenues avec un caméscope amateur, puis numérisées avec le matériel du laboratoire.

Cent soixante deux points ont été détectés et poursuivis avec l'algorithme décrit dans la section 2.2 ; ici, tous les points sont présents dans les trois images (voir figure 3.6).

Dans ce cas, nous n'avons pas la possibilité d'avoir des coordonnées euclidiennes. Nous avons alors utilisé une base projective de 5 points (ces 5 points sont marqués par des croix plus épaisses que celles des autres points sur "image 3" de la figure 3.6). A ces 5 points nous avons affecté les coordonnées homogènes suivantes : $(0, 0, 0, 1)$, $(1, 0, 0, 1)$, $(0, 1, 0, 1)$, $(0, 0, 1, 1)$, $(1, 1, 1, 1)$.

Une telle reconstruction est projective. Cependant puisque les coordonnées fournies correspondent approximativement à leur structure euclidienne, la reconstruction obtenue est proche en apparence d'une reconstruction euclidienne. La figure 3.7 présente le résultat de la reconstruction de cette scène sous trois angles de vue différents.

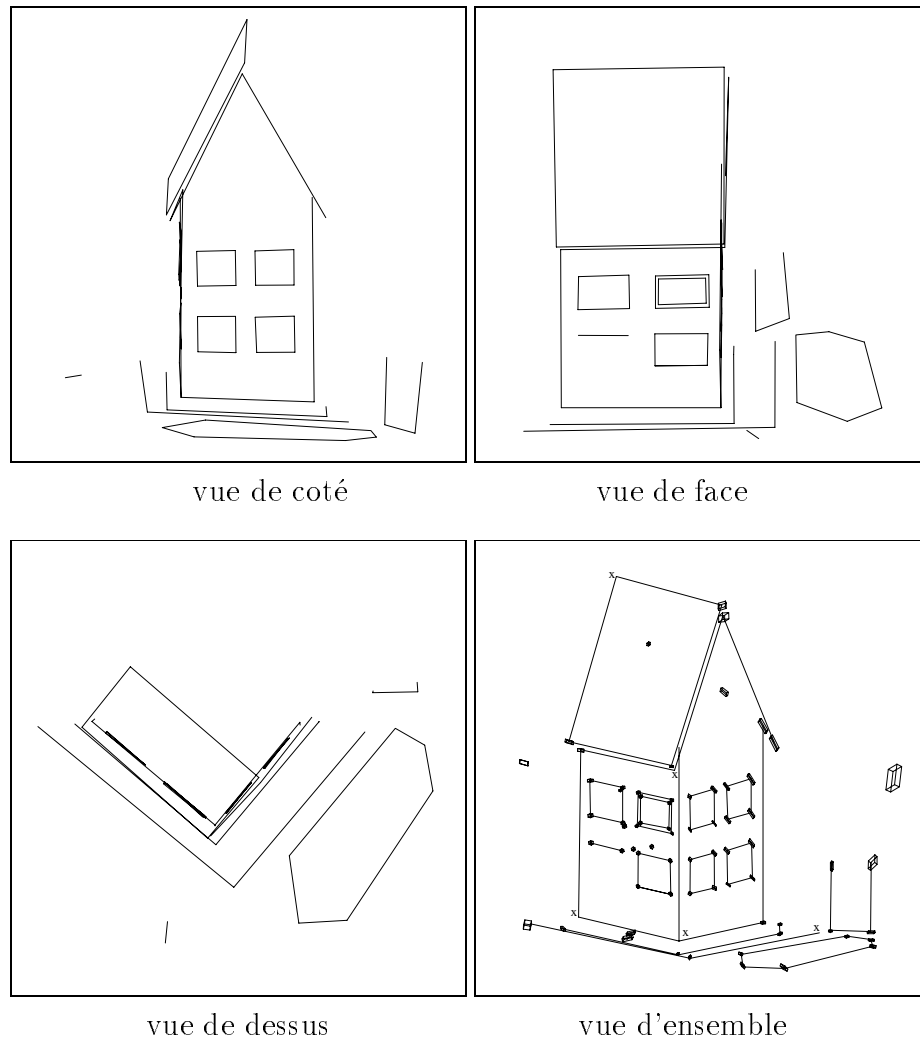


FIG. 3.5 - Les différentes vues de la scène reconstruite illustrant la bonne qualité du résultat. Les parallélépipèdes de la “vue d'ensemble” représentent les régions de confiance (éllipsoïdes) des points reconstruits, ces dernières sont calculées avec tous les σ_{ij} égale à 1, elles correspondent à une confiance de 68,3%



image 1

image 2

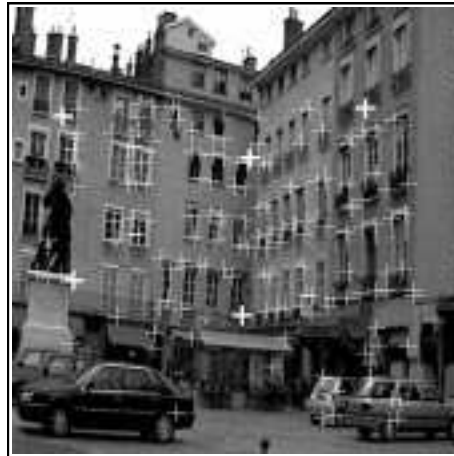


image 3

FIG. 3.6 - Les trois images de la séquence “Bayard” avec les 162 points d’intérêt détectés et mis en correspondance. Sur “image 3”, les croix tracées avec des traits plus épais désignent les 5 points choisis comme base projective. Notons qu’ici le mouvement de la caméra n’a pas été régulier comme pour la scène “maison”

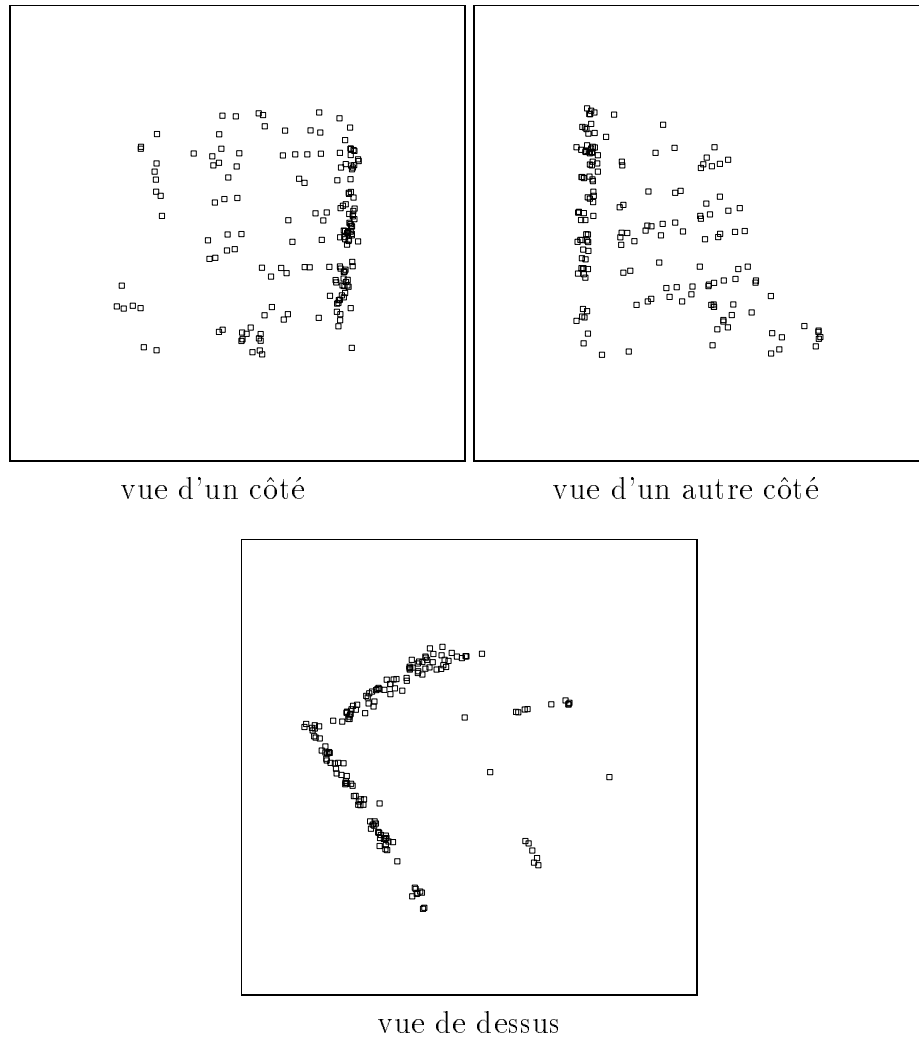


FIG. 3.7 - *Le résultat de la reconstruction de la scène “Bayard” : la vue de dessus est la plus parlante, on retrouve les deux plans presque perpendiculaires correspondants aux deux immeubles.*

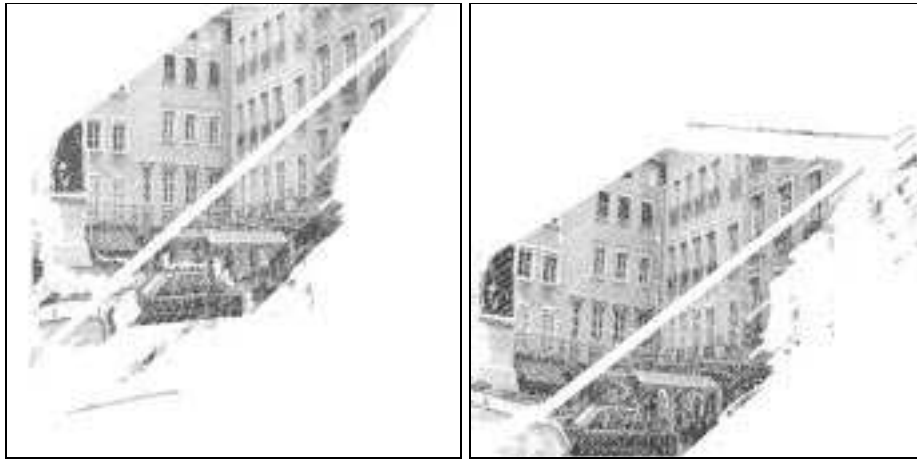


FIG. 3.8 - *Reconstruction dense d'une partie de la scène "Bayard" : ces deux images ne sont pas des images réelles ; elles sont obtenues par projection des points reconstruits sur l'image 1 (en utilisant la première matrice de projection) et sur l'image 2 (en utilisant la deuxième matrice de projection). Les parties blanches représentent les points non reconstruits (échec de la mise en correspondance pour ces points). Notons, que les valeurs des niveaux de gris utilisées ici sont celles de la première image réelle.*

Nous avons essayé de faire une mise en correspondance de tous les points de la première image avec les points de la seconde image (mise en correspondance dense). En d'autres termes, pour chaque pixel de la première image on cherche son correspondant dans la seconde image (la mise en correspondance n'est plus limitée à des points d'intérêt). Sans avoir obtenu toutes les mises en correspondance, nous avons néanmoins réussi pour une partie de la scène (pour environ 80000 points).

Les matrices de projection pour ces images sont connues (calculées en même temps que la reconstruction des 162 points d'intérêt avec notre méthode non linéaire). Ces matrices de projection avec les mises en correspondance ont permis de reconstruire les 80000 points par triangulation (méthode utilisée en stéréovision quand les caméras sont étalonnées). Le résultat est présenté sur la figure 3.8.

- *Scène de la mire* (figure 3.9) : Cette mire, en forme de trois plans, comporte 160 cibles circulaires couvrant un volume de $250mm \times 250mm \times 250mm$. Les coordonnées euclidiennes des centres des cibles (définies dans un repère lié à la mire) sont connues avec une précision de $0,02mm$ (fournie par le constructeur).

Quatre images de cette mire sont utilisées ici (figure 3.9). Les points d'intérêt sont les centres des cibles de la mire dont la détection et la mise en correspondance utilisent d'autres techniques (par rapport à la section 2.2).

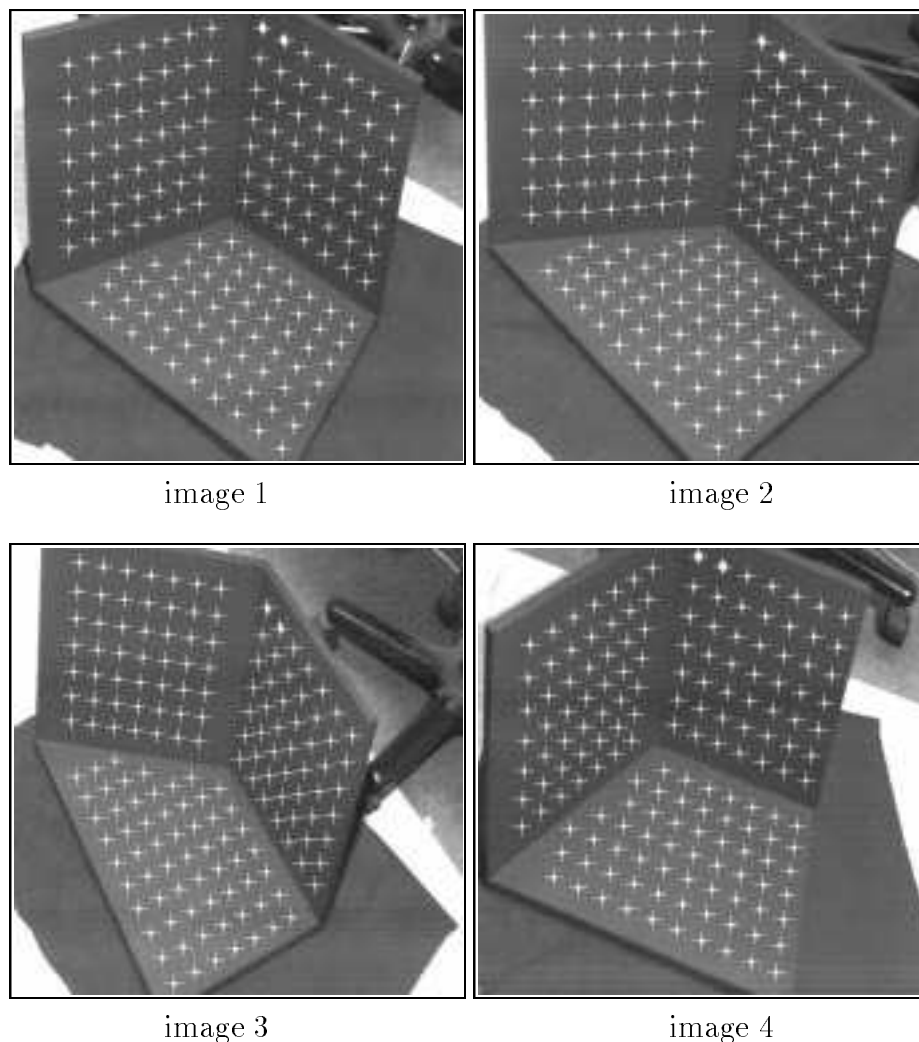


FIG. 3.9 - *La séquence de la mire : 4 images sont utilisées. Les points d'intérêt sont les centres des cibles de la mire*

En particulier, la détection du centre d'une cible dans une image se fait en estimant la meilleure transformation affine entre une cible théorique et la cible réelle. La précision du centre de la cible ainsi détecté est de l'ordre du dixième de pixel.

La figure 3.10 montre le résultat de la reconstruction des points de la mire. Ce résultat est qualitativement parfait, mais ne donne pas d'indication quantitative sur la précision des points reconstruits. Comme les coordonnées euclidiennes de tous les points de la mire sont connus (fournies par le constructeur de la mire), nous avons pu calculer les erreurs de reconstruction. En résumé, nous avons obtenu les erreurs moyennes suivantes :

- sur la coordonnée : X

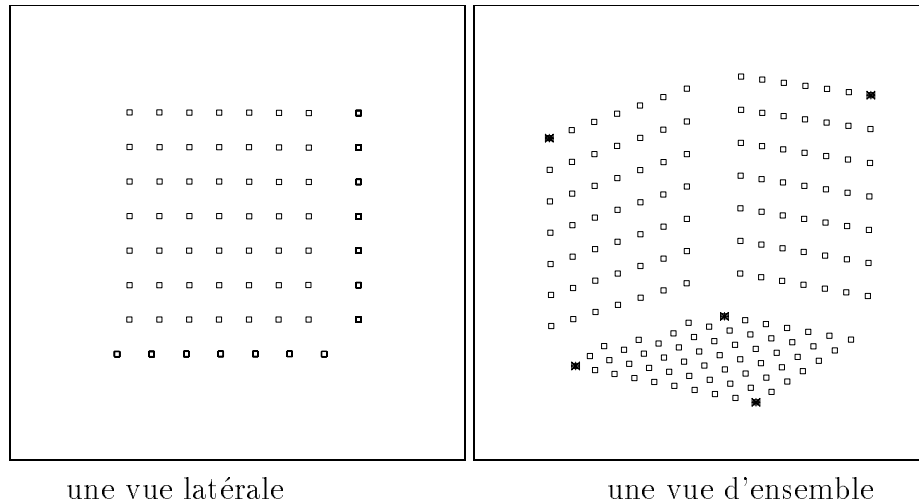


FIG. 3.10 - *Les points reconstruits de la mire: les 5 points utilisés comme base sont marqués par des carrés noirs sur la “vue d’ensemble”*

erreur moyenne $\Delta X_{moy} = 0,3344mm$

erreur relative $\Delta X_{rel} = 0,1338\%$

– sur la coordonnée: Y

erreur moyenne $\Delta Y_{moy} = 0,1311mm$

erreur relative $\Delta Y_{rel} = 0,0525\%$

– sur la coordonnée: Z

erreur moyenne $\Delta Z_{moy} = 0,2157mm$

erreur relative $\Delta Z_{rel} = 0,0863\%$

Dans ce paragraphe, nous avons testé notre méthode de reconstruction sur trois scènes différentes. Avec la première scène (“maison”), nous avons validé la méthode; avec la deuxième (“Bayard”) nous avons montré sa robustesse et enfin, avec la dernière scène (“mire”) la qualité de la reconstruction est mise en valeur.

Cependant, cette méthode est non linéaire. Le problème des valeurs initiales pour les paramètres est donc posé; cette question est traitée dans le paragraphe 3.4.2.

3.4 Comparaison et coopération des deux méthodes

Nous présentons dans cette section une étude comparative des deux méthodes de reconstruction développées dans les deux sections précédentes. Un modèle affine est généralement utilisé pour simplifier les calculs de reconstruction, tandis qu’un

modèle projectif, plus proche d'une caméra réelle, est utilisé pour une meilleure précision.

Les deux méthodes sont comparées en utilisant trois séquences d'images réelles prises sous des conditions correspondant à un *rapport de profondeur* faible, à un *rapport de profondeur* moyen, et enfin à un grand *rapport de profondeur*.

Nous présenterons à la fin de cette section un algorithme de reconstruction permettant la coopération des deux méthodes. La reconstruction avec un modèle affine donne une première reconstruction rapide (méthode linéaire). Cette dernière est peu précise ; elle servira à initialiser la méthode de reconstruction itérative qui utilise un modèle projectif.

3.4.1 Comparaison des deux méthodes

Afin de comparer la qualité et le comportement des deux méthodes présentées dans les sections 3.2 et 3.3, nous avons comparé les reconstructions de trois scènes correspondant aux trois cas suivant :

- **Scène 1 (Rapport de profondeur faible)** : une scène constituée d'un objet de taille $12\text{cm} \times 12\text{cm} \times 12\text{cm}$ et située à une distance d'environ 140cm de la caméra (voir figure 3.11). Le *rapport de profondeur* est inférieur à 0,1. Dans ce cas, la projection perspective faible est alors une bonne approximation pour la projection scène-image. Nous nous attendons donc à un résultat de reconstruction plutôt bon pour la méthode utilisant le modèle de projection perspective faible.
- **Scène 2 (apport de profondeur moyen)** : une scène constituée d'un objet de taille $15\text{cm} \times 15\text{cm} \times 15\text{cm}$ située à une distance d'environ 60cm de la caméra (voir figure 3.12). Le *rapport de profondeur* est d'environ 0,25. Ici, nous nous attendons à un résultat de reconstruction moins bon pour la méthode utilisant le modèle de projection perspective faible.
- **Scène 3 (Rapport de profondeur grand)** : une scène constituée de deux objets s'étalant sur un volume de taille $15\text{cm} \times 15\text{cm} \times 60\text{cm}$ située à une distance d'environ 60cm de la caméra (voir figure 3.13). Dans ce cas, le *rapport de profondeur* est de 1 ; l'approximation de la projection scène-image par une projection perspective faible n'est plus correcte. Nous nous attendons donc à un résultat plutôt mauvais pour la méthode utilisant le modèle de projection perspective faible.

Pour chacun des trois cas cités ci-dessus, nous avons utilisé une séquence de 5 images prises avec une même caméra dont la distance focale est de 12.5mm . Les coins sont extraits et poursuivis automatiquement avec l'algorithme décrit dans la section 2.2.

