Le mouvement projectif : théorie et applications pour l'autocalibrage et la segmentation du mouvement

David Demirdjian

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

THÈSE

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité : Imagerie, Vision et Robotique

préparée au laboratoire GRAVIR - IMAG - INRIA
dans le cadre de l'Ecole Doctorale «Mathématiques, Sciences et Technologie de l'Information»

présentée et soutenue publiquement par

David Demirdjian

le 12 Juillet 2000

LE MOUVEMENT PROJECTIF

Théorie et applications pour l'autocalibrage et la segmentation du mouvement

Directeur de thèse : **Radu Horaud**

JURY

Mme. Marie-Paule Cani	Présidente
M. Patrick Gros	Rapporteur
M. Thierry Viéville	Rapporteur
M. Radu Horaud	Directeur de thèse
M. Andrew Zisserman	Examinateur
M. Michael Lindenbaum	Examinateur

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

THÈSE

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité : Imagerie, Vision et Robotique

préparée au laboratoire GRAVIR - IMAG - INRIA
dans le cadre de l'Ecole Doctorale «Mathématiques, Sciences et Technologie de l'Information»

présentée et soutenue publiquement par

David Demirdjian

le 12 Juillet 2000

LE MOUVEMENT PROJECTIF

Théorie et applications pour l'autocalibrage et la segmentation du mouvement

Directeur de thèse : **Radu Horaud**

JURY

Mme. Marie-Paule Cani	Présidente
M. Patrick Gros	Rapporteur
M. Thierry Viéville	Rapporteur
M. Radu Horaud	Directeur de thèse
M. Andrew Zisserman	Examinateur
M. Michael Lindenbaum	Examinateur

Remerciements

Je tiens tout d'abord à remercier Radu Horaud pour m'avoir permis de faire mes « premiers pas » dans le domaine de la vision, il y a maintenant déjà quelques années, puis pour m'avoir fait confiance tout au long de ma présence dans l'équipe movi. Je remercie également, et de façon aussi enthousiaste, Patrick Gros pour ses encouragements, sa disponibilité et ses relexions pertinentes et constructives pendant ces années ainsi que pour son amitié et l'intérêt constant qu'il a porté à mon travail.

Je remercie vivement les personnes qui m'ont fait l'honneur d'avoir participé à mon jury : mes rapporteurs Mm. Patrick Gros et Thierry Viéville pour leurs commentaires constructifs sur le manuscrit, Mme. Marie-Paule Cani pour l'avoir présidé, ainsi que Mm. Andrew Zisserman et Michael Lindenbaum d'avoir été examinateurs.

Roger Mohr a également été un bon chef et collègue et je le remercie pour la confiance qu'il m'a accordée, sur le plan scientifique d'une part, et sur le plan technique et administratif d'autre part, lorsqu'il m'a permis de prendre de réelles responsabilités dans son équipe.

Merci aussi au personnel de l'inria Rhône-Alpes d'avoir assuré nos exceptionnelles conditions de travail, tout particulièrement notre assistante Véronique Roux, pour son efficacité et sa bonne humeur à toute épreuve.

Il est difficile de citer toutes les personnes que j'ai côtoyées et qui ont pu m'aider. Je tiens à remercier toute l'équipe movi pour l'ambiance chaleureuse qui règne dans son sein. Je la remercie plus particulièrement de m'avoir pardonné mes sauts d'humeur quand mes charges d'administration était lourdes, et pour l'échange scientifique constant et enrichissant qui la caractérise.

Merci en particulier à mon ami et co-trans-bureau – *c.-a.-d.* voisin d'aile de bâtiment ou, plus précisément, occupant du bureau situé juste en face du mien – Bart Lamiroy pour nos nombreux fous rires. Il ne manquera pas de remarquer le clin d'œil que je lui fais dans ces remerciements. J'espère que l'amitié qui s'est construite pendant ces années perdurera maintenant que nos chemins se sont séparés.

Merci également à Frédérick Martin pour la collaboration fructueuse et les échanges scientifiques et épistémologiques qui ont beaucoup contribué à ce travail, et à Yves Dufournaud qui, avec ses questions et rélexions techniques, a su me surprendre et qui m'a incité à aller toujours plus loin dans la réflexion scientifique.

À ma mère,

Table des matières

Chapitre 1

Introduction

L'image se trouve aujourd'hui au cœur d'une véritable révolution scientifique et technique. La puissance de calcul mise à disposition et le faible coût des ordinateurs et des caméras ont démocratisé l'usage de la vision dans le monde industriel (contrôle de qualité, fabrication) mais également dans notre quotidien (logiciels de retouches photographiques, vidéophone).

L'apparition et le développement d'internet – et ses avatars liés au téléphone portable et à la télévision interactive – sont en train de transformer cette révolution en phénomène culturel. Qu'on le veuille ou non, une ère du multimédia est en train de se mettre en place, installant les maître mots *image* et *interactivité* dans notre quotidien.

La vision par ordinateur est là pour fournir les outils techniques et scientifiques nécessaires à la construction de cet édifice technologique. Poussée par la demande des applications de surveillance et de robotique, l'analyse de scènes dynamiques est aujourd'hui un des domaines de recherche les plus actifs. Les avancées proposées dans cette thèse entrent dans ce cadre. Plus précisément, elles concernent l'analyse du mouvement dans le cas de la vision stéréoscopique faiblement étalonnée.

Dans de nombreuses applications, la vision stéréoscopique apparaît comme le moyen le plus évident pour obtenir des informations tridimensionnelles à partir d'images. Dans le cadre de la robotique, l'utilisation d'un système stéréoscopique permet d'obtenir des reconstructions tridimensionnelles qui peuvent servir pour des tâches de planification, d'asservissement ou d'évitement d'obstacles. Les applications de surveillance ou d'assistance dans des environnements intelligents s'intéressent plus à la détection et à l'estimation du mouvement d'objets présents dans la scène. Ces applications reposent généralement sur des modèles euclidiens. Les systèmes stéréoscopiques employés doivent alors être fortement

étalonnés, ce qui implique que les paramètres internes des caméras ainsi que la position relative entre les caméras doivent être connus. Or un étalonnage fort et précis nécessite généralement une intervention humaine qui est souvent difficile, voire impossible, dans la plupart de ces applications.

L'utilisation de systèmes faiblement étalonnés (systèmes dont seule la géométrie épipolaire est connue) pourrait être une alternative permettant d'accroître la flexibilité et l'autonomie de ces applications. Un étalonnage faible est très facile à obtenir, mais la difficulté est qu'alors les informations tridimensionnelles obtenues sont projectives et non plus euclidiennes.

Ce document s'inscrit dans une approche basée sur un étalonnage faible et s'intéresse à l'étude d'un système stéréoscopique faiblement étalonné évoluant dans un environnement *a priori* inconnu. Il montre comment, en pratique, on peut tirer partie du mouvement d'un système stéréoscopique pour remonter à la structure métrique de la scène (par auto-étalonnage) et détecter des objets en mouvement. L'espace projectif est utilisé ici pour représenter l'information visuelle issue du système. En particulier, on étudie les transformations projectives *3–D*– appelées également *homographies 3–D* – qui relient les reconstructions projectives d'une scène rigide. On s'intéresse au problème d'estimation de ces homographies *3–D* et on montre comment celles-ci entrent en jeu dans des applications telles que l'auto-étalonnage ou la segmentation du mouvement. Les contributions de cette thèse sont les suivantes.

- Nous proposons un algorithme de poursuite de points dans une séquence d'images stéréoscopiques qui permet de mettre des points en correspondance entre les deux images du système stéréoscopique et également entre des paires d'images consécutives. L'approche ne suppose pas que la scène observée est *a priori* connue ou rigide, et utilise la géométrie épipolaire du système stéréoscopique afin de contraindre les appariements.

- Nous étudions le problème de l'estimation des homographies *3–D* à partir de deux paires d'images. On propose différents critères à minimiser et une réflexion sur le lien entre critères linéaire et géométrique donne lieu à une méthode itérative dont les propriétés de convergence et de précision sont remarquables.

- Nous proposons différentes méthodes d'auto-étalonnage à partir d'homographies *3–D* estimées à partir de mouvements rigides d'un système stéréoscopique. Le cas de mouvements planaires est décrit en détail. On propose également une approche originale, à base de décomposition de Jordan des homographies à l'infini, pour estimer l'étalonnage métrique à partir de l'étalonnage affine. Cette approche permet d'identifier les mouvements critiques correspondant à cette étape et montre comment les résoudre lorsque c'est possible. Une méthode d'auto-étalonnage à partir de la décomposition de Jordan des homographies *3–D* est également proposée.

- Nous démontrons des résultats nouveaux sur l'auto-étalonnage d'un système stéréoscopique à partir de plans. L'approche qui en découle permet d'autoétalonner un système stéréoscopique à partir de 3 ou 4 vues d'une scène plane.

- Nous proposons un cadre de travail complet pour détecter et segmenter le mouvement à l'aide d'un système stéréoscopique faiblement étalonné. L'approche montre bien que les tâches de détection et de segmentation ne nécessitent aucune information *3–D* euclidienne.
- Enfin, une application de détection du regard d'un observateur, réalisée au cours de ces recherches est présentée.

Organisation

Cette thèse repose, pour l'essentiel, sur une série d'articles publiés dans des journaux et conférences. Lorsque ma contribution à ces articles n'a pas été prépondérante, celle-ci a été clairement spécifiée.

Le chapitre 2 introduit les différentes notations et notions géométriques qui sont utilisées tout au long du manuscrit. Comme on le verra par la suite, le calcul des structures et mouvements repose sur des mises en correspondance de points entre plusieurs paires d'images. Cette tâche est étudiée dans le chapitre 3 qui décrit en détail un algorithme de poursuite de points dans une séquence d'images stéréoscopique. L'estimation des homographies *3–D* est abordée dans le chapitre 4 où plusieurs estimateurs sont étudiés et comparés. Le chapitre 5 s'intéresse, dans la cadre de l'auto-étalonnage, au passage de l'affine à l'euclidien. Au cours de cette étude, on montre comment une décomposition de Jordan des homographies à l'infini entre différentes vues permet de résoudre le problème d'étalonnage et d'identifier les mouvements critiques. Le chapitre 6 propose deux approches originales pour l'auto-étalonnage stéréo. La première approche repose sur la décomposition de Jordan des homographies *3–D*. La seconde concerne l'auto-étalonnage à partir de scènes planes et repose sur des considérations géométriques. Enfin le chapitre 7 décrit un cadre de travail permettant de faire la détection et segmentation automatiques du mouvement à l'aide d'un système stéréo faiblement étalonné. On trouvera également en annexe une synthèse sur la détection du regard à l'aide d'un système stéréo réalisée, au cours de la thèse, à l'université de Gênes.

Chapitre 2

Notations et éléments de base

Dans tout le rapport, les notations suivantes sont utilisées :

Typographie

Les vecteurs ($\boldsymbol{m}$, $\boldsymbol{M}$, ...) ..... gras italique

Les matrices ($\mathbf{K}$, $\mathbf{G}_\infty$, ...) ..... gras

Les entités géométriques (points, lignes, plans, ...) ..... italique

Opérateurs mathématiques

$\mathbf{I}_n$ ..... matrice identité d'ordre n

$\bar{\boldsymbol{X}}$ ..... conjugué du vecteur complexe $\boldsymbol{X}$

$\Re(\boldsymbol{X})$ et $\Im(\boldsymbol{X})$ ..... parties réelle et imaginaire du vecteur complexe $\boldsymbol{X}$

$\mathbf{H}^{-1}$ ..... inverse de $\mathbf{H}$

$\mathbf{H}^{\top}$ ..... transposée de $\mathbf{H}$

$\mathbf{H}^{-\top} = (\mathbf{H}^{\top})^{-1}$ ..... inverse de la transposée de $\mathbf{H}$

$[\boldsymbol{u}]_\times$ ..... matrice antisymétrique associée au vecteur $\boldsymbol{u}$

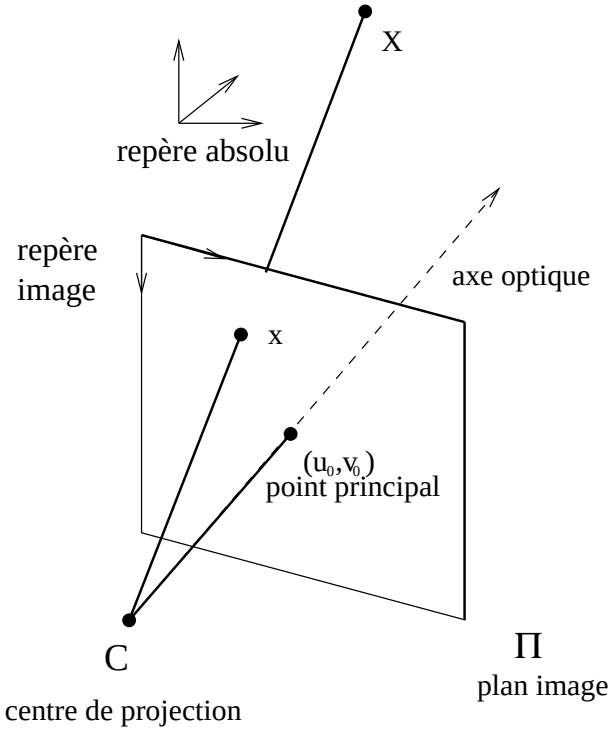

FIG. 2.1: *Modèle de caméra sténopé.*

La géométrie projective est un sujet aujourd'hui bien établi et on suppose dans cette thèse que les notions de base s'y rapportant sont connues du lecteur (à défaut, celui-ci pourra consulter [Fau 93]).

Les points de l'espace sont notés en *coordonnées homogènes*. Un point image (x, y) a comme coordonnées $(x\ y\ 1)^\top$. Un point de l'espace 3-D (X, Y, Z) a comme coordonnées $(X\ Y\ Z\ 1)^\top$. Les coordonnées homogènes sont définies à un facteur multiplicatif près. On note "$\simeq$" l'égalité à un facteur multiplicatif près. Enfin, on note $\mathcal{P}^n$ l'espace projectif de dimension n.

2.1 Formation des images

Le processus de formation des images est une transformation qui projette un point de l'espace 3-D (l'espace physique) dans un espace 2-D (l'image). Cette transformation dépend naturellement du système d'acquisition (caméra, carte d'acquisition, optique) utilisé. Cependant, dans cette thèse, on ne traitera que le cas de la *projection perspective* en laissant de côté les problèmes des distorsions non-linéaires dues aux optiques.

La projection perspective, ou sténopé, modélise une caméra avec un centre de projection C et un plan image Π. Dans un tel modèle, les rayons optiques provenant des points

3-D passent tous par le point C et intersectent le plan Π, formant ainsi les images. Soient $\boldsymbol{X}$ et $\boldsymbol{x}$ les coordonnées homogènes respectives d'un point 3-D et de sa projection dans l'image. La projection s'écrit :

$$\boldsymbol{x} \simeq \mathbf{P}\boldsymbol{X}$$

où $\mathbf{P}$ est une matrice 3×4 appelée *matrice de projection* de la caméra.

Si le centre optique C de la caméra se trouve à l'infini, on parle de *caméra affine*. Dans le cas contraire, on parle de *caméra perspective* et la matrice de projection $\mathbf{P}$ s'écrit alors sous la forme :

$$\mathbf{P} = \mathbf{K}(\ \mathbf{R} \quad \boldsymbol{t}\ )$$

$\mathbf{R}$ et $\boldsymbol{t}$ sont les *paramètres extrinsèques* et correspondent à l'orientation et la position de la caméra par rapport au repère absolu. $\mathbf{K}$ est une matrice 3×3 triangulaire supérieure, appelée matrice des *paramètres intrinsèques* :

$$\mathbf{K} = \begin{pmatrix} \alpha_u & r\alpha_u & u_0 \\ 0 & k\alpha_u & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

- α_u et $\alpha_v = k\alpha_u$ sont les facteurs d'échelles horizontal et vertical ;
- k est le rapport d'aspect ;
- (u_0, v_0) sont les coordonnées du point principal ;
- r est le coefficient de distorsion des axes, ou *skew* en anglais.

En pratique, les axes des caméra sont souvent orthogonaux (ou presque) et l'approximation $r = 0$ se vérifie très bien. Par ailleurs, le rapport d'aspect k est un paramètre très stable de la caméra. Une bonne approximation de k peut alors être obtenue à partir des données-constructeur de la caméra.

Le modèle de projection perspective ne prend pas en compte les distorsions non-linéaires dues aux optiques (zoom, courtes focales). Celles-ci nécessitent l'utilisation d'un modèle de caméra plus complet. Un modèle couramment utilisé contient, outre les paramètres intrinsèques du modèle de projection perspective, des coefficients de distorsion radiale [Bro 71, Tsa 87, Bra 95].

2.2 Reconstruction à partir de paires d'images

2.2.1 Matrices de projection

On s'intéresse, dans cette thèse, à des systèmes composés de deux caméras faisant une acquisition synchronisée d'images que nous appellerons dans la suite *systèmes stéréoscopiques* ou encore *systèmes stéréo*.