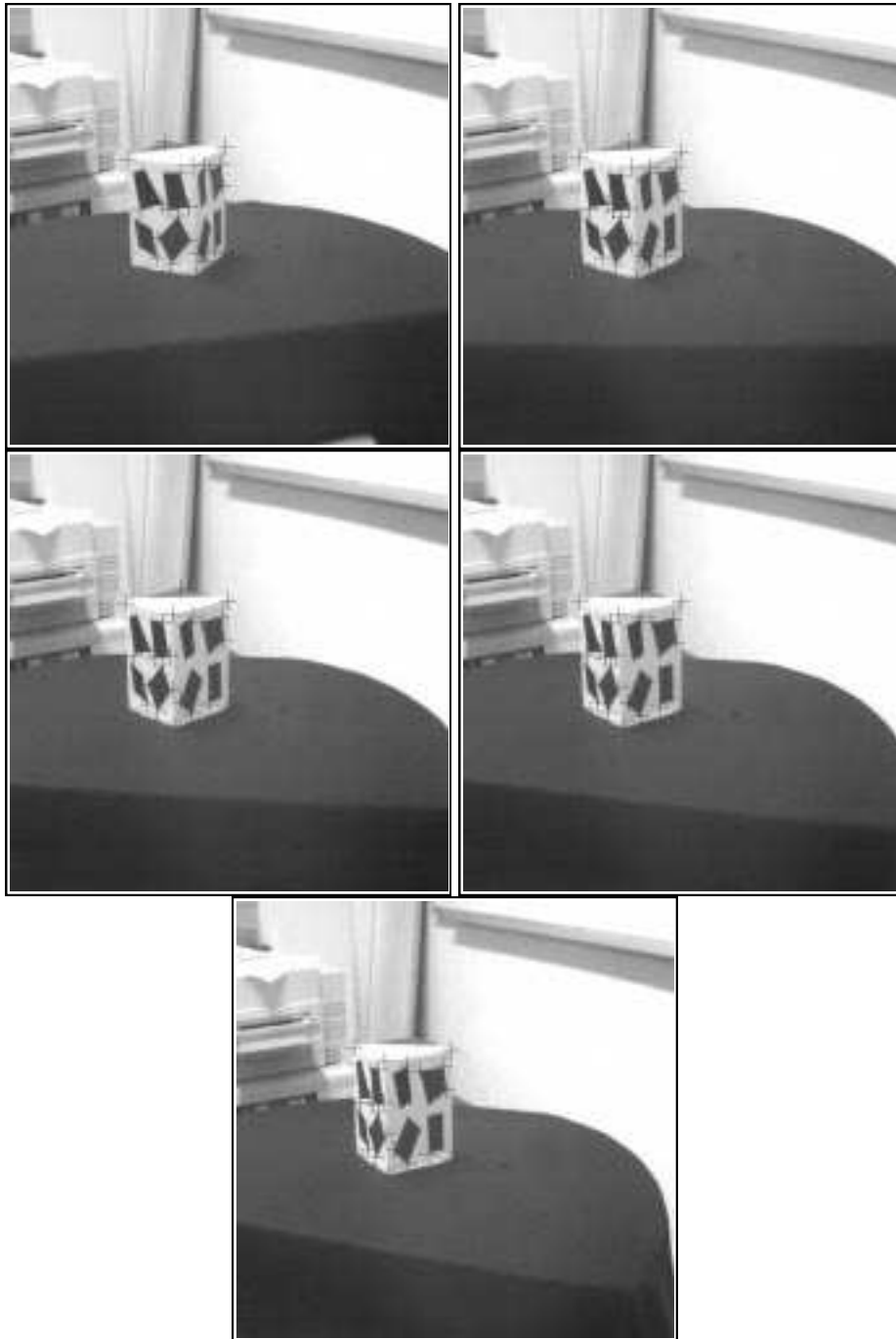


FIG. 3.11 - *Scène 1: les 5 images utilisées pour la reconstruction correspondant au cas d'un rapport de profondeur faible (10%); le nombre de points reconstruits est de 23*

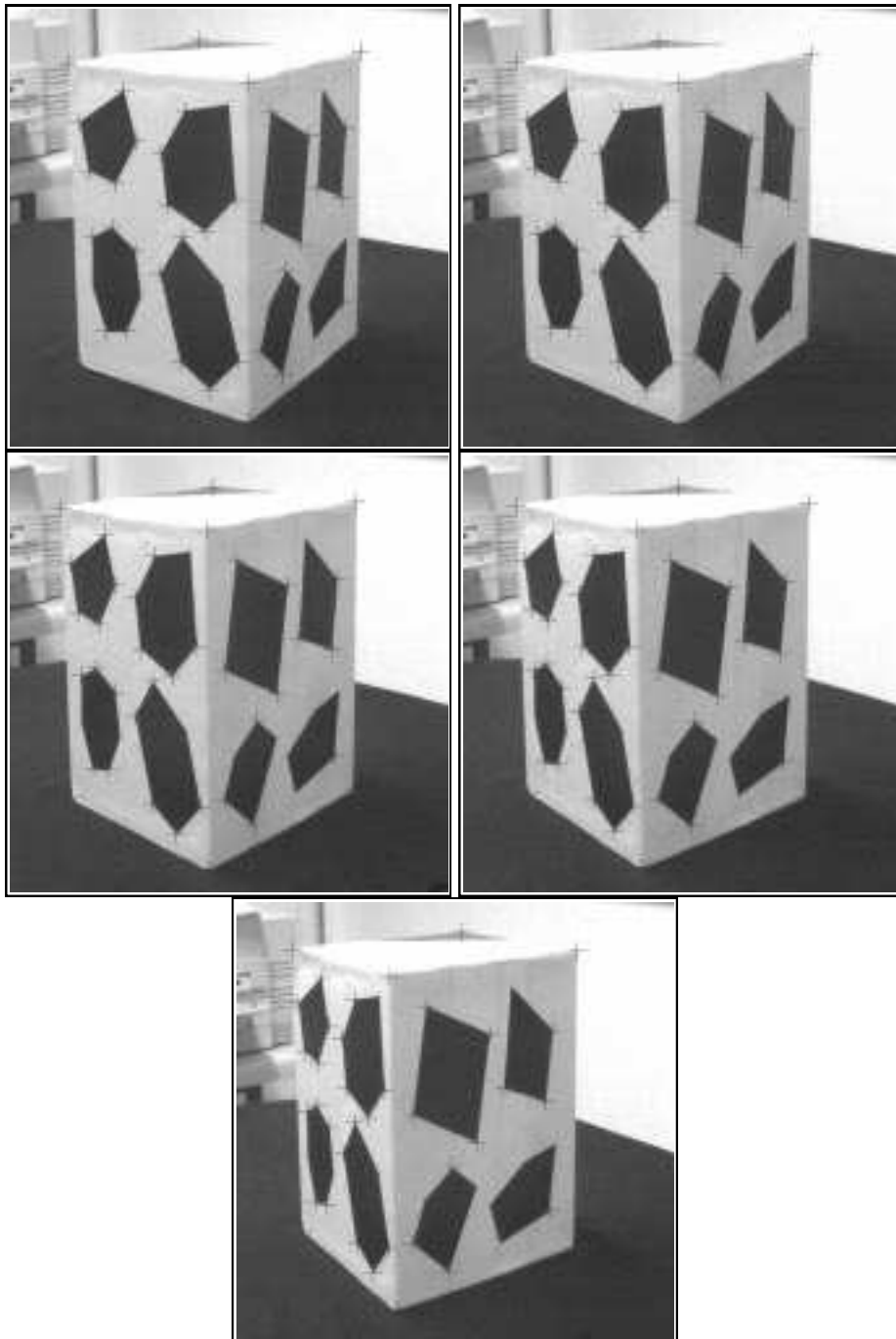


FIG. 3.12 - *Scène 2: les 5 images utilisées pour la reconstruction correspondant au cas d'un rapport de profondeur moyen (25%); l'objet à reconstruire est plus grand et la caméra plus proche; ici le nombre de points reconstruits est de 38*

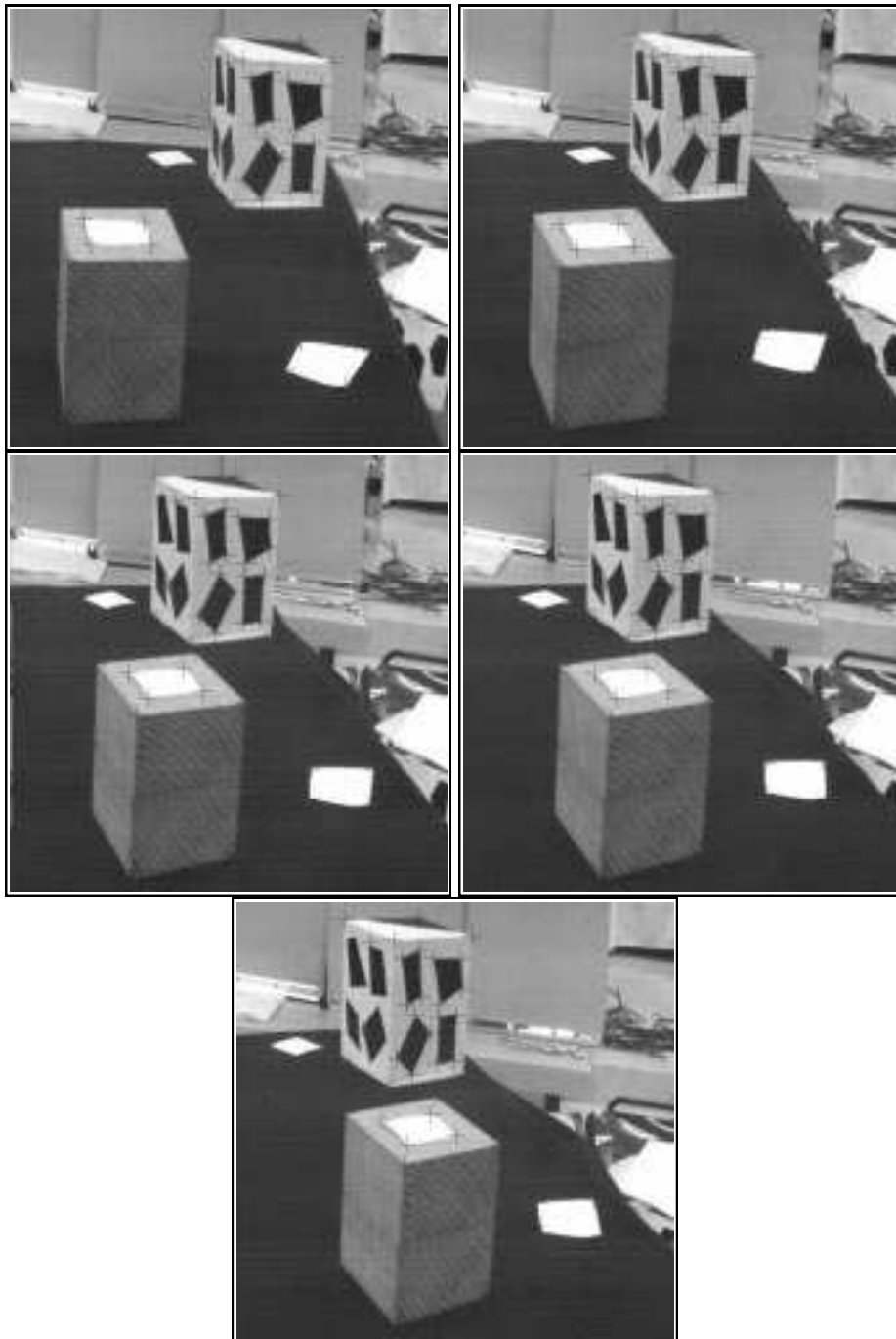


FIG. 3.13 - *Scène 3* : les 5 images utilisées pour la reconstruction correspondant au cas d'un rapport de profondeur grand (100%), la scène comporte plusieurs objets, elle est plus grande ; ici le nombre de points reconstruits est de 37

coordonnées mesurées(cm)			erreurs: perspective faible(cm)			erreurs: perspective(cm)		
X	Y	Z	ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
0,00	12,00	12,50	0,43	0,10	0,18	0,02	1,77	0,27
12,00	0,00	12,50	0,25	0,44	0,13	0,64	0,58	0,37
6,10	0,00	11,00	0,02	0,16	0,03	0,47	0,15	0,21
10,50	0,00	10,80	0,20	0,35	0,06	0,53	0,42	0,41
11,00	0,00	8,10	0,05	0,24	0,04	0,51	0,24	0,56
0,00	5,70	7,70	0,17	0,08	0,05	0,03	0,44	0,25
0,00	2,45	11,40	0,28	0,04	0,15	0,01	0,33	0,01
0,00	4,80	7,10	0,08	0,05	0,04	0,11	0,29	0,30
0,00	3,00	6,60	0,24	0,03	0,08	0,01	0,18	0,29
0,00	4,40	1,00	0,09	0,10	0,18	0,15	0,22	0,03
erreurs moyennes			0,14	0,16	0,13	0,21	0,34	0,24

TAB. 3.2 - Les erreurs de reconstruction pour la scène 1: les deux méthodes donnent des résultats similaires.

Pour pouvoir comparer les différentes reconstructions, nous avons mesuré les coordonnées tridimensionnelles (dans un repère relatif à la scène) avec une règle millimétrique.

Résultats des reconstructions

Les coordonnées reconstruites avec les deux méthodes sont comparées aux coordonnées mesurées. Les erreurs de reconstructions sont tout simplement les différences entre les coordonnées reconstruites et les coordonnées mesurées. Même si ces dernières ne sont pas très précises, elles restent suffisantes pour nos comparaisons dont le but est qualitatif. Pour des résultats quantitatifs voir paragraphe 3.2.4 pour la reconstruction sous l'hypothèse d'une projection perspective faible, ou paragraphe 3.3.3 pour la reconstruction sous l'hypothèse d'une projection perspective.

Les résultats sont donnés par les tableaux 3.2, 3.3 et 3.4 pour les trois cas du *rapport de profondeur*: faible, moyen et grand respectivement. Chaque tableau ne présente que les résultats de 10 points de la scène correspondante. Cependant, les erreurs moyennes sont calculées sur la totalité des points reconstruits.

Les résultats obtenus avec ces trois scènes étaient plus ou moins prévisibles. Cependant quelques remarques sont à noter pour la première et la troisième scène respectivement :

- **scène 1** : d'après le tableau 3.2 il apparaît surprenant que la reconstruction sous l'hypothèse d'une projection perspective faible donne de meilleurs résultats que la reconstruction sous l'hypothèse d'une projection perspective. Les conditions ici (*rapport de profondeur* très faible) font que l'erreur

coordonnées mesurées(cm)			erreurs: perspective faible(cm)			erreurs: perspective(cm)		
X	Y	Z	ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
0,00	0,00	15,00	0,92	1,16	0,36	0,19	0,09	0,22
0,00	15,00	15,00	1,48	0,53	0,16	0,05	0,04	0,08
14,90	15,00	15,00	1,28	0,96	0,31	0,20	0,22	0,24
14,90	0,00	15,00	0,48	1,91	0,22	0,05	0,17	0,03
12,80	0,00	11,40	0,27	1,43	0,03	0,04	0,10	0,02
12,20	0,00	8,00	0,17	1,31	0,08	0,01	0,04	0,04
8,60	0,00	9,00	0,09	0,49	0,05	0,07	0,02	0,04
0,00	9,90	14,10	0,47	0,25	0,01	0,07	0,01	0,03
0,00	10,80	9,70	0,51	0,16	0,02	0,07	0,03	0,04
0,00	13,25	10,00	1,01	0,33	0,03	0,06	0,05	0,06
erreurs moyennes			0,55	0,56	0,22	0,07	0,07	0,06

TAB. 3.3 - Les erreurs de reconstruction pour la scène 2: la reconstruction sous l'hypothèse perspective faible est moins bonne mais elle reste qualitativement acceptable

coordonnées mesurées(cm)			erreurs: perspective faible(cm)			erreurs: perspective(cm)		
X	Y	Z	ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
12,00	0,00	12,50	1,25	0,08	0,29	0,56	0,35	0,05
0,00	12,00	12,50	3,59	0,36	1,92	0,37	0,58	0,14
4,40	0,00	7,60	0,43	0,57	0,09	0,08	0,05	0,20
2,20	0,00	7,10	1,15	1,02	0,13	0,01	0,03	0,19
1,25	0,00	11,15	2,51	1,98	0,55	0,11	0,05	0,02
6,10	0,00	11,00	0,56	0,77	0,16	0,32	0,07	0,08
7,40	0,00	6,80	0,96	0,10	0,43	0,13	0,20	0,17
11,00	0,00	8,10	2,05	0,49	0,49	0,24	0,32	0,02
10,50	0,00	10,80	1,14	0,07	0,18	0,35	0,39	0,09
-20,00	-31,00	9,00	1,19	5,37	1,14	3,85	1,10	0,14
erreurs moyennes			1,89	1,34	0,67	0,58	0,31	0,07

TAB. 3.4 - Les erreurs de reconstruction pour la scène 3: la reconstruction sous l'hypothèse perspective faible se dégrade avec un rapport de profondeur égale à 1

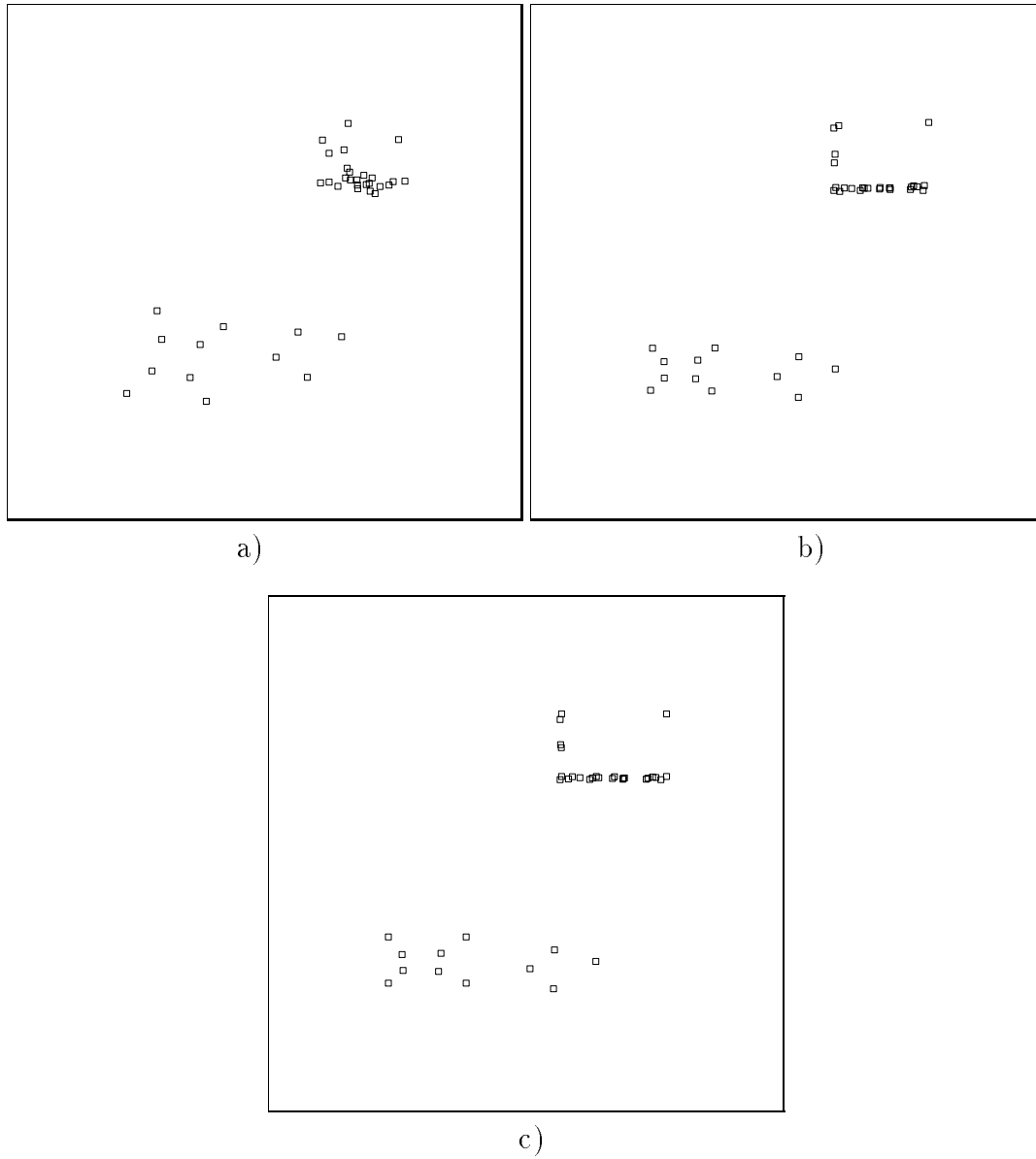


FIG. 3.14 - Vues de dessus des points tridimensionnels de la scène 3 : a) reconstruction sous l'hypothèse d'une projection perspective faible, b) reconstruction sous l'hypothèse d'une projection perspective et c) les points mesurés avec la règle millimétrique

due à l'approximation de la projection scène-image par un modèle affine est moins importante que celle due aux erreurs de localisation des points dans les images. Sous ces mêmes conditions, le modèle projectif pour une caméra est plus sensible aux erreurs de localisation des points dans les images qu'un modèle affine car plus complexe. Ces dernières erreurs ont donc contribué pour ce cas particulier à favoriser la méthode de reconstruction sous l'hypothèse d'une projection perspective faible.

- **scène 3** : le résultat de la reconstruction sous l'hypothèse d'une projection perspective présente des erreurs particulièrement importantes pour certains points. Ces derniers sont isolés et situés loin des points utilisés comme repère relatif, phénomène déjà observé au paragraphe 3.3.3.

3.4.2 Un algorithme de coopération

Les résultats des paragraphes 3.2.4 et 3.3.3 montrent que la qualité des reconstructions sous l'hypothèse d'une projection perspective faible se dégrade avec la croissance du *rapport de profondeur*. D'un autre côté, la reconstruction sous l'hypothèse d'une projection perspective n'est pas sensible à ce *rapport de profondeur*. Cependant elle est basée sur un algorithme non linéaire. Ce dernier, comme tout algorithme non linéaire, n'a aucune garantie de convergence. Les deux méthodes (linéaire et non linéaire) peuvent être alors complémentaires : la reconstruction obtenue sous l'hypothèse d'une projection perspective faible sera utilisée pour initialiser l'algorithme de reconstruction non linéaire. Nous combinons ainsi la simplicité d'un algorithme de reconstruction linéaire (modèle affine) avec la robustesse et la précision d'un algorithme de reconstruction non linéaire (modèle projectif).

Il est à noter que sous certaines conditions l'algorithme de reconstruction sous l'hypothèse d'une projection perspective faible ne peut être utilisé. Ces conditions correspondent à un *rapport de profondeur* assez grand, dépassant l'unité pour nos expérimentations avec les données simulées du paragraphe 3.2.4.

Nous avons testé cet algorithme de coopération sur la scène 3 (Figure 3.13). La convergence de l'algorithme non linéaire est plus rapide. Avec une initialisation aléatoire, l'algorithme de Levenberg-Marquardt utilisé pour la reconstruction non linéaire a convergé après 9 itérations ; mais en initialisant ce même algorithme avec le résultat de la reconstruction linéaire il a convergé après 3 itérations.

3.5 Conclusion

Dans ce chapitre nous avons présenté deux méthodes différentes de reconstruction tridimensionnelle.

La première méthode permet l'obtention de la structure euclidienne à partir d'au moins trois images. Elle suppose un modèle de projection simplifié pour la

caméra (modèle affine), ce qui rend le calcul de la structure linéaire. Cependant, cette méthode ne garantit pas une bonne précision de la reconstruction. Une étude de son comportement quand les conditions géométriques de prise de vues lui sont défavorables a été faite. La précision de la reconstruction sous l'hypothèse affine se dégrade d'une manière linéaire en fonction du *rapport de profondeur*. Nos expérimentations sur les données simulées et les données réelles ont montré qu'avec un *rapport de profondeur* inférieur à 1 la reconstruction restait toujours possible.

La deuxième méthode permet d'obtenir la structure tridimensionnelle d'une scène à partir de deux images avec un modèle de projection perspective. Ce dernier est plus proche de la projection effectuée par une caméra réelle ; la méthode est donc plus robuste. Cependant, le calcul de la structure est non linéaire et la connaissance des coordonnées euclidiennes de 5 points de la scène est nécessaire pour obtenir une reconstruction euclidienne.

Les deux méthodes sont ensuite comparées dans des conditions similaires. Quand le *rapport de profondeur* est assez petit, la précision de la reconstruction sous l'hypothèse d'une projection affine peut être suffisante pour certaines applications. Cependant, une amélioration de la précision peut être obtenue en utilisant la méthode non linéaire comme deuxième étape. Le calcul d'une reconstruction en deux étapes a au moins deux avantages : obtenir une reconstruction précise et garantir la convergence de la méthode non linéaire. La reconstruction par la méthode linéaire donne en général une bonne estimation de la solution. Cette dernière rend la convergence de la méthode non linéaire plus rapide : le nombre d'itérations est très limité car l'initialisation est bonne.

Bien qu'il soit rare de se trouver dans des conditions où le *rapport de profondeur* est très grand (supérieur à 1 pour nos expériences) ce cas peut se produire lorsqu'on observe des scènes d'extérieur avec une grande profondeur de champ. Ceci mettra en échec la méthode linéaire et par conséquent la convergence de la méthode non linéaire ne sera plus garantie.

Une solution dans ce cas serait de calculer une estimation initiale pour la reconstruction à l'aide d'une méthode linéaire utilisant le modèle projectif [Fau92] [HGC92] [Sha92] [PMC94]. Toutes ces méthodes linéaires utilisent la géométrie épipolaire et dépendent donc des résultats de celle-ci. La reconstruction obtenue avec de telles méthodes est peu précise, mais suffisante pour faire converger la méthode non linéaire proposée ici.

Il faut noter que la convergence de notre méthode non linéaire a été obtenue avec des initialisations aléatoires pour toutes nos expériences. Le gain réalisé en initialisant la méthode avec une solution approchée s'est limité à rendre la convergence plus rapide.

Chapitre 4

D'une reconstruction projective à une reconstruction euclidienne

Nous avons présenté dans la section 3.2 une méthode linéaire de reconstruction tridimensionnelle sous l'hypothèse d'un modèle de projection perspective faible. Cette méthode permet d'avoir la structure euclidienne d'une scène. Cependant, elle a des limites : elle donne des résultats peu précis, voire faux, quand certaines conditions géométriques ne sont pas respectées (quand le *rapport de profondeur* devient grand).

Nous avons alors proposé une autre méthode dans la section 3.3. Cette deuxième méthode est itérative et utilise un modèle de projection perspective (plus proche d'une caméra réelle). La structure euclidienne ainsi obtenue est plus précise (voir section 3.4), mais nécessite la connaissance des coordonnées euclidiennes de 5 points de la scène. En absence de coordonnées euclidiennes ou affines, nous pouvons toujours affecter aux 5 points (utilisés comme base) des coordonnées projectives. Nous pouvons, par exemple, leur affecter les coordonnées de la base canonique de \mathbb{P}^3 . La reconstruction ainsi calculée est définie à une transformation projective près. Cette reconstruction est pauvre en information ; en particulier, l'information métrique est absente (voir figure 4.1).

En d'autres termes, sous l'hypothèse d'un modèle projectif pour la caméra (plus correct qu'un modèle affine), il est possible d'obtenir la reconstruction projective d'une scène en utilisant seulement les points appariés dans deux images au moins. De plus, cette reconstruction est reliée à la reconstruction euclidienne par une collinéation de l'espace. En particulier, quand les coordonnées euclidiennes (respectivement affines) de 5 points de la scène sont connues, nous pouvons calculer la collinéation qui fait passer d'une reconstruction projective à une reconstruction euclidienne (respectivement affine).

Dans la suite, on utilisera pour simplifier le terme "reconstruction projective" pour désigner une reconstruction définie à une transformation projective de l'espace près.

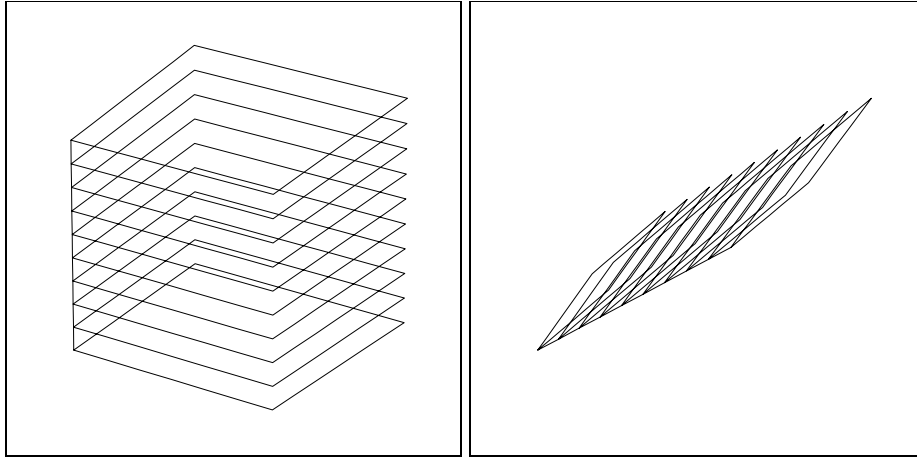


FIG. 4.1 - *A gauche une cube défini dans un espace euclidien, à droite le même cube défini dans un espace projectif: les deux cubes sont liés par une transformation projective de l'espace*

Nous nous posons, dans ce chapitre, le problème suivant : étant donnée une reconstruction projective obtenue avec la méthode itérative de la section 3.3 (ou éventuellement obtenue avec une méthode linéaire [Fau92][HGC92][Sha92][PMC94]), comment calculer la collinéation qui, appliquée à cette reconstruction, donnerait une reconstruction euclidienne (reconstruction définie à une similitude près). La solution est directe si les coordonnées euclidiennes de 5 points (non 4 à 4 coplanaires) de la scène sont connues, car 5 points suffisent pour définir une collinéation de l'espace. Cependant, il n'est pas toujours possible de connaître les coordonnées euclidiennes de 5 points d'une scène donnée. Nous nous proposons alors d'utiliser des connaissances géométriques sur la scène pour déterminer la collinéation de l'espace appropriée. Cette collinéation sera notée W dans toute la suite de ce chapitre. Pour déterminer W , des connaissances géométriques sont traduites en contraintes sur W . Comme toute collinéation de \mathbb{P}^3 dans \mathbb{P}^3 , W possède 15 paramètres indépendants. Aussi, au moins 15 contraintes indépendantes sont alors nécessaires pour le calcul de W .

La section suivante présente les aspects théoriques du passage d'une reconstruction projective à une reconstruction euclidienne. En particulier, nous montrons qu'il est possible de décomposer ce passage en deux étapes : passage projectif-affine puis passage affine-euclidien. Ensuite, dans la section 4.2, nous présentons les aspects pratiques du passage projectif-euclidien ; nous montrons que le problème peut se transformer en un problème d'optimisation non linéaire sur les 15 paramètres de W . Enfin, la section 4.3 est consacrée aux résultats expérimentaux.

4.1 D'une reconstruction projective à une reconstruction euclidienne : aspects théoriques

La structure (géométrique) d'une scène peut être décrite dans une géométrie donnée. Passer d'une structure projective à une structure euclidienne revient à changer de géométrie et donc de description.

La géométrie euclidienne est une restriction de la géométrie affine qui, elle-même, est une restriction de la géométrie projective (chapitre 1) ; on peut alors décomposer le passage d'une reconstruction projective à une reconstruction euclidienne en deux étapes : la reconstruction projective est d'abord transformée en une reconstruction affine, puis en une reconstruction euclidienne. Comme cela est expliqué dans la suite, le passage projectif-affine est plus simple que le passage affine-euclidien. Pour passer de la géométrie projective à la géométrie affine, on doit fixer l'hyperplan de l'infini (trois paramètres). Ceci se traduit pour la collinéation W par la détermination de 3 de ses 15 paramètres. Par contre, pour passer de la géométrie affine à la géométrie euclidienne, on doit fixer *la conique absolue* (cinq paramètres). Cela revient à déterminer 5 autres paramètres de la collinéation W . Enfin, puisque la reconstruction euclidienne recherchée est définie à une similitude près, les 7 paramètres de W qui restent, correspondent au choix d'un repère relatif et d'un facteur d'échelle.

4.1.1 Du projectif à l'anne

La géométrie affine se déduit de la géométrie projective en fixant l'hyperplan de l'infini (plan de l'infini pour le cas tridimensionnel).

Il s'agit donc pour nous de trouver le plan de l'infini adéquat : le plan qui représente effectivement la notion d'infini [Moh93]. Une fois ce plan fixé, il est alors facile de passer d'une reconstruction projective à une reconstruction affine.

Notons que si on choisit n'importe quel autre plan comme celui de l'infini, la géométrie qui en découle ne sera pas affine.

Trouver le plan de l'infini

Trouver le plan de l'infini, appelé Π_∞ , revient à trouver au moins trois points non colinéaires appartenant à Π_∞ . En d'autres termes, il faut identifier parmi les points de l'espace projectif (points définis par leurs coordonnées projectives) au moins 3 points possédant l'information affine : points se trouvant à l'infini.

Plusieurs méthodes peuvent être proposées pour trouver des points de l'infini. La plus commune de ces méthodes consiste à trouver les points d'intersection de droites parallèles : deux ou plusieurs droites parallèles de l'espace se coupent en un point à l'infini [SK52]. De plus, ces droites parallèles se projettent sur une

image en un faisceaux de droites concourantes en un point appelé point de fuite. Ce dernier correspond à la projection du point à l'infini, point d'intersection des droites parallèles de l'espace.

Ainsi, pour reconstruire un point à l'infini, il suffit de localiser un point de fuite et son correspondant dans au moins une autre image. Ceci est possible pour le cas des scènes comportant des droites parallèles puisque des techniques existent pour trouver les points de fuite dans une image [Bar83][QM89][BO90].

Une alternative à cette méthode est d'utiliser l'invariant affine de base : le rapport de trois points colinéaires A , B , et C . Ce rapport n'est autre que le birapport des quatre points colinéaires A , B , C et D , où D est un point à l'infini.

$$[A, B; C, D] = \frac{AC}{BC} \times \frac{BD}{AD} = \frac{AC}{BC} \times \frac{\infty}{\infty} = \frac{AC}{BC} \quad (4.1)$$

Connaissant les projections dans une image des points A , B et C , la projection dans la même image du point D peut être localisée si la valeur de ce rapport est connue. Une connaissance sur la scène permet de déduire la valeur de ce rapport. Par exemple, si nous savons que C est le milieu de AB alors ce rapport vaudra -1.

Comme le birapport est invariant par transformation projective, le point D peut ainsi être localisé dans l'espace projectif dès lors que les points A , B et C sont connus (par leurs coordonnées projectives).

Il est évident, que sans aucune connaissance a priori sur la scène, le plan de l'infini ne peut être fixé, car l'infini est une information affine. En particulier, les deux méthodes présentées ci-dessus pour calculer le plan de l'infini supposent des connaissances affines a priori : présence de droites parallèles dans la scène ou connaissance de rapports de triplets de points colinéaires.

D'une reconstruction projective à une reconstruction affine

Supposons à présent que le plan de l'infini Π_∞ a été fixé par une des deux méthodes citées ci-dessus. En utilisant cette information, comment trouver la reconstruction affine de notre scène dont la reconstruction projective est déjà connue ? Ou encore, comment calculer une collinéation W qui transformerait notre reconstruction projective en une reconstruction affine ?

Pour ce faire, nous pouvons utiliser 5 points de l'espace pour calculer cette collinéation : les 3 points de l'infini déjà trouvés (ou plus généralement n'importe quels 3 points non colinéaires mais appartenant au plan de l'infini), un point de la scène qui peut servir d'origine pour notre repère affine et enfin un autre point quelconque de la scène. Cependant, il faut que les 5 points ainsi choisis ne soient en aucun cas 4 à 4 coplanaires. Le dernier point peut prendre n'importe quelles coordonnées non nulles. Il permet de fixer les trois facteurs d'échelle (pour les trois axes) ; en particulier nous pouvons prendre le point unitaire.

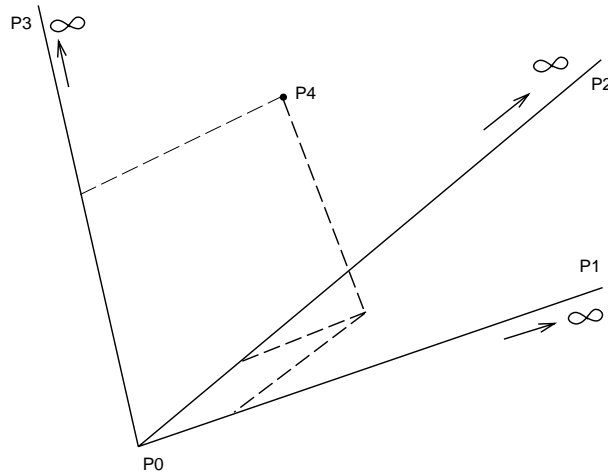


FIG. 4.2 - repère affine construit avec 5 points dont 3 sont à l'infini.

D'une manière plus formelle, soient 5 points non 4 à 4 coplanaires : P_1 , P_2 et P_3 trois points à l'infini, P_0 le point origine et P_4 un cinquième point. Soient (x_1, y_1, z_1, t_1) , (x_2, y_2, z_2, t_2) , (x_3, y_3, z_3, t_3) , (x_4, y_4, z_4, t_4) et (x_0, y_0, z_0, t_0) les coordonnées projectives de P_1 , P_2 , P_3 , P_4 et P_0 respectivement. Ces coordonnées sont connues puisque nous avons déjà une reconstruction projective de la scène.

Les coordonnées (homogènes) affines de ces 5 points peuvent être choisis de la manière suivante : $(1, 0, 0, 0)$ pour P_1 , $(0, 1, 0, 0)$ pour P_2 , $(0, 0, 1, 0)$ pour P_3 , $(1, 1, 1, 1)$ pour P_4 et enfin, $(0, 0, 0, 1)$ pour P_0 .

Avec ce choix de coordonnées, nous avons défini un repère affine $OXYZ$ tel que P_0 est l'origine, P_1 définit l'axe OX et se trouve à l'infini, P_2 définit l'axe OY et se trouve à l'infini, P_3 définit l'axe OZ et se trouve à l'infini, et enfin P_4 fixe les trois facteurs d'échelle selon les trois axes ainsi définis (voir figure 4.2).

Calcul de la collinéation W

Soit la matrice W représentant la collinéation de passage projectif-affine à déterminer :

$$W = \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{pmatrix}$$

W permet de passer pour un point, des coordonnées projectives aux coordonnées affines. En écrivant les équations pour les 5 points P_1 , P_2 , P_3 , P_4 et P_0 , donnés ici par leurs coordonnées projectives, nous obtenons le système d'équations suivant :

$$\begin{cases} WP_1 = \lambda_1 {}^t(1, 0, 0, 0) \\ WP_2 = \lambda_2 {}^t(0, 1, 0, 0) \\ WP_3 = \lambda_3 {}^t(0, 0, 1, 0) \\ WP_4 = \lambda_4 {}^t(1, 1, 1, 1) \\ WP_0 = \lambda_0 {}^t(0, 0, 0, 1) \end{cases} \quad (4.2)$$

où λ_i , $i = 0 \dots 4$, sont des facteurs multiplicatifs quelconques (en géométrie projective, l'égalité entre deux points est toujours définie à un facteur multiplicatif près).

Chaque ligne du système (4.2) comporte quatre équations linéaires en les coefficients de W . Ces équations étant homogènes, seules trois d'entre elles sont donc indépendantes. Nous pouvons éliminer le facteur multiplicatif λ_i et obtenir trois équations indépendantes en utilisant des égalités de rapports. Développons par exemple la première égalité $W {}^tP_1 = \lambda_1 {}^t(1, 0, 0, 0)$; ce qui donne les 3 équations non homogènes suivantes :

$$\begin{cases} \frac{w_{21}x_1 + w_{22}y_1 + w_{23}z_1 + w_{24}t_1}{w_{11}x_1 + w_{12}y_1 + w_{13}z_1 + w_{14}t_1} = \frac{0}{1} = 0 \\ \frac{w_{31}x_1 + w_{32}y_1 + w_{33}z_1 + w_{34}t_1}{w_{11}x_1 + w_{12}y_1 + w_{13}z_1 + w_{14}t_1} = \frac{0}{1} = 0 \\ \frac{w_{41}x_1 + w_{42}y_1 + w_{43}z_1 + w_{44}t_1}{w_{11}x_1 + w_{12}y_1 + w_{13}z_1 + w_{14}t_1} = \frac{1}{0} = 0 \end{cases}$$

Ce système d'équations particulier peut s'écrire aussi sous la forme :

$$\begin{cases} w_{21}x_1 + w_{22}y_1 + w_{23}z_1 + w_{24}t_1 = 0 \\ w_{31}x_1 + w_{32}y_1 + w_{33}z_1 + w_{34}t_1 = 0 \\ w_{41}x_1 + w_{42}y_1 + w_{43}z_1 + w_{44}t_1 = 0 \end{cases}$$

Ainsi, le système (4.2) fournit 15 équations linéaires indépendantes en les coefficients de W qui est donc complètement définie.

4.1.2 De l'affine à l'eulidien

En supposant que la reconstruction affine pour une scène donnée ait été obtenue à l'étape précédente, nous voulons à présent déterminer la reconstruction euclidienne de cette même scène. La géométrie euclidienne est une restriction de la géométrie affine; elle se déduit de cette dernière en fixant l'invariant tridimensionnel euclidien : *la conique absolu* Ω (voir chapitre 1). Pour passer de la géométrie affine à la géométrie euclidienne nous devons donc fixer 5 paramètres qui correspondent aux 5 degrés de liberté de *la conique absolue*. Rappelons ici que l'espace euclidien considéré est "l'espace euclidien étendu" de dimension 3 (voir section 1.1.5). Par conséquent, la reconstruction euclidienne recherchée sera définie à une similitude près.