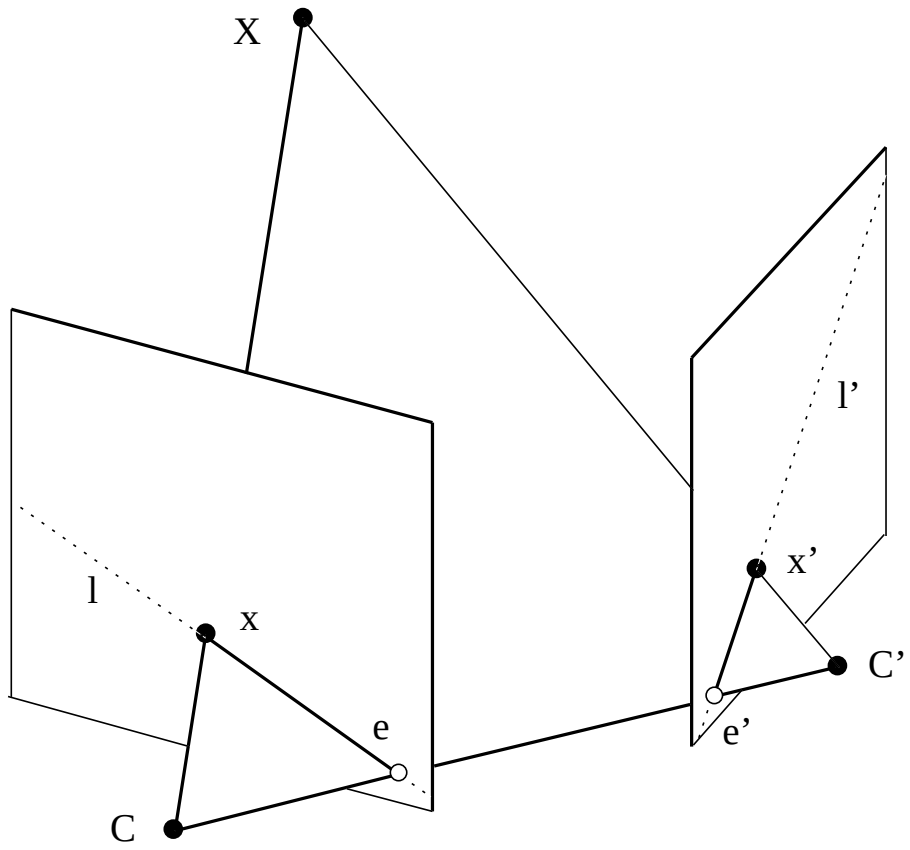

FIG. 2.2: *Modèle de système stéréoscopique.*

Le problème de la reconstruction tridimensionnelle à partir de deux images acquises avec un tel système est maintenant bien connu.

Soient $\mathbf{K}$ et $\mathbf{K}'$ les matrices de paramètres intrinsèques des caméras du système stéréo. Soit $\mathbf{R}$ et $\boldsymbol{t}$ les rotation et translation entre les repères liés aux caméras. Soient $\mathbf{P}$ et $\mathbf{P}'$ les matrices de projection des caméras. Soient $\boldsymbol{x}$ et $\boldsymbol{x}'$ les points correspondant aux images d'un point 3-D physique $\boldsymbol{X}$. On a alors :

$$\left\{ \begin{array}{rcl} \boldsymbol{x} & \simeq & \mathbf{P}\boldsymbol{X} \\ \boldsymbol{x}' & \simeq & \mathbf{P}'\boldsymbol{X} \end{array} \right. \tag{2.1}$$

Le processus de reconstruction stéréo consiste à estimer $\boldsymbol{X}$ à partir de (2.1). La nature de la reconstruction dépend cependant du type d'étalonnage du système stéréo.

- **Cas projectif.** Lorsque seule la matrice fondamentale entre les deux caméras est connue, on parle d'*étalonnage faible*. Un tel étalonnage est facile à obtenir car la matrice fondamentale peut être estimée, en général, à partir de la mise en correspondance de points entre deux images. On peut alors estimer les matrices de projection projectives $\mathbf{P}$ et $\mathbf{P}'$ [Har 97a]. Sans perte de généralité, $\mathbf{P}$ et $\mathbf{P}'$ peuvent s'écrire :

$$\left\{ \begin{array}{rcl} \mathbf{P} & \simeq & \left(\begin{array}{cc} \mathbf{I}_3 & 0 \end{array} \right) \\ \mathbf{P}' & \simeq & \left(\begin{array}{cc} \mathbf{H}_\pi & \beta\boldsymbol{e}' \end{array} \right) \end{array} \right. \tag{2.2}$$

  avec

$$\mathbf{H}_\pi = \mathbf{H}_\infty + \boldsymbol{e}'\boldsymbol{a}^\top$$

  $\mathbf{H}_\infty$ est l'homographie à l'infini entre les deux images et $\boldsymbol{e}'$ l'épipôle de la deuxième image. $\boldsymbol{a}$ est un vecteur arbitraire de $\mathcal{R}^3$ et β un scalaire arbitraire. Le vecteur $(\boldsymbol{a}^\top \ \beta)$ représente les coordonnées du plan à l'infini Π_∞. Les expressions de $\mathbf{H}_\infty$ et $\boldsymbol{e}'$ sont :

$$\left\{ \begin{array}{rcl} \mathbf{H}_\infty & \simeq & \mathbf{K}'\mathbf{R}\mathbf{K}^{-1} \\ \boldsymbol{e}' & \simeq & \mathbf{K}'\boldsymbol{t} \end{array} \right.$$

  Les reconstructions à partir du système précédent sont alors obtenues dans un espace projectif et définies à une reconstruction projective près (15 degrés de liberté).

- **Cas affine.** Si en plus, les coordonnées du plan à l'infini $\Pi_\infty \simeq (\boldsymbol{a}^\top \ \beta)^\top$ est connue, on parle d'*étalonnage affine*. Les équations des matrices de projection correspondantes deviennent :

$$\left\{ \begin{array}{rcl} \mathbf{P} & \simeq & \left(\begin{array}{cc} \mathbf{I}_3 & 0 \end{array} \right) \\ \mathbf{P}' & \simeq & \left(\begin{array}{cc} \mathbf{H}_\infty & \boldsymbol{e}' \end{array} \right) \end{array} \right. \tag{2.3}$$

  Les reconstructions sont alors affines et définies à une transformation affine près (12 degrés de liberté).

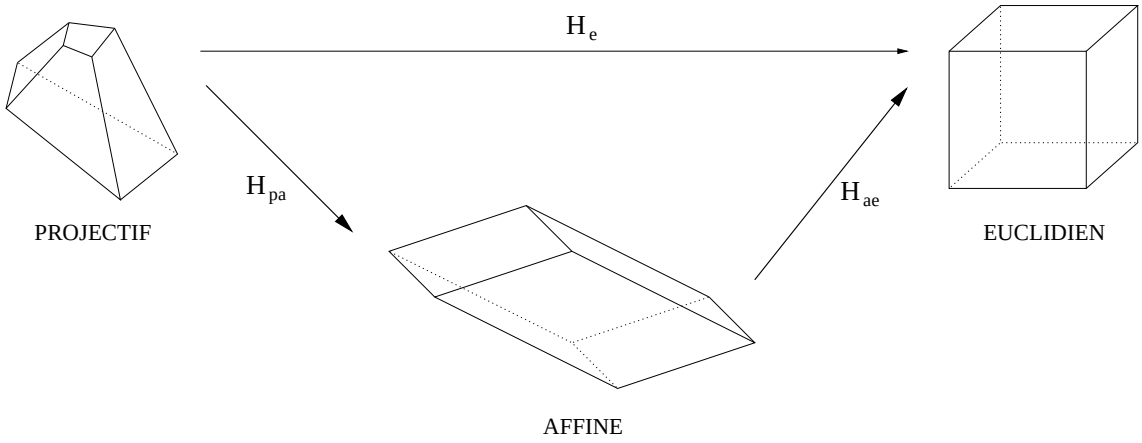

FIG. 2.3: *Reconstructions projective, affine et euclidienne d'un cube.*

– **Cas euclidien.** Enfin, si les paramètres intrinsèques ainsi que la position et orientation relatives des caméras sont connus, on parle d'*étalonnage fort*. Les matrices de projections s'écrivent alors :

$$\left\{ \begin{array}{rcl} \mathbf{P} & \simeq & \mathbf{K} \begin{pmatrix} \mathbf{I}_3 & 0 \end{pmatrix} \\ \mathbf{P}' & \simeq & \mathbf{K}' \begin{pmatrix} \mathbf{R} & \boldsymbol{t} \end{pmatrix} \end{array} \right. \tag{2.4}$$

Les reconstructions sont euclidiennes et donc définies à une transformation euclidienne près (6 degrés de liberté). On peut noter au passage que, contrairement aux cas projectif et affine, le cas euclidien ne permet pas le choix de la matrice canonique $\begin{pmatrix} \mathbf{I}_3 & 0 \end{pmatrix}$ pour $\mathbf{P}$: les matrices $\mathbf{K}$ et $\mathbf{K}'$ apparaissent de façon explicite dans les équations des matrices de projection $\mathbf{P}$ et $\mathbf{P}'$ (dans les cas projectif et affine, $\mathbf{K}$ et $\mathbf{K}'$ étaient "absorbées" par $\mathbf{H}_\pi$ et $\mathbf{H}_\infty$).

Lorsque l'étalonnage est obtenu par une méthode d'auto-étalonnage, la translation $\boldsymbol{t}$ est, en général, connue à un facteur multiplicatif près. A moins de disposer d'informations extérieures (mesures sur la scène, mouvement du capteur) cette incertitude sur la translation ne peut pas être levée. Dans ce cas, la scène est reconstruite à un facteur d'échelle près. On parle alors de reconstruction **métrique**. Pour la plupart des applications, ce facteur d'échelle n'a pas besoin d'être connu et on peut lui donner une valeur arbitraire. C'est pourquoi, par la suite, on confondra les termes *euclidien* et *métrique*.

2.2.2 Reconstruction stratifiée

Soient $\mathbf{H}_{pa}$, $\mathbf{H}_{ae}$ et $\mathbf{H}_{pe}$ les matrices suivantes :

$$\mathbf{H}_{pa} = \begin{pmatrix} \mathbf{I}_3 & 0 \\ \boldsymbol{a}^\top & \beta \end{pmatrix} \qquad \mathbf{H}_{ae} = \begin{pmatrix} \mathbf{K}^{-1} & 0 \\ 0 & 1 \end{pmatrix}$$

et

$$\mathbf{H}_{pe} = \mathbf{H}_{ae}\mathbf{H}_{pa} = \begin{pmatrix} \mathbf{K}^{-1} & 0 \\ \boldsymbol{a}^{\top} & \beta \end{pmatrix}$$

Soient $\boldsymbol{X}_{proj}$, $\boldsymbol{X}_{aff}$ et $\boldsymbol{X}_{eucl}$ les reconstructions projective, affine et euclidienne d'un même point obtenues à partir des systèmes (2.2), (2.3) et (2.4) respectivement. On montre [Hor 98] que l'on a les relations suivantes :

$$\begin{aligned} \boldsymbol{X}_{aff} &\simeq \mathbf{H}_{pa}\boldsymbol{X}_{proj} \\ \boldsymbol{X}_{eucl} &\simeq \mathbf{H}_{ae}\boldsymbol{X}_{aff} \simeq \mathbf{H}_{pe}\boldsymbol{X}_{proj} \end{aligned}$$

Autrement dit, les matrices $\mathbf{H}_{pa}$, $\mathbf{H}_{ae}$ et $\mathbf{H}_{pe}$ sont des matrices de changement de base projectif-affine, affine-euclidien et projectif-euclidien respectivement. Étant donnée une reconstruction projective déterminée à l'aide de (2.2), les reconstructions affines et euclidiennes s'obtiennent par changement de base à l'aide des matrices $\mathbf{H}_{pa}$ et $\mathbf{H}_{pe}$.

2.2.3 Transformations rigides

De même que les reconstructions 3-D, les transformations rigides s'expriment différemment suivant la base dans laquelle elles sont représentées.

- Dans une base euclidienne, une telle transformation prend la forme :

$$\mathbf{D}_{eucl} = \begin{pmatrix} \mathbf{R} & \boldsymbol{t} \\ 0 & 1 \end{pmatrix}$$

  où $\mathbf{R}$ est une matrice 3×3 de rotation et $\boldsymbol{t}$ un vecteur de $\mathcal{R}^3$ arbitraire.

- Dans une base affine, l'expression est alors :

$$\mathbf{D}_{aff} = \begin{pmatrix} \mathbf{A} & \boldsymbol{b} \\ 0 & 1 \end{pmatrix}$$

  où $\mathbf{A}$ et une matrice 3×3 et $\boldsymbol{b}$ un vecteur de $\mathcal{R}^3$ arbitraires.

- Enfin, dans une base projective, une transformation rigide prend la forme d'une matrice $\mathbf{H}$ de dimension 4×4, définie à un facteur multiplicatif près.

Lorsque les reconstructions sont obtenues à l'aide d'un système stéréo rigide, les coordonnées des points 3-D avant et après déplacement rigide peuvent s'exprimer dans la même base (base de reconstruction). En utilisant les matrices de changement de base définies dans la section précédente, on en déduit alors que :

$$\mathbf{D}_{aff} = \mathbf{H}_{ae}\mathbf{D}_{eucl}\mathbf{H}_{ae}^{-1}$$

$$\mathbf{H} = \mathbf{H}_{pe}\mathbf{D}_{eucl}\mathbf{H}_{pe}^{-1}$$

Ce qui est remarquable alors, c'est que les transformations affines $\mathbf{D}_{aff}$ et projectives $\mathbf{H}$ ne sont pas quelconques : celles-ci sont les conjuguées des transformations euclidiennes $\mathbf{D}_{eucl}$ correspondantes. Les groupes des transformations rigides affines et projectives sont isomorphes au groupe des transformations euclidiennes. Ces groupes sont donc de dimension 6 (et non pas 12 et 15 respectivement).

Dans cette thèse, nous nous intéressons à l'utilisation de l'espace projectif comme espace de représentation du mouvement. Dans un tel espace, on ne peut pas obtenir d'informations métriques (en tout cas, pas directement). Cependant, comme on le verra par la suite, cette représentation est suffisante pour de nombreuses applications telles que la segmentation des images à l'aide du mouvement.

Chapitre 3

Suivi de points dans une séquence stéréoscopique

Le problème de mise en correspondance de primitives est une des opérations de bas niveau les plus importantes en traitement d'image et sert de base pour de nombreuses applications telles que la reconstruction, l'estimation de mouvements et l'auto-étalonnage. Ce problème est souvent considéré résolu bien que beaucoup de travail reste à faire dans ce domaine.

Nous nous intéressons ici au cas de la mise en correspondance de points dans une séquence stéréoscopique, appelé aussi **tracking stéréo**. On suppose ici que la fréquence d'acquisition des images est élevée et qu'ainsi le mouvement apparent dans les images est faible. Aucune hypothèse n'est faite sur la nature ou la structure de la scène observée : celle-ci est inconnue et non nécessairement rigide.

3.1 État de l'art

Plusieurs auteurs se sont concentrés sur le problème de tracking dans le cas monoculaire. Les méthodes basées sur le flot optique permettent d'estimer le mouvement d'un ensemble dense de points entre deux images proches et d'assurer une cohérence locale sur le mouvement des points voisins. Plusieurs auteurs ont traité ce problème : Nagel [Nag 83, Nag 87], Anandan [Ana 89] (approche hiérarchique), Barron et al. [Bar 94] (comparaison de différentes approches). L'estimation du mouvement est généralement faite en deux étapes :

1. mise en correspondance d'un ensemble épars de points entre les deux images (par des techniques à base de corrélation) ;

2. régularisation et lissage du mouvement sur l'ensemble des images.

Shi et Tomasi[Shi 94] et Tommasini *et al.*[Tom 98] ont concentré leurs études sur le problème de mise en correspondance éparse. Dans un premier temps, ils modélisent le mouvement d'un point (et de son voisinage) entre deux images consécutives par une translation pure. Une fois le déplacement trouvé (par minimisation d'une erreur résiduelle), ils modélisent le mouvement du point entre la première image où le point apparaît (référence) et l'image courante par une transformation affine afin de ré-estimer plus précisément le déplacement. Un critère est estimé, en même temps que la transformation affine, permettant de décider si le point doit continuer à être suivi: lorsque la déformation affine est trop importante, le point n'est plus suivi.

D'autres approches considère le problème de tracking dans le cadre de la *stéréo étalonnée* [Kel 95, Wan 96, Wen 92, Yi 97]. Elles exploitent la géométrie 3-D de la scène (*a priori* inconnue) en effectuant des reconstructions 3-D explicites. Le suivi des points est fait en utilisant des informations sur la profondeur estimée des points et la géométrie épipolaire du système. Par ailleurs, ces informations sont introduites dans un filtre de Kalman qui permet de prédire la position des points dans les images suivantes. En pratique, ces approches sont assez lourdes dans la mesure où elles supposent que le système stéréo est fortement calibré. De plus, ces approches reposent sur une estimation des profondeurs des points à partir d'images très proches et il n'est pas du tout sûr que la précision sur l'estimation de ces profondeurs soit suffisante [1] pour aider la mise en correspondance.

L'approche que nous proposons est plus souple dans la mesure où elle suppose uniquement une calibration faible du système stéréo. Les mises en correspondance sont établies entre des paires stéréo successives et reposent sur des informations provenant à la fois des images et sur la géométrie épipolaire du système. Dans un premier temps, nous faisons des considérations sur la comparaison de points entre deux images. Nous proposons ensuite un algorithme d'appariement de points entre deux images. Enfin un algorithme de tracking stéréo est décrit en détail.

3.2 Mesures de ressemblance

Le problème du tracking conduit à définir une **mesure de ressemblance** entre deux points (et de leur voisinage) provenant de deux images différentes.

De nombreux critères ont été proposés et sont, pour la plupart, basés sur des mesures de corrélation (voir [Bla 98] pour une revue détaillée de ces critères). Elles mesurent la ressemblance d'un point (u_1, v_1) de l'image I_1 à un point (u_2, v_2) de l'image I_2 sur une fenêtre (ou *patch*) carrée de taille $(2n+1) \times (2n+1)$. Le tableau 3.1 donne les expressions de quatre des mesures les plus utilisées. SAD et SSD correspondent mathématiquement aux norme L_1 et L_2 évaluées entre les fonctions I_1 et I_2 en $(2n+1) \times (2n+1)$ points.

1. il est bien connu que lorsque des points sont reconstruits dans l'espace 3-D à partir d'images proches, la variance sur l'estimation de la profondeur de ces points est très grande.

ZSAD et ZSSD sont des mesures centrées, obtenues par soustraction de la moyenne locale des intensités $\bar{I}_1$ et $\bar{I}_2$ sur les fenêtres considérées.

Mesure	Expression
SAD	$\sum_{du=-n}^{du=+n} \sum_{dv=-n}^{dv=+n} \lvert I_2(u_2+du, v_2+dv) - I_1(u_1+du, v_1+dv)\rvert$ (3.1)
ZSAD	$\sum_{du=-n}^{du=+n} \sum_{dv=-n}^{dv=+n} \lvert I_2(u_2+du, v_2+dv) - \overline{I_2} - I_1(u_1+du, v_1+dv) + \overline{I_1}\rvert$ (3.2)
SSD	$\sum_{du=-n}^{du=+n} \sum_{dv=-n}^{dv=+n} (I_2(u_2+du, v_2+dv) - I_1(u_1+du, v_1+dv))^2$ (3.3)
ZSSD	$\sum_{du=-n}^{du=+n} \sum_{dv=-n}^{dv=+n} (I_2(u_2+du, v_2+dv) - \overline{I_2} - I_1(u_1+du, v_1+dv) + \overline{I_1})^2$ (3.4)

TAB. 3.1: *Quatre mesures usuelles de corrélation.*

Des mesures robustes aux occultations ont également été étudiées. On peut citer par exemple *rank* et *cencus* [Zab 94] ou encore RZSSD [Lan 97].

Les mesures de ressemblance par corrélation (classiques ou robustes) ne s'appliquent pas en cas de rotation ou de changement d'échelle. Des mesures basées sur des invariants [Sch 96, Duf 00] du signal-image doivent alors être utilisées.

3.3 Quels points apparier dans les images?

Les mesures de ressemblance citées précédemment permettent, dans le cadre du tracking, de mettre en correspondance des points entre des images successives d'une séquence.

Il est cependant clair que tous les points ne peuvent pas être appariés avec la même confiance. Le problème est de même nature que le *problème d'ouverture* développé dans le cadre de l'estimation du flot optique. Il est bien connu qu'un point se trouvant dans une zone de niveau de gris homogène a beaucoup moins de chances d'être bien apparié qu'un point se trouvant dans la zone d'un coin. Plus la texture présente autour d'un point est riche et variée, plus l'information image est discriminante.

Des travaux tels que [Har 88, Shi 94] visent à trouver un critère permettant d'évaluer la quantité d'information image présente autour d'un point. Les points les plus discriminants appelés **points d'intérêt** sont les points dont les variations d'intensité lumineuse sont fortes dans deux directions différentes. De nombreux détecteurs de points d'intérêt ont été proposés. La plupart de ces détecteurs sont évalués et comparés dans [Sch 96] et le détecteur de Harris [Har 88] apparaît comme celui possédant les meilleurs performances en terme de *répétabilité* [Sch 96] – critère définissant l'aptitude du détecteur à extraire

le même point dans des images obtenues dans des conditions (point de vue, éclairage) différentes.

Ainsi, les points d'intérêt sont des zones de l'image où l'information est la plus riche. Étant données deux images, la recherche d'appariements conduit naturellement à extraire puis apparier les points d'intérêt. Par opposition aux méthodes de flot optique qui mettent en correspondance des ensembles denses de points, les méthodes qui se basent sur l'appariement des points d'intérêt *seuls* ont deux avantages considérables : la recherche d'appariements est plus rapide – l'ensemble des points à apparier est réduit – et les appariements réalisés sont plus fiables – les informations contenues par ces points étant les plus discriminantes.

Il faut cependant garder à l'esprit que la notion de point d'intérêt est trompeuse. Un point d'intérêt ne représente ni un point anguleux, ni un coin. C'est le centre d'une région de l'image dans laquelle la texture varie beaucoup. Par ailleurs, un point d'intérêt se trouve souvent sur un contour d'occultation et correspond alors à un point *3–D* instable. Sa localisation est donc, par conséquent, sujette à une incertitude dont il faut tenir compte par la suite.

3.4 Appariement de points dans deux images

Les sections précédentes nous ont appris comment évaluer la ressemblance de points dans des images différentes et indiqué les points qui pouvaient, et donc devaient, être comparés. Cette partie propose un algorithme pour effectuer la mise en correspondance *globale* d'un ensemble de points dans deux images. Cet algorithme sera ensuite étendu naturellement dans le cadre du tracking stéréo.

3.4.1 Problème

Soient I_1 et I_2 deux images. Soient $\mathcal{P} = (p_1 ... p_n)$ et $\mathcal{Q} = (q_1 ... q_m)$ deux ensembles de points extraits de I_1 et I_2 respectivement.

L'algorithme de mise en correspondance globale consiste à déterminer une *fonction d'appariement* $r : \mathcal{P} \mapsto \mathcal{Q}$ injective qui attribue à chaque point p_i de $\mathcal{P}$ un correspondant $q_j = r(p_i)$ de $\mathcal{Q}$ tel que p_i et q_j respectent des contraintes de ressemblance. En pratique, rien ne garantit que tous les points de $\mathcal{P}$ ont un correspondant dans $\mathcal{Q}$ et tous les points p_i peuvent ne pas être appariés.

Le problème de mise en correspondance est en général très difficile car :

- lorsque les conditions d'observation (point de vue, éclairage) sont différentes entre les images, il est difficile d'évaluer la ressemblance entre des points ;
- en présence de plusieurs points localement similaires (textures répétitives), les possibilités d'appariement sont multiples et il est impossible de supprimer les ambiguïtés uniquement à l'aide de critères locaux ;

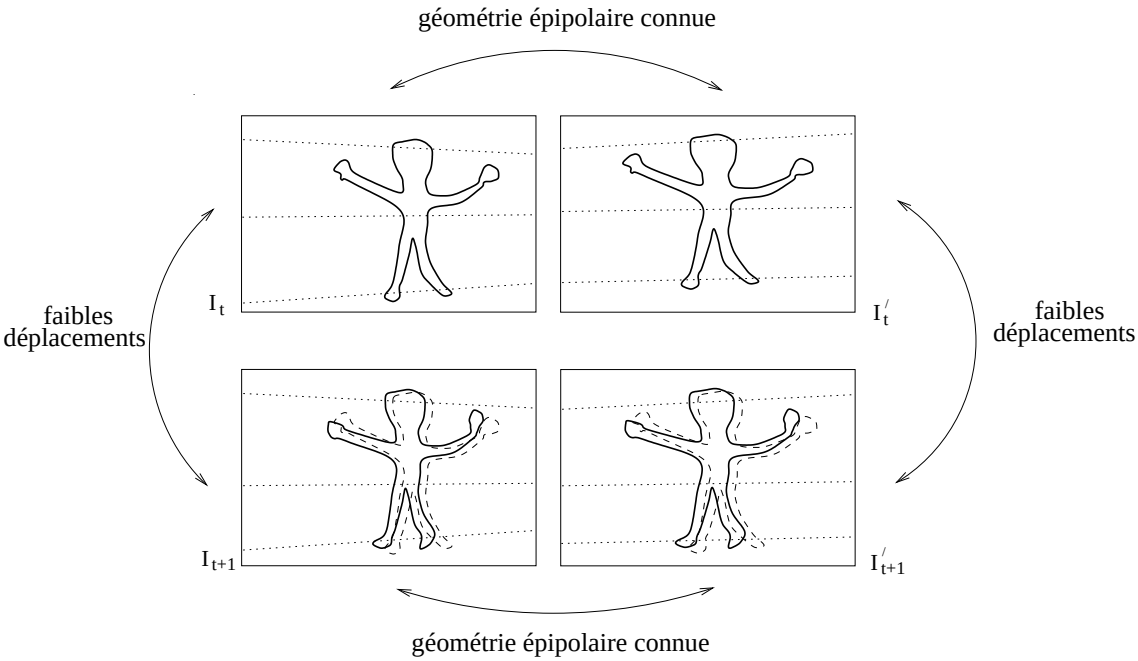

FIG. 3.1: *Deux paires d'images consécutives d'une séquence stéréo. La géométrie épipolaire entre les images de gauche (I_t) et droite (I'_t) est connue. Les mouvements apparents entre deux images consécutives de la séquence (I_t et I_{t+1} par exemple) sont faibles.*

– l'algorithme d'appariement est un problème de complexité combinatoire [Bla 98]. Cependant la complexité du problème peut être réduite lorsque des heuristiques ou informations supplémentaires (*p.e.* géométrie épipolaire, mouvement du capteur) sont utilisées.

3.4.2 Hypothèses pour l'appariement

Dans le cadre du tracking stéréo, on considère que les déformations entre les images sont peu importantes entre les images consécutives du même capteur mais également, dans une moindre mesure, entre les images des deux capteurs prises au même instant. Les points en correspondance se trouvent dans des zones d'image semblables. L'utilisation d'une mesure de corrélation standard comme critère de ressemblance des points se justifie alors pleinement. En pratique, on utilise la mesure SSD avec une taille de masque 5×5 ou 7×7. Il faut cependant admettre que dans le cas d'une paire d'images stéréo, les déformations des images sont plus importantes que dans le cas d'images consécutives d'une séquence monoculaire : la mise en correspondance est alors plus difficile.

Par ailleurs, il existe des relations particulières entre les images de deux paires stéréo consécutives (voir Figure 3.1). Nous allons considérer deux cas : le cas (i) de deux images consécutives relatives au même capteur et le cas (ii) d'une paire stéréo d'images prises au

même instant par les deux capteurs d'un système stéréo dont la géométrie épipolaire est connue. Exploiter ces relation permet de :

- réduire la complexité du problème d'appariement ;
- réduire les ambiguïtés lors de l'appariement.

En reprenant les notations de la section précédente, on note p_i un point de $\mathcal{P}$ et $r(p_i)$ son correspondant dans $\mathcal{Q}$. La Figure 3.1 illustre les relations qui existent entre deux paires consécutives d'une séquence stéréo.

(i) faible déplacements. Ce cas correspond à des images très proches (*p.e.* deux images consécutives I_t et I_{t+1} d'une séquence monoculaire). $\vec{v}_i = r(p_i) - p_i$ représente le mouvement apparent du point p_i. Ce mouvement est supposé faible et, en pratique, on définit un seuil s_{motion} et $||\vec{v}_i||$ doit être inférieure à s_{motion}.

(ii) géométrie épipolaire connue. Lorsqu'on considère deux images d'une même scène, prises au même instant t, il existe une géométrie épipolaire entre les deux images. Soit $\mathbf{F}$ la matrice fondamentale entre les deux images. Les points p_i et $r(p_i)$ doivent se trouver sur des droites épipolaires correspondantes. Soit $d_{epip}(p_i, r(p_i))$ le critère défini [Zha 96] par :

$$d_{epip}(p_i, r(p_i)) = d^2(p_i, \mathbf{F}r(p_i)) + d^2(r(p_i), \mathbf{F}^\top p_i)$$

où $d(x, l)$ représente la distance euclidienne d'un point x à une droite l. Ce critère correspond à la distance des points aux épipolaires. En théorie, cette distance devrait être nulle mais, en pratique, elle ne l'est pas à cause des erreurs de localisation des points p et q et d'estimation de $\mathbf{F}$. On définit alors un seuil s_{epip} et $d_{epip}(p_i, r(p_i))$ doit être inférieur à s_{epip}.

3.4.3 Algorithme "Best Confident First" (BCF)

Il existe différents algorithmes d'appariement. Les plus répandus sont "Winner Takes All" (WTA) et "Cross Check Raw" (CCR).

L'algorithme WTA consiste simplement à chercher, successivement, pour chaque point p de $\mathcal{P}$ le meilleur correspondant q de $\mathcal{Q}$. Si le score d'appariement entre p et q est en dessous d'un seuil défini à l'avance, p et q sont appariés.

L'algorithme CCR [Fua 91] consiste à effectuer une corrélation croisée. Pour chaque point p de $\mathcal{P}$, on cherche le meilleur correspondant q de $\mathcal{Q}$. On cherche alors, dans $\mathcal{P}$, le meilleur correspondant de q. Si le meilleur correspondant de q est p alors p et q sont appariés, sinon p et q n'ont pas de correspondant. L'algorithme CCR, contrairement à WTA, donne un rôle symétrique aux deux images. Cependant il est très strict et ne tolère pas d'ambigüités : si dans une zone de l'image, plusieurs points se ressemblent (*p.e.* cas

d'une texture répétitive), il y a de fortes chances pour que CCR n'apparie aucun des points de la zone.

L'algorithme que nous proposons peut être vu comme une extension de CCR. Dans la suite, nous appellerons cet algorithme "Best Confident First" (BCF). Cet algorithme consiste tout d'abord à créer une liste de tous les appariements possibles entre les deux images (en tenant compte, lorsque c'est possible, des hypothèses faites dans la section précédente). Puis les appariements sont estimées à partir de la liste, un à un, en commençant par ceux qui correspondent aux meilleurs scores.

Création d'une liste d'appariements potentiels

La première étape consiste à définir une liste de correspondances (p, q) potentielles. Pour chaque point p de $\mathcal{P}$, on cherche les correspondants potentiels q de $\mathcal{Q}$.

p et q sont des correspondants potentiels si :

$$\begin{cases} SSD(p,q) & < & s_{corr} & \\ ||p-q|| & < & s_{motion} & \text{si hypothèse "faibles déplacements"} \\ d_{epip}(p,q) & < & s_{epip} & \text{si hypothèse "géométrie épipolaire connue"} \end{cases}$$

où s_{corr}, s_{motion} et s_{epip} sont des seuils définis par avance.

Une liste de correspondances $\mathcal{L}$ est constituée avec tous les appariements potentiels (p, q) et ensuite triée suivant la valeur croissante de $SSD(p, q)$. On note $\mathcal{L}_t$ la liste triée.

Extraction des appariements

L'algorithme consiste ensuite à extraire de $\mathcal{L}_t$ les bons appariements en considérant d'abord les appariements correspondant aux meilleurs scores :

1. le premier couple (p, q) de $\mathcal{L}_t$ (*c.-a.-d.* le couple (p, q) dont la SSD est la plus faible) est considéré comme apparié ;
2. tous les autres couples dans lesquels p ou q apparaît sont supprimés de $\mathcal{L}_t$;
3. si $\mathcal{L}_t$ n'est pas vide, retour en 1.

On constate que les appariements (p, q) vérifiant les hypothèses de la corrélation croisée CCR sont retrouvés par l'algorithme BCF et que l'on obtient, en plus, d'autres appariements. D'autre part, l'algorithme donne un rôle symétrique aux deux images et ne dépend pas de l'ordre dans lequel les points sont considérés.

Le principal avantage de cette méthode, c'est qu'elle est globale. La liste $\mathcal{L}$ contient la liste de tous les appariements potentiels. A partir de là, on peut imaginer plusieurs modifications de l'algorithme. On pourrait par exemple filtrer la liste $\mathcal{L}$ en éliminant les

appariements qui ne vérifieraient pas certaines contraintes semi-locales (contraintes de voisinage, contraintes géométriques). Si l'on voulait imposer que des points voisins dans une image aient leurs correspondants voisins dans l'autre image, on pourrait parcourir la liste $\mathcal{L}$ et supprimer les appariements (p, q) tels que p a un voisin dont le correspondant dans l'autre image n'est pas voisin de q.

3.5 Poursuite de points dans une séquence d'images

L'algorithme de tracking stéréo consiste à poursuivre dans une séquence, des paires de points appariés entre les deux images prises par un système stéréo. Dans la suite, on note (I_t, I'_t) les paires d'image stéréo prises à l'instant t.

Dans le cas des séquences stéréoscopiques que l'on considère, les hypothèses sont les suivantes :

- les séquences sont acquises avec un système stéréoscopique composé de deux caméras similaires et de *baseline* standard (distance entre les deux caméras inférieure à 50 cm). **Les déformations entre les images de gauche et de droite (*i.e.* entre I_t et I'_t), à un instant donné, sont peu importantes** ;
- les deux caméras qui composent le système stéréoscopique sont rigidement liées et leurs paramètres intrinsèques sont constants. **La géométrie épipolaire du système est constante dans le temps**. On a donc la même géométrie épipolaire entre toutes les paires (I_t, I'_t).[2]

On suppose que la géométrie épipolaire entre les deux capteurs est connue et on note $\mathbf{F}$ la matrice fondamentale correspondante. L'algorithme de poursuite que l'on propose peut alors être résumé de la façon suivante.

Dans la première paire stéréo (I_0, I'_0), des points sont appariés à l'aide de l'algorithme BCF décrit précédemment. Les appariements réalisés correspondent alors à des paires de points qui sont poursuivis dans les images suivantes.

Soit (p, p') une paire de points appariés à l'instant t dans la paire stéréo (I_t, I'_t). La poursuite consiste à trouver les correspondants (q, q') de (p, p') dans la paire (I_{t+1}, I'_{t+1}). Les considérations que l'on fait sont les suivantes : (i) q et q' doivent vérifier la géométrie épipolaire du système et (ii) p, p', q et q' doivent correspondre à des textures similaires.

Le couple (q, q') est trouvé à partir de l'algorithme BCF adapté à la stéréo et décrit ci-dessous.