Si nous raisonnons en terme de transformation projective, celle que nous cherchons n'a plus que 12 paramètres à déterminer étant donné que nous en avons déjà fixé 3 à l'étape précédente. Ces 12 paramètres correspondent à la transformation affine qui permet de passer d'une reconstruction affine à une reconstruction euclidien. Puisque notre reconstruction euclidienne est définie à une similitude près, fixer cette dernière permet de définir 7 des 12 paramètres de la transformation affine recherchée. Ces 7 paramètres correspondent aux 7 degrés de liberté d'une similitude de l'espace tridimensionnel. Une façon simple de les fixer consiste à choisir une orientation du repère euclidien dans lequel sera définie la reconstruction, et un facteur d'échelle. Pour ce faire, il suffit de prendre parmi les points de la scène un point comme origine (ce point aura les coordonnées non homogènes $(0, 0, 0)$), un autre point comme définissant un des trois axes du repère euclidien (par exemple l'axe des X et ainsi ce point aura les coordonnées non homogènes $(x_1, 0, 0)$) et enfin, un troisième point est choisi pour fixer un des trois plans du repère euclidien (par exemple le plan OXY , ce troisième point aura les coordonnées non homogènes $(x_2, y_2, 0)$). Le facteur d'échelle est défini en fixant une distance quelconque de la scène; le plus simple, est de considérer $x_1 = 1$.

A présent, nous avons défini 7 des 12 paramètres de la transformation affine recherchée. Il reste exactement 5 paramètres à déterminer. Nous retrouvons ainsi le nombre de paramètres de *la conique absolue* qui définit la géométrie euclidienne (voir paragraphe 1.1.5).

Il y a donc une équivalence entre le calcul de *la conique absolue* et celui des 5 paramètres qui restent à déterminer de la transformation affine recherchée.

La géométrie euclidienne possède la notion d'angle, notion absente en géométrie affine. Il est alors nécessaire d'introduire cette notion pour passer de la géométrie affine à la géométrie euclidienne; ou d'une manière équivalente pour passer d'une reconstruction affine à une reconstruction euclidienne. Puisque nous sommes en déficit de 5 paramètres, au moins 5 connaissances euclidiennes sont à introduire.

Dans le paragraphe qui suit, nous allons voir comment traduire la connaissance d'un angle en contrainte sur la *conique absolue* et sur la transformation affine recherchée. Il est à noter que l'angle est l'invariant de base de la géométrie euclidienne. Toute autre connaissance euclidienne peut être traduite en contrainte sur la *conique absolue* ou sur la transformation affine. La section 4.2 présente quelques exemples de connaissances euclidiennes utilisées en pratique pour calculer W .

Contraintes liées aux angles

La connaissance de la valeur d'un angle entre deux segments de droites coplanaires donne une contrainte sur *la conique absolue*. Cette contrainte est donnée par la formule de Laguerre (1.4) (voir aussi la figure 1.3 pour plus de détails).

D'un autre côté, la connaissance de la valeur d'un angle entre deux segments de droites coplanaires donne une contrainte sur la transformation affine. Pour sim-

plifier, soient A , B et C 3 points coplanaires de l'espace avec $\alpha = (\vec{AB}, \vec{AC})$ connu et soit W la transformation affine recherchée :

$$W = \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Pour éviter toute confusion, les points sont notés avec des indices pour indiquer s'ils sont donnés par leurs coordonnées affines ou par leurs coordonnées euclidiennes. Par exemple, A_a représente le point physique A avec ses coordonnées affines tandis que A_e représente le même point avec ses coordonnées euclidiennes. Il faut noter que les coordonnées affines sont connues puisque la reconstruction affine est connue à cette étape. Avec ces notations nous avons l'égalité :

$$A_e = W A_a \text{ ou } {}^t A_e = {}^t A_a {}^t W \quad (4.3)$$

La connaissance de l'angle α est liée au produit scalaire $\vec{AB} \cdot \vec{AC}$ par la relation suivante :

$$A_e \vec{B}_e \cdot A_e \vec{C}_e = |A_e B_e| |A_e C_e| \cos \alpha \quad (4.4)$$

où \cdot dénote le produit scalaire de deux vecteurs.

En utilisant la relation (4.3), le produit scalaire (4.4) s'écrit :

$$A_e \vec{B}_e \cdot A_e \vec{C}_e = |W A_a \ W B_a| |W A_a \ W C_a| \cos \alpha \quad (4.5)$$

avec $|W A_a \ W B_a|$ le module du vecteur défini par les deux points : $W A_a$ et $W B_a$.

En combinant l'égalité (4.3) avec la relation (4.5) nous obtenons une équation non linéaire de degré 4 en les éléments de la matrice W . Nous avons donc obtenu une contrainte sur la transformation affine recherchée.

4.2 D'une reconstruction projective à une reconstruction euclidienne : aspect pratique

Pour les cas réels, il est souvent difficile d'utiliser les étapes précédentes pour passer d'une reconstruction projective à une reconstruction euclidienne. En particulier, il est très difficile de faire le passage de la reconstruction affine à la reconstruction euclidienne en utilisant les contraintes liées aux angles données par les équations (4.5) dont le degré est 4.

Cependant, l'étude théorique du passage projectif-euclidien nous permet de comprendre quelles sont les informations nécessaires pour obtenir une reconstruction euclidienne à partir d'une reconstruction projective ; ce qui permet de vérifier

que les contraintes utilisées pour calculer la collinéation de passage W contiennent toutes les informations requises.

Supposons, par exemple, qu'en plus d'une reconstruction projective, nous avons les informations suivantes : un ensemble de points appartenant à un plan (appelons ce plan Π), un faisceau de droites parallèles ; de plus nous savons que ces droites sont perpendiculaires au plan Π . Il est clair qu'avec ces informations nous ne pouvons même pas passer à une reconstruction affine. En effet, la coplanarité des points est une information projective, elle ne nous apporte rien de plus ; savoir que plusieurs droites sont parallèles à une même direction nous donne une seule information affine (un seul point à l'infini) et enfin, savoir que ces mêmes droites sont perpendiculaires à un plan nous donne une seule information euclidienne contraignant *la conique absolue*. Le plan Π_∞ n'est pas défini. Par conséquent ces informations sont insuffisantes pour passer du projectif à l'euclidien. Nous sommes en déficit de 2 informations affines (pour fixer le plan de l'infini) et de 4 informations euclidiennes (pour fixer *la conique absolue*).

Si nous considérons par exemple la scène donnée par la figure 4.4 : les trois directions principales peuvent être calculées ; de plus, elles sont orthogonales. Ces trois directions dont chacune est définie par un faisceau de droites parallèles permettent de trouver trois points de fuite, et ainsi le calcul du plan de l'infini est possible. L'orthogonalité de ces trois directions (connaissance de trois angles) apporte trois contraintes sur *la conique absolue*. Cette dernière n'est pas encore définie, deux autres contraintes euclidiennes sont nécessaires.

4.2.1 Contraintes euclidiennes utilisées en pratique

Ce paragraphe montre comment utiliser en pratique certaines connaissances a priori sur la scène pour dériver des contraintes sur la matrice W . Ces connaissances sont de nature géométrique, et ne sont pas exclusivement affines ou euclidiennes : elles peuvent être affines et euclidiennes en même temps. Nous allons donc traiter le passage d'une reconstruction projective à une reconstruction euclidienne en une seule étape en contraignant globalement la collinéation W .

Le problème, ici, consiste à transformer une reconstruction projective en une reconstruction euclidienne définie à une similitude près (hypothèse de départ). En d'autres termes, nous sommes libre de choisir un repère euclidien quelconque et éventuellement un facteur d'échelle. Par conséquent, nous pouvons choisir un repère lié à notre scène. Par exemple, pour notre application sur la scène de la maison (voir figure 4.4), le sol forme le plan OXY (plan d'équation $Z = 0$) et les deux murs verticaux et perpendiculaires entre eux sont utilisés comme plan OXZ et OYZ .

Un tel choix peut paraître assez restrictif. Cependant, nous pensons que dans beaucoup de scènes d'intérieur ou d'extérieur la présence d'un plan horizontal est souvent une conséquence naturelle liée à la présence du sol. De plus, deux plans verticaux et perpendiculaires entre eux peuvent être facilement trouvés (murs d'un

immeuble, murs d'un bureau...). De tels murs seront utilisés comme plan OXZ et OYZ .

Dans ce qui va suivre, nous devons supposer qu'un repère euclidien est déjà défini, mais les coordonnées tridimensionnelles des points restent inconnues.

Nous utilisons les notations suivantes : $A' = (x_{A'}, y_{A'}, z_{A'}, t_{A'})$ est un point donné par ses coordonnées projectives (avant l'utilisation des contraintes euclidiennes), et $WA' \sim A = (x_A, y_A, z_A, 1)$ le même point physique donné par ses coordonnées euclidiennes. Avec (w_{ij}) les éléments de la matrice 4×4 qui représente la collinéation de passage projectif-euclidien à déterminer.

Traduction de connaissances en contraintes sur W

Nous présentons dans ce paragraphe quelques exemples illustrant l'obtention des contraintes à partir de connaissances géométriques sur la scène.

- *Connaissance des coordonnées 3D d'un point* : quand les coordonnées euclidiennes d'un point A sont connues, par exemple en choisissant un point comme origine, ceci donne 3 contraintes linéaires sur W . Cela revient à écrire les équations résultant de l'égalité projective $WA' \sim A$ (on rappelle ici que \sim est l'égalité définie à un facteur multiplicatif près).

Pour obtenir les égalités usuelles, il faut passer par les rapports.

$$\begin{cases} x_A = \frac{w_{11}x_{A'} + w_{12}y_{A'} + w_{13}z_{A'} + w_{14}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} \\ y_A = \frac{w_{21}x_{A'} + w_{22}y_{A'} + w_{23}z_{A'} + w_{24}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} \\ z_A = \frac{w_{31}x_{A'} + w_{32}y_{A'} + w_{33}z_{A'} + w_{34}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} \end{cases} \quad (4.6)$$

Les coordonnées euclidiennes d'un point de l'espace sont les connaissances les plus difficiles à obtenir; cependant, les contraintes qui en résultent sont les plus simples.

- *Un point appartenant à un des trois plans du repère* : quand un point A de la scène se trouve sur un des trois plans du repère orthonormé choisi (OXY , OXZ et OYZ sont les 3 plans du repère), une de ses coordonnées est alors nulle. Par exemple, si A appartient au plan horizontal (plan OXY), ceci se traduit par $z_A = 0$, ce qui donne une contrainte sur W :

$$z_A = \frac{w_{31}x_{A'} + w_{32}y_{A'} + w_{33}z_{A'} + w_{34}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} = 0 \quad (4.7)$$

- *Deux points formant un segment parallèle avec un axe du repère* : quand un segment de droite AB est parallèle à un des 3 axes du repère orthonormé,

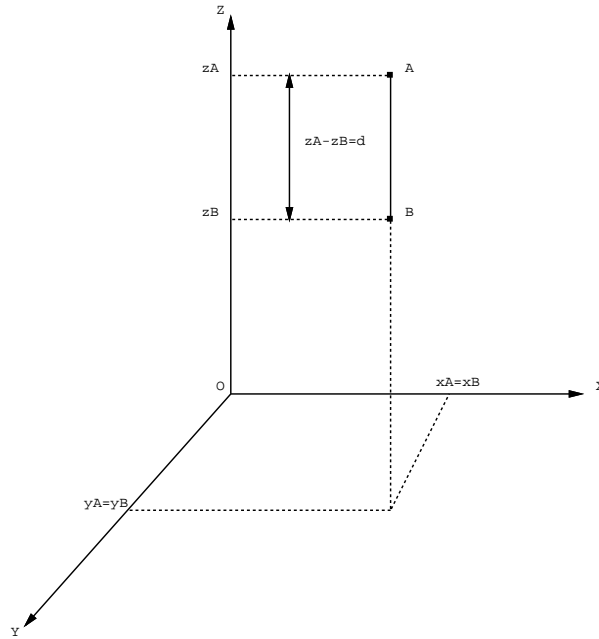


FIG. 4.3 - Le segment AB est vertical (parallèle à l'axe OZ).

il en résulte que les deux points ont deux de leurs coordonnées euclidiennes égales.

Par exemple, soient deux points A et B tels que le segment AB soit vertical, comme le montre la figure 4.3 (AB est perpendiculaire au sol). Sans que les coordonnées euclidiennes des deux points A et B ne soient connues, nous avons les égalités suivantes :

$$x_A = x_B \quad \text{et} \quad y_A = y_B$$

Ce qui donne deux contraintes non linéaires sur W :

$$\begin{cases} \frac{w_{11}x_{A'} + w_{12}y_{A'} + w_{13}z_{A'} + w_{14}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} = \frac{w_{11}x_{B'} + w_{12}y_{B'} + w_{13}z_{B'} + w_{14}t_{B'}}{w_{41}x_{B'} + w_{42}y_{B'} + w_{43}z_{B'} + w_{44}t_{B'}} \\ \frac{w_{21}x_{A'} + w_{22}y_{A'} + w_{23}z_{A'} + w_{24}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} = \frac{w_{21}x_{B'} + w_{22}y_{B'} + w_{23}z_{B'} + w_{24}t_{B'}}{w_{41}x_{B'} + w_{42}y_{B'} + w_{43}z_{B'} + w_{44}t_{B'}} \end{cases} \quad (4.8)$$

- *Distance entre deux points* : la connaissance d'une distance entre deux points de la scène entraîne une contrainte non linéaire. Soient deux points A et B telle que la distance du segment AB est connue et vaut d . Dans un espace euclidien ceci se traduit par l'équation :

$$(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 = d^2$$

En remplaçant dans l'équation ci-dessus les coordonnées euclidiennes en fonction de leurs coordonnées projectives et de W , nous obtenons l'équation non

linéaire suivante :

$$\begin{aligned}
& \left(\frac{w_{11}x_{A'}+w_{12}y_{A'}+w_{13}z_{A'}+w_{14}t_{A'}}{w_{41}x_{A'}+w_{42}y_{A'}+w_{43}z_{A'}+w_{44}t_{A'}} - \frac{w_{11}x_{B'}+w_{12}y_{B'}+w_{13}z_{B'}+w_{14}t_{B'}}{w_{41}x_{B'}+w_{42}y_{B'}+w_{43}z_{B'}+w_{44}t_{B'}} \right)^2 + \\
& \left(\frac{w_{21}x_{A'}+w_{22}y_{A'}+w_{23}z_{A'}+w_{24}t_{A'}}{w_{41}x_{A'}+w_{42}y_{A'}+w_{43}z_{A'}+w_{44}t_{A'}} - \frac{w_{21}x_{B'}+w_{22}y_{B'}+w_{23}z_{B'}+w_{24}t_{B'}}{w_{41}x_{B'}+w_{42}y_{B'}+w_{43}z_{B'}+w_{44}t_{B'}} \right)^2 + \\
& \left(\frac{w_{31}x_{A'}+w_{32}y_{A'}+w_{33}z_{A'}+w_{34}t_{A'}}{w_{41}x_{A'}+w_{42}y_{A'}+w_{43}z_{A'}+w_{44}t_{A'}} - \frac{w_{31}x_{B'}+w_{32}y_{B'}+w_{33}z_{B'}+w_{34}t_{B'}}{w_{41}x_{B'}+w_{42}y_{B'}+w_{43}z_{B'}+w_{44}t_{B'}} \right)^2 = d^2
\end{aligned} \tag{4.9}$$

- *Deux vecteurs orthogonaux* : si 2 vecteurs \vec{AB} et \vec{CD} sont orthogonaux alors leur produit scalaire est nul. En utilisant les coordonnées euclidiennes, le produit scalaire $\vec{AB} \cdot \vec{CD}$ s'écrit :

$$(x_B - x_A)(x_D - x_C) + (y_B - y_A)(y_D - y_C) + (z_B - z_A)(z_D - z_C) = 0$$

En remplaçant dans l'équation ci-dessus les coordonnées euclidiennes en fonction de leurs coordonnées projectives et de W , nous obtenons l'équation non linéaire suivante :

$$\begin{aligned}
& \left(\frac{w_{11}x_{A'}+w_{12}y_{A'}+w_{13}z_{A'}+w_{14}t_{A'}}{w_{41}x_{A'}+w_{42}y_{A'}+w_{43}z_{A'}+w_{44}t_{A'}} - \frac{w_{11}x_{B'}+w_{12}y_{B'}+w_{13}z_{B'}+w_{14}t_{B'}}{w_{41}x_{B'}+w_{42}y_{B'}+w_{43}z_{B'}+w_{44}t_{B'}} \right) \\
& \left(\frac{w_{11}x_{C'}+w_{12}y_{C'}+w_{13}z_{C'}+w_{14}t_{C'}}{w_{41}x_{C'}+w_{42}y_{C'}+w_{43}z_{C'}+w_{44}t_{C'}} - \frac{w_{11}x_{D'}+w_{12}y_{D'}+w_{13}z_{D'}+w_{14}t_{D'}}{w_{41}x_{D'}+w_{42}y_{D'}+w_{43}z_{D'}+w_{44}t_{D'}} \right) \\
& + \\
& \left(\frac{w_{21}x_{A'}+w_{22}y_{A'}+w_{23}z_{A'}+w_{24}t_{A'}}{w_{41}x_{A'}+w_{42}y_{A'}+w_{43}z_{A'}+w_{44}t_{A'}} - \frac{w_{21}x_{B'}+w_{22}y_{B'}+w_{23}z_{B'}+w_{24}t_{B'}}{w_{41}x_{B'}+w_{42}y_{B'}+w_{43}z_{B'}+w_{44}t_{B'}} \right) \\
& \left(\frac{w_{21}x_{C'}+w_{22}y_{C'}+w_{23}z_{C'}+w_{24}t_{C'}}{w_{41}x_{C'}+w_{42}y_{C'}+w_{43}z_{C'}+w_{44}t_{C'}} - \frac{w_{21}x_{D'}+w_{22}y_{D'}+w_{23}z_{D'}+w_{24}t_{D'}}{w_{41}x_{D'}+w_{42}y_{D'}+w_{43}z_{D'}+w_{44}t_{D'}} \right) \\
& + \\
& \left(\frac{w_{31}x_{A'}+w_{32}y_{A'}+w_{33}z_{A'}+w_{34}t_{A'}}{w_{41}x_{A'}+w_{42}y_{A'}+w_{43}z_{A'}+w_{44}t_{A'}} - \frac{w_{31}x_{B'}+w_{32}y_{B'}+w_{33}z_{B'}+w_{34}t_{B'}}{w_{41}x_{B'}+w_{42}y_{B'}+w_{43}z_{B'}+w_{44}t_{B'}} \right) \\
& \left(\frac{w_{31}x_{C'}+w_{32}y_{C'}+w_{33}z_{C'}+w_{34}t_{C'}}{w_{41}x_{C'}+w_{42}y_{C'}+w_{43}z_{C'}+w_{44}t_{C'}} - \frac{w_{31}x_{D'}+w_{32}y_{D'}+w_{33}z_{D'}+w_{34}t_{D'}}{w_{41}x_{D'}+w_{42}y_{D'}+w_{43}z_{D'}+w_{44}t_{D'}} \right) = 0
\end{aligned} \tag{4.10}$$

L'équation (4.10) est hautement non linéaire. Cependant, dans certains cas particuliers elle peut devenir plus simple. Par exemple si nous savons que le segment AB est vertical l'équation (4.10) se réduit à la forme :

$$\begin{aligned}
& \left(\frac{w_{31}x_{A'}+w_{32}y_{A'}+w_{33}z_{A'}+w_{34}t_{A'}}{w_{41}x_{A'}+w_{42}y_{A'}+w_{43}z_{A'}+w_{44}t_{A'}} - \frac{w_{31}x_{B'}+w_{32}y_{B'}+w_{33}z_{B'}+w_{34}t_{B'}}{w_{41}x_{B'}+w_{42}y_{B'}+w_{43}z_{B'}+w_{44}t_{B'}} \right) \cdot \\
& \left(\frac{w_{31}x_{C'}+w_{32}y_{C'}+w_{33}z_{C'}+w_{34}t_{C'}}{w_{41}x_{C'}+w_{42}y_{C'}+w_{43}z_{C'}+w_{44}t_{C'}} - \frac{w_{31}x_{D'}+w_{32}y_{D'}+w_{33}z_{D'}+w_{34}t_{D'}}{w_{41}x_{D'}+w_{42}y_{D'}+w_{43}z_{D'}+w_{44}t_{D'}} \right) = 0
\end{aligned}$$

D'autres contraintes sur W peuvent être trouvées, par exemple celles déjà présentées dans la section 4.1, l'égalité de deux distances etc. En général, toute connaissance euclidienne ou affine peut se traduire en une ou plusieurs contraintes sur W . Cependant, certaines contraintes peuvent être si compliquées que leur utilisation pratique devient impossible (degré des équations trop élevé).

4.2.2 Résolution numérique

En vérifiant que les contraintes utilisées sont suffisantes pour passer d'une reconstruction projective à une reconstruction euclidienne, la matrice W peut être estimée. Les connaissances sur la scène se traduisent en contraintes sous forme d'équations liant les éléments de W . Les équations sont soit linéaires (4.6)(4.7) soit non linéaires (4.8)(4.9)(4.10); la résolution d'un tel système d'équations peut être faite en utilisant une minimisation au sens des moindres carrés. L'algorithme de Levenberg-Marquardt [PFTW88] est bien adapté à notre problème, comme dans le cas de la reconstruction non linéaire. Cependant, ici, le problème est plus simple, le nombre de paramètres à estimer étant de 15 seulement (les éléments de W).