Création de la liste des correspondants potentiels. Pour chaque paire (p, p'), on cherche les correspondants potentiels (q, q'). On cherche une liste $\{q_k\}$ (*resp.*

2. Dans la mesure où la scène observée n'est pas nécessairement rigide, il existe, bien sûr, une géométrie épipolaire entre les images I_t et I_{t+1} ou entre I'_t et I'_{t+1} mais celle-ci n'est pas exploitée puisque les points observés ne la vérifient pas.

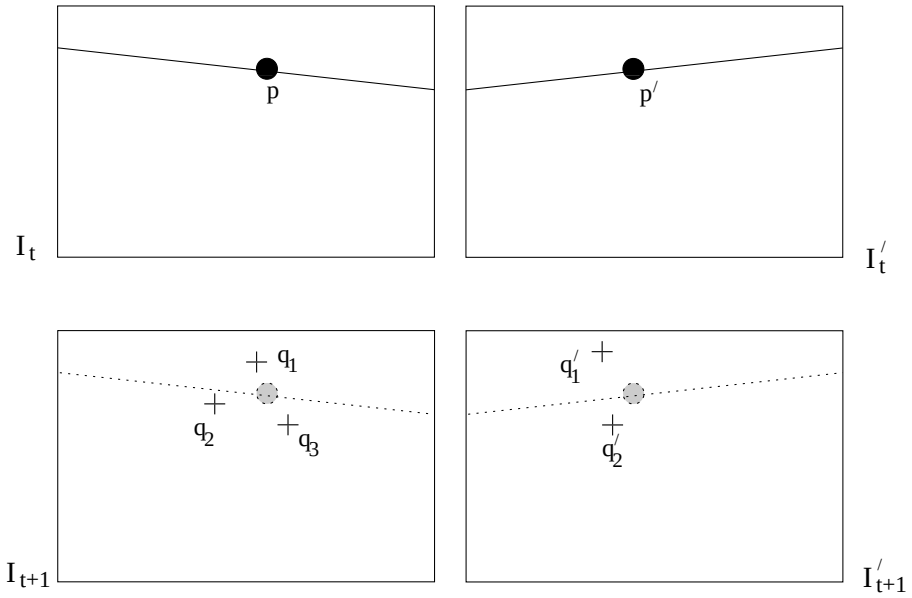

FIG. 3.2: *Algorithme de poursuite stéréo : recherche des correspondants potentiels (q,q') de (p,p') entre deux paires d'images consécutives.*

$\{q'_l\}$) de correspondants potentiels de p (*resp.* p') dans I_{t+1} (*resp.* I'_{t+1}) vérifiant les hypothèses «faibles déplacements». Parmi toutes les combinaisons (q_k, q'_l), on ne considère comme correspondant potentiel que les paires de points ayant des profils de niveaux de gris similaires et vérifiant au mieux la géométrie épipolaire :

$$\begin{cases} SSD(q_k, q'_l) & < & s_{corr} \\ d_{epip}(q_k, q'_l) & < & s_{epip} \end{cases}$$

Une liste $\mathcal{L}$ de tous les appariements $((p, p'), (q_k, q'_l))$ potentiels est créée. Soit $J(p, p', q_k, q'_l)$ le critère défini par :

$$J(p, p', q_k, q'_l) = SSD(p, q_k) + SSD(p', q'_l) + SSD(q_k, q'_l)$$

Les valeurs de $J(p, p', q_k, q'_l)$ faibles correspondent à des points p, p',q_k, q'_l de textures voisines. On trie alors $\mathcal{L}$ suivant l'ordre croissant des valeurs du critère $J(p, p', q_k, q'_l)$. On note $\mathcal{L}_t$ la liste triée correspondante.

Extraction des appariements. L'extraction des correspondances est obtenue comme dans le cas d'appariements entre deux images.

1. le premier quadruplet (p, p', q_k, q'_l) de $\mathcal{L}_t$, *c.-a.-d.* le quadruplet correspondant à la valeur $J(p, p', q_k, q'_l)$ la plus faible est considéré comme apparié ;
2. tous les autres quadruplets dans lesquels p, p',q_k ou q'_l apparaît sont supprimés de $\mathcal{L}_t$;
3. si $\mathcal{L}_t$ n'est pas vide, retour en 1.

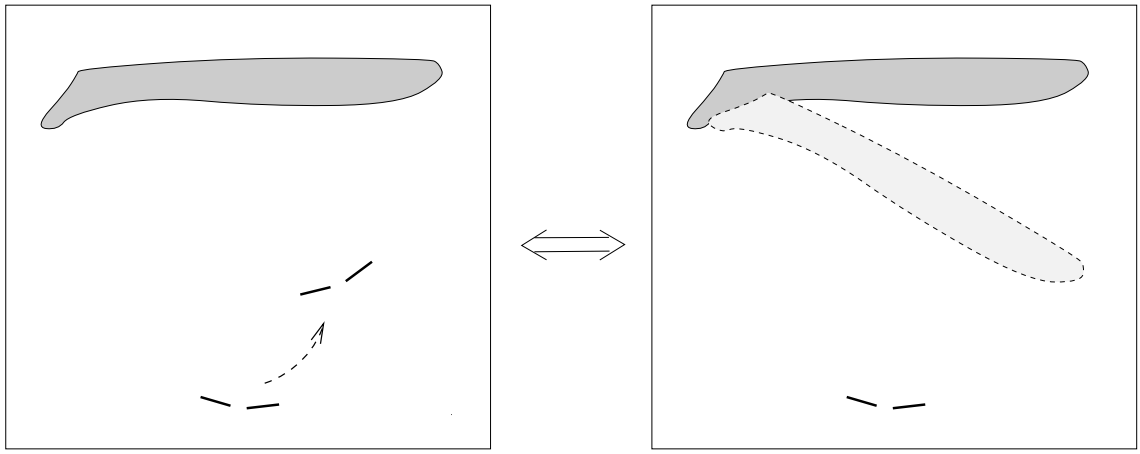

FIG. 3.3: *Estimation de la géométrie épipolaire avec deux paires d'images. Un système stéréo en mouvement observant une scène quasi-planaire (à gauche). Si une seule des deux paires d'images est utilisée pour l'estimation de la géométrie épipolaire, le calcul sera probablement instable. Si les deux paires sont utilisées, cela revient à utiliser une paire d'image correspondant à la structure 3-D (à droite) et l'estimation sera meilleure.*

3.6 Système stéréo rigide

L'avantage d'utiliser un système stéréo *rigide* – mais néanmoins mobile – c'est que la géométrie épipolaire associée aux deux capteurs du système est constante.

Ceci implique que cette géométrie épipolaire peut être estimée à partir de plusieurs paires d'images. Son estimation est bien connue pour être sensible aux dégénérescences (*p.e.* mouvements critiques, scènes planaires). L'utilisation de plusieurs paires d'images permet de supprimer plusieurs de ces dégénérescences et rend par ailleurs l'estimation plus stable et précise (voir Figure 3.3).

Ainsi, dans le cadre du tracking stéréo, même si la géométrie épipolaire donnée à l'algorithme est approximative, sa mise à jour itérative dans la boucle de poursuite permet d'obtenir au bout de quelques paires d'images une très bonne estimation (en pratique, on considère qu'une dizaine de paires d'images suffit pour avoir une estimation correcte).

3.7 Résumé de l'algorithme de tracking stéréo

- **Initialisation.** Des appariements de points sont effectués entre les images I_0 et I'_0 à l'aide de l'algorithme BCF (géométrie épipolaire connue).
- **Poursuite.** On suppose que des paires (p, p') ont été poursuivies jusqu'au temps t :
  - **Mises en correspondance.** On cherche alors les correspondants de toutes les

paires (p, p') dans la paires d'images (I_{t+1}, I'_{t+1}). Si aucune paire correspondante n'a été trouvée pour une paire (p, p'), celle-ci n'est plus poursuivie.

- **Estimation de la géométrie épipolaire (facultatif).** Toutes les paires de points appariés dans la séquence $(I_0, I'_0)...(I_{t+1}, I'_{t+1})$ sont utilisées pour mettre à jour l'estimation de la géométrie épipolaire.
- **Introduction de nouveaux points.** S'il reste des points dans la paire d'images (I_{t+1}, I'_{t+1}) qui ne sont pas appariés, on utilise l'algorithme BCF pour les apparier. Les appariements constituées sont considérés comme des points apparaissant dans la scène. Ces points viennent s'ajouter aux autres et sont poursuivis dans le reste de la séquence.

3.8 Résultats expérimentaux

L'algorithme décrit dans la section précédente a été appliqué sur différentes séquences. Les Figures 3.4, 3.5 et 3.6 montrent des résultats obtenus avec des séquences, de différentes natures, comportant chacune plusieurs dizaines de paires d'images.

Sur une séquence comme 3.5 l'algorithme stéréo donne des résultats similaires à ceux obtenus par un algorithme de tracking monoculaire classique car ce genre de scène est assez structurée : il y a peu de points voisins similaires et donc peu d'ambigüités potentielles.

L'efficacité de notre approche se révèle surtout sur des séquences plus "réelles" comme 3.4 ou 3.6. La géométrie épipolaire permet, sur ces séquences, d'identifier des correspondances correctes sur des zones difficiles comme les textures répétitives (*p.e.* pantalons à carreaux de l'homme de 3.4, boutons de la veste de l'homme de 3.6). De telles mises en correspondance seraient très difficiles à obtenir avec un algorithme monoculaire.

3.9 Discussion

Nous avons présenté ici un algorithme de poursuite de points dans une séquence d'images stéréo. Cet algorithme fait des hypothèses sur les images (faibles déformations entre les images, faibles mouvements) qui sont généralement vérifiées en pratique.

Nous avons appliqué l'algorithme à différentes scènes et les résultats obtenus sont très satisfaisants. L'utilisation de la géométrie épipolaire comme contrainte globale pour les appariements semble pertinente. Une des objections qui pourrait être faite à cette approche, c'est que celle-ci nécessite l'extraction de points d'intérêt dans *toutes* les images et que cette extraction est reconnue pour être lente. Ceci est vrai si l'on n'utilise que des moyens logiciels. Cependant une façon de résoudre ce problème pourrait être d'utiliser du matériel (DSP) ou logiciel spécialisé pour l'extraction de points d'intérêt (cette opération ne correspond finalement qu'à l'application de différents filtres sur l'image).

Un des points sur lesquels l'algorithme pourrait être amélioré concerne la comparaison entre les images d'une même paire. En effet, lorsque les paramètres des deux caméras

FIG. 3.4: *Séquence stéréoscopique 1. Ligne 1 : 306 points extraits et mis en correspondance dans la paire stéréo 1. Ligne 2 : points poursuivis jusqu'à la paire stéréo 7. Ligne 3 : points poursuivis jusqu'à la paire stéréo 20.*

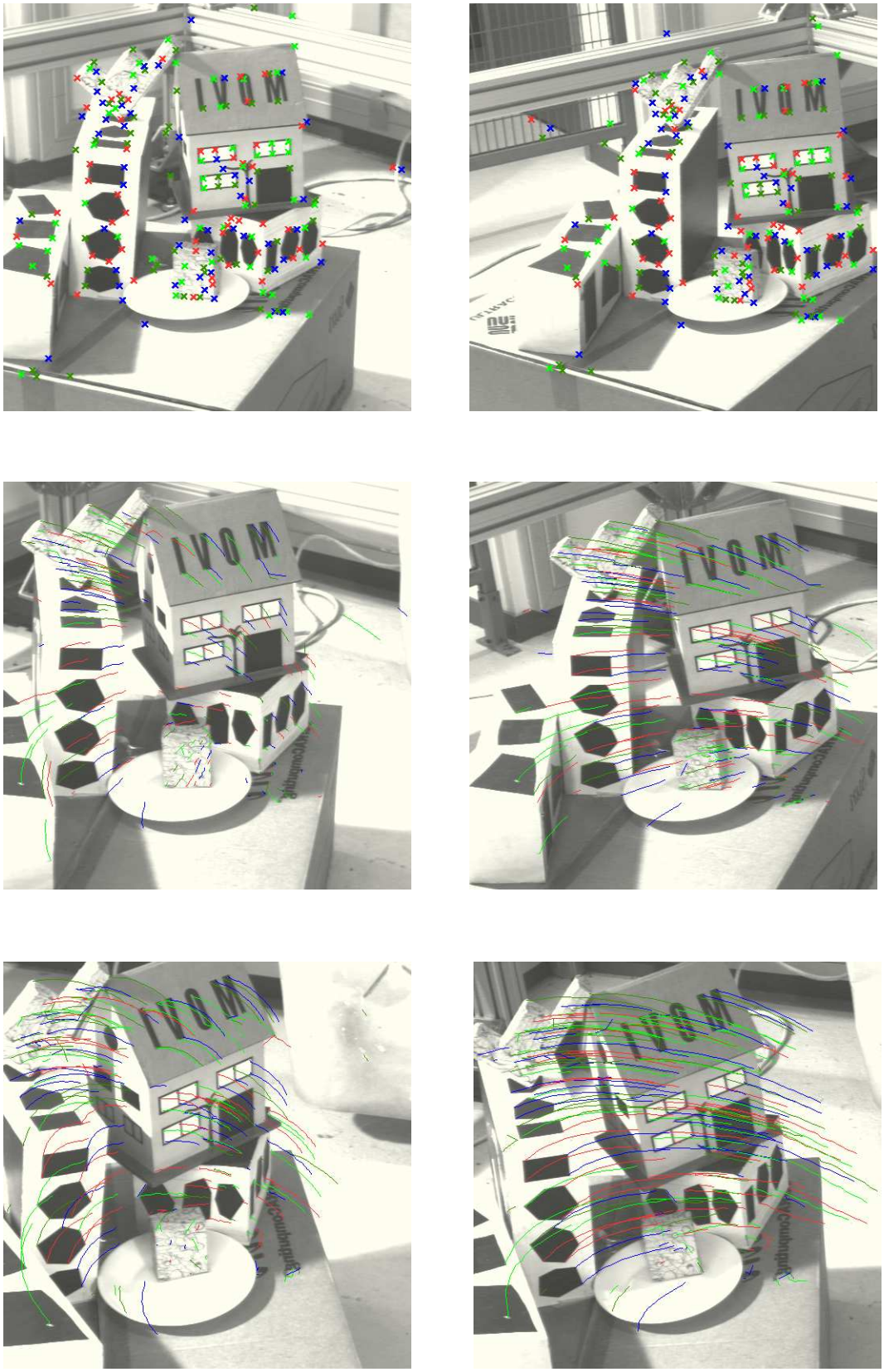

FIG. 3.5: *Séquence stéréoscopique 2. Ligne 1 : 256 points extraits et mis en correspondance dans la paire stéréo 1. Ligne 2 : points poursuivis jusqu'à la paire stéréo 10. Ligne 3 : points poursuivis jusqu'à la paire stéréo 25.*

FIG. 3.6: *Séquence stéréoscopique 3. Ligne 1 : 193 points extraits et mis en correspondance dans la paire stéréo 1. Ligne 2 : points poursuivis jusqu'à la paire stéréo 6. Ligne 3 : points poursuivis jusqu'à la paire stéréo 23.*

(focales, luminosité) ne sont pas exactement les mêmes, la mesure SSD n'est plus vraiment adéquate. Dans un tel cas, une rectification des images à l'aide de la géométrie épipolaire et une normalisation des intensités lumineuses pourraient être effectuées.

Chapitre 4

Estimation de Transformations Projectives 3–D

Dans ce chapitre, nous nous intéressons au problème de l'estimation des transformations projectives de $\mathcal{P}_3$ dans $\mathcal{P}_3$. Ces transformations sont au coeur de toute cette thèse.

Considérons un ensemble de points de l'espace *3–D* en mouvement rigide par rapport à un repère fixe. Selon que ce repère est muni d'une base euclidienne, affine ou projective, la transformation entre deux positions quelconques de l'ensemble de points est euclidienne, affine ou projective.

Dans le cadre de la vision stéréoscopique, les coordonnées des points *3–D* sont en général obtenues, pour chaque position, par triangulation à partir de deux images. Suivant que le système stéréoscopique est fortement ou faiblement étalonné, la base de reconstruction est euclidienne ou projective.

Si le système stéréo est **fortement étalonné**, la base de reconstruction est euclidienne. La transformation entre deux positions de l'ensemble de points est euclidienne et se décrit alors en termes de rotation $\mathbf{R}$ et translation $\boldsymbol{t}$. De nombreuses méthodes d'estimation de $\mathbf{R}$ et $\boldsymbol{t}$ ont été proposées : méthodes basées sur la SVD (Décomposition en Valeurs Singulières) [Aru 87, Ume 91], méthodes utilisant les quaternions [Hor 88], méthodes prenant en compte l'hypothèse d'hétéroscédasticité du bruit (bruit non homogène et non isotrope) sur les données *3–D* [Oht 98, Mat 99].

Si le système stéréo est **faiblement étalonné**, la base de reconstruction est projective [Fau 92, Har 92]. La transformation entre deux positions de l'ensemble de points est projective et peut être alors représentée par une matrice 4×4 définie à un facteur multiplicatif près. Contrairement au cas euclidien, le cas projectif a été très peu traité : en dehors de

[Bea 95a] qui aborde brièvement le sujet, on ne trouve aucune publication sur le problème d'estimation des transformations projectives *3–D*.

Les articles ci-dessous exposent et comparent différentes méthodes d'estimation de transformations projectives reliant les reconstructions projectives *3–D* d'une scène en mouvement observée par un capteur stéréoscopique faiblement étalonné.

4.1 Résumé de «Finding the Collineation between Two Projective Reconstructions» - CVIU

Cet article, écrit avec Gabriella Csurka et Radu Horaud, traite en détail de l'estimation des transformations projectives *3–D*, que nous appelons par la suite *homographies 3–D*. Ma contribution dans cet article a principalement portée sur la section 4 (estimation robuste). J'ai contribué dans une moindre mesure aux sections 1, 2 et 3 et pas du tout à la section 5.

On suppose ici qu'un capteur stéréoscopique observe un ensemble rigide de m points en mouvement. On note $\boldsymbol{x}_i$ et $\boldsymbol{x}'_i$ (*resp.* $\boldsymbol{y}_i$ et $\boldsymbol{y}'_i$) les images de ces points avant (*resp.* après) mouvement. On note $\boldsymbol{X}_i$ (*resp.* $\boldsymbol{Y}_i$) les coordonnées homogènes des points reconstruits projectivement avant (*resp.* après) mouvement à partir des points image. Il existe alors une homographie $\mathbf{H}$ telle que, pour tout i, il existe un scalaire μ_i, tel que :

$$\mu_i \boldsymbol{Y}_i = \mathbf{H}\boldsymbol{X}_i \tag{4.1}$$

$\mathbf{H}$ est une matrice 4×4. $\mathbf{H}$ est une matrice de changement de base projective et est donc définie de façon unique – à un facteur multiplicatif près – à partir de cinq points non coplanaires.

La difficulté dans l'estimation de $\mathbf{H}$ réside en l'absence de métrique dans l'espace projectif *3–D* et donc de modèle d'erreur sur les données. Le choix d'un critère à minimiser (ainsi que la méthode pour minimiser ce critère) est alors crucial. Deux approches sont proposées.

La première approche est linéaire et ne tient pas compte des incertitudes liées aux données. La linéarité de (4.1) est alors exploitée et permet de définir un **critère linéaire**.

$$E_{lin} = \sum_i ||\mathbf{B}_i \boldsymbol{h}||^2 \tag{4.2}$$

où $\boldsymbol{h}$ est un vecteur de $\mathcal{R}^{16}$ dont les composantes sont les éléments de $\mathbf{H}$. Ce critère, norme d'une fonction linéaire en les éléments de $\mathbf{H}$, peut être minimisé par une méthode linéaire classique comme la SVD.

La deuxième approche est non-linéaire et utilise les incertitudes liées aux données. Les erreurs de reconstruction sont dues aux erreurs de localisation des points dans les images et on fait l'hypothèse que ces erreurs sont gaussiennes, homogènes et isotropiques. On

introduit alors un **critère géométrique**. Ce critère est défini comme la distance entre les reprojections estimées $\tilde{\boldsymbol{y}}_i$ et $\tilde{\boldsymbol{y}}'_i$ des points $\tilde{\boldsymbol{Y}}_i = \mathbf{H}\boldsymbol{X}_i$ dans les deux images du capteur stéréoscopique et les projections mesurées $\boldsymbol{y}_i$ et $\boldsymbol{y}'_i$ des points $\boldsymbol{Y}_i$.

$$E_{geom} = \sum_i d(\boldsymbol{y}, \tilde{\boldsymbol{y}}_i)^2 + d(\boldsymbol{y}', \tilde{\boldsymbol{y}}'_i)^2 \tag{4.3}$$

où $d(.,.)$ désigne la distance euclidienne dans l'image. Ce critère est non-linéaire et sa minimisation au sens des moindres carrés nécessite l'utilisation d'une méthode itérative comme Levenberg-Marquardt (initialisée avec une solution donnée par la méthode linéaire).

Des variantes sont proposées pour chacun des critères et les méthodes d'estimation de $\mathbf{H}$ sont comparées. Ce que le papier montre, c'est que, comme on s'y attend logiquement, la minimisation du critère géométrique donne de meilleurs résultats que la minimisation du critère linéaire.

Ce que le papier ne montre pas, par contre, c'est le comportement instable de la méthode linéaire. On s'aperçoit que, en pratique, même avec un nombre de points important (≥ 50) et une précision sur la localisation des points raisonnable (≤ 0.5 *pix.*) la méthode linéaire peut donner des résultats aberrants. Dans de tels cas, la méthode non-linéaire, converge généralement dans un minimum local qui n'est pas le minimum global attendu.

La normalisation [Har 95] permet d'améliorer le conditionnement des données mais ne résout malheureusement pas les problèmes d'instabilité liés à la méthode linéaire.

4.2 Estimateur quasi-linéaire

Pour remédier aux problèmes soulevés dans la section précédente, on a introduit un **estimateur quasi-linéaire** pour le calcul de $\mathbf{H}$. Cet estimateur est décrit dans le papier [Dem 00a] qui sera présenté dans le chapitre 7 dans le cadre plus général de la détection de mouvement.

L'idée principale est que le critère géométrique défini en (4.3) peut se mettre sous la forme d'un critère linéaire, semblable à (4.2). Plus exactement, ce critère s'écrit sous la forme de la norme d'une fonction linéaire en les éléments de $\mathbf{H}$, pondérés par des poids w_i :

$$E_{geom} = \sum_i ||(w_i \mathbf{A}_i)\boldsymbol{h}||^2 \tag{4.4}$$

où $\boldsymbol{h}$ est le vecteur dont les composantes sont les éléments de $\mathbf{H}$, les $\mathbf{A}_i$ sont des matrices ne dépendant que de quantités connues du problème (coordonnées des points, matrices de projection) et les w_i sont des fonctions de $\boldsymbol{h}$. Les poids w_i sont des facteurs de pondérations qui permettent d'exprimer le problème sous une forme linéaire mais en donnant à chaque équation l'influence optimale.

En utilisant une méthode itérative, on peut alors mettre à jour les poids $w_i(\boldsymbol{h})$ en fonction de l'estimation courante de $\boldsymbol{h}$ et minimiser le critère E_{geom} par rapport à $\boldsymbol{h}$ (les poids w_i étant fixés) avec une méthode linéaire.

L'idée de relier un critère géométrique à un critère linéaire rejoint des travaux tels que [Bea 94, Har 95, Zha 96]. Outre sa simplicité d'implémentation, l'estimateur quasi-linéaire possède les qualités de chacun des deux critère cités précédemment :

- **précision.** Les expériences montrent que l'estimateur quasi-linéaire donne une estimation de $\mathbf{H}$ sensiblement équivalente à celle donnée par une méthode non-linéaire (en terme d'erreur de reprojection dans les images). La raison principale est que dans les deux cas, c'est le même critère E_{geom} qui est minimisé. Ce n'est que la technique de minimisation utilisée qui change.

- **convergence.** La principale qualité de l'estimateur est de converger quasi systématiquement. En effet, bien que la méthode employée ici soit itérative, la minimisation ne correspond pas à une descente de gradient et n'est pas soumise aux problèmes de minima locaux. En fait, à chaque itération, ce n'est pas l'estimation $\boldsymbol{h}$ que l'on raffine, c'est le système (4.4) que l'on fait passer progressivement de "algébrique" à "géométrique". Par ailleurs, on constate que la convergence de l'estimateur quasi-linéaire est très rapide (2 à 3 itérations suffisent).

  Néanmoins il est nécessaire d'ajouter ici que ces remarques sur la convergence de l'estimateur ne sont pas démontrées : elles ont juste été observées au cours de nombreuses expérimentations. D'ailleurs, on constate que l'estimateur quasi-linéaire seul ne suffit pas à obtenir de bons résultats. La normalisation des données s'avère indispensable car elle permet de mieux conditionner les équations : sans cette normalisation, l'estimateur quasi-linéaire donne des résultats très décevants.

Perspectives

On a proposé et étudié différentes méthodes d'estimation des homographies *3–D* à partir de points appariés dans deux paires d'images.

Une autre approche consiste à maximiser la vraisemblance *a posteriori* de $\mathbf{H}$ et d'une structure *3–D* étant donnés les points $\boldsymbol{x}_i$, $\boldsymbol{x}'_i$, $\boldsymbol{y}_i$ et $\boldsymbol{y}'_i$. Ici une telle approche revient à considérer comme inconnues $\mathbf{H}$ et les reconstructions projectives $\tilde{\boldsymbol{X}}_i$ et $\tilde{\boldsymbol{Y}}_i$ avant et après mouvement, et minimiser le critère :

$$E_{map} = \sum_i d(\boldsymbol{x}_i, \tilde{\boldsymbol{x}}_i)^2 + d(\boldsymbol{x}'_i, \tilde{\boldsymbol{x}}'_i)^2 + d(\boldsymbol{y}_i, \tilde{\boldsymbol{y}}_i)^2 + d(\boldsymbol{y}'_i, \tilde{\boldsymbol{y}}'_i)^2$$

sous la contrainte que pour tout i, on a : $\tilde{\boldsymbol{Y}}_i \simeq \mathbf{H}\tilde{\boldsymbol{X}}_i$ où $(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}'_i)$ et $(\tilde{\boldsymbol{y}}_i, \tilde{\boldsymbol{y}}'_i)$ sont les projections respectives de $\tilde{\boldsymbol{X}}_i$ et $\tilde{\boldsymbol{Y}}_i$.

Cette approche est statistiquement optimale et fournit, en plus de l'homographie $\mathbf{H}$, une reconstruction *3–D* projective de la scène optimale. On remarque, qu'en pratique, on gagne finalement peu en précision sur l'estimation de $\mathbf{H}$ par rapport à une méthode non-linéaire ou quasi-linéaire. De plus, cette approche est lourde car implique un plus grand nombre d'inconnues. C'est pourquoi on ne l'utilise que dans le cadre d'estimation globale de structures et de mouvements (voir chapitre 6 pour une application au *bundle-adjustment stratifié*).

La méthode quasi-linéaire donne des résultats très satisfaisants en pratique. Cependant, d'un point de vue théorique, on peut regretter d'être obligé de travailler dans les images et non pas dans l'espace *3–D*. Le critère E_{geom} est un critère de reprojection et donc par nature non-linéaire et tous les problèmes de sa minimisation (recours à une méthode itérative) viennent de là ! Si l'on pouvait travailler dans l'espace *3–D* et que les équations étaient linéaires (comme dans le cas euclidien) les choses seraient plus faciles. Bien que l'espace projectif ne soit pas métrique, je reste convaincu qu'il existe un moyen de représenter le bruit des images dans l'espace projectif *3–D* et d'effectuer une minimisation directement dans l'espace *3–D* sans avoir à passer par des erreurs de reprojection dans les images.

Enfin, dans le cas d'une homographie *3–D* induite par un système stéréo *rigide*, la matrice $\mathbf{H}$ est conjuguée à une transformation euclidienne. La contrainte qui en résulte n'est pas utilisée dans les méthodes citées précédemment. Une solution pourrait être de trouver, dans ce cas, une paramétrisation adéquate pour les homographies (*c.-a.-d.* prenant en compte la relation de conjugaison). Ruf [Ruf 99] a proposé une telle paramétrisation dans le cas de mouvements planaires. Cette paramétrisation implique l'utilisation de contraintes qui sont difficiles à imposer en pratique mais l'approche semble être la bonne.

Chapitre 5

Auto-étalonnage : De l'affine à l'euclidien

Le processus de formation des images dans un système d'acquisition classique (appareils photographiques, caméras) est une transformation de projection des points de l'espace physique 3-D sur une ou plusieurs images. Dans le cadre de la reconstruction multi-images, c'est cette transformation qu'on désire retrouver. Une fois déterminée, cette transformation peut être inversée et permet alors d'obtenir la structure tridimensionnelle des points observés à partir de leurs projections dans les images.

Déterminer cette transformation, c'est **étalonner** le système d'acquisition. Étalonnage et reconstruction apparaissent alors comme deux problèmes duaux. Suivant l'espace dans lequel on désire reconstruire (projectif, affine ou euclidien), on parle alors d'étalonnage projectif, affine ou euclidien. Dans la mesure où la plupart des applications s'intéressent à des reconstructions euclidiennes, l'étalonnage désigne par défaut l'étalonnage euclidien.

L'étalonnage a beaucoup été étudié : son estimation repose généralement sur des contraintes liées à l'étalonnage interne des caméras [May 92, Har 93, Zis 95, Tri 97], à leur mouvement [Dro 93, Bas 93, Du 93] ou à la structure de la scène observée [Fau 93, Stu 99, Zha 99].

Lorsqu'on ne fait d'hypothèses que sur l'étalonnage interne des caméras – par exemple, constance des paramètres internes – on parle alors d'**auto-étalonnage**. Toutes les méthodes d'auto-étalonnage, à l'exception de [Tri 98], consiste à partir de structures projectives (reconstruction, matrices de projection), à utiliser des contraintes liées à l'étalonnage interne des caméras (*p.e.* constance de tout ou partie des paramètres internes) pour déterminer l'étalonnage et la structure euclidienne.

La relation entre structures projective et euclidienne est exprimée soit directement – avec les équations de Kruppa [Luo 94, Har 97b] par exemple – soit indirectement en

établissant un lien intermédiaire avec une structure affine. Dans ce dernier cas, on parle alors d'**approche stratifiée** [Fau 95].

Les problèmes majeurs rencontrés dans la mise en pratique de l'auto-étalonnage sont l'**instabilité numérique** des équations et les **mouvements critiques** [May 93], *i.e.* les mouvements du système d'acquisition qui ne permettent pas d'obtenir une solution unique pour l'étalonnage.

Travailler avec une approche stratifiée permet d'aborder plus simplement le problème d'auto-étalonnage. D'une part, les équations qui interviennent dans le problème sont linéaires et donnent des résultats beaucoup plus stables que des approches directes. D'autre part, le découpage de l'auto-étalonnage en deux étapes projectif-affine et affine-euclidien permet de comprendre beaucoup mieux les mouvements critiques.

Ce sont sur ces problèmes que les articles présentés ci-dessous se positionnent. Empruntant l'approche stratifiée, ces articles traitent, pour une part, du problème général de l'auto-étalonnage mais leur contribution la plus importante concerne le passage de l'affine à l'euclidien. En particulier, on montre comment on peut obtenir l'étalonnage d'une caméra à partir de la décomposition de Jordan d'homographies à l'infini entre plusieurs vues. L'approche est à la fois pratique et thórique puisqu'elle permet de mettre en évidence les mouvements critiques qui apparaissent lors du passage de l'affine à l'euclidien.

5.1 Résumé de «Autocalibration d'un Capteur Stéréoscopique en Mouvement Planaire» - ORASIS'97

Ce papier, publié à ORASIS'97, décrit l'auto-étalonnage d'un système stéréoscopique en mouvement planaire.

Auto-étalonnage à partir d'homographies 3-D

Des travaux antérieurs [Bea 95b, Zis 95, Arm 96] montrent comment l'auto-étalonnage d'un système stéréoscopique peut être réalisé à partir d'homographies 3-D (dont le chapitre précédent a donné les clefs pour leur estimation). Ces travaux reposent sur le fait que, lorsque le système stéréo a une géométrie constante, les matrices des homographies 3-D ont une forme particulière : elles sont les conjuguées de transformations euclidiennes. Plus précisément, on montre qu'étant donné un système stéréo, il existe une matrice $\mathbf{H}_e$ telle que, pour toute homographie 3-D $\mathbf{H}$ lié au système, il existe une transformation euclidienne $\mathbf{D} = \begin{pmatrix} \mathbf{R} & \boldsymbol{t} \\ 0 & 1 \end{pmatrix}$ telle que :

$$\mathbf{H} = {\mathbf{H}_e}^{-1}\mathbf{D}\mathbf{H}_e \tag{5.1}$$

La matrice $\mathbf{H}_e$ est une matrice de passage d'une base projective à une base euclidienne et ne dépend que de la géométrie du système stéréo. On montre qu'avec un choix judicieux

de base projective, la matrice $\mathbf{H}_e$ s'écrit :

$$\mathbf{H}_e = \begin{pmatrix} \mathbf{K}^{-1} & 0 \\ \boldsymbol{a}^\top & \beta \end{pmatrix}$$

où $\mathbf{K}$ est la matrice des paramètres intrinsèques d'une des caméras et $(\boldsymbol{a}^\top \ \beta)$ l'équation du plan à l'infini Π_∞.

L'auto-étalonnage consiste alors à utiliser la relation (5.1) pour déterminer $\mathbf{H}_e$ à partir de $\mathbf{H}$ – la transformation euclidienne $\mathbf{D}$ étant bien sûr, inconnue. La forme particulière de $\mathbf{H}_e$ avec une partie affine $(\boldsymbol{a}^\top \ \beta)$ et une partie euclidienne $\mathbf{K}^{-1}$ permet d'utiliser une approche stratifiée et de retrouver ces différentes parties affine et euclidienne en deux étapes successives.

L'étape d'étalonnage affine a été étudiée en détail dans [Hor 98]. L'étude repose sur une analyse algébrique de la relation (5.1). On montre, en particulier, que $\mathbf{H}^{-\top}$ et $\mathbf{D}^{-\top}$ ont les mêmes valeurs propres et que leurs vecteurs propres sont reliés entre eux par la matrice $\mathbf{H}_e$. Les valeurs propres de $\mathbf{H}^{-\top}$ sont donc celles de $\mathbf{D}^{-\top}$, *c.a.d.* $\lambda \in \{e^{-i\theta}, e^{i\theta}, 1, 1\}$ où θ est l'angle de la rotation $\mathbf{R}$. La multiplicité algébrique de la valeur propre $\lambda = 1$ est donc égale à 2 à moins que le mouvement soit une translation pure, *i.e.* $\theta = 0$, auquel cas sa multiplicité algébrique est 4 (le cas des rotations pures a été traité dans [Ruf 98]).