Le problème peut s'écrire comme une minimisation de la somme

$$\sum_{k=1}^n \left(\frac{f_k(W_{11}, \dots, W_{44})}{\sigma_k} \right)^2 \quad (4.11)$$

où $f_k(\cdot)$ sont les équations obtenues par traduction des connaissances et les σ_k^2 sont les variances sur les contraintes.

Comme toute méthode de résolution d'équations non linéaires, l'algorithme de Levenberg-Marquardt nécessite des valeurs initiales pour les inconnues. Bien que pour toutes nos expériences avec des scènes réelles et des scènes simulées la convergence a été obtenue sans estimation préalable de W , l'estimation de cette dernière reste le seul garant pour la convergence de l'algorithme de Levenberg-Marquardt.

Obtenir une première estimation pour W revient à obtenir une première estimation de la structure euclidienne de la scène. Cette dernière peut être obtenue avec la méthode de reconstruction sous l'hypothèse d'un modèle de perspective faible pour la caméra (voir chapitre 3). La reconstruction euclidienne ainsi calculée permet de trouver une première estimation de W en résolvant un système d'équations linéaires du type (4.6).

4.3 Expérimentations

Ce paragraphe contient deux types de résultats. Le premier, avec une scène réelle, a pour but de valider la méthode proposée; le second, avec une scène simulée, vise à étudier la précision de la reconstruction euclidienne calculée. Une

comparaison de la reconstruction euclidienne utilisant 5 points connus de la scène avec celle utilisant des contraintes géométriques est résumée sous forme de courbe.

La reconstruction euclidienne utilisant des contraintes géométriques est obtenue en deux étapes :

1. calcul de la reconstruction projective (section 3.3) : le problème de la reconstruction est spécifié par les équations (3.14). Il faut en outre définir une base de 5 points pour pouvoir reconstruire. Dans ce cas nous n'avons pas les coordonnées euclidiennes des points de la base. Nous avons utilisé 5 points de la scène (non quatre à quatre coplanaires) avec les coordonnées projectives : $(0, 0, 0, 1)$, $(1, 0, 0, 1)$, $(0, 1, 0, 1)$, $(0, 0, 1, 1)$ et $(1, 1, 1, 1)$. La reconstruction obtenue en minimisant la somme (3.15) est définie dans cette dernière base ; c'est une reconstruction projective.
2. calcul de W (paragraphe 4.2.2) : des contraintes géométriques sont utilisées pour calculer W . Cette dernière est une collinéation qui transforme la reconstruction projective en une reconstruction euclidienne. Les coordonnées euclidiennes des points de la scène sont obtenues en calculant pour chaque point de la scène de coordonnées projectives (x', y', z', t') ses coordonnées euclidiennes $(x, y, z, 1)$ données par ${}^t(x, y, z, 1) \sim W \cdot {}^t(x', y', z', t')$.

Résultats

- Scène réelle (figure 4.4) : 3 images sont utilisées dans cette expérience. 49 points sont extraits et mis en correspondance dans les trois images. En utilisant une base projective (voir figure 4.4), la reconstruction projective est alors calculée.

Ensuite, pour estimer la collinéation W , un repère euclidien doit être choisi dans la scène. Dans ce dernier sont exprimées les contraintes géométriques utilisées pour le calcul de W . Le plus simple est d'utiliser le repère naturel formé par les deux murs perpendiculaires et le sol. Les axes de ce repère sont les intersections des trois plans (les deux murs et le sol).

Dans ce repère où aucune coordonnée n'est connue, nous avons utilisé les contraintes géométriques suivantes :

1. Deux points qui forment un segment parallèle à un des trois axes du repère euclidien entraîne des contraintes du type (4.8). Un total de douze couples de points est utilisé.
Ces contraintes contiennent l'information affine de parallélisme ; elles permettent de fixer le plan de l'infini.
2. Une distance entre deux points entraîne une contrainte du type (4.9). Nous avons utilisé 12 distances. Ces contraintes sont de type euclidienne

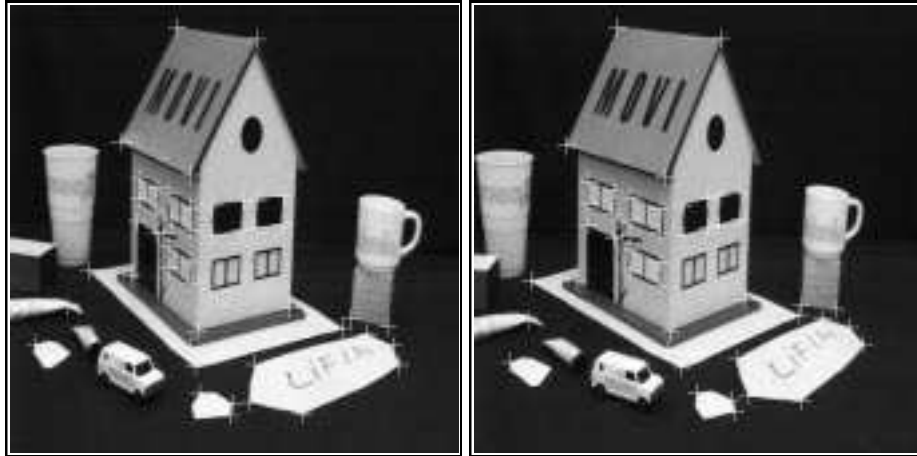


image 1

image 2



image 3

FIG. 4.4 - Scène réelle : les trois images utilisées pour la reconstruction projective avec les points d'intérêts détectés et mis en correspondance. Les 5 points utilisés comme base projective sont marqués par des gros points blancs dans la troisième image

coordonnées 3D exactes (cm)			coordonnées calculées (cm)			les erreurs (cm)		
X	Y	Z	X_c	Y_c	Z_c	$X - X_c$	$Y - Y_c$	$Z - Z_c$
0,000	0,000	0,500	0,000	0,030	0,478	0,000	0,030	0,022
12,000	0,000	0,500	12,041	0,059	0,504	0,041	0,059	0,004
0,000	0,000	14,500	0,000	0,025	14,458	0,000	0,025	0,042
6,000	18,000	22,500	5,946	17,982	22,431	0,054	0,018	0,069
-1,500	-1,500	0,500	-1,511	-1,526	0,523	0,011	0,026	0,023
-1,500	19,500	0,500	-1,462	19,501	0,505	0,038	0,001	0,005
0,000	6,000	0,500	0,000	6,007	0,554	0,000	0,007	0,054
7,000	0,000	2,500	7,143	0,070	2,440	0,143	0,070	0,060
10,000	0,000	6,500	10,164	0,095	6,466	0,164	0,095	0,034
21,000	-8,000	8,500	21,086	-8,040	8,500	0,086	0,040	0,000

TAB. 4.1 - Les coordonnées reconstruites avec les erreurs correspondantes pour 10 points parmi les 60 points de la scène simulée.

(la distance est une notion euclidienne). Elles sont suffisantes pour fixer la conique absolue.

- Un point qui appartient à un des trois plans du repère que nous avons choisi entraîne une contrainte linéaire du type (4.7) sur les éléments de W . Douze points ont été utilisés dans ce cas, trois pour chaque plan. Ces dernières contraintes fixent le repère euclidien choisi.

Ces contraintes géométriques donnent un total de 48 équations sur les 15 inconnues de W . Une fois W calculée, la structure projective est alors transformée en une structure euclidienne. La figure 4.5 donne le résultat de la reconstruction de 49 points de la scène réelle.

- Scène simulée : la scène simulée comporte 60 points dont la forme géométrique est similaire à la scène réelle avec un volume de $30cm \times 40cm \times 30cm$ (voir figure 4.6). Quatre images de cette scène sont utilisées pour la reconstruction projective. Les coordonnées images ont été perturbées avec un bruit uniforme d'amplitude 1 pixel. Les mêmes types et le même nombre de contraintes que pour le cas de la scène réelle sont utilisés. Qualitativement, le résultat est parfait (voir figure 4.6).

Le tableau 4.1 donne pour 10 points les erreurs sur les coordonnées euclidiennes calculées.

Un deuxième type d'expérience avec des données simulées consiste à comparer la reconstruction euclidienne utilisant les contraintes géométriques avec celle utilisant 5 points connus de la scène.

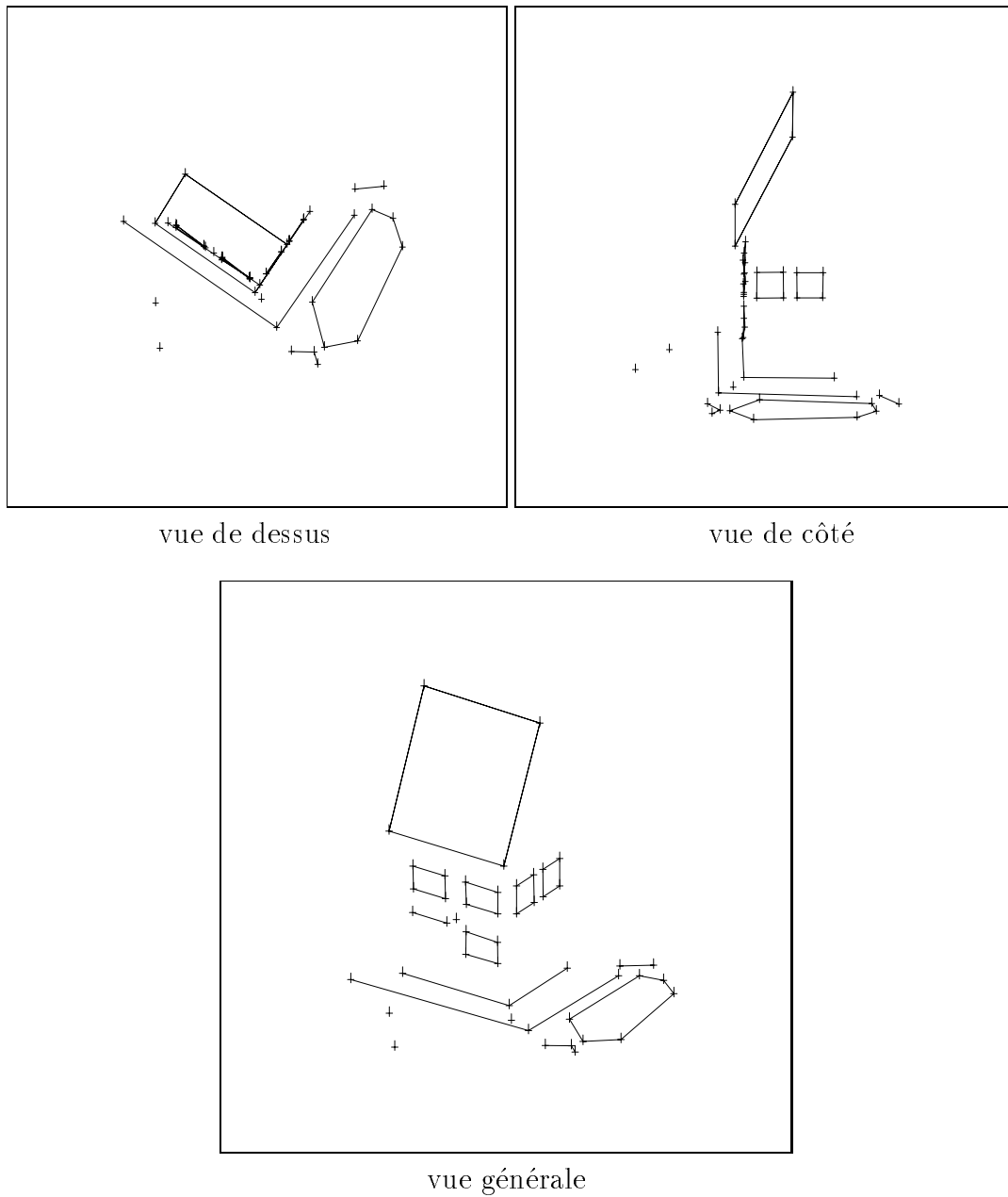


FIG. 4.5 - *Reconstruction euclidienne des points de la scène réelle en utilisant les contraintes géométriques. Nous avons relié les points par des segments de droite pour mieux illustrer le résultat*

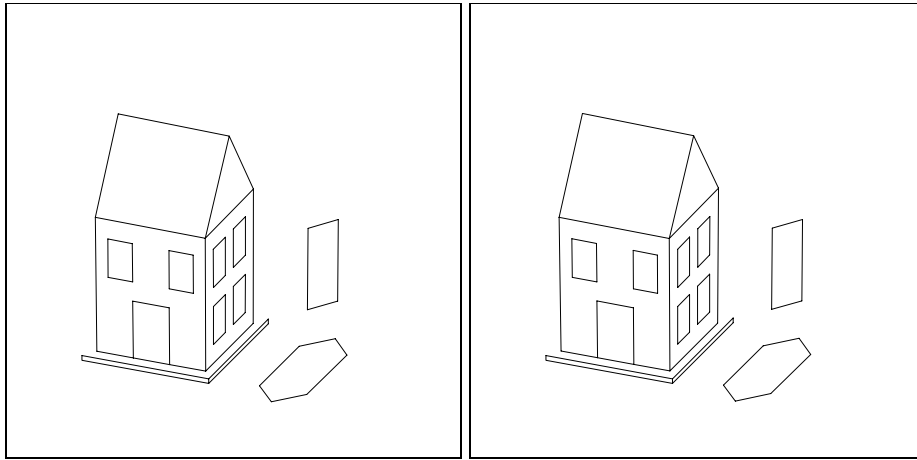


FIG. 4.6 - *La scène simulée : à gauche la scène exacte et à droite la même scène reconstruite en utilisant les contraintes géométriques*

En utilisant la même scène (figure 4.6), nous avons perturbé les coordonnées images avec un bruit (uniforme) de différentes amplitudes. Ensuite, nous avons calculé les coordonnées euclidiennes de la scène avec les deux méthodes : celle utilisant 5 points avec des coordonnées euclidiennes et permettant d'obtenir directement la reconstruction euclidienne et celle utilisant des contraintes géométriques. La figure 4.7 donne l'erreur moyenne sur les positions tridimensionnelles en fonction de l'amplitude du bruit ajouté. Sans bruit les deux méthodes donnent une reconstruction parfaite. Mais dès que l'amplitude du bruit augmente, l'utilisation des contraintes euclidiennes redondantes prend l'avantage au niveau de la précision. Cependant, il faut noter que les deux méthodes n'utilisent pas les mêmes informations euclidiennes ou affines. En particulier, la méthode utilisant 5 points connus de la scène est sensible à leur structure géométrique. Par exemple si les 5 points connus occupent une petite zone de la scène la qualité de la reconstruction se dégrade.

Pour la méthode utilisant les contraintes géométriques, nous avons plus de liberté pour choisir les 5 points de la base projective, car la connaissance de leurs coordonnées euclidiennes n'est pas nécessaire. En pratique, nous choisissons cinq points entourant la scène comme sur la figure 4.4. Un autre avantage pour la deuxième méthode est la facilité d'obtenir des contraintes géométriques. Il est plus facile d'avoir des informations telles que : distance, parallélisme... les coordonnées euclidiennes des points étant difficiles à obtenir. Par exemple, la taille d'une fenêtre sur le mur d'une maison peut être connue plus facilement que les coordonnées euclidiennes des coins de cette fenêtre.

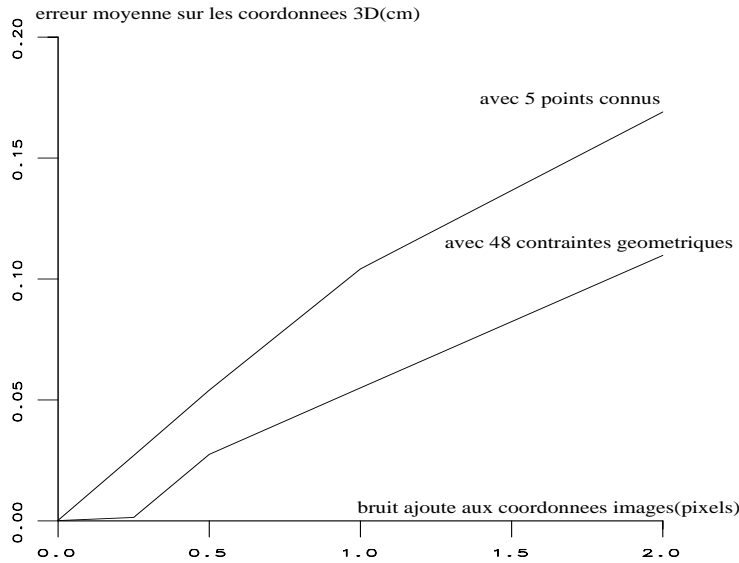


FIG. 4.7 - Chaque graphe donne pour chacune des deux méthodes les erreurs de la reconstruction en fonction du bruit (uniforme) ajouté aux coordonnées dans les images

4.4 Conclusion

Ce chapitre montre qu'il est possible d'obtenir une reconstruction euclidienne à partir d'une reconstruction projective étant donné un ensemble de connaissances géométriques a priori.

Pour faire ce travail, nous avons été motivé par le fait que la reconstruction projective peut être obtenue uniquement à partir des points mis en correspondance dans les images (chapitre 3).

Cependant, la reconstruction projective est pauvre en information (voir figure 4.1). La reconstruction la plus riche en information et la plus utilisée en vision et en robotique est de type euclidien. Pour une scène donnée, la reconstruction projective est liée à la reconstruction euclidienne par une collinéation de l'espace. Cette dernière peut être simplement déterminée si les positions euclidiennes d'au moins 5 points de la scène dans un repère euclidien sont données. Cependant, il n'est pas toujours possible de connaître de telles informations. Nous avons donc proposé une alternative permettant de se passer des informations de type position (coordonnées euclidiennes d'un point) et d'utiliser des informations plus faciles à obtenir.

La première partie de ce chapitre se consacre à l'étude théorique concernant le passage d'une reconstruction projective à une reconstruction euclidienne. Ce passage est décrit par une matrice inversible de taille 4×4 qui représente une col-

linéation de l'espace.

La deuxième partie du chapitre concerne les aspects pratiques de ce passage. Nous avons montré comment utiliser des connaissances géométriques a priori sur la scène pour les traduire en contraintes sur la collinéation recherchée. Ces contraintes s'écrivent sous forme d'équations dont les inconnues sont les paramètres de la collinéation recherchée; le problème devient alors un problème d'optimisation non linéaire résolu au sens des moindres carrés.

Les résultats expérimentaux valident l'approche proposée. De plus, par rapport à la reconstruction utilisant 5 points connus de l'espace, la reconstruction avec des contraintes géométriques redondantes est moins sensible au bruit.

Chapitre 5

Calcul du mouvement et reconstruction

Dans la section 3.3, la reconstruction est calculée à partir d'images issues d'une ou de plusieurs caméras non étalonnées. En particulier, les paramètres intrinsèques sont supposés différents pour chaque caméra.

Avec cette formulation, le problème de la reconstruction comporte deux types de paramètres : les positions des points dans l'espace et les matrices de projection dont chacune possède 11 paramètres. Ces derniers correspondent aux paramètres extrinsèques et intrinsèques d'une caméra. Les paramètres extrinsèques sont au nombre de 6 ; ils correspondent à la position et à l'orientation d'une caméra par rapport à un repère de référence, le repère de la scène par exemple.

Rappelons qu'à une caméra est attaché un repère (voir figure 1.5) ; l'orientation d'une caméra n'est autre que l'orientation de son repère (rotation). Les paramètres intrinsèques sont au nombre de 5 dont 4 seulement sont pris en compte généralement (voir paragraphe 1.2.1).

Quand les images sont issues d'une ou de plusieurs caméras dont les paramètres intrinsèques sont connus, le problème de reconstruction se voit réduire le nombre de ses paramètres. En effet, à chaque nouvelle image le nombre des paramètres du problème de reconstruction augmente de 6 (paramètres extrinsèques) au lieu de 11 (paramètres extrinsèques et intrinsèques).

Ce chapitre traite ce cas particulier, cependant, nous ne supposons pas la connaissance exacte des paramètres intrinsèques mais seulement une approximation de ceux-ci. Dans ce chapitre nous adoptons le modèle sténopé pour la caméra et nous ne considérons que le cas de la structure euclidienne. Nous supposons aussi que les mises en correspondance ont déjà été établies (voir section 2.2).