La multiplicité géométrique d'une valeur propre est égale à la dimension de l'espace propre associé [Hor 85]. Lorsque le mouvement considéré n'est pas une translation pure, la multiplicité géométrique de $\lambda = 1$ dépend du type de mouvement :

- elle est égale à 1 pour un mouvement général et
- elle est égale à 2 pour un mouvement planaire (axe de la rotation orthogonal au vecteur translation)

Dans le cas d'un mouvement général, on montre alors que le vecteur propre associé à la valeur propre 1 de $\mathbf{H}^{-\top}$ est l'équation du plan à l'infini $\Pi_\infty = (\boldsymbol{a}^\top \ \beta)$ qui est alors identifié à partir d'une unique homographie 3-D.

Dans le cas d'un mouvement planaire, l'espace propre associé à la valeur propre 1 de $\mathbf{H}^{-\top}$ correspond à un faisceau de plans. Le plan Π_∞ appartient à ce faisceau et ne peut pas être identifié à partir d'une unique homographie 3-D.

L'approche

Notre papier étudie le cas particulier d'un mouvement planaire. Il décrit dans un premier temps l'étalonnage affine et montre comment le plan à l'infini Π_∞ peut être estimé à partir de deux homographies. Bien qu'il n'apporte rien de nouveau sur l'étalonnage affine par rapport à [Bea 95b], il donne cependant de nombreux détails techniques (*p.e.* comment éviter le passage par des quantités imaginaires).

La contribution la plus significative de ce papier concerne l'introduction de la **décomposition de Jordan** pour l'estimation de l'étalonnage euclidien d'un capteur monoculaire à partir de son étalonnage affine. L'approche utilise la relation de conjugaison [Zis 95, Har 94] qui lie une homographie à l'infini $\mathbf{G}_\infty$ – associée à des vues prises avec un capteur de paramètres intrinsèques $\mathbf{K}$ – à une rotation $\mathbf{R}$:

$$\mathbf{G}_\infty = \mathbf{K}\mathbf{R}\mathbf{K}^{-1} \tag{5.2}$$

$\mathbf{G}_\infty$ est estimée après étalonnage affine. L'étalonnage euclidien consiste alors à estimer $\mathbf{K}$ à partir de la relation (5.2), la rotation $\mathbf{R}$ étant bien sûr inconnue. On montre que la relation (5.2) implique que $\mathbf{G}_\infty$ admet une décomposition de Jordan, *i.e.* il existe une matrice $\mathbf{S}$ inversible et un réel θ tels que :

$$\mathbf{G}_\infty = \mathbf{S}\mathbf{J}_\theta\mathbf{S}^{-1} \tag{5.3}$$

$$\text{avec } \mathbf{J}_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$\mathbf{G}_\infty$ peut être interprétée comme une *rotation projective*. La matrice $\mathbf{S}$ représente alors une matrice de changement de base permettant d'exprimer $\mathbf{G}_\infty$ sous une forme canonique $\mathbf{J}_\theta$.

La décomposition (5.3) n'est cependant pas unique. $\mathbf{J}_\theta$ est déterminée uniquement (au signe de θ près) mais on montre qu'il existe une famille à 3 degrés de liberté de matrices $\mathbf{S}$ qui vérifie (5.3). Une matrice $\mathbf{S}$ étant donnée – le papier donne un algorithme de décomposition – on montre qu'il existe un réel γ tel que :

$$\mathbf{K}\mathbf{K}^\top \simeq \mathbf{S} \begin{pmatrix} \gamma & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{S}^\top \tag{5.4}$$

La résolution de (5.4), d'inconnue γ, permet d'obtenir $\mathbf{K}$ et donc l'étalonnage euclidien. Cette approche se révèle être particulièrement efficace à différents points de vue :

- **Conditionnement numérique.** En présence de données bruitées, les homographies $\mathbf{G}_\infty$ estimées ne sont jamais exactement conjuguées à des rotations. Le calcul de la décomposition (5.3) effectue une correction en approchant $\mathbf{G}_\infty$ par la conjuguée d'une rotation ;

- **Utilisation de contraintes.** L'équation (5.4) fait apparaître clairement les liens entre les paramètres intrinsèques, les éléments de $\mathbf{S}$ et l'inconnue γ. Elle permet ainsi d'imposer facilement des contraintes sur certains paramètres intrinsèques (rapport d'aspect, skew).

5.2 Résumé de «Autocalibration in the Presence of Critical Motions» - BMVC'98

Cet article, publié à BMVC'98 s'inscrit dans la continuité du papier précédent. Il se consacre à l'étude de l'auto-étalonnage euclidien d'un capteur monoculaire en mouvement à partir de données affines. La méthode générale est donnée et les cas conduisant à une ambiguïté dans l'estimation de $\mathbf{K}$ sont identifiés et discutés.

Plus précisément, on suppose que l'étalonnage affine d'un capteur de paramètres intrinsèques $\mathbf{K}$ est connue et que les homographies à l'infini $\mathbf{G}_\infty$ sont estimées entre plusieurs vues. On étudie le calcul de $\mathbf{K}$ à partir des différentes matrices $\mathbf{G}_\infty$.

Cette étude utilise le formalisme développé précédemment à base de la décomposition de Jordan. La résolution générale de (5.4) est ici détaillée pas à pas. La forme des équations de (5.4) permet très facilement d'identifier les matrices $\mathbf{S}$ (et donc les mouvements du système monoculaire) qui rendent la résolution ambiguë. Ces différents *mouvements critiques* sont alors explicités et les hypothèses à faire sur les paramètres du système permettant de lever ces ambiguïtés sont discutées.

Cette étude fait apparaître que :

- les mouvements critiques correspondent à des situations courantes (rotations autour d'axes parallèles aux axes horizontal, vertical et normal de la caméra) ;
- la connaissance du rapport d'aspect permet de lever l'ambiguïté dans les cas de rotations autour d'axes parallèles aux axes horizontal et vertical ;
- les hypothèses sur la position du point principal ou de l'angle entre les axes de l'image ne servent en rien à la résolution de ces cas critiques.

Cette étude retrouve des résultats similaires à ceux publiés dans [Zis 98]. L'avantage d'utiliser la décomposition de Jordan n'est pas que d'ordre esthétique ! Cette approche fait clairement apparaître les paramètres intrinsèques dans les équations et leur influence est plus facilement identifiable.

On montre, dans ce papier, comment les équations peuvent être résolues dans le cas de mouvements critiques. Un travail intéressant pourrait être d'étudier la sensibilité de ces équations dans ces cas-là. En d'autres termes, lors d'un mouvement critique, est-ce que certains paramètres sont estimés avec plus de précision que d'autres ? Je pense que la réponse est positive mais l'étude reste à faire.

Conclusions

L'étude décrite dans les papiers ci-dessus permet de comprendre de façon plus intuitive les mouvements critiques pour l'étalonnage que des approches globales comme [Stu 97].

Ici on peut distinguer deux niveaux d'ambigüités :

- **Projectif-affine.** A ce niveau, les ambigüités concernent l'estimation de Π_∞. Celles-ci apparaissent si le mouvement est un mouvement planaire mais disparaissent dès lors que l'on dispose de deux mouvements planaires différents (*i.e.* d'axes de rotation différents). Il est important de noter qu'à cette étape les translations pures ne correspondent pas à des mouvements critiques [Ruf 98].
- **Affine-euclidien.** Ici les ambigüités concernent l'estimation de la conique absolue. Ces ambigüités correspondent à des orientations particulières de l'axe de la rotation du mouvement ou à une absence de rotation (à cette étape les translations correspondent à des mouvements critiques).

Étant donnée une séquence de mouvements et des hypothèses sur la connaissance de paramètres intrinsèques, on est maintenant en mesure de dire rapidement si la séquence correspond à une configuration critique pour l'étalonnage ou pas.

La contribution la plus importante concerne l'introduction de la décomposition de Jordan pour l'auto-étalonnage. Cette décomposition représente un moyen très naturel d'aborder le problème de l'auto-étalonnage. Elle fait apparaître $\mathbf{S}$ comme une matrice de changement de base entre une base affine dans laquelle une rotation s'écrit $\mathbf{G}_\infty$ et une base euclidienne dans laquelle cette même rotation s'écrit sous une forme canonique $\mathbf{J}_\theta$. En fait, lorsque la décomposition est estimée, le problème d'auto-étalonnage est presque résolu. La décomposition permet d'extraire de $\mathbf{G}_\infty$ ce qui est intéressant pour l'étalonnage, *c.-a.-d.* la matrice $\mathbf{S}$ (en fait θ n'apporte rien). Finalement l'approche utilisée ici consiste uniquement à appliquer des contraintes sur $\mathbf{S}$ pour pallier aux ambiguïtés dues à la décomposition.

Comme on va le voir dans le chapitre suivant, la décomposition de Jordan permet également d'exprimer le problème d'auto-étalonnage stéréo de façon très élégante.

Chapitre 6

Auto-étalonnage Stéréo

Le chapitre précédent étudie l'auto-étalonnage en insistant plus particulièrement sur le passage de l'affine à l'euclidien. Ce chapitre porte sur l'auto-étalonnage d'un système stéréoscopique appelé aussi **auto-étalonnage stéréo**.

Lorsque la géométrie d'un système stéréo est constante – paramètres intrinsèques et position relative des deux capteurs du système stéréo constants – il existe des contraintes sur les structures projectives qui permettent d'estimer l'étalonnage du système. Ces contraintes correspondent à l'invariance affine du plan à l'infini Π_∞ et à l'invariance euclidienne de la conique absolue $\mathbf{\Omega}_\infty$.

Les deux papiers ci-dessous exploitent ces contraintes de façon radicalement opposée. Le premier utilise une approche **algébrique** en utilisant la relation de conjugaison (5.1) des homographies *3–D*. Le second exploite directement la caractérisation **géométrique** de Π_∞ et de $\mathbf{\Omega}_\infty$.

6.1 Résumé de «Closed-form Solutions for the Euclidean Calibration of a Stereo Rig» - ECCV'98

Cet article, écrit avec Gabriella Csurka, Andréas Ruf et Radu Horaud décrit une méthode d'auto-étalonnage à partir d'une *unique* homographie *3–D*. Il s'agit ici d'un travail collectif dont le principal instigateur est Gabriella Csurka. Cependant, la méthode décrite repose essentiellement sur l'approche par décomposition de Jordan introduite dans le chapitre précédent. Ce travail ne constitue finalement qu'une extension de la décomposition de Jordan aux transformations de l'espace *3–D* et c'est pourquoi il a été intégré dans cette thèse.

L'idée principale vient du lien qu'on a mis en évidence entre relation de conjugaison (à une transformation euclidienne) et décomposition de Jordan. En effet, le chapitre précédent [Dem 97, Dem 98] se base sur la relation de conjugaison liant une homographie à l'infini $\mathbf{G}_\infty$ à une matrice de rotation $\mathbf{R}$ par la relation :

$$\mathbf{G}_\infty = \mathbf{K}\mathbf{R}\mathbf{K}^{-1} \tag{6.1}$$

On montre qu'une décomposition de Jordan de $\mathbf{G}_\infty$ implique une matrice $\mathbf{S}$ qui intervient alors dans l'équation :

$$\mathbf{K}\mathbf{K}^\top \simeq \mathbf{S} \begin{pmatrix} \gamma & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{S}^\top \tag{6.2}$$

Or, dans le cadre de l'auto-étalonnage stéréo, une autre relation de conjugaison apparaît également. Elle a d'ailleurs été citée dans le chapitre précédent : c'est la relation qui lie une homographie *3–D* $\mathbf{H}$ à une transformation euclidienne $\mathbf{D}$:

$$\mathbf{H} = \mathbf{H}_e \mathbf{D} {\mathbf{H}_e}^{-1} \tag{6.3}$$

Vues les analogies évidentes entre (6.1) et (6.3) on s'est demandé si une approche par décomposition de Jordan de $\mathbf{H}$ pouvait conduire à une équation analogue à (6.2).

Dans ce cas, la décomposition de Jordan de $\mathbf{H}$ prend la forme suivante :

$$\mathbf{H} = \mathbf{\Lambda} \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & \varepsilon \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{\Lambda}^{-1} = \mathbf{\Lambda}\mathbf{J}_{\varepsilon,\theta}\mathbf{\Lambda}^{-1} \tag{6.4}$$

où $\mathbf{\Lambda}$ est une matrice 4×4 inversible et ε est un scalaire prenant les valeurs 0 ou 1.

Le papier utilise la même méthodologie que celle développée dans [Dem 97, Dem 98]. En particulier, celui-ci :

- donne un algorithme pour calculer la décomposition de Jordan de $\mathbf{H}$;
- étudie les ambiguïtés inhérentes à la décomposition ;
- introduit les ambiguïtés de la décomposition dans l'équation de conjugaison (6.3) et dérive les contraintes pour faire le calcul de l'étalonnage.

Cette approche permet d'exprimer les contraintes sur $\mathbf{K}$ en fonction de la décomposition de $\mathbf{H}$. Elle est directe et évite le calcul intermédiaire du plan à l'infini.

Cependant cette étude ne fait que confirmer ce qui était annoncé dans le chapitre précédent. Les données de mouvement affine ($\mathbf{G}_\infty$) ou projectif ($\mathbf{H}$) contiennent des informations sur l'étalonnage du système d'acquisition que la décomposition de Jordan permet de faire ressortir – à quelques ambiguïtés près.

Une des autres propriétés de la décomposition (6.4) est qu'elle fournit, grâce à $\mathbf{J}_{\varepsilon,\theta}$, des informations sur la nature du mouvement. En effet $\mathbf{J}_{\varepsilon,\theta}$ correspond à la transformation $\mathbf{H}$ mais exprimée dans un repère canonique. De ce fait, $\mathbf{J}_{\varepsilon,\theta}$ et $\mathbf{H}$ sont de même nature.

- θ correspond à l'angle de la rotation du mouvement (si $\theta = 0$, le mouvement est une translation) ;
- Par ailleurs, la dernière colonne de $\mathbf{J}_{\varepsilon,\theta}$ correspond à la translation du mouvement.
  - Si $\varepsilon = 0$ la translation est $(0\ 0\ 0\ 1)^\top$ et le mouvement est planaire ;
  - Si $\varepsilon = 1$ le mouvement est général (vissage).

D'un point de vue pratique, on voit qu'il est donc possible d'obtenir des informations qualitatives sur le mouvement sans avoir à utiliser de représentation métrique.

6.2 Résumé de «Stereo Autocalibration from One Plane» - ECCV'2000

Cet article, publié à ECCV'2000, concerne l'auto-étalonnage stéréo avec une scène plane.

Motivations

Pourquoi travailler avec des scènes planes? Les motivations tiennent autant de l'expérience acquise en auto-étalonnage que de la curiosité scientifique. En effet, on s'aperçoit que, même si beaucoup de progrès ont été accomplis ces dernières années, les algorithmes d'auto-étalonnage sont toujours très difficiles à utiliser en pratique. Aucun algorithme d'auto-étalonnage ne peut prétendre rivaliser avec une méthode d'étalonnage usuelle avec une mire, qu'il s'agisse de :

- la **mise en oeuvre.** Une image et quelques *clicks* de souris pour l'étalonnage contre plusieurs dizaines d'images, de nombreuses mises en correspondance et une lourde procédure de bundle-adjustment pour l'auto-étalonnage ;
- la **précision.** Excellente pour l'étalonnage alors que très variable[1] pour l'auto-étalonnage.

Les échecs des méthodes d'auto-étalonnage ne sont pas seulement dus aux algorithmes utilisés. Ils proviennent principalement de l'insuffisante précision de localisation des points dans les images. Les structures projectives (transformations projectives, matrices de projection) obtenues à partir de ces points sont alors trop imprécises pour permettre un étalonnage correcte.

1. Souvent moins bonne que les données constructeur des caméras.

Pour réduire ces imprécisions sur ces structures projectives, il faut :

- soit augmenter la précision des données, *c.-a.-d.* augmenter la précision de détection des points ;
- soit ajouter des contraintes :
  - en augmentant le nombre de données (vues, primitives images) ;
  - en contraignant le système d'acquisition (mouvements contrôlés, modèles de capteurs simplifiés) ;
  - en contraignant la scène.

L'auto-étalonnage avec des scènes planes permet d'utiliser des **données précises** et de **contraindre la scène**. En effet, la localisation des points dans les images peut alors être obtenue par des techniques utilisant des fenêtres de corrélation déformables [Rem 94], avec une précision de l'ordre de 0.05 *pix.* (contre 0.5 *pix.* pour des images standards). D'autre part, la scène peut naturellement être contrainte par un plan et permet donc d'introduire une contrainte globale sur les reconstructions *3–D*.

L'approche

Au moment où a été écrit ce papier, le cas de scènes planes avaient déjà été étudié dans le cadre de l'étalonnage [Lie 99, Stu 99, Zha 99] et de l'auto-étalonnage [Tri 98] mais aucune étude n'avait été faite sur la stéréo.

L'auto-étalonnage stéréo avec des scènes planes est difficile. L'utilisation de techniques basées sur les homographies *3–D* est impossible (les homographies *3–D* ne peuvent pas être estimées car il faut au moins 5 points **non-coplanaires** pour estimer une homographie *3–D*) et il est alors nécessaire de retourner aux sources de la géométrie projective pour aborder le problème correctement.

Ce papier donne les bases du problème et démontre les résultats suivants :

- l'étalonnage affine peut être estimée *de façon unique* à partir de 3 vues stéréo ;
- l'étalonnage euclidien peut être estimée *de façon unique* à partir de :
  - 3 vues stéréo avec un modèle de capteur simplifié (*skew* et rapport d'aspect connus) ;
  - 4 vues stéréo avec un modèle de capteur général.

La stratification de l'espace en projectif-affine-euclidien se révèle (encore une fois) très adaptée.