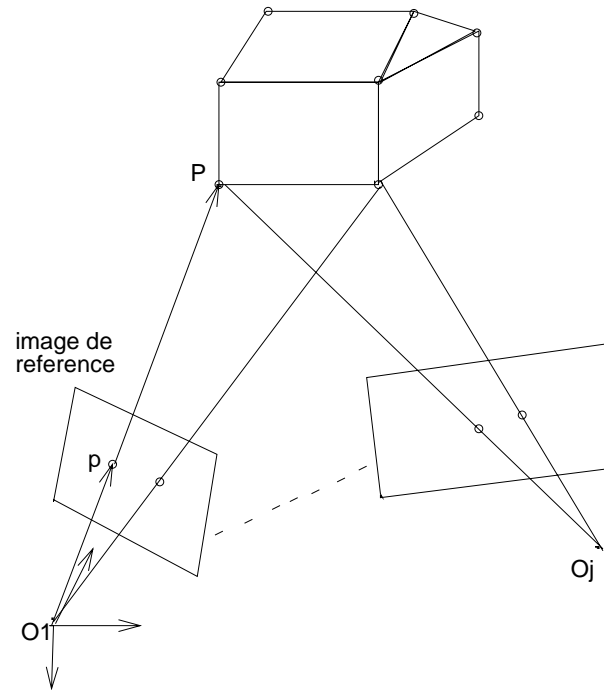


FIG. 5.1 - Le repère de la première caméra est utilisé comme repère de référence, nous avons la relation: $\vec{O_1P} = \lambda \vec{O_1p}$, (λ : nombre réel strictement positif) pour tout point se trouvant devant la caméra

5.1 Principes des méthodes classiques

Les approches actuelles pour résoudre le problème de la reconstruction à partir d'images non étalonnées peuvent être classées en deux catégories : celles qui calculent la structure définie dans un repère relatif, généralement attaché à la scène (chapitre 3) et celles qui calculent la structure définie dans le repère attaché à une caméra utilisée comme référence (voir figure 5.1). Dans ce dernier cas, l'information calculée est appelée *profondeur*.

Nous nous intéressons dans ce chapitre à cette deuxième catégorie de méthodes où la position et l'orientation de chaque caméra sont définies relativement au repère d'une caméra choisie comme référence. Ces méthodes sont en majorité basées sur la matrice essentielle due à Longuet-Higgins [LH81]. Nous rappelons donc cette notion ici.

Matrice essentielle

Quand une caméra effectue un mouvement (déplacement) entre deux positions, celui-ci peut être décomposé en deux composantes : une rotation et une translation. Ou d'une manière équivalente, les deux repères de deux caméras différentes sont liés par un mouvement composé d'une rotation et d'une translation.

Pour simplifier, nous supposons l'existence de deux caméras que nous appelons première et deuxième caméra (voir figure 5.2).

Soient O et O' les deux origines des repères attachés à deux caméras, et soit R la rotation qui fait passer du repère de la première caméra au repère de la seconde (voir figure 5.2).

Supposons que les paramètres intrinsèques de la caméra sont connus, ce qui nous permet d'utiliser les coordonnées normalisées des points dans les images.

Soit P un point de l'espace qui se projette dans les deux images en q et q' respectivement, avec q et q' donnés par leurs coordonnées normalisées. Il est facile de voir depuis la figure 5.2 que les trois vecteurs $\vec{O'P}$, \vec{OP} et $\vec{O'O}$ sont coplanaires. Aussi, les trois vecteurs $\vec{O'q'}$, \vec{Oq} et $\vec{O'O}$ sont coplanaires si et seulement si q et q' sont les projections du même point de l'espace P . En d'autres termes, pour un couple d'appariements (q, q') on peut écrire :

$$\vec{O'q'} \cdot (\vec{O'O} \wedge \vec{Oq}) = 0 \quad (5.1)$$

où \cdot dénote le produit scalaire et \wedge le produit vectoriel.

Les vecteurs de la relation ci-dessus doivent être définis dans un même repère.

Si on choisit le repère attaché à la deuxième caméra pour écrire la relation (5.1), le vecteur \vec{Oq} connu dans le repère de la première caméra (coordonnées normalisées) a pour coordonnées dans la seconde caméra $R\vec{Oq}$, avec R la rotation qui fait passer du repère de la première caméra à celui de la deuxième caméra.

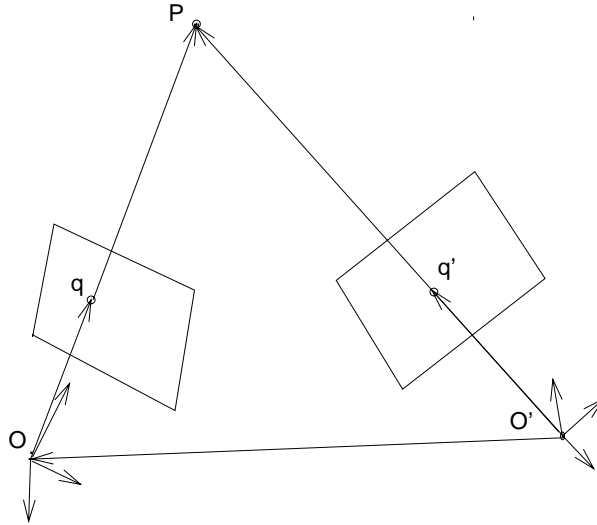


FIG. 5.2 - Deux caméras (ou une caméra en mouvement considérée à deux instants différent) observant la même scène : la contrainte de Longuet-Higgins stipule que les points O , O' , q et q' sont coplanaires.

D'où la nouvelle relation définie dans le repère attaché à la deuxième caméra :

$$O'\vec{q}' \cdot (O'\vec{O} \wedge R\vec{O}q) = 0$$

Notons qu'on peut définir la matrice antisymétrique V telle que $V\vec{x} = \vec{v} \wedge \vec{x}$ pour tout vecteur \vec{x} avec :

$$V = \begin{pmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{pmatrix} \quad (5.2)$$

où (v_x, v_y, v_z) sont les coordonnées du vecteur \vec{v} .

En notant ${}^tO'\vec{q}'$ par q' , $\vec{O}q$ par q et $O'\vec{O}$ par t nous obtenons :

$${}^tq'(t \wedge Rq) = 0 \quad (5.3)$$

Soit T la matrice antisymétrique associée à t selon (5.2). En posant $E = TR$ nous obtenons à partir de la relation 5.3 la contrainte de Longuet-Higgins :

$${}^tq'Eq = 0 \quad (5.4)$$

La matrice E , produit d'une matrice antisymétrique par une matrice de rotation, est appelée *matrice essentielle*.

Quand elle n'est pas nulle, cette matrice est de rang 2 et est définie à un facteur multiplicatif près. De plus, elle possède une autre propriété [HF89] : ses deux valeurs singulières non nulles sont égales.

La relation (5.4) est similaire à celle de la matrice fondamentale (2.1) à la différence qu'ici les coordonnées normalisées sont utilisées. Le lien entre la matrice fondamentale et la matrice essentielle a été clairement établi par Luong [LF92]. Soient p et p' les projections dans les deux images du point P de l'espace donnés par leurs coordonnées pixels et soit A la matrice contenant les paramètres intrinsèques (ici, ils sont supposés identiques pour les deux caméras). Nous avons les deux égalités $q = pA^{-1}$ et $q' = p'A^{-1}$. En remplaçant dans (5.4) nous obtenons :

$${}^t p' A^{-1} E A^{-1} p = 0$$

D'où le lien de la matrice fondamentale avec la matrice essentielle

$$F = {}^t A^{-1} E A^{-1}$$

Les méthodes classiques basées sur la matrice essentielle supposent la connaissance des paramètres intrinsèques de la caméra. Ces méthodes (voir par exemple [LH81] [TH84] [TF87] [WHA89]) calculent d'abord la matrice essentielle E puis décomposent celle-ci en un produit d'une matrice antisymétrique par une matrice de rotation.

Même si des méthodes plus récentes [Luo92] [Fau93] [WAH93] permettent de prendre en compte le bruit dans le calcul de E , la décomposition de celle-ci en un produit d'une matrice antisymétrique par une matrice de rotation reste problématique. En particulier, en pratique E ne vérifie pas les propriétés nécessaires et suffisantes d'une matrice essentielle (le rang et l'égalité des deux valeurs singulières non nulles). Ceci rend la décomposition de E difficile et très sensible aux erreurs, car elle comporte plusieurs minimisations non linéaires. Si de plus les valeurs paramètres intrinsèques utilisées pour obtenir les coordonnées normalisées sont peu précises, le résultat est voué à l'échec.

D'autres méthodes n'utilisant pas la matrice essentielle ont été proposées [AHP93] [SPFP93]. Ces dernières donnent de bons résultats selon leurs auteurs, mais elles sont récursives et nécessitent un grand nombre d'images, typiquement quelques dizaines.

Nous proposons ici une méthode qui n'est pas basée sur le calcul de la matrice essentielle et qui permet de calculer à la fois le mouvement et la structure. Dans la suite, nous supposons que les paramètres intrinsèques de la caméra ne changent pas d'une image à l'autre et sont approximativement connus.

5.2 Les équations du problème

Rappelons que dans le cas d'une caméra dont les paramètres intrinsèques restent inchangés, une matrice de projection perspective M_j (correspondant à la j ème

image) est définie par :

$$M_j = AID_j$$

avec

$$A = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$D_j = \begin{pmatrix} r_{11}^j & r_{12}^j & r_{13}^j & t_x^j \\ r_{21}^j & r_{22}^j & r_{23}^j & t_y^j \\ r_{31}^j & r_{32}^j & r_{33}^j & t_z^j \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Notons que si la première caméra est utilisée comme référence alors D_1 est une matrice identité et donc sa matrice de projection ne possède que 4 inconnues :

$$M_1 = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

En utilisant le modèle sténopé, un point de l'espace P_i se projette dans la j ème image en un point p_{ij} selon la relation :

$$p_{ij} = \mu M_j P_i \quad (5.5)$$

où M_j est la matrice de projection de la j ème image, P_i et p_{ij} sont donnés par leurs coordonnées homogènes respectives.

Les coordonnées non homogènes des points peuvent être utilisées, dans ce cas nous obtenons les équations usuelles qui résultent du modèle sténopé :

$$\begin{cases} x_{ij} = \frac{(\alpha_u r_{11}^j + u_0 r_{31}^j)X_i + (\alpha_u r_{12}^j + u_0 r_{32}^j)Y_i + (\alpha_u r_{13}^j + u_0 r_{33}^j)Z_i + \alpha_u t_x + u_0 t_z}{r_{31}^j X_i + r_{32}^j Y_i + r_{33}^j Z_i + t_z} \\ y_{ij} = \frac{(\alpha_v r_{21}^j + v_0 r_{31}^j)X_i + (\alpha_v r_{22}^j + v_0 r_{32}^j)Y_i + (\alpha_v r_{23}^j + v_0 r_{33}^j)Z_i + \alpha_v t_y + v_0 t_z}{r_{31}^j X_i + r_{32}^j Y_i + r_{33}^j Z_i + t_z} \end{cases} \quad (5.6)$$

(X_i, Y_i, Z_i) et (x_{ij}, y_{ij}) sont respectivement les coordonnées non homogènes du point P_i de l'espace et de sa projection p_{ij} dans la j ème image.

Les équations ci-dessus sont similaires à celles données dans (3.10) à la différence qu'ici nous avons utilisé l'égalité: $M_j = AID_j$.

Pour une scène de p points observés dans v images, nous avons un total de $2 \times p \times v$ équations. Quand une caméra est utilisée comme référence, le nombre des inconnues est de $3 \times p$ pour les points et $6 \times (v - 1)$ pour les paramètres extrinsèques. Notons que le facteur d'échelle ne peut être déterminé à cause de l'ambiguïté distance-taille. Nous avons donc un total de $3 \times p + 6 \times (v - 1) - 1$ inconnues.

En théorie, le problème ainsi posé admet une solution dès que le nombre d'équations égale ou dépasse le nombre d'inconnues, c'est à dire, dès que $2 \times p \times v \geq 3 \times p + 6 \times (v - 1) - 1$.

Si les paramètres intrinsèques sont inconnus alors une autre condition est à ajouter : au moins trois images sont nécessaires. Cette condition ($v \geq 3$) vient du fait que les paramètres intrinsèques d'une caméra ne peuvent être déterminés à partir de deux images seulement [May93][LF92]. Un couple d'images donne deux contraintes indépendantes sur les paramètres intrinsèques. Donc au moins trois images sont nécessaires.

Malheureusement, les équations (5.6) sont hautement non linéaires, même avec une bonne initialisation, nous n'avons pas pu obtenir des résultats avec les méthodes d'optimisation standard.

5.3 Paramétrer autrement

Généralement, les paramètres intrinsèques ne sont pas complètement inconnus. Leurs valeurs sont fournies par le constructeur de la caméra, mais ces dernières sont peu précises.

Utiliser ces valeurs peu précises rend les méthodes basées sur la matrice essentielle inutilisables. Nous proposons ici une autre façon de paramétrer le problème du calcul de la structure et du mouvement. Les équations qui en résultent sont simples, ce qui permet d'obtenir des résultats qui montrent que même avec des paramètres intrinsèques connus avec la précision du constructeur, la structure et le mouvement calculés sont très proches de la solution exacte.

Dans la suite nous supposons que des valeurs approchées pour les paramètres intrinsèques sont disponibles (à partir des données constructeur ou d'une estimation connue en pratique). Comme repère de référence, nous choisissons celui attaché à la première caméra.

Soit un point p dans une image donné par ses coordonnées pixels (x, y) . Ses coordonnées normalisées (voir paragraphe 1.2.1) sont notées (u, v) . Ces dernières peuvent être calculées si les paramètres intrinsèques de la caméra sont connus. En utilisant les valeurs approximatives des paramètres intrinsèques, nous pouvons

obtenir une approximation des valeurs de u et de v :

$$\begin{cases} u = \frac{x-u_0}{\alpha_u} \\ v = \frac{y-v_0}{\alpha_v} \end{cases} \quad (5.7)$$

où (u_0, v_0) sont les coordonnées du centre de l'image, α_u et α_v sont les deux facteurs d'échelle.

Puisque nous utilisons le repère de la première caméra comme repère de référence, les coordonnées tridimensionnelles des points de la scène sont donc calculées dans ce repère (voir figure 5.1).

Dans ce cas, les coordonnées (X_i, Y_i, Z_i) d'un point de l'espace P_i peuvent s'écrire en fonction des coordonnées normalisées (u_{i1}, v_{i1}) du point p_{i1} , projection de P_i dans la première image.

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \lambda_i \begin{pmatrix} u_{i1} \\ v_{i1} \\ 1 \end{pmatrix} \quad (5.8)$$

En supposant que u_{i1} et v_{i1} sont connus (5.7), nous réduisons ainsi le nombre de paramètres de la structure. Nous passons de $3 \times p$ à p inconnues pour tous les points de la scène. En effet, la relation (5.8) montre que chaque point de l'espace P_i possède une seule inconnue λ_i (la profondeur).

Les équations des projections des points de la scène dans la j ème image ($j \geq 2$) s'écrivent :

$$\begin{cases} u_{ij} = \frac{\lambda_i r_{11}^j u_{i1} + \lambda_i r_{12}^j v_{i1} + \lambda_i r_{13}^j + t_x^j}{\lambda_i r_{31}^j u_{i1} + \lambda_i r_{32}^j v_{i1} + \lambda_i r_{33}^j + t_z^j} \\ v_{ij} = \frac{\lambda_i r_{21}^j u_{i1} + \lambda_i r_{22}^j v_{i1} + \lambda_i r_{23}^j + t_y^j}{\lambda_i r_{31}^j u_{i1} + \lambda_i r_{32}^j v_{i1} + \lambda_i r_{33}^j + t_z^j} \end{cases} \quad (5.9)$$

où (t_x^j, t_y^j, t_z^j) est le vecteur translation, (r_{kl}^j) la matrice rotation (voir figure 5.3).

En divisant par λ_i les numérateurs et dénominateurs des équations (5.9) et en posant $\beta_i = \frac{1}{\lambda_i}$ nous obtenons :

$$\begin{cases} u_{ij} = \frac{r_{11}^j u_{i1} + r_{12}^j v_{i1} + r_{13}^j + \beta_i t_x^j}{r_{31}^j u_{i1} + r_{32}^j v_{i1} + r_{33}^j + \beta_i t_z^j} \\ v_{ij} = \frac{r_{21}^j u_{i1} + r_{22}^j v_{i1} + r_{23}^j + \beta_i t_y^j}{r_{31}^j u_{i1} + r_{32}^j v_{i1} + r_{33}^j + \beta_i t_z^j} \end{cases} \quad (5.10)$$

On peut facilement vérifier que les dénominateurs des équations ci-dessus ne s'annulent jamais (sauf pour les points du plan d'équation $Z = 0$, mais ces points ne sont pas devant la caméra), ce qui nous permet de transformer (5.10) en :

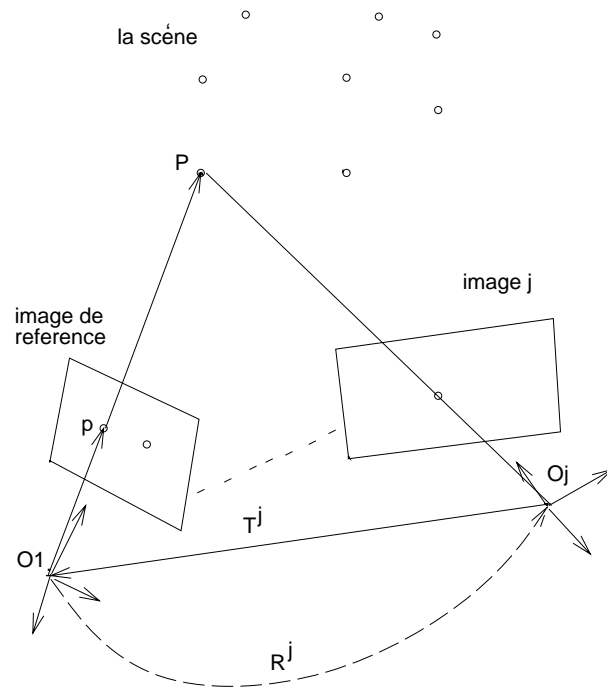


FIG. 5.3 - Le vecteur translation T^j est le vecteur $O_j \vec{O}_1$ défini dans le repère de la jème caméra, R^j est la rotation qui fait passer du repère de la première caméra au repère de la jème caméra

$$\begin{cases} u_{ij}(r_{31}^j u_{i1} + r_{32}^j v_{i1} + r_{33}^j + \beta_i t_z^j) = r_{11}^j u_{i1} + r_{12}^j v_{i1} + r_{13}^j + \beta_i t_x^j \\ v_{ij}(r_{31}^j u_{i1} + r_{32}^j v_{i1} + r_{33}^j + \beta_i t_z^j) = r_{21}^j u_{i1} + r_{22}^j v_{i1} + r_{23}^j + \beta_i t_y^j \end{cases} \quad (5.11)$$

Ces équations, comparées avec celles données par (5.6), sont beaucoup plus simples. Mais elles restent non linéaires. Une nouvelle fois nous sommes donc en présence d'un problème décrit par un ensemble d'équations non linéaires. Une minimisation au sens des moindres carrés permet de résoudre un tel problème. Aussi nous avons utilisé l'algorithme de Levenberg-Marquardt [PFTW88].

Cependant, une initialisation du processus d'optimisation avec une solution approchée est nécessaire pour garantir la convergence de celui-ci. Le paragraphe qui suit présente notre méthode pour calculer une solution approchée pour le mouvement et pour la structure.

5.3.1 Calcul d'une solution approchée

Pour simplifier l'exposé, considérons le cas de deux images où la première image est utilisée comme référence.

Il est évident depuis la figure 5.3 que la projection du point O_1 dans la jème image n'est autre que l'épipole de celle-ci par rapport à la première image. Ce dernier peut être calculé avec une méthode linéaire qui est éventuellement suivi par un calcul non linéaire plus précis (section 2.1). Soit $e = (e_x, e_y, e_t)$ cet épipole donné par ses coordonnées homogènes dans l'image. Le point O_1 est l'origine du repère de référence (repère de la première caméra), il a donc comme coordonnées homogène $(0, 0, 0, 1)$. O_1 se projette en e dans la deuxième image selon le modèle sténopé (5.5), nous avons donc :

$$\begin{pmatrix} e_x \\ e_y \\ e_t \end{pmatrix} = \mu M_2 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad (5.12)$$

avec μ un facteur multiplicatif inconnu.

En développant (5.12), on peut déduire la relation suivante :

$$\begin{pmatrix} e_x \\ e_y \\ e_t \end{pmatrix} = \mu A \begin{pmatrix} t_x^2 \\ t_y^2 \\ t_z^2 \end{pmatrix} \quad (5.13)$$

où A est la matrice 3×3 des paramètres intrinsèques et (t_x^2, t_y^2, t_z^2) sont les coordonnées du vecteur translation T^2 (voir figure 5.3).

En supposant que A est connu, la relation (5.13) comporte deux équations linéaires indépendantes en les 3 inconnues t_x^2 , t_y^2 et t_z^2 . Puisque le facteur d'échelle ne peut être déterminé, nous pouvons donc fixer le module de la translation, par exemple en mettant ce dernier à 1, c'est à dire, ajouter l'équation :

$$(t_x^2)^2 + (t_y^2)^2 + (t_z^2)^2 = 1$$

Ainsi, la translation peut être calculée séparément. Les équations (5.11) deviennent maintenant linéaires (translation supposée connue) ; ce qui permet de calculer une solution approchée de tous les paramètres du problème.

Cette solution approchée garantit la convergence du processus non linéaire basé sur les équations (5.11).