L'**étalonnage projectif** du système stéréo est assez immédiate. L'estimation de la géométrie épipolaire est bien connue pour être impossible à partir d'une paire d'images

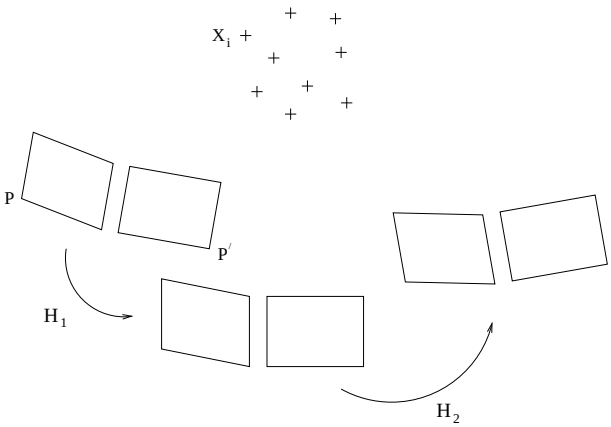

FIG. 6.1: *Système stéréo en mouvement observant un ensemble de points 3–D fixes.*

d'une scène plane. Cependant, comme on l'a vu dans le chapitre 3, cette géométrie épipolaire peut être estimée à partir de plusieurs paires d'images. Dans ce cas, le mouvement de la scène plane induit une structure *3–D* non planaire (voir Figure 3.3) et l'estimation de la matrice fondamentale $\mathbf{F}$ n'est plus ambiguë.

Soient Π_i les différents plans correspondant aux différentes positions de la scène planaire. Une fois $\mathbf{F}$ estimée, les corrdonnées des différents plans Π_i peuvent alors être obtenues dans l'espace projectif.

L'**étalonnage affine** revient à déterminer l'équation du plan à l'infini Π_∞. Cette étape est effectuée par considération des droites à l'infini L_i correspondant aux plans Π_i. Ces droites L_i sont *coplanaires* (par définition, elle sont contenues dans Π_∞) et *correspondent à la même droite physique* de la scène. En écrivant ces contraintes sur les projections de ces droites dans les images on obtient un système quadratique que l'on peut résoudre avec 3 paires d'images et donne finalement Π_∞.

Une fois les droites à l'infini L_i connues, l'**étalonnage euclidien** se fait par considération des points cycliques I_i et $\bar{I}_i$ de chaque plan Π_i. Par définition, les points I_i et $\bar{I}_i$ *sont sur les droites* L_i et l'ensemble de tous les points cyclique *se situe sur une conique* (la conique absolue Ω_∞). En écrivant ceci sous forme de contraintes, on obtient un système cubique que l'on peut résoudre avec 4 paires d'images. Cependant, lorsqu'on connaît le *rapport d'aspect* et le *skew* d'un des capteurs du système stéréo on obtient des contraintes supplémentaires et le système obtenu est alors quadratique et peut se résoudre avec 3 paires d'images. La conique Ω_∞ peut alors être estimée.

6.3 Bundle-adjustment stratifié

La méthode d'auto-étalonnage proposée dans la section précédente fonctionne plutôt bien. En pratique, l'utilisation d'une procédure d'*ajustement de faisceaux* ou *bundle-*

adjustment à chaque étape de la stratification projective/affine/euclidien permet d'obtenir des résultats d'étalonnage très précis. Dans la mesure où cette procédure n'a pas été décrite dans le papier, nous allons le faire ici.

Le bundle-adjustment est une méthode maintenant classique en vision *3–D*. Elle consiste, à partir des projections de points *3–D* dans plusieurs images, à trouver les matrices de projection et les reconstructions *3–D* qui correspondent le mieux à ces projections. L'idée géométrique qu'il y a derrière tout cela, c'est qu'on tente à la fois d'orienter les caméras et de déterminer les points *3–D* tels que les rayons de projection des points image correspondants passent au plus près de ces points *3–D*.

Dans le cadre de la stéréo, le problème du bundle-adjustment peut s'exprimer de la façon suivante. Supposons qu'un système stéréo rigide en mouvement (voir Figure 6.1) observe une scène composée de n points X_i $(1 \leq i \leq n)$. Soit m le nombre de paires d'images de la séquence correspondante. Soit $\mathbf{P}$ et $\mathbf{P}'$ les matrices de projection du système. On note $\boldsymbol{x}_{iv}$ et $\boldsymbol{x}'_{iv}$ les points de la paire v de la séquence $(1 \leq v \leq m)$ correspondant au point $\boldsymbol{X}_i$. Enfin on note $\mathbf{H}_v$ le déplacement correspondant au mouvement du système entre les paires 1 et v.

Soit $d(.,.)$ la distance euclidienne dans l'image entre deux points donnés en coordonnées homogènes. Soit E le critère défini par :

$$E = \sum_{v=1}^{m} \sum_{i=1}^{n} d(\boldsymbol{x}_{iv}, \mathbf{P}\mathbf{H}_v \boldsymbol{X}_i)^2 + d(\boldsymbol{x}'_{iv}, \mathbf{P}'\mathbf{H}_v \boldsymbol{X}_i)^2 \tag{6.5}$$

Le critère E correspond aux erreurs de reprojection des reconstruction estimées $\boldsymbol{X}_i$ dans toutes les paires d'images de la séquence. La procédure de bundle-adjustment correspond alors à la minimisation de ce critère par rapport à $\mathbf{P}$, $\mathbf{P}'$, $\boldsymbol{X}_1 ... \boldsymbol{X}_n$ et $\mathbf{H}_1 ... \mathbf{H}_m$. Cette approche est statistiquement optimale dans la mesure où elle correspond à la maximisation de la vraisemblance *a posteriori* des mouvements et structures étant donnés leurs projections $\boldsymbol{x}_{iv}$ et $\boldsymbol{x}'_{iv}$.

Suivant la nature de l'espace (projectif, affine ou euclidien) dans lequel on se trouve, différentes paramétrisations sont utilisées pour modéliser la structure (reconstructions, matrices de projection) et le mouvement :

- **projectif.** Les reconstructions $\boldsymbol{X}_i$ sont projectives. $\mathbf{P}$ et $\mathbf{P}'$ sont données par (2.2). $\mathbf{H}_v$ est une homographie *3–D* quelconque. Pour supprimer toute ambiguïté due aux "facteurs multiplicatifs près", on impose les contraintes $||\mathbf{H}_v||^2 = 1$ et $||\boldsymbol{X}_i||^2 = 1$ où $||.||$ désigne la norme $\mathcal{L}_2$ sur les vecteurs et les matrices.
- **affine.** Les reconstructions $\boldsymbol{X}_i$ sont affines. $\mathbf{P}$ et $\mathbf{P}'$ sont données par (2.3). $\mathbf{H}_v$ est une transformation affine, soit :

  $$\mathbf{H}_v = \begin{pmatrix} \mathbf{A} & \boldsymbol{b} \\ 0 & 1 \end{pmatrix}$$

  où $\mathbf{A}$ et une matrice 3×3 et $\boldsymbol{b}$ un vecteur de $\mathcal{R}^3$ arbitraires.

– **euclidien.** Les reconstructions $\boldsymbol{X}_i$ sont euclidiennes. $\mathbf{P}$ et $\mathbf{P}'$ sont données par (2.4). $\mathbf{H}_v$ est une transformation euclidienne, soit :

$$\mathbf{H}_v = \begin{pmatrix} \mathbf{R} & \boldsymbol{t} \\ 0 & 1 \end{pmatrix}$$

où $\mathbf{R}$ est une matrice 3×3 de rotation et $\boldsymbol{t}$ un vecteur de $\mathcal{R}^3$ arbitraires.

Dans le cadre de l'auto-étalonnage, le bundle-adjustment stratifié est utilisé de la façon suivante : après étalonnage de chacune des strates, une procédure de bundle-adjustment est initialisée avec les données estimées par étalonnage. La minimisation du critère E (6.5) est ensuite lancée en utilisant les modèles de structures et de mouvements adéquats (voir ci-dessus).

Cette utilisation d'un bundle-adjustment stratifié semble totalement adaptée à l'étalonnage stéréo (et pas seulement au cas de scènes planes traité ici) car elle permet pour chaque strate (i) d'obtenir un étalonnage optimal et (ii) de raffiner la calcul des structures qui sont utilisées pour l'étalonnage de la strate suivante. Le schéma complet de l'auto-étalonnage stéréo est alors :

– estimation de la matrice fondamentale $\mathbf{F}$ à partir de toutes les paires d'images, puis calcul de $\mathbf{P}$ et $\mathbf{P}'$;

– estimation des reconstructions projectives $\boldsymbol{X}_i$ associées à chaque paire ;

– estimation des transformations projectives $\mathbf{H}_v$ associées aux mouvements entre des paires consécutives ;

– *bundle-adjustment projectif* ;

– auto-étalonnage affine, puis mise à jour de $\mathbf{P}$ et $\mathbf{P}'$, $\boldsymbol{X}_i$ et $\mathbf{H}_v$ [2] ;

– *bundle-adjustment affine* ;

– auto-étalonnage euclidien, puis mise à jour de $\mathbf{P}$ et $\mathbf{P}'$, $\boldsymbol{X}_i$ et $\mathbf{H}_v$ [2] ;

– *bundle-adjustment euclidien.*

Cette approche permet ainsi d'éviter l'accumulation d'erreurs d'étalonnage dues au passage intermédiaire par une structure affine. En effet, dans la plupart des méthodes d'auto-étalonnage stratifié, l'étalonnage affine est obtenu à l'aide de méthodes non optimales. Cet étalonnage affine est utilisé pour initialiser des données affines (*p.e.* homographies à l'infini) qui sont utilisées à leur tour pour l'estimation de l'étalonnage euclidien.

2. On peut remarquer que dans le cas de scènes planaires, il existe une ambiguïté sur l'estimation des transformations $\mathbf{H}_v$ projectives et affines puisque celles-ci ne peuvent être déterminées qu'à partir de points en configuration non planaire. Dans ce cas, les transformations estimées ne correspondent pas aux *vraies* transformations. Ceci n'a cependant aucune influence dans le calcul des structure et dans l'auto-étalonnage.

Un problème qui arrive souvent en pratique est que, lorsque l'étalonnage affine est trop mal estimé, les données affines sont inutilisables pour permettre l'étalonnage euclidien.

Le recours à un bundle-adjustment affine permet d'obtenir des données affines optimales : l'étalonnage euclidien est alors sensiblement meilleur.

Chapitre 7

Détection et Segmentation du Mouvement

Le chapitre précédent décrit des approches pour auto-étalonner un système stéréo. Ainsi étalonné, un système stéréo permet d'obtenir une reconstruction tridimensionnelle métrique, voire euclidienne, de la scène. Cette reconstruction peut alors être utilisée pour des applications de vision telles que la réalité virtuelle ou augmentée, l'odométrie ou la robotique.

Cependant il existe de nombreuses autres applications pour lesquelles une structure métrique n'est pas indispensable. C'est le cas, par exemple, de la navigation de véhicules et de l'asservissement visuel pour lesquelles des structures affine [Bea 94] ou projective [Rob 95] peuvent suffire.

Les détection et segmentation du mouvement sont également des applications qui, comme le présente l'article ci-dessous, ne nécessitent qu'une structure projective.

7.1 Résumé de «Motion-Egomotion Discrimination and Motion Segmentation from Image-pair Streams» - CVIU

Ce travail, publié dans CVIU et CVPR'99 (sous une forme plus réduite), décrit une approche de détection/segmentation du mouvement avec un système stéréo faiblement étalonné. On utilise ici les homographies *3–D* pour représenter le mouvement du système stéréo. Contrairement aux deux chapitres précédents, les homographies *3–D* sont ici utilisées sans arrières pensées de reconstructions métriques : on reste dans l'espace projectif du début à la fin.

Ce papier propose, dans le cadre de la stéréo faiblement étalonnée, des solutions pour des tâches élémentaires comme la poursuite de primitives, la détection du mouvement et la segmentation des images.

La **poursuite de primitives** ou *tracking* consiste ici à mettre en correspondance et poursuivre des points dans une séquence d'images stéréo. L'approche utilisée est celle décrite dans le chapitre 3. Cette approche permet de raffiner, au cours du temps, l'estimation de la géométrie épipolaire du système et de déterminer, à chaque instant t, les reconstructions projectives $\boldsymbol{X}_i(t)$ des points suivis.

La **détection de mouvement** consiste à déterminer le mouvement due au déplacement du capteur stéréo, appelé également *egomotion*. Ce mouvement correspond au mouvement rigide dominant observé dans la séquence stéréo. Le papier propose une estimation de ce mouvement dominant par estimation de l'homographie *3–D* dominante entre deux instants successifs à l'aide de l'algorithme robuste RANSAC. Un critère est alors attribué à chaque point $\boldsymbol{X}_i$ sur *toute* la séquence. Suivant la valeur de ce critère, le point est considéré comme fixe – appartenant à la scène – ou en mouvement.

La **segmentation** des images consiste à partitionner les images en différentes zones, chacune étant constituée de points similaires, *c.-a.-d.* possédant des propriétés communes. Différents critères ont été utilisés depuis les premières heures de la vision. Ainsi, sur les images fixes, on a beaucoup utilisé la texture et la couleur pour faire de la segmentation en région. L'apparition de séquences vidéo a permis d'intégrer le temps et le mouvement comme composantes de l'image.

La plupart des travaux en segmentation du mouvement considèrent que tous les objets de la scène sont rigides. La segmentation se fait alors suivant un critère de similarité lié à une entité géométrique utilisée pour caractériser le mouvement (matrice fondamentale [Tor 94], tenseur trifocal [Tor 95] et autres caractérisations du mouvement rigide [Ira 98, Kel 95, Yi 97]). Indépendamment des algorithmes utilisés, certaines critiques peuvent être faites à l'égard de ces travaux. Ils limitent, bien entendu, la segmentation aux objets rigides et ne peuvent être étendus naturellement à des objets non rigides. D'un point de vue pratique, ces méthodes posent de nombreux problèmes lorsque les objets mobiles sont petits. Les points détectés sur ces objets sont alors peu nombreux et mal distribués dans l'espace. Dans de tels cas, la segmentation est en général mauvaise car le calcul des entités utilisées pour caractériser le mouvement devient instable.

L'approche développée dans le papier n'effectue la segmentation que sur les points mobiles trouvés à l'étape de détection du mouvement décrite précédemment. Le critère de similarité entre deux points est alors défini comme une fonction de la distance entre ces deux points dans les images de la séquence. En dépit de sa simplicité, ce critère se montre, en pratique, suffisant pour différencier des objets indépendants. Les raisons principales sont que les informations concernant la proximité des points est très redondante (séquence, stéréo).

7.2 Notes sur RANSAC/Moindres Carrés Médians

Dans le cadre de la vision, RANSAC [Fis 81] et MCM (Moindres Carrés Médians) [Rou 87] sont des méthodes robustes qui ont connu beaucoup de succès en raison de leurs simplicité et efficacité. Les notes ci-dessous complètent les explications données dans l'article introduit précédemment.

Qu'est ce qu'une méthode robuste?

Les méthodes d'optimisation usuelles font l'hypothèse que le bruit lié aux données est connu. La solution (statistiquement) optimale du problème est alors donnée par minimisation d'un critère au sens des moindres carrés (pondérés par les variances des données).

Cependant il est bien connu que la minimisation au sens des moindres carrés n'est pas robuste à la sous-estimation de la variance du bruit sur les données (*i.e.* cas de données aberrantes). Dans un problème pour lequel on suppose, par exemple, que le bruit sur les données est homogène (*c.-a.-d.* de même variance pour toutes les données), la présence de données aberrantes (correspondant à un bruit de variance très élevée et donc sous-estimée) influe de façon néfaste sur la solution estimée qui est alors très éloignée de la solution désirée. Pire, dans le cadre des moindre carrés, l'influence d'une donnée est proportionnelle à l'erreur sur cette donnée. Ainsi, même si une seule donnée est aberrante, l'erreur correspondante est très élevée et l'influence de cette donnée devient prépondérante par rapport aux données exactes, conduisant alors à des résultats absurdes.

Dans de tels cas, on est amené à utiliser des méthodes **robustes**. Ces méthodes optimisent des critères robustes et ré-estiment la distribution du bruit sur les données (en pratique, elles classifient juste les données en *inliers* et *outliers*).

Cas de la détection de mouvements

Dans la plupart des applications, l'usage de RANSAC et de MCM s'avèrent équivalents car les outliers correspondent à des erreurs aberrantes et donc très élevées. Les valeurs des critères robustes entre *bons* et *mauvais* modèles sont alors suffisamment importantes pour permettre d'identifier les *bons* modèles.

Dans ce papier, les outliers correspondent à des objets de faibles mouvements et donc à de petites erreurs (seulement deux à trois fois plus élevées que les erreurs sur les inliers). Dans un tel cas, MCM est trop grossière pour permettre une estimation robuste fiable. En effet, l'écart des valeurs de la médiane des erreurs (critère de MCM) entre *bons* et *mauvais* modèles n'est pas assez important pour permettre d'identifier sûrement les *bons* modèles.

Le critère utilisé dans RANSAC prend en considération la variance théorique des inliers (information très riche non utilisée par MCM) et permet une estimation robuste beaucoup plus fine: on constate que l'écart des valeurs de ce critère entre *bons* et *mauvais* modèles est beaucoup plus important que dans le cas de MCM et permet d'identifier plus sûrement les bons modèles.