Remarque

Le vecteur translation ne peut être déterminé qu'à un signe près, c'est dire que si (t_x^2, t_y^2, t_z^2) est solution alors $(-t_x^2, -t_y^2, -t_z^2)$ l'est aussi. Ceci correspond à l'existence de deux solutions compatibles avec un ensemble de points mis en correspondance dans deux images [May93]. Ces deux solutions correspondent à deux cas de reconstruction : dans un cas les points se trouvent devant la caméra alors que dans l'autre cas ils se trouvent derrière celle-ci. Une seule de ces deux solutions est acceptable ; c'est celle où les points reconstruits se trouvent devant la caméra, c'est à dire que tous les β_i de (5.11) sont strictement positifs. Pour choisir la bonne solution, nous calculons les β_i des points avec (t_x^2, t_y^2, t_z^2) et avec $(-t_x^2, -t_y^2, -t_z^2)$. La solution retenue est celle où les β_i sont positifs. Notons que le calcul est linéaire car les équations (5.11) sont linéaires quand la translation est connue.

5.3.2 Résumé de la méthode de reconstruction

En supposant que les paramètres intrinsèques sont connus (une approximation de ceux-ci), le problème de la reconstruction est défini par les équations non linéaires (5.11).

En considérant le cas de deux images pour simplifier, soit un ensemble de points de l'espace (une scène) observés par deux caméras. Dans ce cas, les paramètres du problème de la reconstruction sont : une rotation, une translation et enfin, la structure des points donnée par les β_i .

La méthode de reconstruction proposée dans ce chapitre peut se résumer aux deux étapes suivantes :

1. Calcul d'une solution approchée : la translation est d'abord calculée grâce à la connaissance de l'épipole, ensuite une estimation des autres paramètres du problème (la rotation et les β_i) est calculé avec un processus linéaire.

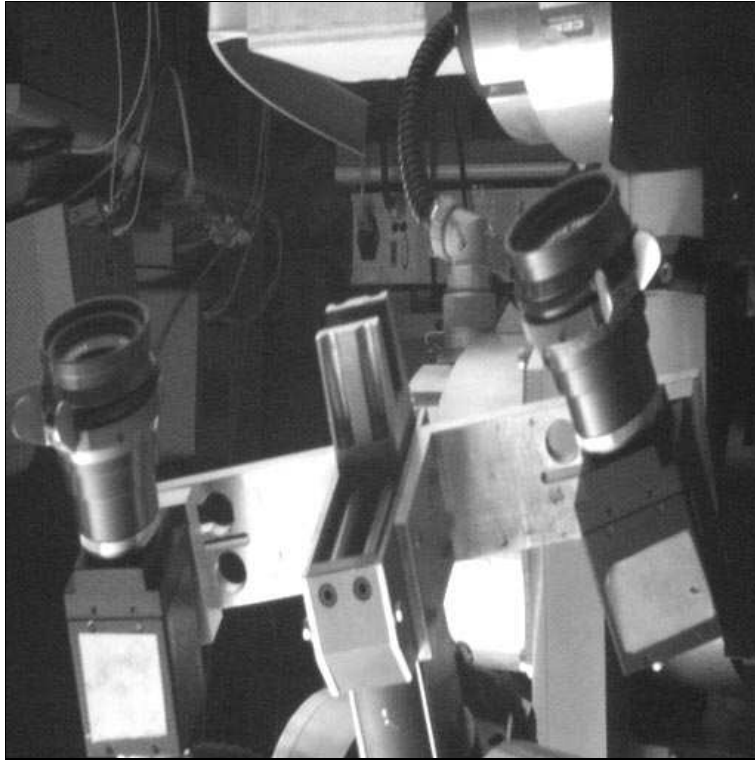


FIG. 5.4 - *La tête stéréo du LIFIA : les deux caméras sont à environ 35 cm l'une par rapport à l'autre.*

2. A présent, nous avons une estimation de tous les paramètres du problème de la reconstruction. Cette estimation est utilisée comme solution approchée au calcul non linéaire basé sur les équations (5.11). Dans cette dernière étape, les inconnues du problèmes sont la translation, la rotation et les β_i .

5.4 Expérimentation

Deux expériences ont été menées. La première dont le but est de valider notre approche et de montrer que les résultats obtenus sont qualitativement corrects. La deuxième, plutôt quantitative, s'intéresse à la précision des paramètres du mouvement calculé avec notre méthode.

Pour chaque expérience, nous avons utilisé deux images seulement. Celles-ci sont issues de la tête stéréo du LIFIA (voir figure 5.4). Pour chaque paire d'images, les mises en correspondances sont calculées automatiquement.

- première expérience : la scène est composée de deux cubes blancs sur lesquels nous avons collé des motifs polygonaux de couleur noire (voir figure 5.5). Un total de 49 points ont été extraits et mis en correspondance dans les deux

	Angles de rotation			Translation		paramètres intrinsèques			
	α	β	γ	t_x/t_z	t_y/t_z	α_u	α_v	u_0	v_0
<i>Étalonnage</i>	-23,56	4,24	-9,83	-0,995	-3,135	-1466,22	997,70	234,38	242,03
<i>Test1</i>	-23,57	6,01	-8,88	-0,884	-2,908	-1500	1000	256	256
<i>Test2</i>	-25,75	7,58	-7,37	-0,820	-2,881	-1600	1100	256	256
<i>Test3</i>	-24,03	2,37	-10,43	-0,846	-2,824	-1500	1000	280	280

TAB. 5.1 - *Comparaison des résultats : en supposant que les valeurs des paramètres intrinsèques obtenues par étalonnage sont une référence ; on voit bien que même avec 10% d'erreur sur ceux-ci, le mouvement calculé avec notre méthode reste d'une bonne qualité*

images. Les valeurs des paramètres intrinsèques utilisées sont données par la matrice A tel que :

$$A = \begin{pmatrix} -1500 & 0 & 256 \\ 0 & 1000 & 256 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.14)$$

La figure 5.6 montre une vue de dessus et une vue générale des points reconstruits avec notre méthode. La reconstruction est qualitativement bonne, en particulier, la vue de dessus montre bien les différents plans des deux cubes.

- deuxième expérience : ici nous avons utilisé la mire du LIFIA, les points d'intérêt sont les sommets des carrés noirs (voir figure 5.7). Une table micrométrique, sur laquelle est fixée la mire, permet de déplacer celle-ci en translation pure. Grâce à cette table micrométrique, nous pouvons obtenir des scènes non planes en utilisant plusieurs images de la mire à des positions différentes.

Pour cette expérience, nous avons utilisé une scène composée de 3 plans de la mire, soit un total de 276 points.

En utilisant notre méthode de reconstruction avec les valeurs données dans (5.14) pour les paramètres intrinsèques, nous obtenons la reconstruction de la figure 5.8.

Comme les points de la mire sont connus, l'étalonnage des deux caméras est alors possible. Nous avons pu comparer les paramètres du mouvement obtenu par notre méthode avec ceux obtenus par étalonnage des caméras. Le tableau 5.1 résume cette comparaison, où trois tests ont été réalisés en modifiant à chaque fois les valeurs des paramètres intrinsèques.

Les résultats obtenus ci-dessus montrent que la reconstruction et le mouvement calculés sont qualitativement bons. La précision de ces derniers dépend des erreurs sur les valeurs des paramètres intrinsèques.

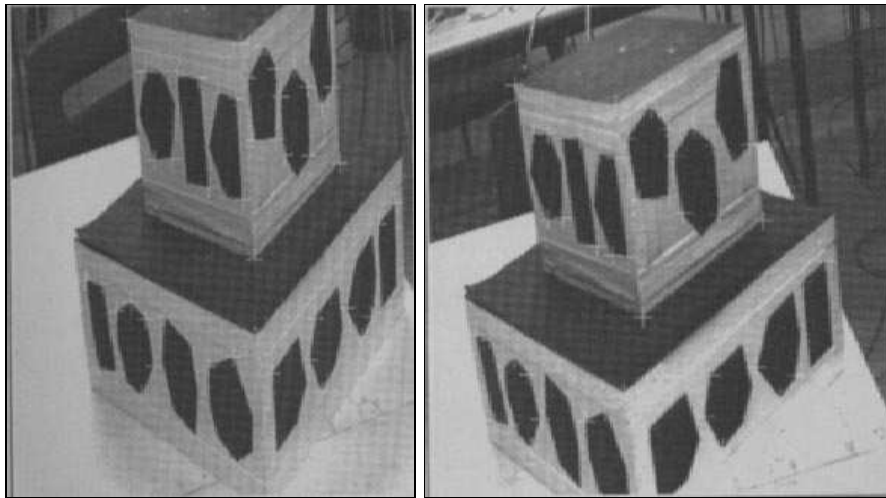


FIG. 5.5 - *La paire stéréo des cubes*

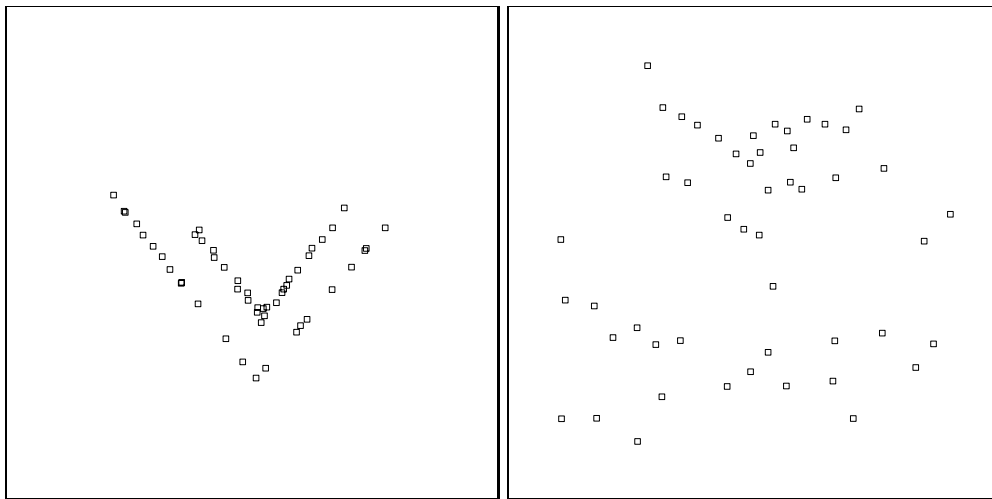
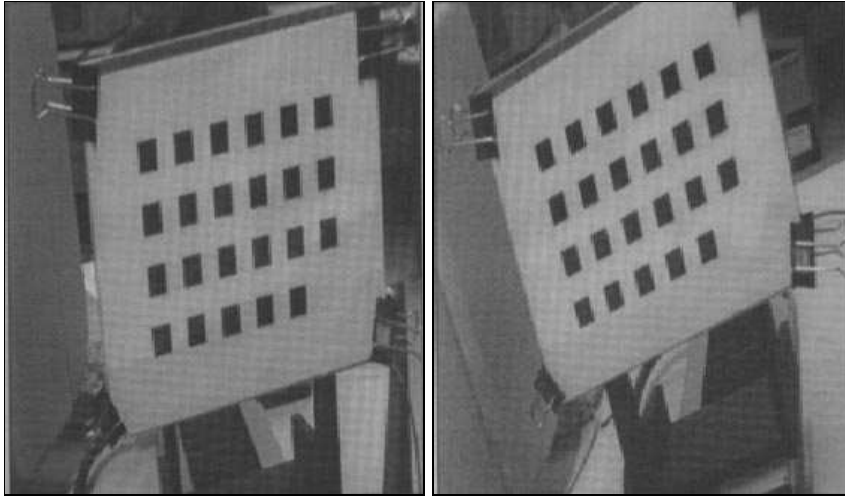
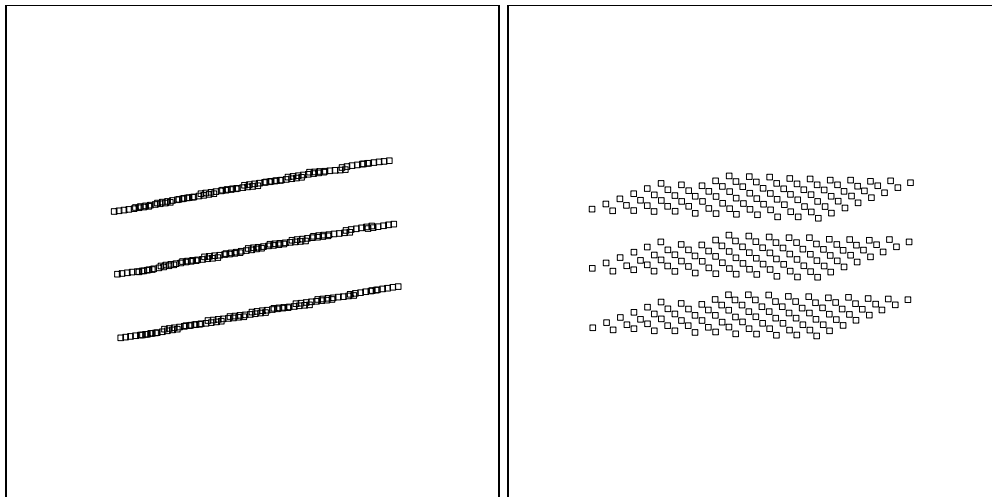


FIG. 5.6 - *Vue de dessus (à gauche) et vue de face (à droite) de la reconstruction des cubes*

FIG. 5.7 - *La mire observée par la tête stéréo*FIG. 5.8 - *Les 3 plans de la mire reconstruits*

5.5 Conclusion

Nous avons présenté dans ce chapitre une méthode qui permet de calculer la structure et le mouvement d'une caméra en supposant la connaissance approximative des paramètres intrinsèques. Ces derniers sont toujours approximativement connus pour une caméra donnée, soit depuis les données du constructeur, soit depuis une estimation.

Les méthodes classiques, basées sur la matrice essentielle, supposent la connaissance exacte des paramètres intrinsèques, obtenues par étalonnage. Quand ces derniers sont peu précis, ces méthodes deviennent inutilisables en pratique.

Dans la méthode proposée, nous avons utilisé un paramétrage particulier du problème de la reconstruction. Ce paramétrage a permis d'utiliser un nombre minimal d'inconnues. Les équations décrivant le problème sont devenues simples, conduisant à de bons résultats qualitatifs, même quand les paramètres intrinsèques sont peu précis.

La précision des résultats (reconstruction et mouvement) pourra être améliorée en utilisant une autre étape d'optimisation où les paramètres intrinsèques ne seront plus supposés connus. Ces derniers feront partie des paramètres à estimer. Les équations deviendront plus complexes, car leur degré augmente ; aussi, nous n'avons pas réussi à améliorer les résultats en utilisant l'algorithme de Levenberg-Marquardt.

Une voie qui nous paraît prometteuse et qui reste à explorer est l'utilisation des techniques d'optimisation sous contraintes où nous supposerons que chaque paramètre se trouve dans un intervalle limitant son domaine. L'utilisation de telles techniques est justifiée par les résultats obtenus dans ce chapitre. Ces derniers montrent que les valeurs calculées sont une bonne approximation des inconnues. Ces valeurs permettent de déduire des intervalles de confiance qui contraignent le domaine des inconnues.

Conclusion

Dans cette thèse nous avons traité principalement le problème de la reconstruction tridimensionnelle dans un cadre très général où les images sont issues de caméras non étalonnées. Nous avons évité l'étalonnage des caméras, qui est un processus délicat et parfois impossible à réaliser.

La reconstruction d'une scène contient l'information géométrique tridimensionnelle de celle-ci. Cette information peut être de différents types. En particulier, elle peut être projective, affine ou euclidienne.

La reconstruction projective est la plus facile à obtenir ; elle ne nécessite aucune information supplémentaire sauf les mises en correspondance des points dans les images. Cependant, une telle reconstruction est pauvre en information ; elle reste inutilisable en robotique car on ne dispose pas encore d'une description projective pour les commandes de déplacements des robots. L'utilisation de l'information projective reste limitée à quelques applications de reconnaissance.

La reconstruction affine est plus riche que la reconstruction projective, mais elle est moins facile à obtenir que celle-ci.

Enfin, la reconstruction euclidienne est la plus utilisée dans le domaine de vision et de la robotique ; elle est la plus riche en information mais la plus difficile à calculer. Cette reconstruction contient les informations dont l'homme maîtrise la manipulation. La robotique actuelle utilise exclusivement ce type d'information, l'information euclidienne.

Des méthodes permettant d'obtenir ces trois types de reconstruction ont été proposées dans cette thèse. En particulier, nous avons clairement défini les relations liant ces trois types de reconstruction. En présence d'une reconstruction projective d'une scène donnée, nous avons énuméré les informations nécessaires pour obtenir la reconstruction affine ou euclidienne de cette même scène.

Dans cette thèse, la reconstruction euclidienne, vu son importance, a eu plus d'intérêt que les autres reconstructions. Nous avons présenté plusieurs méthodes pour la calculer :

- elle peut être obtenue à partir d'une reconstruction projective en injectant des informations euclidiennes dans celle-ci,
- elle peut être obtenue, si les conditions géométriques le permettent, en sim-

plifiant le modèle de la caméra : en utilisant un modèle affine de projection scène-image,

- elle peut être obtenue directement sans passer par une reconstruction projective et sans simplifier le modèle de la caméra si les paramètres intrinsèques des caméras sont connus.

Par ailleurs, nous avons apporté des contributions à deux sujets liés à la reconstruction :

- la géométrie épipolaire : nous avons proposé une nouvelle méthode pour calculer la matrice fondamentale, matrice qui décrit la géométrie épipolaire. Cette méthode permet de réduire le nombre des paramètres à estimer, les équations utilisées pour résoudre le problème sont plus simples que celles utilisées dans la méthode classique. Comme conséquence, les calculs sont plus stables et un calcul linéaire permet d'obtenir des résultats corrects.
- la mise en correspondance de points dans une séquence d'images : ce problème étant une étape nécessaire pour résoudre les problèmes de reconstruction ou de calcul de mouvement, nous avons donc réalisé un algorithme robuste basé sur la géométrie épipolaire. Cette algorithme traite aussi bien les images de scènes polyédriques que les images complexes, typiquement les images d'extérieur.

Un problème, déjà soulevé au chapitre 5, qui mérite d'être approfondi concerne le cas de la reconstruction à partir d'images issues d'une caméra dont les paramètres intrinsèques restent inchangés. Nous n'avons traité que le cas où ces paramètres sont supposés connus. Cependant, même si ces derniers sont inconnus, théoriquement une solution existe dès que le nombre d'images utilisées est supérieur à 2. Ce cas est très intéressant, d'autant plus que nous avons fourni une méthode qui permet d'obtenir une bonne approximation de la solution. Cette voie sera explorée dans un avenir proche.

Ce travail ouvre plusieurs perspectives d'application ; en particulier dans les domaines de la robotique et de la synthèse d'images :

- en robotique l'information relative est souvent la seule information requise pour certaines tâches. L'utilisation des méthodes de reconstruction relative sans étalonnage peut servir à guider et à étalonner un robot. En particulier, pour saisir un objet on a uniquement besoin de sa position par rapport à la pince, donc d'une information relative. Nous avons déjà commencé à explorer cette voie en collaboration avec R. Horaud [HDBM94].
- les techniques de reconstruction sans étalonnage peuvent être utilisées en synthèse d'images. Deux images d'une scène réelle permettent de déduire

différentes informations tridimensionnelle sur celle-ci. La fabrication d'images virtuelles correspondant à des points de vue imaginaires de la caméra peut servir à donner l'illusion de naviguer dans une scène. Une thèse traitant de ce sujet vient de démarrer au LIFIA.

Annexe A

Détecteur de points d'intérêt

Nous avons implanté le détecteur de points d'intérêt développé initialement par Harris et Stephens [HS88] auquel nous avons apporté une amélioration au niveau de la localisation. Ce travail a été réalisé au sein du laboratoire LIFIA par J. C. Cottier, un stagiaire que nous avons encadré durant son stage de magistère.

Nous commençons par la présentation du détecteur de points d'intérêt de Moravec [Mor77], puis nous présentons celui de Harris et Stephens [HS88], appelé détecteur de plessey, qui est une amélioration de celui de Moravec.

A.1 Détecteur de Moravec

Le détecteur de Moravec travaille directement sur l'image à niveaux de gris. Il modélise les différentes régions (régions uniformes, frontières et points caractéristiques) en fonction de la mesure de ressemblance locale de la région. C'est à dire qu'il considère une région (fenêtre) de l'image, et vérifie si elle ressemble ou non aux régions qui l'entourent (la fenêtre effectue un faible déplacement).

Nous pouvons identifier les régions de la façon suivante :

- Pour une région uniforme, l'intensité à l'intérieur de la fenêtre reste à peu près constante pour tout déplacement de la fenêtre.
- Pour une frontière (contour), quand la fenêtre est déplacée le long de la frontière, les intensité à l'intérieur de la fenêtre ne varient pas beaucoup. Par contre, quand la fenêtre est déplacée orthogonalement à la frontière, alors l'intérieur de la fenêtre varie beaucoup.
- Pour un point d'intérêt, l'intensité à l'intérieur de la fenêtre change, pour tout déplacement.