Améliorations de RANSAC

Les méthodes robustes itératives comme RANSAC et MCM sont très performantes et d'implémentation facile. Cependant elles nécessitent généralement de très nombreuses itérations. En effet, le nombre théorique d'itérations à faire afin de trouver une solution acceptable est déjà élevé, mais en pratique ce nombre théorique doit encore être multiplié par un facteur 5-10 car celui-ci ne tient pas compte du bruit sur les inliers. En effet, à une itération donnée, même si l'échantillon tiré aléatoirement contient uniquement des inliers, l'estimation du modèle (homographie *3–D* dans notre cas) à partir de cet échantillon est souvent trop imprécise pour permettre de retrouver une grande partie des inliers : on doit alors faire suffisamment de tirages de façon à être sûr d'avoir obtenu, au moins une fois, un échantillon (i) composé uniquement d'inliers, et (ii) qui donne une estimation du modèle assez bonne pour permettre d'identifier les autres inliers.

L'estimation du modèle à partir de l'échantillon tiré aléatoirement est une étape à laquelle il faut apporter beaucoup de soins. Il faut que cette estimation soit précise et qu'elle permette, lorsqu'elle correspond à un *bon* échantillon, de retrouver le plus grand nombre d'inliers.

Dans le papier, l'utilisation de l'estimateur quasi-linéaire pour l'estimation d'homographie *3–D* permet d'obtenir une estimation précise. Une des solutions que nous avons trouvé pour retrouver le plus grand nombre d'inliers dans le cas d'un bon échantillon, c'est d'utiliser, dès qu'un *bon* modèle semble avoir été trouvé (*c.-a.-d.* dès qu'un nombre suffisant d'inliers a été détecté), un M-estimateur à l'intérieur de la boucle de RANSAC en utilisant comme initialisation le modèle estimé à partir de l'échantillon. Avec cette modification de RANSAC, on constate que le nombre d'itérations à faire est considérablement réduit car dès qu'un *bon* échantillon est tiré, les inliers détectés permettent de ré-estimer le modèle avec une plus grande précision. Le nouveau modèle permet alors de retrouver de nouveaux inliers, etc... On pourrait penser que l'utilisation d'un M-estimateur impliquerait que plus d'itérations soient à faire (puisque un M-estimateur est lui-même un algorithme itératif). Néanmoins, une mise à jour du nombre théorique d'itérations en fonction du pourcentage d'outliers courant (estimé dans la boucle de RANSAC à partir du plus grand ensemble d'inliers trouvé à un instant donné) aboutit finalement à un nombre total d'itérations moindre (2 à 3 fois plus petit que dans le cas d'un RANSAC standard).

Beaucoup de travail reste à faire dans cette direction pour améliorer les performances de RANSAC. Plus exactement, il semble naturel de penser que lors de l'estimation d'un modèle à partir d'un échantillon, l'incertitude liée à cette estimation (provenant de l'incertitude des données de l'échantillon) pourrait être utilisée afin de détecter les inliers éventuels.

Chapitre 8

Conclusions et perspectives

Nous avons étudié dans cette thèse les moyens d'estimer et d'utiliser les homographies *3–D* induites par le déplacement d'un système stéréo faiblement étalonné. Une chaîne de traitement complète a été proposée :

- mise en correspondance et poursuite de points dans une séquence stéréo;
- évaluation d'homographies *3–D* dominantes et, par effet de bord, détection et segmentation de la scène;
- auto-étalonnage du système stéréo.

On a montré que l'espace projectif se prêtait très bien à une représentation du mouvement. On a vu que les homographies *3–D* estimées à partir du mouvement d'un système stéréo contenaient des informations pertinentes sur le mouvement lui-même (angle de la rotation θ, nature du mouvement) mais également sur l'étalonnage des capteurs. On peut, dès lors, remonter à des données métriques mais on peut aussi choisir de rester à un niveau projectif.

Les applications sont nombreuses. Dans le cadre de la navigation de robots, cette thèse fournit les outils pour utiliser un système stéréo pour détecter des obstacles en mouvement, faire une reconstruction *3–D* de la scène et estimer le mouvement du robot. Une application originale consisterait d'ailleurs à fusionner les informations provenant d'une centrale inertielle classique (bien connue pour sa dérive temporelle) avec les données image pour estimer précisément le mouvement du robot.

Dans le cadre d'applications de surveillance, notre système fournit une approche pour détecter des personnes ou véhicules en mouvement et éventuellement asservir le système

stéréo sur un objet particulier en utilisant, par exemple, un algorithme d'asservissement stéréo projectif [Ruf 97].

La trame d'étude fournit dans cette thèse peut être approfondie sur plusieurs points.

Concernant l'auto-étalonnage stéréo, des développements théoriques sont possibles dans différentes directions (paramètres intrinsèques variables, utilisation de primitives autres que les points, ...). Une voie intéressante serait l'utilisation de scènes planes. L'approche que nous proposons dans cette thèse donne de très bons résultats et s'avère en pratique facile d'utilisation – la mise en correspondance de points est plus facile, la méthode donne des résultats satisfaisants avec 5-6 paires de vues. La méthode pourrait être étendue au cas de scènes dans lesquelles on a identifié plusieurs plans, ce qui permettrait d'utiliser encore moins de vues pour obtenir un étalonnage précis. Par ailleurs, lorsqu'on regarde de près la méthode d'auto-étalonnage stéréo par scènes planaires, on s'aperçoit que finalement, les contraintes liées à la stéréo sont peu utilisées. Ceci laisse penser que la méthode pourrait être adaptée à l'étalonnage d'un système monoculaire.

Enfin, le sujet abordé dans cette thèse qui me semble le plus intéressant à développer est la détection/segmentation du mouvement par stéréo. L'approche développée ici fournit les outils géométriques pour exploiter la rigidité du système stéréo (géométrie épipolaire, homographies *3–D*) dans les tâches de tracking et de détection de mouvement. Ce qui n'est pas montré dans la thèse mais qui pourrait être obtenu avec peu d'efforts, c'est une *segmentation dense* de la scène. En effet, une fois l'homographie *3–D* dominante estimée à partir de la mise en correspondance de points d'intérêt, on dispose de suffisamment de contraintes entre deux paires d'images (géométrie épipolaire, homographie *3–D*) pour apparier un ensemble dense de points. Une relation entre les champs des vitesses apparentes (flot optique) des points dans les deux images de la paire stéréo peut être estimée [Wan 96] et utilisée pour contraindre le calcul robuste de ces champs de vitesse dans les deux images, donnant ainsi une segmentation dense du mouvement (l'approche serait similaire au cas monoculaire [Odo 94]).

Par ailleurs, je pense que la géométrie n'est qu'un des aspects de la segmentation du mouvement. Elle correspond à des critères très objectifs. Cependant, suivant le type d'objets que l'on est amené à détecter, plusieurs autres critères peuvent être utilisés. Dans le cas de la détection de personnes (en mouvement), par exemple, la couleur et la texture (peau, habits) sont des indices caractéristiques qui peuvent être utilisés en plus du mouvement *3–D* pour la segmentation. La fusion d'informations de différentes natures (géométrie, caractéristiques images) s'avère indispensable dans ce cas.

Il ne fait aucun doute que le succès de l'analyse de scènes dynamiques tient en l'utilisation de nombreuses informations, de natures différentes (temporelles, spatiales, photométriques, ...). Dans cette perspective, l'utilisation d'outils statistiques (réseaux bayésiens, modèles de Markov cachés) s'avère indispensable pour pouvoir gérer la masse importante de connaissances issues des images.

Annexe A

Détection du regard par stéréo

Le document suivant est un rapport technique présentant une approche pour déterminer le regard d'un observateur (*c.a.d.* le point ou la partie de l'écran qu'un observateur fixe du regard) à l'aide d'un système stéréo. Ce travail a été effectué, dans le cadre du projet VIMINI (projet LTR n.27603), au *Vision Lab* de l'université de Gênes (Italie).

Bibliographie

[Ana 89] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2: 283–310, 1989.

[Arm 96] M.N. Armstrong. *Self-Calibration from Image Sequences*. PhD thesis, Department of Engineering Science, University of Oxford, UK, December 1996.

[Aru 87] K.S. Arun, T.S. Huang, and S.D. Blostein. Least-squares fitting of two 3-d points sets. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 9: 698–700, 1987.

[Bar 94] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1): 43–77, 1994.

[Bas 93] A. Basu. Active calibration: Alternative strategy and analysis. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 495–500. IEEE, IEEE Computer Society Press, June 1993.

[Bea 94] P. Beardsley, A. Zisserman, and D. Murray. Navigation using affine structure from motion. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 85–96, 1994.

[Bea 95a] P.A. Beardsley, I.D. Reid, A. Zisserman, and D.W. Murray. Active visual navigation using non-metric structure. In E. Grimson, editor, *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 58–64. IEEE Computer Society Press, June 1995.

[Bea 95b] P.A. Beardsley and A. Zisserman. Affine calibration of mobile vehicles. In R. Mohr and C. Wu, editors, *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, Xian, China*, pages 214–221. Xidan University Press, April 1995.

[Bla 98] J. Blanc. *Synthèse de nouvelles vues d'une scène 3D à partir d'images existantes*. Thèse de doctorat, Institut National Polytechnique de Grenoble, January 1998.

[Bra 95] P. Brand. *Reconstruction tridimensionnelle d'une scène à partir d'une caméra en mouvement: de l'influence de la précision.* Thèse de doctorat, Université Claude Bernard, Lyon I, October 1995.

[Bro 71] D.C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8): 855–866, 1971.

[Csu 98] G. Csurka, D. Demirdjian, A. Ruf, and R. Horaud. Closed-form solutions for the euclidean calibration of a stereo rig. In *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, 1998.

[Csu 99] G. Csurka, D. Demirdjian, and R. Horaud. Finding the collineation between two projective reconstructions. *Computer Vision and Image Understanding*, 75(3): 260–268, September 1999.

[Dem 97] D. Demirdjian, G. Csurka, and R. Horaud. Autocalibration d'un capteur stereoscopique en mouvement planaire. In *Journées ORASIS 1997,La Colle sur Loup, France*, pages 7–16, 1997.

[Dem 98] D. Demirdjian, G. Csurka, and R. Horaud. Autocalibration in the presence of critical motions. In Paul H.Lewis and Mark S. Nixon, editors, *Proceedings of the ninth British Machine Vision Conference, Southampton, England*, volume 2, pages 751–759. British Machine Vision Association, September 1998.

[Dem 99] D. Demirdjian and R. Horaud. A projective framework for scene segmentation in the presence of moving objects. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, 1999. to appear.

[Dem 00a] D. Demirdjian and R. Horaud. Motion-egomotion discrimination and motion segmentation from image-pair streams. *Computer Vision and Image Understanding*, 78(1): 53–68, april 2000.

[Dem 00b] D. Demirdjian, A. Zisserman, and R. Horaud. Stereo autocalibration from one plane. In *Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland (to appear)*, 2000.

[Dro 93] L. Dron. Dynamic camera self-calibration from controlled motion sequences. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 501–506. IEEE Computer Society Press, 1993.

[Du 93] F. Du and M. Brady. Self-calibration of the intrinsic parameters of cameras for active vision systems. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 477–482. IEEE Computer Society Press, 1993.

[Duf 00] Y. Dufournaud, C. Schmid, and R. Horaud. Appariement d'images à des échelles différentes. In *12ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, volume 2, pages 327–336, February 2000.

[Fau 92] 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.

[Fau 93] O. Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint*. Artificial intelligence. The MIT Press, Cambridge, MA, USA, Cambridge, MA, 1993.

[Fau 95] O. Faugeras. Stratification of three-dimensional vision: Projective, affine and metric representations. *Journal of the Optical Society of America*, 12: 465–484, 1995.

[Fis 81] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6): 381 – 395, June 1981.

[Fua 91] P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve discontinuities. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence, Sydney, Australia*, August 1991.

[Har 88] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[Har 92] R.I. 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.

[Har 93] 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.

[Har 94] R.I. Hartley. Self-calibration from multiple views with a rotating camera. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 471–478. Springer-Verlag, May 1994.

[Har 95] R. Hartley. In defence of the 8-point algorithm. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1064–1070, June 1995.

[Har 97a] R. I. Hartley. Self-calibration of stationary cameras. *International Journal of Computer Vision*, 22(1): 5–23, 1997.

[Har 97b] R.I. Hartley. Kruppa's equations derived from the fundamental matrix. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(2): 133–135, February 1997.

[Hor 85] R.A. Horn and C.R. Johnson. *Matrix analysis.* Cambridge University Press, 1985.

[Hor 88] B.K.P. Horn, H.M. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America*, 5(7): 1127–1135, July 1988.

[Hor 98] R. Horaud, G. Csurka, and D. Demirdjian. Stereo calibration using rigid motions. Technical Report 3467, INRIA, Machine IntelligenceAugust 1998. Paper accepted to appear in IEEE Transactions on Pattern Analysis and.

[Ira 98] M. Irani and P. Anandan. A unified approach to moving object detection in 2d and 3d scenes. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 20(6): 577–589, June 1998.

[Kel 95] P. J. Kellman and M. K. Kaiser. Extracting object motion during observer motion: Combining constraints from optic flow and binocular disparity. *Journal of the Optical Society of America A*, 12(3): 623–625, 1995.

[Lan 97] Z.D. Lan. *Méthodes robustes en vision : application aux appariements visuels.* Thèse de doctorat, Institut National Polytechnique de Grenoble, 1997.

[Lie 99] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architectural models from images. In *Proc. EuroGraphics*, volume 18, pages 39–50, September 1999.

[Luo 94] Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 589–599, May 1994.

[Mat 99] B. Matei and P. Meer. Optimal rigid motion estimation and performance evaluation with bootstrap. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pages 339–345, 1999.

[May 92] 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.

[May 93] S. Maybank. *Theory of Reconstruction from Image Motion.* Springer-Verlag, 1993.

[Nag 83] H.H. Nagel. Displacement vectors derived from second order intensity variations in image sequences. *Computer Vision, Graphics and Image Processing*, 21: 85–117, 1983.

[Nag 87] H.H. Nagel. On the estimation of optical flow: Relations between different approches and some new results. *Artificial Intelligence*, 33: 299–324, 1987.

[Odo 94] J.M. Odobez and P. Bouthemy. Estimation robuste multi-échelle de modèles paramétrés de mouvement sur des scènes complexes. *International Journal of Computer Vision*, 11: 419–430, 1994.

[Oht 98] N. Ohta and K. Kanatani. Optimal estimation of three-dimensional rotation and reliability evaluation. In *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, pages 175–187, 1998.

[Rem 94] P. Remagnino, P. Brand, and R. Mohr. Correlation techniques in adaptative template matching with uncalibrated cameras. In *Vision Geometry III, SPIE's international symposium on photonic sensors & control for commercial applications*, volume 2356, pages 252–253, October 1994.

[Rob 95] L. Robert, M. Buffa, and M. Hebert. Weakly-calibrated stereo perception for rover navigation. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 46–51, June 1995.

[Rou 87] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*, volume XIV. John Wiley & Sons, Ltd., New York, 1987.

[Ruf 97] Andreas Ruf and Radu Horaud. Visual trajectories from uncalibrated images. In Radu Horaud and Francois Chaumette, editors, *Proceedings of Workshop on New Trends in Image-based Robot Servoing*, pages 83 – 91, Grenoble, France, m7 1997. IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 97,.

[Ruf 98] A. Ruf, G. Csurka, and R. Horaud. Projective translations and affine stereo calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, USA*, pages 475–481, 1998.

[Ruf 99] A. Ruf and R. Horaud. Projective rotations applied to non-metric pan-tilt head. *IEEE Computer Society*, June 1999.

[Sch 96] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris.* Thèse de doctorat, Institut National Polytechnique de Grenoble, GRAVIR – IMAG – INRIA Rhône–Alpes, July 1996.

[Shi 94] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pages 593–600, 1994.

[Stu 97] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 1100–1105, June 1997.

[Stu 99] P. Sturm and S. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, 1999.

[Tom 98] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto. Making good features track better. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, USA*, pages 178–183, june 1998.

[Tor 94] P.H.S. Torr and D.W. Murray. Stochastic motion clustering. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, volume 801 of *Lecture Notes in Computer Science*, pages 328–337, May 1994.

[Tor 95] P.H.S. Torr. *Motion Segmentation and Outlier Detection.* PhD thesis, University of Oxford, England, Department of Engineering Science, Parks Road, Oxford, 1995.

[Tri 97] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 609–614. IEEE Computer Society Press, June 1997.

[Tri 98] B. Triggs. Autocalibration from planar scenes. In *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, 1998.

[Tsa 87] 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, August 1987.

[Ume 91] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 13: 376–380, 1991.

[Wan 96] W. Wang and J.H. Duncan. Recovering the three-dimensional motion and structure of multiple moving objects from binocular image flows. *Computer Vision and Image Understanding*, 63(3): 430–446, May 1996.

[Wen 92] J. Weng, P. Cohen, and N. Rebibo. Motion and structure estimation from stereo image sequences. IEEE *Transactions on Robotics and Automation*, 8(3): 362–382, June 1992.

[Yi 97] J.W. Yi and J.H. Oh. Recursive resolving algorithm for multiple stereo and motion matches. *Image and Vision Computing*, 15(3): 181–196, March 1997.

[Zab 94] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondance. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 151–158. Springer-Verlag, May 1994.

[Zha 96] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. Technical Report RR 2927, INRIA, July 1996.

[Zha 99] Z. Zhang. A flexible new technique for camera calibration. In *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, September 1999.

[Zis 95] A. Zisserman, P.A. Beardsley, and I.D. Reid. Metric calibration of a stereo rig. In *Workshop on Representation of Visual Scenes, Cambridge, Massachusetts, USA*, pages 93–100, June 1995.

[Zis 98] A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. *Philosophical Transactions of the Royal Society of London, SERIES A*, 356(1740): 1193–1211, 1998.