Moravec a donné une spécification mathématique à l'idée précédente. Soit I représentant l'intensité de l'image ; soit E le changement produit par un déplacement

(x, y) . Alors,

$$E_{x,y} = \sum_{u,v} w_{u,v} [I_{x+u,y+v} - I_{u,v}]^2$$

où w représente un masque de l'image à 1 sur une certaine région rectangulaire, et à 0 ailleurs (le rectangle rempli de 1 représente évidemment la région que nous sommes en train d'analyser). Les déplacements (x, y) envisagés par Moravec font partie de l'ensemble $\{(1,0),(1,1),(0,1),(-1,1)\}$. Moravec calcule les valeurs de E pour les quatre déplacements précédents. Ensuite, il cherche les points dont $\min(E)$ dépasse un certain seuil. Ces points sont les points d'intérêt de Moravec. Effectivement, si nous considérons une région uniforme, E est faible pour les quatre déplacements. Donc nous obtenons une valeur très faible pour $\min(E)$. Pour un point de frontière, le déplacement le long de celle ci donnera une faible valeur pour E , donc ce point aussi obtiendra une faible réponse au coin. Par contre, pour un point riche en information, nous savons que, quelque soit le déplacement, le changement est important. Donc même $\min(E)$ aura une valeur importante. Ainsi, les points caractéristiques sont identifiables.

Ce détecteur fournit des résultats globalement satisfaisants. Cependant plusieurs imperfections dégradent les résultats obtenus. C'est pour cela que Harris et Stephens ont cherché à l'améliorer.

A.2 Le détecteur de points d'intérêt de Plessey

Le détecteur de Plessey est une amélioration du détecteur de Moravec. L'idée directrice de Moravec est reprise tout en corrigeant les défauts de son détecteur. Ces principaux défauts sont les suivants :

1. La réponse est anisotropique car seules les directions définies tous les 45 degrés sont considérées.
2. La réponse est bruitée car la fenêtre est rectangulaire et binaire.
3. Certains points situés sur des frontières plus ou moins linéaires ont des réponses aux coins trop élevées. La principale cause est que seul le minimum de E est pris en compte dans la réponse.

Pour éliminer ces problèmes, Harris et Stephens [HS88] proposent les corrections respectives suivantes :

1. Tous les petits déplacements possibles doivent être considérés pour ne pas favoriser certaines directions. Ce problème peut être résolu en considérant un développement limité de la mesure de ressemblance.

$$E_{xy} = \sum_{u,v} w_{u,v} [I_{x+u,y+v} - I_{u,v}]^2$$

s'écrit pour de faibles déplacements x, y :

$$E_{xy} = \sum_{u,v} w_{u,v} \left[x \frac{\partial I}{\partial x} + y \frac{\partial I}{\partial y} + O(x^2, y^2) \right]^2$$

x et y étant petits, nous négligeons le terme $O(x^2, y^2)$. De plus, les gradients peuvent être approchés par

$$X = I \otimes (-1, 0, 1) \approx \frac{\partial I}{\partial x}$$

$$Y = I \otimes (-1, 0, 1)^T \approx \frac{\partial I}{\partial y}$$

La formule s'écrit donc

$$E_{xy} = \sum_{u,v} w_{u,v} [xX + yY]^2$$

finalelement :

$$E_{xy} = Ax^2 + By^2 + 2Cxy$$

avec

$$A = X^2 \otimes w$$

$$B = Y^2 \otimes w$$

$$C = (XY) \otimes w$$

2. Il est préférable d'utiliser une fenêtre circulaire plutôt que rectangulaire. Pour lisser les réponses et atténuer les sources de bruits, il est proposé de filtrer le calcul des dérivées par une Gaussienne :

$$w_{u,v} = e^{-\frac{u^2+v^2}{2\sigma^2}}$$

3. Il faut re-formuler la mesure du coin en considérant la variation de E en fonction de la direction du déplacement. La variation de E pour de faibles déplacements x, y peut être réécrite de façon plus concise :

$$E_{x,y} = (x, y)M(x, y)^T$$

où M est une matrice 2x2 symétrique :

$$M = \begin{pmatrix} A & C \\ C & B \end{pmatrix}$$

Considérons α, β : les valeurs propres de M . Alors, α et β sont proportionnelles aux courbures principales de la fonction de ressemblance locale E . A ce stade, nous pouvons considérer trois cas :

4. A. Si les deux courbures sont petites, alors quelque soit le déplacement envisagé, l'image ne changera pas beaucoup. Donc la région que nous sommes en train de traiter possède approximativement une intensité constante : c'est une région pauvre en information.
5. B. Si une courbure est grande et l'autre petite, alors la fonction de ressemblance locale E a une forme en crête. Ainsi, seuls les déplacements le long de l'arrête de la crête (le long de la frontière) causeront de faibles changements de l'image. Cela indique une frontière.
6. C. Si les deux courbures sont grandes, alors quelque soit le déplacement considéré, E produira une valeur importante. C'est la signature d'un point de type coin.

Jusqu'ici, nous avons classé les régions comme étant des coins ou des frontières, mais il faudrait aussi avoir une mesure de la qualité des coins et des régions. Les auteurs proposent un opérateur intéressant car il évite le calcul explicite des valeurs propres α et β de M . Ceci est possible en utilisant :

$$Tr(M) = \alpha + \beta = A + B$$

$$Det(M) = \alpha\beta = AB - C^2$$

L'opérateur R proposé pour la réponse aux coins est le suivant :

$$R = Det - kTr^2$$

k étant une variable à déterminer expérimentalement.

1. Si R est proche de zéro, alors c'est que nous sommes dans une région uniforme.
2. Si R a une grande valeur, c'est que nous sommes en présence d'un coin.
3. Si R est négative alors, c'est que nous sommes sur une frontière.

A.3 Amélioration du détecteur de Plessey

Ce détecteur possède tout de même un défaut important. La localisation des points n'est pas très bonne. Il arrive trop souvent que certains points soient décalés de trois ou quatre pixels par rapport à leur position idéale. Cela risque de nuire à la précision des calculs ultérieurs en introduisant un bruit important.

Nous savons que le détecteur de Plessey n'est pas optimal en ce qui concerne la finesse de la localisation. Il n'est pas rare qu'il détecte des coins à plusieurs pixels de leurs positions idéales. Notre objectif étant de réaliser des outils stables et robustes, nous ne pouvons accepter une telle imprécision.

Notre solution à ce problème est une détection à deux échelles dont le principe est décrit dans le paragraphe qui suit.

Une détection à deux échelles

Pour calculer la réponse d'un point d'intérêt, nous considérons la région centrée autour de celui-ci. C'est par la mesure de ressemblance de cette région avec celles qui l'entourent qu'un point est considéré comme point d'intérêt ou non.

Quand les dimensions de cette région diminuent, alors la localisation est améliorée. A ce stade, cette remarque ne serait en fait qu'un simple réglage de paramètres. Or ce n'est pas le cas. Effectivement, si la taille de la fenêtre diminue, alors de nombreux points sont détectés comme points d'intérêt alors qu'ils n'en sont pas. Dans ce cas, nous aurions bien amélioré la localisation des points mais nous aurions fortement perdu en ce qui concerne la qualité de détection.

Il faut donc faire deux passages :

- A Le premier passage détecte les points d'intérêt avec une fenêtre de grande taille, le résultat est un ensemble de points qui sont bien des points d'intérêt mais dont la localisation n'est pas encore optimale¹.
- B Dans le second passage, nous diminuons la taille de la fenêtre de travail. Par contre, la recherche des points d'intérêt se limite cette fois aux voisinages des points détectés précédemment (en pratique, plus ou moins 2 pixels).

Ainsi, la qualité de détection n'est pas changée et nous augmentons la précision de localisation.

1. Même si nous contraignons la recherche qu'aux points de contours

Annexe B

Méthode robuste pour trouver la matrice fondamentale

Etant donné un ensemble d'appariements obtenus par corrélation, typiquement une centaine pour nos expériences, nous pouvons calculer la matrice fondamentale avec la méthode donnée dans la section 2.1. Cependant, la présence de faux appariements peut déformer la solution, conduisant dans certains cas à des résultats complètement faux. Pour remédier à ce problème, nous avons utilisé une méthode basée sur les statistiques proposée par Z. Zhang [ZDFL94]. Si le calcul est fait sans faux appariements alors la matrice fondamentale trouvée est correcte. Donc au lieu de faire un seul calcul, il propose d'en faire plusieurs. Ainsi, nous pouvons espérer qu'au moins une des matrices trouvées est correcte. Dans la pratique voici comment nous faisons. Nous répétons 300 fois l'opération suivante:

1. nous choisissons dix¹ appariements au hasard,
2. avec ces dix appariements, nous calculons la matrice fondamentale.

Dans nos appariements initiaux, nous n'avions qu'un faible pourcentage d'erreurs. Un calcul simple de probabilité montre que si 80 % des appariements sont corrects alors il est quasiment certain qu'au moins un des tirages ne comporte aucun faux appariement. Auquel cas, il existe parmi les 300 matrices que nous avons calculées une matrice correcte. Le problème est donc maintenant de trouver quelle est la meilleure matrice parmi les 300. Il faut pouvoir quantifier la qualité d'une matrice fondamentale. Voici comment nous pouvons faire :

Pour chaque appariement (p_i, p'_i) , l'erreur est la distance entre le point p'_i et sa droite épipolaire $F \hat{p}_i$. Par exemple l'erreur pour le point p'_i est donnée par :

¹ La théorie indique que huit appariements suffisent. Mais en pratique, nous obtenons de meilleurs résultats avec quelques appariements de plus

$$d(p'_i, F {}^t p_i) = \frac{|p'_i F {}^t p_i|}{\sqrt{(F {}^t p_i)_1^2 + (F {}^t p_i)_2^2}}$$

Si nous accumulons cette erreur pour tous les appariements dans les deux images, la formule générale devient:

$$\sum_i d(p'_i, F {}^t p_i) + d(p_i, {}^t F p'_i)$$

Ainsi plus cette somme est faible plus la matrice est considérée comme correcte. Cette évaluation est valide s'il n'existe pas de faux appariement. Or, ce n'est malheureusement pas notre cas. Il nous faut donc utiliser une méthode plus robuste prenant en compte les faux appariements. C'est le domaine des statistiques numériques qui donne la solution. La méthode des moindres carrés médians [PFTW88] permet de trouver la meilleure matrice même si dans l'évaluation, nous utilisons de faux appariements. Pour chaque matrice, nous calculons le carré des erreurs vues précédemment et nous déterminons quelle est leur valeur médiane. La meilleure matrice est celle qui aura la valeur médiane la plus faible.

Effectivement, l'évaluation de la valeur médiane des carrés des erreurs avec quelques faux appariements, n'est pas très différente de celle faite sans faux appariements.

Bibliographie de l'auteur

Conférences internationales sans comité de lecture

- **Roger Mohr, Boubakeur Boufama, et Pascal Brand**
Accurate projective reconstruction. Dans *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision*, Azores, Portugal, octobre 1993, pages 203–227.

Conférences internationales avec comité de lecture et actes

- **Stéphane Paris et Boubakeur Boufama**
Concerning the Effect of the Indexing on the Recognition. Dans *Proceedings of the workshop "Advances in Structural and Syntactic Pattern Recognition"*, Bern, Suisse, août 1992, publié dans *Series on Machine Perception and Artificial Intelligence*, volume 5, pages 451–460.
- **Boubakeur Boufama, Roger Mohr et Françoise Veillon**
Euclidian constraints for uncalibrated reconstruction. Dans *Proceedings of the 4th International Conference on Computer Vision*, Berlin, Germany, mai 1993, pages 466–470.
- **Boubakeur Boufama, Daphna Weinshall et Mike Werman**
Shape from motion algorithms: a comparative analysis of scaled orthography and perspective. Dans *Proceedings of the 3rd European Conference on Computer Vision*, Stockholm, Sweden, mai 1994, pages 199–204.
- **Radu Horaud, Fadi Dornaika, Boubakeur Boufama et Roger Mohr**
Self calibration of a stereo head mounted onto a robot arm. Dans *Proceedings of the 3rd European Conference on Computer Vision*, Stockholm, Sweden, mai 1994, pages 455–462.
- **David Sinclair, Boubakeur Boufama et Roger Mohr**
Independent motion segmentation and collision prediction. Dans *Proceedings*

of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA, juin 1994, pages 958–961.

Articles de Revues

- **C.S. Zhao, R. Mohr et B. Boufama**
Reconstruction de surfaces B-splines régularisées et tridimensionnelles à partir d'une caméra en mouvement. Dans *Techniques et Sciences Informatiques*, à paraître.

Conférences nationales avec comité de lecture et actes

- **Chang Zhao Roger Mohr et Boubakeur Boufama**
Reconstruction de surfaces tridimensionnelles à partir d'une caméra en mouvement. Dans *Actes du congrès Reconnaissance des Formes et Intelligence Artificielle*, AFCET-RFIA, Paris, janvier 1994, Tome 1, pages 407–418.

Rapports techniques

- **Roger Mohr, Long Quan, Françoise Veillon et Boubakeur Boufama**
Relative 3D reconstruction using multiples uncalibrated images. Rapport technique RT 84 IMAG 12 LIFIA, mars 1992.
- **Boubakeur Boufama Roger Mohr et Françoise Veillon**
Euclidian constraints for uncalibrated reconstruction. Rapport technique RT 84 IMAG 12 LIFIA, mars 1993.
- **David Sinclair, Boubakeur Boufama et Roger Mohr**
Independent motion segmentation and collision prediction for road vehicles. Rapport technique RT 110 IMAG 22 LIFIA, février 1994
- **Roger Mohr, Boubakeur Boufama et Pascal Brand**
Accurate Relative Positioning From Multiple Images. Rapport technique RT 102 IMAG 20 LIFIA, novembre 1993.

Bibliographie

- [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.
- [AHP93] A. Azarbayejani, B. Horowitz, and A. Pentland. Recursive estimation of structure and motion using relative orientation constraints. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 294–299, June 1993.
- [Aya89] N. Ayache. *Vision Stéréoscopique et Perception Multisensorielle: Application à la Robotique Mobile*. Science Informatique. InterEditions, 1989.
- [Bar83] S.T. Barnard. Interpreting perspective images. *Artificial Intelligence*, 21(3):435–462, 1983.
- [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.
- [BM94] P. Brand and R. Mohr. Accuracy in image measure. In Sabry F. El-Hakim, editor, *Videometrics III*, volume 2350, pages 218–228. SPIE, November 1994.
- [BO90] B. Brillault-O'Mahony. New Method for Vanishing Point Detection. Technical report, EDF-Direction des Etudes et Recherches, Paris, France, 1990.

- [BWR90] J.B. Burns, R. Weiss, and E.M. Riseman. View variation of point set and line segment features. In *Proceedings of DARPA Image Understanding Workshop, Pittsburgh, Pennsylvania, USA*, pages 650–659, 1990.
- [Cou94] B. Couapel. *Stéréovision par Ordinateur, Géométrie et Expérimentation*. Thèse de doctorat, Université de Rennes 1, March 1994.
- [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.
- [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.
- [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.
- [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.
- [FMZ90] D. Forsyth, J. Mundy, and A. Zisserman. Transformational invariance - a primer. In *Proceedings of the British Machine Vision Conference, Oxford, England*, pages 1–6, September 1990.
- [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.
- [Gro93] P. Gros. *Outils Géométriques pour le modélisation et la reconnaissance d'objets polyédriques*. PhD thesis, Institut National Polytechnique de Grenoble, July 1993.
- [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.

- [HDBM94] R. Horaud, F. Dornaika, B. Boufama, and R. Mohr. Self calibration of a stereo head mounted onto a robot arm. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 455–462. Springer-Verlag, 1994.
- [HF89] T. S. Huang and O. D. Faugeras. Some properties of the E matrix in two-view motion estimation. *IEEE Transactions on PAMI*, 11(12):1310–1312, December 1989.
- [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.
- [HM93] R. Horaud and O. Monga. *Vision par Ordinateur: Outils Fondamentaux*. Editions Hermès, Paris, 1993.
- [HS88] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.
- [KvD89] J.J. Koenderink and A. J. van Doorn. Affine structure from motion. Technical report, Utrecht University, Utrecht, The Netherlands, October 1989.
- [LC94] J. M. Lawn and R. Cipolla. Robust egomotion estimation from affine motion parallax. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*. Springer-Verlag, May 1994.
- [LF92] Q.T. Luong and O.D. Faugeras. Self-calibration of a camera using multiple images. In *Proceedings of the 11th International Conference on Pattern Recognition, The Hag, Netherland*, pages 9–12, 1992.
- [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.
- [Luo92] Q.T. Luong. *Matrice Fondamentale et Autocalibration en Vision par Ordinateur*. Thèse de doctorat, Université de Paris-Sud, Orsay, France, December 1992.
- [MA90] R. Mohr and E. Arbogast. It Can Be Done without Camera Calibration. Technical Report RR 805-I- IMAG 106 LIFIA, LIFIA–IMAG, February 1990.
- [Mar78] D. Marr. Representing visual information — A computational approach. In *Computer Vision Systems*, pages 61–80, 1978.

- [May90] S.J. Maybank. The projective geometry of ambiguous surfaces. Technical Report 1623, Long Range Laboratory, GEC, Wembley, Middlesex, England, July 1990.
- [May93] S. Maybank. *Theory of Reconstruction from Image Motion*. Springer-Verlag, 1993.
- [MBB93] R. Mohr, B. Boufama, and P. Brand. Accurate projective reconstruction. In *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 203–227, October 1993.
- [MF92] S.J. Maybank and O.D. Faugeras. A theory of self calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.
- [MM91] R. Mohr and L. Morin. Relative positioning from geometric invariants. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, USA*, pages 139–144, June 1991.
- [Moh93] R. Mohr. Projective geometry and computer vision. In C.H. Chen, L.F.Pau, and S.P. Wang, editors, *Handbook of Pattern Recognition and Computer Vision*. World Scientific Pub. Co., 1993.
- [Mor77] H.P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, USA*, page 584, August 1977.
- [Mor93] L. Morin. *Quelques Contributions des Invariants Projectifs à la Vision par Ordinateur*. PhD thesis, Institut National Polytechnique de Grenoble, January 1993.
- [MQVB92] R. Mohr, L. Quan, F. Veillon, and B. Boufama. Relative 3D reconstruction using multiple uncalibrated images. Technical Report RT 84-I-IMAG LIFIA 12, LIFIA-IRIMAG, 1992.
- [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.
- [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.
- [PFTW88] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling W.T. *Numerical Recipes in C*. Cambridge University Press, 1988.

- [PK94] C. J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 97–108, May 1994.
- [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.
- [QM89] L. Quan and R. Mohr. Determining perspective structures using hierarchical Hough transform. *Pattern Recognition Letters*, 9(4):279–286, 1989.
- [Roh92] K. Rohr. Recognizing corners by fitting parametric models. *International Journal of Computer Vision*, 9(3):213–230, December 1992.
- [Sha92] A. Shashua. Projective Structure from Two Uncalibrated Images: Structure from Motion and Recognition. Technical Report A.I. Memo No. 1363, Massachusetts Institute of Technology, September 1992.
- [Sha94] A. Shashua. Trilinearity in visual recognition by alignment. In Jan-Olof Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 479–484. Springer Verlag, May 1994.
- [SK52] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford Science Publication, 1952.
- [SN90] G. Sparr and L. Nielsen. Shape and mutual cross-ratios with applications to the interior, exterior and relative orientation. In *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 607–609. Springer-Verlag, April 1990.
- [Spa91] G. Sparr. Projective invariants for affine shapes of point configurations. In *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Reykjavik, Iceland*, pages 151–170, March 1991.
- [SPFP93] S. Soatto, P. Perona, R. Frezza, and G. Picci. Recursive motion and structure estimation with complete error characterization. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 428–433, June 1993.
- [Stu69] R. Sturm. Das Problem der Projektivität und seine Anwendung auf die Flächen zweiten Grades. *Math. Ann.*, 1:533–574, 1869.

- [TF87] G. Toscani and O.D. Faugeras. Mouvement par reconstruction et re-projection. In *11ème Colloque sur le Traitement du signal et des images (GRETSI), Nice, France*, number Tome 2, pages 535–538, 1987.
- [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 PAMI*, volume 6, pages 13–27. IEEE, 1984.
- [TK91] C. Tomasi and T. Kanade. Factoring image sequences into shape and motion. In *Proceedings of IEEE Workshop on Visual Motion, Princeton, New Jersey*, pages 21–28, Los Alamitos, California, USA, October 1991. IEEE Computer Society Press.
- [Tom91] C. Tomasi. *Shape and Motion from Image Streams: a Factorization Method*. PhD thesis, University of Carnegie Mellon, USA, 1991.
- [Tsa87] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.
- [WAH93] J. Weng, N. Ahuja, and T.S. Huang. Optimal motion and structure estimation. *IEEE Transactions on PAMI*, 15(9):864–884, September 1993.
- [WCH92] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accurate evaluation. *IEEE Transactions on PAMI*, 14(10):965–980, October 1992.
- [Wei93] D. Weinshall. Model-based invariants for 3-D vision. *International Journal of Computer Vision*, 10(1):27–42, 1993.
- [WHA89] J. Weng, T.S. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Transactions on PAMI*, 11(5):451–476, May 1989.
- [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.